I6T-MC-AMAN Statistical Analysis Plan Version 4.0

A Phase 3, Multicenter, Randomized, Double-Blind, Parallel, Placebo-Controlled Induction Study of Mirikizumab in Conventional-Failed and Biologic-Failed Patients with Moderately to Severely Active Ulcerative Colitis LUCENT 1

NCT03518086

Approval Date: 14-Jan-2021

Statistical Analysis Plan: I6T-MC-AMAN: A Phase 3, Multicenter, Randomized, Double-Blind, Parallel, Placebo-Controlled Induction Study of Mirikizumab in Conventional-Failed and Biologic-Failed Patients with Moderately to Severely Active Ulcerative Colitis
LUCENT 1

Mirikizumab (LY3074828) Ulcerative Colitis

Study I6T-MC-AMAN is a Phase 3, multicenter, randomized, double-blind, parallel, placebo-controlled induction study of Mirikizumab in conventional-failed and biologic-failed patients with moderately to severely active ulcerative colitis. The study consists of a 12-week induction period where subjects will receive 1 of 2 treatment arms (300 mg mirikizumab or placebo) every 4 weeks.

Eli Lilly and Company
Indianapolis, Indiana USA 46285
Protocol I6T-MC-AMAN
Phase 3
Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 26 October 2018
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 13 December 2019
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly: 28 August 2020
Statistical Analysis Plan Version 4 electronically signed and approved by Lilly on date provided below.

Approval Date: 14-Jan-2021 GMT

1. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

Table of Contents

Appendix **Page**

2. Revision History

Statistical analysis plan (SAP) Version 1 was approved on 30 October 2018 prior to the first unblinding and was based on protocol approved on 13 March 2018.

Statistical analysis plan Version 2 was approved on 13 Dec 2019 and was based on protocol amendment (a) approved 12 Sep 2019. The following updates were made in Version 2 after the first external safety DMC analysis but before the first unblinded analysis by the sponsor:

1. Made the following changes to the primary and major secondary endpoints:
    a. Included "The change from baseline in the Urgency Numeric Rating Scale score" as a major secondary endpoint.
    b. Included "Histologic-Endoscopic mucosal improvement" as a major secondary endpoint.
    c. Included "Alternate clinical remission" as a major secondary endpoint.
2. Updated the sample size calculations and assumptions based on the changed endpoints in in the protocol amendment (a) for Study I6T-MC-AMBG.
3. Clarified the study period definitions in Table AMAN.5.2.
4. For general methods Section 5.1.4:
    a. Clarified that the relative risk and risk difference will be adjusted for stratification factors.
    b. The odds ratio will not be presented for binary efficacy endpoints.
5. The graphical testing scheme (Section 5.7) was updated to reflect additional major secondary endpoints.
6. Several modifications to the categories and calculations for baseline characteristics were made to Table AMAN.3.1.
7. The specific summary table for prior and concomitant therapy were clarified (Section 5.11).
8. Made the following changes to Table AMAN.5.4 and Table AMAN.5.5:
    a. "Alternate Clinical Remission" endpoint was added.
    b. "Alternate Symptomatic Remission" endpoint was added.
    c. "Total Mayo Clinical Remission" and "Total Mayo Clinical Response" were added.
    d. "UCEIS endoscopic remission" endpoint was added.
    e. EQ-5D-5L items and population-based index will no longer be in the primary SAP document.
    f. "Urgency NRS >= 3 Point Improvement" was added.
    g. For histology:
        i. The description of the Geboes grades and calculation for the Geboes Score were updated.
        ii. Description of the "Robarts Histology index (RHI)" and "Nancy index" were added.
        iii. The endpoints "Primary Histologic Remission," "RHI < 3," "Nancy Index < 1," "Histologic Improvement," and "Alternative Histologic Improvement" were added.

   h. The endpoints "Histologic-Endoscopic Improvement" and "Mucosal Healing" were added.
   i. Analysis of extraintestinal manifestations was added.
9. Analysis of UC surgeries and hospitalization was added to Section 5.13.
10. Removed by patient listing of exposure in Section 5.15.1.
11. Clarified the analysis AESIs including the analysis of infections (Section 5.15.8.2), hypersensitivity (Section 5.15.8.3) and suicidal ideation/behavior and depression (Section 5.15.8.7).
12. Altered the calculation of visit date and clarified the calculations for daily diary in Appendix 1.
13. Added region definitions in Appendix 2.

Statistical analysis plan Version 3 was approved on 28 August 2020 and was based on protocol amendment (a) approved on 12 September 2019. The following updates were made in Version 3 after the first external safety DMC analysis but before the first unblinded analysis by the sponsor:

1. Added the following sections concerning the coronavirus disease 2019 (COVID-19) pandemic:
   a. A description of the COVID-19 addendum for handling patients during the COVID-19 pandemic was added in Section 4.2.
   b. A description of analysis considerations related to COVID-19 mitigations was added in Section 5.5, and sensitivity analysis added to Table AMAN.5.5.
   c. Clarified that even some mitigations approved under the COVID-19 addendum will cause patients to be excluded from the per-protocol population in Section 5.17
2. Added or edited the following sections based on an error found in some of the electronic clinical outcome assessment (eCOA) devices:
   a. A description of the error was added in Section 4.3 and Appendix 3.
   b. The definition of the modified intent-to-treat (mITT) population found in Table AMAN.5.1 was amended to exclude patients impacted by the eCOA transcription error. The safety population will include patients with the eCOA transcription error. Concurrence with the Food and Drug Administration (FDA) on this analyses was obtained based on the FDA Type C written response received on 20 July 2020.
   c. A general description of considerations for the primary analysis and sensitivity analysis related to the eCOA error was added Section 5.4.
   d. Additional ITT summary analysis in several sections, including Section 5.9.1, Section 5.9.2 and Section 5.11, was added.
3. Wording in Section 5.3.1 on estimands was clarified, and an alternative estimand was added.
4. A multiple imputation approach was added as a sensitivity analysis in Section 5.3.4 and Table AMAN.5.5. The multiple imputation approach will be used for ITT analysis instead of the previously described nonresponder imputation (NRI) approach.

5. A summary of sensitivity analysis was added in Section 5.12.2 and Section 5.12.3.2.
6. Added the definition and analysis of "Alternate Clinical Remission 2" in Table AMAN.5.4 and Table AMAN.5.5; added "Alternate Clinical Remission" to Table AMAN.5.5.
7. Changed the definition of 'Primary histologic remission' in Table AMAN.5.4.
8. Clarified what populations would be used for analysis of covariance (ANCOVA) analysis of Urgency numeric rating scale (NRS) in Table AMAN.5.5.
9. Edited categories and analysis description for extraintestinal manifestations in Table AMAN.5.4 and Table AMAN.5.5.
10. In the Table AMAN.5.3 baseline characteristics table, changed the description of several variables, removed/added some categories, and edited subgroup analysis.
11. Added additional details about Week 12 and final database lock (DBL) in Section 5.18. Moved information about potential pharmacokinetics DBL from Section 6 to Section 5.18 with wording edits.
12. Clarified that changes to the subgroup analysis will not require an amendment to the SAP, and added a section on subgroup analysis for the Urgency NRS endpoint in Section 5.16; also added the section on the Japan subgroup analysis.
13. Aligned the calculation of the time window for determining the Mayo score with the protocol in Appendix 1; stated that when eCOA duplicates exist, the first of the duplicates will be used (Appendix 1 and Section 5.12). Added that baseline will be imputed if not available for eCOA daily diary assessments.
14. Clarified several places in the SAP that "biologic failed" refers to "biologic (or tofacitinib) failed."
15. Added wording about exploratory endpoints to Section 3.2.
16. Fixed typographical errors in Section 5.14 changing Week 40 to Week 12.
17. Minor edits that did not change meaning:
    a. Induction period described as Treatment/Induction period
    b. Changed wording for Mantel-Haenszel estimator.
    c. "Categorical" changed to "binary."
    d. Changed wording "Mucosal Healing" to "Histologic-Endoscopic Mucosal Remission"

Statistical analysis plan Version 4 was approved on the date provided on the title page and was based on protocol amendment (a) approved on 12 September 2019. The following updates were made in Version 4 after the first external safety DMC analysis but before the first unblinded analysis by the sponsor:

1. Clarified in Section 5.1.4 that Fishers exact test would only be used as a supportive analysis for binary efficacy endpoints.
2. Added a note to Section 5.2 about adjusting for covariates when patients are inadvertent randomizations.
3. Added in Section 5.3 and 5.5 that endoscopies performed outside of the window from study days 71 to 113 will be considered missing for analysis purposes. This also impacted the diary window in Appendix 1.
4. Changed how the multiple imputation will be performed in the modified NRI (mNRI) analysis in Section 5.3.4.

5. Clarified in Section 5.3.2 that the estimand associated with the mixed-effects model for repeated measures (MMRM) analysis could be justified as consistent with the treatment policy strategy.
6. Added the tipping point analysis in Section 5.3.5.
7. Added specific safety sensitivity analysis in Section 5.4 in the mITT population.
8. Changed the definition of treatment compliance to state that a partial dose does not count towards compliance (Section 5.10).
9. An analysis of "Urgency Remission" was added to Section 5.12.
10. The MMRM subgroup analysis description was altered in Section 5.16.1.
11. Fixed typos and made other minor edits.

3. Study Objectives

3.1. Primary and Major Secondary Objectives

Table AMAN.3.1 shows the protocol defined primary and major secondary objectives and endpoints of the study. *In addition, the analysis of other secondary endpoints is described in Section 5.12 to provide supportive evidence of efficacy.*

The estimand (ICH E9 R1 2017) associated with each endpoint/analysis is documented in the following places:

- The population of interest is described in the protocol inclusion/exclusion criteria and in this document Table AMAN.5.1 and Table AMAN.5.5.
- The endpoint/variables may be found in Table AMAN.3.1, Table AMAN.5.4, and Table AMAN.5.5.
- The handling of intercurrent events and missing data may be found in Sections 5.3 and Table AMAN.5.5.
- Population summary measures are described in Section 5.1.4 and Table AMAN.5.5.

Table AMAN.3.1. Primary and Major Secondary Objectives and Endpoints

Objectives	Endpoints
Primary[a]	
• To test the hypothesis that mirikizumab is superior to placebo in inducing **clinical remission** at Week 12 in patients with moderately to severely active ulcerative colitis (UC)	• The proportion of patients in **clinical remission** at Week 12. Clinical remission is based on the modified Mayo score (MMS) and is defined as: o Stool frequency (SF) subscore = 0, or SF = 1 with a ≥1-point decrease from baseline, and o Rectal bleeding (RB) subscore = 0, and o Endoscopic subscore (ES) = 0 or 1 (excluding friability)
Major Secondary[a,b]	
• To test the hypothesis that mirikizumab is superior to placebo in **inducing alternate clinical remission** at Week 12 in patients with moderately to severely active ulcerative colitis (UC) [c]	• The proportion of patients in **alternate clinical remission** at Week 12. Clinical remission is based on the modified Mayo score (MMS) and is defined as [c]: o Stool frequency (SF) subscore = 0, or SF = 1, and o Rectal bleeding (RB) subscore = 0, and o Endoscopic subscore (ES) = 0 or 1 (excluding friability)
• To evaluate the efficacy of treatment with mirikizumab compared to placebo in inducing a **clinical response** at Week 12	• The proportion of patients in **clinical response** at Week 12. Clinical response is based on the MMS and is defined as: o A decrease in the MMS of ≥2 points and ≥30% decrease from baseline, and • A decrease of ≥1 point in the RB subscore from baseline or an RB score of 0 or 1

Objectives	Endpoints
• To evaluate the efficacy of treatment with mirikizumab compared to placebo in inducing **endoscopic remission** at Week 12	• The proportion of patients with **endoscopic remission** at Week 12, defined as: o ES = 0 or 1 (excluding friability)
• To evaluate the efficacy of treatment with mirikizumab compared to placebo in inducing **symptomatic remission** at Week 4	• The proportion of patients in **symptomatic remission** at Week 4, defined as: o SF= 0, or SF = 1 with a ≥1-point decrease from baseline, and o RB = 0
• To evaluate the efficacy of treatment with mirikizumab compared to placebo in inducing **symptomatic remission** at Week 12	• The proportion of patients in **symptomatic remission** at Week 12, defined as: o SF = 0, or SF = 1 with a ≥1-point decrease from baseline and o RB = 0
• To evaluate the efficacy of treatment with mirikizumab compared to placebo in inducing **clinical response** in the biologic-failed population at Week 12	• The proportion of patients in the biologic-failed population in **clinical response** at Week 12. Clinical response is based on the MMS and is defined as: o A decrease in the MMS of ≥2 points and ≥30% decrease from baseline, and o A decrease of ≥1 point in the RB subscore from baseline or an RB score of 0 or 1
• To evaluate the efficacy of treatment with mirikizumab compared to placebo in inducing **bowel movement urgency improvement** at Week 12[d]	• The change from baseline in the Urgency Numeric Rating Scale score[e]
• To evaluate the efficacy of treatment with mirikizumab compared to placebo in inducing histologic-endoscopic mucosal improvement at Week 12[f]	• The proportion of patients with histologic-endoscopic mucosal improvement at Week 12, defined as achieving both[f]: o **Histologic improvement**, defined using Geboes scoring system with neutrophil infiltration in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue. o **Endoscopic remission**, defined as ES = 0 or 1 (excluding friability)

Abbreviation: FDA = Food and Drug Administration.

[a] All primary and major secondary endpoints will be evaluated for mirikizumab versus placebo. All primary and major secondary endpoint analyses will utilize the multiplicity control approach based on 'graphical multiple testing procedure' to control the overall family-wise type I error rate at a 2-sided alpha level of 0.00125. The graphical multiple testing procedure described in Bretz et al. (2009) will be used.

[b] The order of testing of the major secondary endpoints will be determined from the results of the statistical simulations. Therefore, the order of the secondary endpoints does not reflect the order of the statistical testing.

[c] Alternate clinical remission was not included as an objective/endpoint in protocol amendment (a). However, this objective/endpoint is designated as "major secondary" (i.e., multiplicity controlled) in the SAP and will supersede the protocol amendment (a). The alternate definition of clinical remission is added based on the FDA's feedback on the mirikizumab pediatric program proposal.

[d] The statistical analysis plan (SAP) language for this objective supersedes the protocol language, which states "To evaluate the efficacy of treatment with mirikizumab compared to placebo in inducing bowel movement urgency improvement at Week 12 in patients with bowel urgency symptoms at baseline."

[e] The SAP language for this endpoint supersedes the protocol language, which states "The proportion of patients with bowel movement urgency improvement at Week 12 as defined in the study SAP."

[f] Histologic-endoscopic mucosal improvement was not included as an objective/endpoint in protocol amendment (a). However, this objective/endpoint is designated as "major secondary" (i.e., multiplicity controlled) in the SAP and will supersede the protocol amendment (a).

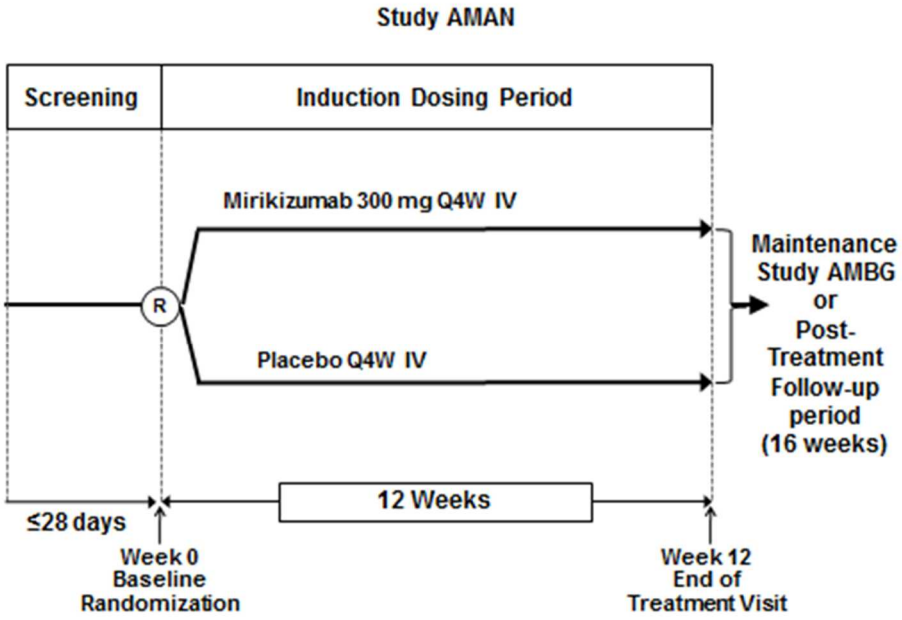
3.2. Other Secondary and Exploratory Objectives

Other Secondary Endpoints are described in the I6T-MC-AMAN (AMAN) Protocol Amendment (a), Section 4. The details for these objectives and endpoints can be found in Section 5.12 and in Table AMAN.5.4 and Table AMAN.5.5 below, along with additional exploratory endpoints not described in the protocol.

4. Study Design

4.1. Summary of Study Design

Study AMAN is a multicenter, randomized, double-blind, parallel-arm, placebo-controlled study designed to evaluate the safety and efficacy of mirikizumab, compared with placebo, over a 12-week induction period (approximately 1160 randomized patients). The study population includes patients with moderately to severely active ulcerative colitis (UC) who have an inadequate response to, loss of response to, or are intolerant to conventional (nonbiologic) therapy for UC (conventional-failed), and those who have an inadequate response to, loss of response to, or are intolerant to biologic therapy or tofacitinib for UC (biologic-failed). Throughout this document, the term biologic-failed will be used to refer to failure of either a biologic or tofacitinib unless otherwise specified. Complete definitions of the "conventional-failed" and "biologic-failed" terms are given in Section 6.1, Inclusion Criterion [8] of the protocol. Patients will be randomized with a 3:1 ratio to receive blinded intravenous (IV) administration of 300 mg mirikizumab or placebo every 4 weeks (Q4W) at Weeks 0, 4, and 8. Randomization will be stratified by (a) biologic-failed status (yes/no), (b) baseline corticosteroid use (yes/no), (c) baseline disease activity (modified Mayo score [MMS]: [4-6] or [7-9]), and (d) region (North America/Europe/Other).



Abbreviations: IV = intravenous; Q4W = every 4 weeks; R = randomization.

Figure AMAN.4.1. Illustration of study design for Clinical Protocol I6T-MC-AMAN.

4.2. COVID-19 Addendum

Study AMAN was ongoing during the global coronavirus disease 2019 (COVID-19) pandemic, during which many patients were unable or unwilling to conduct on-site clinic visits and/or have some study procedures performed. Mitigations for COVID-19 were initially implemented as emergency measures which have been formalized in an addendum to the primary protocol. Mitigations to allow these patients to continue in the Phase 3 mirikizumab UC program included, but were not limited to:

- Extending the window for the Week 12 endoscopy assessment
- Allowing a patient missing the Week 12 endoscopy, despite the window extension, to continue into the maintenance study
- Extending the window for investigational product (IP) administration
- Use of local laboratories if central safety laboratory testing could not be performed
- Use of virtual telephone visits if patients were unable to attend in-office visits for assessments (e.g., adverse events [AEs] and concomitant medications)

Additional details of the mitigations and planned analysis can be found in **Protocol Addendum I6T-MC-AMAN(15)** and in analysis Section 5.5 below.

4.3. Electronic Clinical Outcomes Assessment Transcription Error Protocol Addenda

Questions to assess patient-reported outcome (PRO) measures of stool frequency (SF) and rectal bleeding (RB) are recorded using electronic clinical outcome assessment (eCOA) devices. It was discovered that the devices for daily diary assessment of RB and SF Mayo subscores contained errors in the wording for patients in Poland and Turkey, respectively, as described in Appendix 3. We will refer to these errors in wording as "transcription" errors, and patients who were enrolled into the trial based on this incorrect eCOA assessment will be referred to as "impacted by the eCOA transcription errors." The wording on the devices was corrected after discovery of the issues. Any endpoints which make use of the SF and RB data will be difficult to interpret in the impacted patients from Poland and Turkey, respectively. An addendum was created to handle these patients by allowing them to roll over to the long-term extension study. Additional details of the issue and planned analysis can be found in the addendum, Appendix 3, and the analysis section, 5.4, below.

4.4. Determination of Sample Size

The study will randomize approximately 1160 patients with a 3:1 ratio of 300 mg mirikizumab to placebo, with an assumption that approximately 1044 patients will complete the study.

The power calculations for this study assume the following:

1. The randomized study population will include approximately 50% biologic-failed patients and approximately 50% conventional-failed patients.

2. The predicted clinical remission rates at Week 12 for mirikizumab versus placebo are expected to be 23% versus 7.8% (biologic-failed patients: 16% versus 3.5%; conventional-failed patients: 30% versus 12%).

The primary endpoint of this study is to test the hypothesis that mirikizumab is superior to placebo in inducing clinical remission at Week 12 in patients with moderately to severely active UC. Given the assumptions described above, a sample size of 1160 patients are expected to provide >90% power to demonstrate that mirikizumab is superior to placebo in achieving this endpoint, as assessed using a chi-square test with a 2-sided significance level of 0.00125.

Patients who complete Study AMAN may be eligible to participate in Study I6T-MC-AMBG (AMBG), a 40-week maintenance study. The primary objective of Study AMBG is to test the hypothesis that mirikizumab is superior to placebo in achieving clinical remission at Week 40 of Study AMBG (Week 52 of continuous therapy) amongst patients induced into clinical response with mirikizumab at Week 12 of Study AMAN. A sample size of 1160 patients in Study AMAN is predicted to ensure that there will be a sufficient number of biologic-failed patients in clinical remission at Week 12 of Study AMAN who will enter Study AMBG, as well as >90% power to demonstrate that mirikizumab is superior to placebo in achieving the primary endpoint in Study AMBG, as assessed using a chi-square test with a 2-sided significance level of 0.05.

4.5. Method of Assignment to Treatment

Patients who meet all criteria for enrollment will be randomized to double-blind treatment at Visit 1. Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web-response system (IWRS). To achieve between-group comparability, patients will be stratified to these arms based upon (a) biologic-failed status (yes/no), (b) baseline corticosteroid use (yes/no), (c) baseline disease activity (MMS: [4-6] or [7-9]), and (d) region (North America/Europe/Other). This stratification will be controlled by IWRS.

Patients will be randomized 3:1 to receive blinded IV administration of 300 mg mirikizumab, or placebo Q4W at Weeks 0, 4, and 8. Duration of treatment with the IP is 12 weeks.

5. A Priori Statistical Methods

5.1. General Considerations

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (hereafter Lilly) or its designee. The statistical analyses will be performed using SAS® Version 9.4 or higher. The latest version of the Medical Dictionary for Regulatory Activities (MedDRA®) will be used.

Not all displays and analyses described in this statistical analysis plan (SAP) will necessarily be included in the clinical study report (CSR). Not all displays will necessarily be created as a "static" display. Some displays may be incorporated as interactive display tools such as Spotfire instead of or in addition to a static display. Any display described in this SAP and not provided in the CSR would be available upon request.

Any change to the data analysis methods described in the protocol will require a protocol amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol and the justification for making the change will be described in the CSR.

Additional exploratory analyses of the data may be conducted as deemed appropriate. Some of these additional supplementary analyses may be prespecified in a separate supplemental SAP.

5.1.1. Analysis Populations

Patient populations are defined in Table AMAN.5.1 along with the analysis they will be used to conduct. The treatment groups and inferential comparisons described in Table AMAN.5.1 will be used unless otherwise specified. Also, unless otherwise specified, for all populations/analysis, patients will be analyzed according to the treatment to which they were assigned.

Table AMAN.5.1. Patient Populations for Analysis

Population	Description
Screening Population	**Definition:** All patients who signed informed consent. **Purpose:** Used for disposition analysis. **Treatment Groups:** None **Inferential Comparisons:** None
Modified Intent-to-Treat (mITT) Population	**Definition:** All randomized patients who received any amount of study treatment *excluding patients impacted by the eCOA transcription error in Poland and Turkey* (regardless if the patient does not receive the correct treatment, or otherwise does not follow the protocol). **Purpose:** Used for efficacy and health outcomes analysis. **Treatment Groups (Short Label):** 300 mg mirikizumab Q4W IV (MIRI), placebo Q4W IV (PBO) **Inferential Comparisons:** 300 mg mirikizumab Q4W IV vs. placebo Q4W IV
Safety Population	**Definition:** All randomized patients who received any amount of study treatment (regardless of whether the patient does not receive the correct treatment, or otherwise does not follow the protocol). **Purpose:** Used for safety-related analysis.

	Treatment Groups (Short Label): 300 mg mirikizumab Q4W IV (MIRI), placebo Q4W IV (PBO) **Inferential Comparisons:** 300 mg mirikizumab Q4W IV vs. placebo Q4W IV
Intent-to-Treat (ITT) Population	**Definition:** All randomized patients. Patients will be analyzed according to the treatment to which they were assigned. **Purpose:** Used as a sensitivity analysis for the primary and major secondary efficacy endpoints. **Treatment Groups (Short Label):** 300 mg mirikizumab Q4W IV (MIRI), placebo Q4W IV (PBO) **Inferential Comparisons:** 300 mg mirikizumab Q4W IV vs. placebo Q4W IV
Per-Protocol Population (PP)	**Definition:** All mITT patients who are not deemed noncompliant with treatment, who do not have significant protocol deviations, and whose investigator site does not have significant GCP deviations that require a report to regulatory agencies (regardless of study period). Qualifications and identification of the specific significant protocol deviations that result in exclusion from the PP population will be determined while the study remains blinded, prior to the database lock (See Section 5.17). **Purpose:** Used as a sensitivity analysis for the primary and major secondary efficacy endpoints. **Treatment Groups (Short Label):** 300 mg mirikizumab Q4W IV (MIRI), placebo Q4W IV (PBO) **Inferential Comparisons:** 300 mg mirikizumab Q4W IV vs. placebo Q4W IV

Abbreviations: eCOA = electronic clinical outcomes assessment; GCP = good clinical practice; IV = intravenous; MIRI = mirikizumab; PBO = placebo; Q4W = every 4 weeks.

5.1.2. Study Time Intervals

Table AMAN.5.2 displays a list of study periods along with the definition of which patients will be considered to have entered the study period and when the individuals start and end the study period. The table shows both a date and a time.

To calculate the length of any time interval or time period in this study the following formula will be used:

$$Length\ of\ interval\ (days) = End\ Date - Interval\ Start\ Date + 1$$

To convert any time length from days to years, the following formula will be used:

$$Length\ of\ interval\ (years) = Length\ of\ interval\ (days)/365.25$$

To convert any time length from days to weeks, the following formula will be used:

$$Length\ of\ interval\ (weeks) = Length\ of\ interval\ (days)/7$$

Only for the purpose of calculating the length of study period time intervals, the words “prior to” in Table AMAN.5.2 should be understood to mean “the day before” while the words “after” should be understood to mean “the day after.” For the purpose of determining whether a date/time lies within an interval these words are intended to convey whether the start or end of the period is inclusive of the specified date.

Table AMAN.5.2. Definition of Study Period Time Intervals

Study Period	Interval Start Definition	Interval End Definition
Screening: All patients who sign informed consent are considered as entering the Screening Period.	Informed consent date	Prior to the start of Induction Period.
Treatment/Induction: All patients who are randomized to the study are considered as entering the Induction Period.	At the first study drug administration date/time[a] following randomization. For patients who are randomized but not dosed, the Induction Period starts on the date of randomization.	The maximum of treatment discontinued date or last treatment visit date.
Post-Treatment Follow-Up: All patients who had a Visit 801 or 802 are considered as entering follow-up period.	After the Induction Period ends.	The maximum of the last study visit date or study disposition date.

[a] Missing dose time will be imputed as the earliest time that is consistent with available data about dose time. For example, suppose the minutes are missing but hour is present. In this case, we would impute the minutes to be 0.

5.1.3. Definition of Study Baseline

The baseline for variables collected as part of the daily diary (including the PRO components SF and RB of the Modified Mayo) will be calculated from valid daily diary entries obtained prior to baseline endoscopy preparation (see Appendix 1). The baseline endoscopy component of the Mayo will use the endoscopic appearance of the mucosa at the screening endoscopy. For other efficacy, health outcome and safety assessments, baseline is defined as the last nonmissing assessment recorded on or prior to the date of the first study drug administration at Visit 1 (Week 0).

Baseline for safety analysis is described in the safety section.

Change from baseline will be calculated as the visit value of interest minus the baseline value. If a baseline values or the value at the visit is missing for a particular variable, then the change from baseline is defined as missing.

5.1.4. Analysis Methods

Unless otherwise specified, variables will be analyzed in the original scale on which they are measured. The parametric approach will be employed by default for statistical analysis except when nonparametric analysis, such as by a rank-based method, is assessed to be more fitting. Additional exploratory analyses of the data will be conducted as deemed appropriate. All hypothesis tests will be 2-sided, and the family-wise type I error rate (FWER) will be controlled at an α level of 0.00125 for primary and major secondary endpoints using a pre-specified graphical procedure (see Section 5.7).

Unless otherwise specified, for the analyses of hypotheses with multiplicity control at a family wise significance level of 0.00125, a 2-sided 99.875% confidence interval (CI) will be provided along with the p-value. For other analyses of the hypotheses without multiplicity control, the

tests will be conducted using a 2-sided significance level of 0.05. The corresponding p-value along with its 95% 2-sided CI will be provided.

For assessments of the primary endpoints and other binary efficacy and health outcomes endpoints, the following will be provided unless otherwise specified:

- Crude proportions for each treatment group along with the 2-sided asymptotic (i.e., not continuity corrected) CIs will be provided.
- The estimated common risk difference along with 2-sided CIs. The common risk difference (Agresti 2013, pp231) is the difference in proportions adjusted for the stratification factors. SAS PROC FREQ will be used for the estimates and CIs, where the CIs are calculated by using Mantel-Haenszel estimator of risk differences with standard error calculated as described by Sato (1989).
- Cochran-Mantel-Haenszel (CMH) test will be used to compare the treatment groups while adjusting for the stratification factors. The CMH p-value will be reported.
- As a secondary measurement of efficacy, the relative risk along with its 2-sided CI will be provided, adjusting for the stratification factors using the Mantel-Haenszel estimator.
- If deemed necessary as a supportive analysis, additional analyses of binary efficacy variables may be conducted to address sparse data or small sample sizes. For example, a Fisher's exact test may be utilized.
- The stratification factors used for common risk difference, relative risk and CMH test are: (a) previous biologic therapy failure (yes/no), (b) baseline corticosteroid use (yes/no), (c) baseline disease activity (MMS: [4-6] or [7-9]), and (d) region (North America/Europe/Other). For the major secondary endpoint, which includes only biological failed patients, the stratification factors are: (b) baseline corticosteroid use (yes/no), (c) baseline disease activity (MMS: [4-6] or [7-9]), and (d) region (North America/Europe/Other).

When specified as a sensitivity analysis for binary endpoints, logistic regression with a Firth penalized likelihood will be used (Firth 1993). The model will include the treatment groups and the covariates described in Section 5.2. The Firth correction can be implemented in PROC Logistic by including *'firth'* as an option in the model statement. The odds ratio and the corresponding CIs, as well as the treatment differences and the corresponding CIs, will be reported.

Treatment comparisons of continuous efficacy and health outcome variables will be made using mixed-effects model for repeated measures (MMRM) analysis. When the MMRM is used, it includes: (a) treatment group, (b) previous biologic therapy failure status (yes/no), (c) baseline corticosteroid use (yes/no), (d) baseline disease activity (MMS: [4-6] or [7-9]), (e) region (North America/Europe/Other), (f) baseline value in the model, (g) visit, and (h) the interactions of treatment-by-visit and baseline-by-visit as fixed factors. The covariance structure to model the within-patient errors will be unstructured. If the unstructured covariance matrix results in a lack of convergence, the heterogeneous Toeplitz covariance structure, followed by the heterogeneous autoregressive covariance structure will be used. The first structure to yield convergence will be

used for inference. The Kenward-Roger method will be used to estimate the denominator degrees of freedom. Type III sums of squares for the least-squares (LS) means will be used for the statistical comparison; the 95% CI will also be reported. Unless otherwise specified, for MMRM, reported data from only planned visits will be used as the primary analysis.

Treatment comparisons of continuous efficacy and health outcome variables with a single post-baseline timepoint will be made using analysis of covariance (ANCOVA) with: (a) treatment group, (b) previous biologic therapy failure status (yes/no), (c) corticosteroid use (yes/no), (d) baseline disease activity (MMS: [4-6] or [7-9]), (e) region (North America/Europe/Other), and (f) baseline value in the model. Type III sums of squares for least square (LS) means will be used for statistical comparison between treatment groups. The LS mean difference, standard error, p-value, and 95% CI, unless otherwise specified, will also be reported. Missing data imputation method for the ANCOVA model is specified in Section 5.3.

The Kaplan-Meier (KM) product limit method will be used to estimate the survival for several time-to-event analyses. The hazard ratio and log-rank test stratified by covariates mentioned in Section 5.1.4 will be reported. Time for all analyses will be described in units of weeks.

5.2. Adjustments for Covariates

The randomization is stratified by (a) biologic failed patient (yes/no), (b) baseline corticosteroid use (yes/no), (c) baseline disease activity (MMS: [4-6] or [7-9]) and (d) region (North America/ Europe/Other). These factors will be adjusted for as described in Section 5.1.4. Note: when adjusting for covariates, the small number of patients (expected to be <1% of randomized patients) who are inadvertently enrolled with a baseline MMS <4 will be pooled with the MMS [4-6] category.

5.3. Handling of Dropouts or Missing Data

Intercurrent events (ICH E9R1) are events which occur after the treatment initiation and make it impossible to measure a variable or influence how it should be interpreted. Examples of such events include treatment discontinuation due to death or AEs, rescue treatment, and loss to follow-up. The missing data methods described below handle intercurrent events in different ways and thus are relevant to different estimands.

The Schedule of Activities outlined in the protocol specifies the allowable windows for assessments. In general, assessments performed outside these windows will not be excluded from the analysis (unless otherwise specified) but will be reported as a protocol deviation (see Section 5.17). However, Week 12 endoscopies that occurred more than 2 weeks early or more than 4 weeks late (i.e., from study days 71 to 113) will be considered missing for analysis purposes.

5.3.1. Nonresponder Imputation (NRI)

For analysis of binary efficacy and health outcomes variables, missing data will be imputed using an NRI method. Patients will be considered nonresponders for the NRI analysis if they do

not meet the binary efficacy criteria, or have missing clinical efficacy data at a time point of interest.

The above NRI method can be justified based on the composite strategy (ICH E9R1) for handling intercurrent events. In this strategy, patients are defined as responders only if they meet the clinical requirements for response at the predefined time AND they complete the study treatment period without missing relevant data. Failing either criteria by definition makes them nonresponders. It should be understood that if the composite strategery for handling intercurrent events is used, then the term "imputation" is not technically correct as the data are not missing. However, with the above understanding, based on historical usage, we continue to use the descriptor "NRI" in this document.

Additionally, we will utilize an alternative sensitivity estimand for the primary endpoint based on the composite strategy with the following conditions considered as treatment failure: failing to meet protocol-defined primary endpoint criteria, missing data to calculate the primary endpoint criteria, initiation of systemic corticosteroid treatment, increasing systemic corticosteroid treatment dose above baseline, switching corticosteroid treatment or having a UC surgery.

5.3.2. Mixed-effects Model for Repeated Measures (MMRM)

For continuous variables, the primary analysis will be MMRM with the missing at random (MAR) assumption for handling missing data. This analysis takes into account both missingness of data and the correlation of the repeated measurements. No additional imputation methods will be applied to the MMRM analysis.

As noted by Jin and Liu (2020), the MMRM method may be used both under a treatment policy strategy for handling intercurrent events (ICH E9R1) and under a hypothetical strategy for handling intercurrent events. The hypothetical strategy would treat patient data after certain intercurrent events as missing, while the treatment policy strategy would use all available data. For the MMRM analysis in Study AMAN, data from planned visits prior to treatment discontinuation will be used in the analysis regardless of whether the patient took prohibited concomitant rescue medication or otherwise violated the protocol during their time in the treatment period. Therefore, our analysis under the missing at random assumption may be justified as consistent with the treatment policy strategy in intent.

5.3.3. Modified Baseline Observation Carried Forward (mBOCF)

For patients discontinuing IP due to an AE, the baseline observation for the endpoint will be carried forward to the corresponding visit for all missing observations after the patient discontinued study treatment. For patients discontinuing IP for any other reason, the last nonmissing postbaseline observation before discontinuation will be carried forward to the corresponding visit for all missing observations after the patient discontinued. For all patients with sporadically missing observations prior to discontinuation, the last nonmissing observation before the sporadically missing observation will be carried forward to the corresponding visit. Randomized patients without at least 1 postbaseline observation will not be included for evaluation with the exception of patients discontinuing study treatment due to an AE.

The modified baseline observation carried forward (mBOCF) method is based on an estimand that handles the intercurrent event of discontinuing study drug due to an AE by defining the patient as not receiving any benefit from study drug after the event. That is the patient is defined as reverting back to baseline regardless of any continuing efficacy benefits they may still have received after the event. For other intercurrent events (e.g., discontinuation due to reasons other than an AE) or sporadic missingness the "while on treatment" strategy is applied. That is, the endpoint is defined as the last observed value at or before the visit of interest before the patient discontinued study treatment.

5.3.4. Modified Nonresponder Imputation (mNRI)

For a sensitivity analysis of the primary endpoint and selected secondary endpoints for patients impacted by the eCOA transcription error, missing data will be imputed using mNRI. Data from patients who discontinued treatment due to COVID-19-related reasons, lost to follow-up, or a protocol deviation will be imputed. Patients who discontinued from the study treatment period for other reasons such as an AE or lack of efficacy will be categorized as nonresponders by definition. Patients with sporadically missing daily diary data (i.e., when a patient was still in the treatment period but forgot to fill out the daily diary) will be imputed. The endoscopic subscore for patient who received an endoscopy outside of the study days 71 to 113 window will be imputed.

The multiple imputation will be implemented as follows:

- The modified mayo subscores for all scheduled visits will be imputed under the multivariate normal assumption. Indicator variables for treatment and for all stratification factors will be included in the model. A total of 50 imputed datasets will be created.
- Imputed continuous scores for Week 12 Mayo subscores will be rounded using calibrated cutoffs to create ordinal scores based on the approach by Yucel et al 2011. In this approach, the data will be duplicated prior to imputation with the second copy of the data having all Mayo scores from baseline to Week 12 set as missing. The imputed missing scores in the duplicate part of the data are used to select the cutoffs. These calibrated cutoffs are used so that the imputed values are similar to what is in the observed data. The remission status will be calculated using the nonadministrative dropout status and the definition of clinical remission for each of 50 imputed datasets.
- The Mentel-Haenzel estimate of common risk differences along with standard errors (Sato 1989) will be calculated for each imputed dataset and combined using Rubin's Rules (Rubin 1996) to calculate estimates and CIs. P-values will be calculated by using the estimate and standard errors from Rubin's rules to derive a Z-score.

5.3.5. Tipping Point Analysis

Tipping point analysis will be conducted as a sensitivity analysis for the primary endpoint.

Within each analysis, the most extreme case will be considered, in which all missing data for patients randomized to mirikizumab will be imputed using the worst possible outcomes and all missing data for patients randomized to placebo will be imputed with the best possible outcomes:

- Missing responses in the mirikizumab group will be imputed with a range of response probabilities (probabilities of 0, 0.2, 0.4, 0.6, 0.8, and 1.0).
- Missing responses in the placebo group will be imputed with a range of response probabilities (probabilities of 0, 0.2, 0.4, 0.6, 0.8, and 1.0).

Multiple imputed datasets will be generated for each response probability. Treatment differences between mirikizumab and placebo will be analyzed for each imputed dataset using CMH test (Section 5.1.4). Results across the imputed datasets will be aggregated using SAS Proc MIANALYZE in order to compute a p-value or 95% CI for the treatment comparisons for the given response probability. If the probability values do not allow for any variation between the multiple imputed datasets (e.g., all missing responses in the placebo and mirikizumab groups are imputed as responders and nonresponders, respectively), then the p-value from the single imputed dataset will be used to assess the treatment effect.

5.4. Analysis Considerations for the eCOA Error in Turkey and Poland

Data from patients impacted by the Poland and Turkey eCOA transcription error addendum (as described in Section 4.3 and Appendix 3) will be analyzed with the following considerations (Concurrence with the Food and Drug Administration [FDA] on below analyses was obtained based on the FDA Type C written response received on 20 July 2020):

- Efficacy Analysis:
    - Impacted patients will be excluded from the primary efficacy analysis for all endpoints.
    - Additional sensitivity analysis for efficacy will be performed on primary and selected key secondary endpoints by including impacted patients using the ITT population with mNRI multiple imputation. The RB score will be imputed for impacted patients in Poland and the SF score will be imputed for impacted patients in Turkey.
- Safety Analysis:
    - Impacted patients will be included in the primary safety analysis (Section 5.15).
    - To assess the potential impact of the eCOA errors, sensitivity analysis will be performed to summarize key safety results in the modified intent-to-treat (mITT) population (excluding impacted patients) for the Biologics License Application (BLA) submission. The planned safety analyses will include the following summary tables: overview of AEs, treatment-emergent adverse events (TEAEs) by Preferred Term (PT) nested within System Organ Class (SOC), serious adverse events (SAEs) by PT nested within SOC, treatment-emergent abnormal lab results, treatment-emergent abnormal vitals, and AEs leading to treatment/study discontinuation.

5.5. Analysis Considerations for COVID-19-Related Mitigations

A listing of all patients impacted by the COVID-19-related study disruptions will be provided by unique subject identifier and investigator site, with a description of how individual participation was altered. COVID-19-related impact will also be summarized by treatment group for the mITT population. Summaries will include patients having a COVID-19-related AE, discontinuations from study treatment due to a COVID-19-related AE, discontinuations from study treatment due to COVID-19-related issues (e.g., due to COVID related site / travel restrictions), and important deviations/mitigations from the protocol related to COVID-19.

Other analysis considerations for efficacy analysis related to COVID-19 mitigations include the following:

- Extended window for endoscopy:
    - The primary analysis for all endpoints will include all patients, regardless of whether the measurement of endoscopy was out of window. However, Week 12 endoscopies which occurred more than 2 weeks early or more than 4 weeks late (i.e., study days 71 to 113) will be considered missing for analysis purposes (e.g., will be considered nonresponders for NRI analysis).
    - Depending upon the number of patients who are unable to meet the protocol-defined endoscopy window, a sensitivity analysis may be performed to assess clinical remission (primary endpoint) and endoscopic remission by treating as nonresponders patients who received their endoscopy outside of the 85 ±14 study day window.
- Missing endoscopies:
    - For the primary analysis of endoscopy related endpoints, missing Week 12 endoscopies due to COVID-19 will be treated the same as missing endoscopies due to other reasons; for example, in the primary endpoint they will be treated as nonresponders.
    - Sensitivity analysis will be performed for the primary endpoint using the mNRI multiple imputation method of the endoscopy score.
- Other mitigations:
    - The primary analysis of all endpoints will include patients in the analysis populations as described in Table AMAN.5.1, regardless of the COVID-19 mitigations.
    - Sensitivity analysis will include an analysis of primary and key secondary endpoints in the per-protocol (PP) population as described in in Section 5.17. No additional sensitivity analyses are planned.

Safety analyses will be performed as planned in Section 5.15, unless COVID-19 impact is considered substantial enough in the assessment of safety to require a change to an existing analysis or a need for additional analyses as outlined in a recent cross industry manuscript

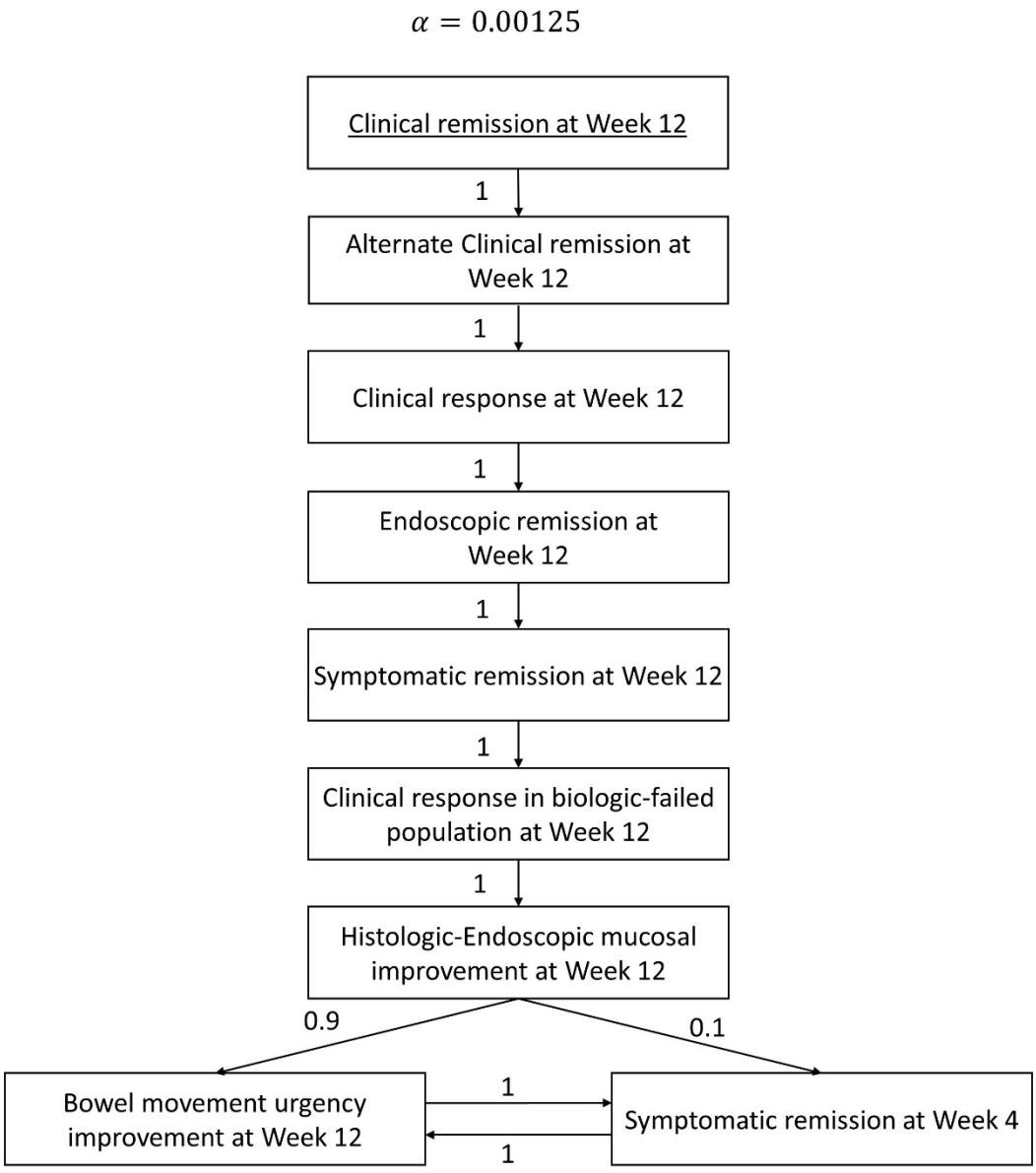
(Nilsson et. al 2020); for example, if the impact due to COVID-19 is different across treatment groups, different or additional analyses will be considered.

5.6. Multicenter Studies

For the analysis of the primary endpoint, treatment-by-region interaction (for regions: North America, Europe, Other) will be added to the logistic regression model as a sensitivity analysis and results from this model will be compared to the model without the interaction effect. If the treatment-by-region interaction is significant at a 2-sided alpha level of 0.05, the nature of this interaction will be inspected as to whether it is quantitative (i.e., the treatment effect is consistent in direction across all regions but not in size of treatment effect) or qualitative (the treatment is beneficial in some but not all regions). If the treatment-by-region interaction effect is found to be quantitative, results from the primary model will be presented. If the treatment-by-region interaction effect is found to be qualitative, further inspection will be used to identify in which regions mirikizumab is found to be more beneficial.

5.7. Multiple Comparisons/Multiplicity

A prespecified graphical multiple testing approach (Bretz et al. 2009, 2011) will be implemented to control the overall Type I error rate at 2-sided alpha of 0.00125, for all primary and major secondary endpoints. More specifically, we will calculate multiple testing adjusted p-values using "Algorithm 2" described by Bretz et al. (2009), and any hypothesis tests with a multiple testing adjusted p-value of less than 0.00125 will be considered statistically significant. This graphical approach is a closed testing procedure; hence, it strongly controls the family-wise error rate across all endpoints (Bretz et al. 2009, 2011; Alosh et al. 2014). Each hypothesis is represented as a node in a graph. Directed arrows between the nodes with associated weights represent how alpha is passed from its initial allocation to other nodes. The testing scheme is fully specified by the graph (including nodes, arrows and weights) along with the initial alpha allocation. Figure AMAN.5.1 describes the graphical scheme, and all of our alpha will be allocated to the primary endpoint initially. The primary and all major secondary endpoints (except Bowel Movement Urgency at Week 12) are binary and will be analyzed using the CMH test with NRI imputation. Bowel Movement Urgency at week 12 will be analyzed using MMRM. Unless otherwise specified, there will be no adjustment for multiple comparisons for any other analyses. The testing scheme will be finalized before the first unblinding of efficacy data.



Note: The underlined endpoint is the primary endpoint.

Figure AMAN.5.1. Graphical approach to control the Type 1 error rate for Study I6T-MC-AMAN.

5.8. Patient Disposition

Screen failures and reason for screen failure will be summarized. The treatment disposition and study disposition will be summarized for the mITT population. Disposition summaries will be by treatment group. Summaries will also include reason for discontinuation from the study tabulated by treatment group.

All patients who are randomized (i.e., the ITT population) and discontinued from study treatment during any period from the study will be listed, and the timing of discontinuing the study will be reported. If known, a reason for their discontinuation will be given.

In addition, a graphical summary (i.e., Kaplan-Meier plot) of time to early permanent discontinuation of study treatment due to AEs may be generated for the Induction Safety, if there are a substantial number of such events. This graphical summary would be by treatment group and include the log-rank test results.

5.9. Patient Characteristics

5.9.1. Demographics and Baseline Characteristics

Patient demographic variables and baseline characteristics will be summarized by treatment and overall for the mITT and ITT populations with the baseline values. The continuous variables will be summarized using descriptive statistics and the categorical variables will be summarized using frequency counts and percentages. No inferential analysis for the comparability of baseline covariates across treatment groups will be performed. By-patient listings of basic demographic characteristics (i.e., age, sex, race, racial subgroup, ethnicity, ethnic subgroup, country, body weight) for the ITT population will be provided.

Table AMAN.5.3 describes the specific variables and how they will be summarized. The final column specifies variables used for the efficacy subgroup analysis described in Section 5.15. Changes to Table AMAN.5.3 including the summary of additional patient characteristics and subgroup analysis will not require an amendment to the SAP.

Table AMAN.5.3. Patient Characteristics (and Variables for Subgroup Analysis)

Variable	Continuous Summary	Categorical Summary	Subgroup Analysis[a]
Demographic Characteristics			
Age[b]	Yes	<65 years, ≥65 years	X
		<40 years, ≥40 years	X
Sex	No	Male, Female	X
Age within Sex	No	Male <40 years, Male ≥40 years, Female <40 years, Female ≥40 years	
Ethnicity	No	Hispanic/Latino, Non-Hispanic/Non-Latino	X
Race	No	American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian or other Pacific Islander, White, or Multiple	X
Geographic Region	No	North America, Europe, Other (Appendix 2)	X
	No	By Country (listed in other documents)	
	No	Asia, North America, Central America/South America, East Europe, West Europe, and ROW (rest of the world) (Appendix 2)	X
Height (cm)	Yes	None	X
Weight (kg)	Yes	<80 kg, ≥80 kg	
		<100 kg, ≥100 kg	X
BMI[c]	Yes	Underweight (<18.5 kg/m2), Normal (≥18.5 and	X

Variable	Continuous Summary	Categorical Summary	Subgroup Analysis[a]
		<25 kg/m2), Overweight (≥25 and <30 kg/m2), Obese (≥30 and <40 kg/m2), Extreme obese (≥40 kg/m2)	
Tobacco use	No	Never, Current, Former	X
Prior UC Therapy			
Prior biologic[d] or tofacitinib exposure	No	Ever, Never	X
Prior biologic[d] or tofacitinib failure[e]	No	Failed, Not failed	X
Prior biologic[d] failure[e] excluding tofacitinib	No	Failed, Not failed	X
Inadequate response or loss of response to a biologic[d] or tofacitinib	No	Ever, Never	X
Inadequate response to a biologic[d] or tofacitinib	No	Ever, Never	
Loss of response to a biologic[d] or tofacitinib	No	Ever, Never	
Intolerance to a biologic[d] or tofacitinib	No	Ever, Never	
Number of prior biologics or tofacitinib used[d]	No	0, 1, 2, >2	
Number of failed[e] biologics[d] or tofacitinib	No	0, 1, 2, >2	
Number of failed[e] biologics[d] excluding tofacitinib	No	0, 1, 2, >2	
Prior biologic[d] or tofacitinib exposure/failure[e]	No	Not exposed, Exposed but not failed, Exposed and failed at least one	X
Prior anti-TNF[f] failure[e]	No	Failed, Not failed	X
Prior anti-TNF[f] failure[e] or vedolizumab failure[e]	No	Failed, Not failed	
Prior anti-TNF[f] failure[e] and prior failure of either vedolizumab or tofacitinib	No	Failed, Not failed	X
Number of failed[e] (unique) prior anti-TNFs[f]	No	0,1, 2, >2	
Prior vedolizumab failure[e]	No	Failed, Not failed	X
Prior tofacitinib failure[e]	No	Failed, Not failed	
Prior systemic corticosteroid[g] failure[e]	No	Failed, Not failed	
Prior systemic immunomodulator[h] failure[e]	No	Failed, Not failed	
Prior systemic corticosteroid[g] or immunomodulator[h] failure[e]	No	Failed, Not failed	
Baseline UC Therapies			
Baseline corticosteroid use[i]	No	Yes, No	X
Baseline prednisone equivalent dose	Yes	None	
Baseline budesonide MMX[j]	No	Yes, No	
Baseline beclomethasone[k]	No	Yes, No	
Baseline immunomodulator	No	Yes, No	X

Variable	Continuous Summary	Categorical Summary	Subgroup Analysis[a]
use[i]			
Baseline corticosteroid and immunomodulator use[i]	No	Corticosteroid only, Immunomodulator, Neither, Both	
Baseline use of oral aminosalicilates[l]	No	Yes, No	
Baseline use of methotrexate[i]	No	Yes, No	
Baseline use of thiopurine[i]	No	Yes, No	
Baseline Disease Characteristics			
Duration of UC[m]	Yes	<1 year, ≥1 to <3 years, ≥3 year to <7 years, ≥7 years	X
Age at Diagnosis of UC[n]	Yes	<6, ≥6 to <10 year, ≥10 to <17 years, ≥17 year to <40 years, ≥40 years	
Baseline Disease Location	No	Proctitis, Left-sided colitis, Pancolitis	X
Baseline Fecal Calprotectin	Yes	≤250 µg/g, >250 µg/g	X
Baseline C-reactive Protein (CRP)	Yes	≤6 mg/L, >6 mg/L	X
Baseline Modified Mayo Score	Yes	Mild (1-3), Moderate (4-6), Severe (7-9)	X
Baseline Total Mayo Score	Yes	Mild (3-5), Moderate (6-9), Severe (10-12)	
Baseline Partial Mayo Score	Yes	None	
Baseline Endoscopic Mayo Subscore	No	Possible values of 4-point scale in Table AMAN.5.4.	
Baseline Stool Frequency Mayo Subscore	No	Possible values of 4-point scale in Table AMAN.5.4.	
Baseline Rectal Bleeding Mayo Subscore	No	Possible values of 4-point scale in Table AMAN.5.4.	
Baseline PGA Mayo Subscore	No	Possible values of 4-point scale in Table AMAN.5.4.	
Baseline UCEIS Score	Yes	None	
Baseline IBDQ Total Score and Domain Scores	Yes	None	
Baseline Urgency NRS	Yes	None	
Baseline Abdominal Pain NRS	Yes	<4, ≥4	
Baseline Patient's Global Rating of Severity (PGRS)	Yes	None	
Baseline Nocturnal Stool	Yes	Yes (≥1), No (0)	
Baseline Fatigue NRS	Yes	Yes (1-10), No (0)	
Baseline Bristol Stool Scale	No	Not Loose Stool (1 – 5), Loose Stool (6 – 7)	
Other Baseline Patient-Reported Outcomes			
Baseline SF-36 PCS, MCS	Yes	None	
Baseline WPAI:UC employment status	No	Yes, No	
Baseline WPAI:UC score	Yes	None	
EQ-5D 5L VAS score	Yes	None	

Abbreviations: ATC = Anatomical Therapeutic Chemical; eCRF = electronic clinical report form; EQ-5D-5L = The European Quality of Life-5 Dimensions-5 Level; IBDQ = Inflammatory Bowel Disease Questionnaire; MCS = mental component score; MMX = Multi Matrix System; NRS = numeric rating scale; PCS = physical component score; PGA = Physician's Global Assessment; SF-36 = 36-Item Short Form Survey; TNF = tumor necrosis factor; UC = ulcerative colitis; UCEIS = Ulcerative Colitis Endoscopic Index of Severity; VAS = visual analog scale; WPAI = Work Productivity and Activity Impairment Questionnaire.

[a] Subgroup analysis will be used for efficacy endpoints only. See Section 5.16 for more details.

[b] Age in years will be calculated as length of the time interval from the imputed date of birth (July 1st in the year of birth collected in the eCRF) to the informed consent date.

[c] Body Mass Index (BMI) will be calculated as: $BMI\ (kg\ /\ m^2) = Weight\ (kg)\ /(Height\ (m))^2$.

[d] Biologic systemic therapies include: adalimumab, adalimumab biosimilar, golimumab, infliximab, infliximab biosimilar, ustekinumab, vedolizumab. For the purpose of counting the number of prior biologics, adalimumab and adalimumab biosimilar will be counted as one biologic. Also, infliximab and infliximab biosimilar will be counted as one biologic.

[e] Failure defined as reasons for prior treatment discontinuation are: loss of response, inadequate response or intolerance to medication.

[f] Anti-TNF alpha biologics include: Infliximab, Infliximab biosimilar, Adalimumab, Adalimumab biosimilar, Golimumab.

[g] Options on the prior med eCRF for Corticosteroids include: prednisone and other corticosteroids. Note that this is not exactly the same as the inclusion criteria defined in the protocol.

[h] Options on the prior med eCRF for Immunomodulator include: 6-mercaptopurine, azathioprine and other thiopurines. Note that this is not exactly the same as the inclusion criteria defined in the protocol.

[i] ATC codes for corticosteroid use (including budesonide MMX and beclomethasone) and immunomodulators (including methotrexate and thiopurines) are listed in the compound level safety standards.

[j] Budesonide MMX will be defined based on a string search of the trade name and reported name with Preferred Term of Budesonide with oral route.

[k] Beclomethasone will be defined based on a string search of the trade name and reported name with Preferred Term of Beclometasone with oral route.

[l] Aminosalicilates will be defined using ATC code A07EC (all members).

[m] Length of the interval from the date of UC diagnosis to the date of informed consent.

[n] Age at diagnosis in years will be calculated as the time interval from the imputed date of birth (July 1st in the year of birth collected in the eCRF) to the date of UC diagnosis.

5.9.2. Preexisting Conditions

Preexisting condition is defined as the condition/event recorded on the Preexisting Conditions and Medical History electronic case report form (eCRF) page with a start date prior to the date of informed consent, and no end date (that is, the event is ongoing) or an end date on or after the date of informed consent. In addition, the AEs occurring prior to first dose are also included. Notice if a preexisting condition worsens in severity on or after the date of informed consent, it will be recorded as an AE on the AE eCRF page with the date of worsening as the start date. The number and percentage of patients with preexisting conditions will be summarized by treatment group using the MedDRA PT nested within SOC. Summaries will be performed for the mITT and ITT population.

5.10. Treatment Compliance

Treatment compliance with IP will be summarized for patients who enter the Induction Period. Treatment compliance for each patient will be calculated as:

$$Treatment\ compliance\ (\%) = 100 \times \frac{Total\ number\ of\ infusions\ administered}{Total\ number\ of\ infusions\ planned\ per\ protocol}$$

Here the planned drug administrations per protocol is based on the number of visits before the patient discontinued study drug. Each patient will be defined as having received a dose on a given date if they received the planned dose (i.e., a partial dose does not count). "Overall compliance" with therapy is defined as missing no doses before discontinuing study treatment. Proportions of patients who meet the definition of *overall compliance* during the Induction Period will be compared between treatment groups using Fisher's exact test.

Patient treatment compliance will be summarized for the mITT population.

5.11. Prior and Concomitant Therapy

Medications will be classified into anatomical therapeutic chemical (ATC) drug classes using the latest version of the World Health Organization (WHO) drug dictionary. Medication start and stop dates will be compared to the date of first dose of treatment in each treatment period to allow medications to be classified as Concomitant for each treatment period.

Prior medications are those medications that start and stop prior to the date of first dose of study treatment. *Concomitant medications* are those medications that start before, on or after the first day of study treatment of the defined treatment period and continue into the treatment period. Concomitant medications are assigned to the treatment period in which they are actually ongoing. For all summary tables of concomitant medications, preferred terms of concomitant medication will be sorted by descending frequency. Also, summaries will be by treatment group and comparisons between treatment groups will use Fisher's exact test for the mITT population.

Summary tables include the following:

- For the mITT and ITT population, summary tables with the number and frequency of patients by treatment group will be presented for:
    - Preferred names of prior therapies ordered by frequency
    - Preferred names of concomitant therapies (use during the induction period) ordered by frequency
    - Prespecified prior therapies in the "Prior Therapy: Indication" eCRF within the categories used in the eCRF. The number and percentage of patients with each reason for discontinuation of previous UC therapy will be summarized by type and therapy.
- A summary of concomitant medications within classes of interest will also be provided for the mITT and ITT populations. This will include: (1) corticosteroid therapy and (2) immunomodulatory therapy. Definition of these two classes of interest will be based on compound level safety standards.

5.12. Efficacy Analyses

Table AMAN.5.4 includes the description and derivation of the efficacy/health outcomes measures and endpoints. Many of these endpoints are collected using a site-facing eCOA device. If duplicate entries are made with different responses, the first nonmissing response will be used.

Table AMAN.5.5 provides the detailed analyses including analysis type, method and imputation, population, time point, and dosing regimen comparisons for efficacy/health outcomes analyses. Note that the details of each analysis will follow the general principles described in Section 5.1.4. For example, the "CMH with NRI" analysis will include descriptive statistics and the common risk differences.

Table AMAN.5.4. Description and Derivation of Efficacy/Health Outcomes Measures and Endpoints

Measure	Description	Variable	Derivation / Comment	Definition of Missing
Mayo Score and components	The Mayo score is a composite instrument to measure Ulcerative Colitis (UC) disease activity. It is comprised of the following 4 subscores: • Stool Frequency (SF): The SF subscore is a patient-reported measure. This item reports the number of stools in a 24-hour period, relative to the normal number of stools for that patient in the same period. The normal reference is collected at baseline/screening. • Rectal Bleeding (RB): The RB subscore is a patient-reported measure. This item reports the most severe amount of blood passed for a given day • Endoscopic Subscore (ES): The ES is a physician-reported measure that reports the worst appearance of the mucosa on flexible sigmoidoscopy or colonoscopy. • Physician's Global Assessment (PGA): The PGA is a physician-reported measure that summarizes the investigator's assessment of the patient's UC disease activity. Each subscore is on a 4-point scale, ranging from 0 to 3.	SF subscore	Calculated by averaging and rounding the 4-point daily SF subscore over 3 days as described in Appendix 1. Possible values are: (0) Normal number of stools for subject; (1) 1 to 2 stools more than normal; (2) 3 to 4 stools more than normal; (3) 5 or more stools than normal.	Missing if fewer than 3 available measurements in the relevant 7 days.
		RB subscore	Calculated by averaging and rounding the 4-point daily RB subscore over 3 days as described in Appendix 1. Possible values are: (0) No blood seen; (1) Streaks of blood with stool less than half of the time; (2) Obvious blood (more than just streaks) or streaks of blood with stool most of the time; (3) Blood alone passed	Missing if fewer than 3 available measurements in the relevant 7 days.
		ES subscore	Possible values are: (0) Normal or inactive disease; (1) Mild disease (erythema, decreased vascular pattern); (2) Moderate disease (marked erythema, absent vascular pattern, friability, erosions); (3) Severe disease (spontaneous bleeding, ulceration)	Single item. Missing if missing. Missing for Week 12 if endoscopy date was outside of the study days 71 to 113 inclusive window.
		PGA subscore	Possible values are: (0) Normal, (1) Mild disease, (2) Moderate disease, (3) Severe disease	Single item. Missing if missing.
		Clinical Remission	○ SF subscore = 0, or SF = 1 with a ≥1-point decrease from baseline, and ○ RB subscore = 0, and ○ ES subscore = 0 or 1 (excluding friability)	Missing if SF, RB or ES subscores are missing.
		Alternate Clinical Remission	○ SF subscore = 0 or 1 ○ RB subscore = 0, and ○ ES subscore = 0 or 1 (excluding friability)	Missing if SF, RB or ES subscores are missing.
		Alternate Clinical Remission 2	○ SF subscore = 0 or 1 ○ RB subscore = 0, and ○ ES subscore = 0 (excluding friability)	Missing if SF, RB or ES subscores are missing.
		Modified Mayo Score (MMS)	Calculated as: SF + RB + ES.	Missing if SF, RB or ES subscores are missing.

Measure	Description	Variable	Derivation / Comment	Definition of Missing
		Total Mayo Score	Calculated as: SF + RB + ES + PGA.	Missing if SF, RB, ES or PGA subscores are missing.
		Total Mayo Clinical Remission	○ Total Mayo Score <=2, and No individual subscore (SF, RB, ES, PGA) >1	Missing if Total Mayo Score is missing.
		Partial Mayo Score	Calculated as: SF + RB + PGA.	Missing if SF, RB or PGA subscores are missing.
		Clinical Response	○ A decrease in the MMS of ≥2 points and ≥30% decrease from baseline, and ○ A decrease of ≥1 point in the RB subscore from baseline or a RB score of 0 or 1	Missing if baseline or Week 12 MMS is missing.
		Total Mayo Clinical Response	○ A decrease in the Total Mayo score of >=3 points and >=30% decrease from baseline, and ○ A decrease of >=1 point in the RB subscore from baseline or a RB score of 0 or 1	Missing if total Mayo Score is missing
		Endoscopic Remission	ES = 0 or 1 (excluding friability).	Missing if ES is missing.
		Symptomatic Remission	○ SF = 0, or SF = 1 with a ≥1-point decrease from baseline and ○ RB = 0	Missing if SF or RB is missing.
		Alternate Symptomatic Remission	○ SF = 0 or 1 ○ RB = 0	Missing if SF or RB is missing.
		Symptomatic Response	≥30% decrease from baseline in the composite clinical endpoint of the sum of SF and RB subscores.	Missing if SF or RB is missing.
		Time to first Symptomatic Remission (Response)	For patients who are observed to meet the remission (or response) criteria during the Induction Period, time will be from the start of the Induction Period to the first measurement date where the patient met the remission (or response) criteria based on weekly averages.	Patients not observed to meet remission (or response) criteria during the Induction Period will be censored after the date of their last measurement

Measure	Description	Variable	Derivation / Comment	Definition of Missing
				during the Induction Period.
		Endoscopic Normalization	ES = 0.	Missing if ES is missing.
		Total Symptomatic Score	Calculated as SF + RB.	Missing if SF or RB is missing.
		Endoscopic Response	A decrease in the ES of ≥1 point compared to baseline.	Missing if ES is missing.
		SF component of clinical remission	SF = 0, or SF = 1 with a ≥1-point decrease from baseline	Missing if SF is missing.
		RB component of clinical remission	RB = 0	Missing if RB is missing.
UCEIS	The Ulcerative Colitis Endoscopic Index of Severity (UCEIS) will be evaluated at the time of endoscopy (Travis 2012). The UCEIS is comprised of the following 3 subscores: • Vascular Pattern • Bleeding • Erosions and Ulcers These subscores are combined to form the UCEIS score which ranges from 0 to 8.	Vascular Pattern	Possible values are: Normal (0), Patchy loss (1) and Obliterated (2).	Single item, missing if missing
		Bleeding	Possible values are: None (0), Mucosal (1), Luminal mild (2) and Luminal severe (3).	Single item, missing if missing.
		Erosion and Ulcers	Possible values are: None (0), Erosion (1), Superficial Ulcer (2) and Deep ulcer (3).	Single item, missing if missing.
		UCEIS score	Calculated as the sum of the subscores: Vascular Pattern + Bleeding + Erosion and Ulcers	Missing if any of the 3 subscores are missing.
		UCEIS endoscopic remission	UCEIS score <=1	Missing if UCEIS score is missing.
Urgency NRS	The Urgency NRS is a single patient-reported item that measures the severity for the urgency (sudden or immediate need) to have a bowel movement in the past 24 hours using an 11-point NRS ranging from 0 (no urgency) to 10 (worst possible urgency).	Urgency NRS Score	Calculated by averaging data from all available daily diary entries of Urgency NRS for a 7-day period as described in Appendix 1.	Missing if fewer than 4 available measurements in the relevant 7 days (see Appendix 1).
		Urgency NRS ≥3-Point Improvement	Decrease from baseline in the NRS Urgency Score is ≥3	Missing if fewer than 4 available measurements in the relevant 7 days (see Appendix 1).

Measure	Description	Variable	Derivation / Comment	Definition of Missing
		Urgency Remission	Urgency NRS = 0 or 1	Missing if urgency NRS score is missing.
Abdominal Pain NRS	The Abdominal Pain NRS is a single patient-reported item that measures the "worst abdominal pain in the past 24 hours" using an 11-point NRS ranging from 0 (no pain) to 10 (worst possible pain).	Abdominal Pain NRS Score	Calculated by averaging data from all available daily diary entries of Abdominal Pain NRS for a 7-day period as described in Appendix 1.	Missing if fewer than 4 available measurements in the relevant 7 days.
PGRS	Patient's Global Rating of Severity (PGRS) is a 1-item patient-rated questionnaire designed to assess the patients' rating of their disease symptom severity over the past 24 hours. Responses are graded on a 6-point scale in which a score of 1 indicates the patient has no symptoms (that is, "none") and a score of 6 indicates that the patient's symptom(s) are "very severe."	PGRS Score	Calculated by averaging data from all available daily diary entries of PGRS for a 7-day period as described in Appendix 1.	Missing if fewer than 4 available measurements in the relevant 7 days.
Nocturnal Stool	The Nocturnal Stool instrument is a single item asking the patient to record the number of stools they had during the night (or day, for shift workers) causing them to wake from sleep.	Nocturnal Stool Score	Calculated by averaging data from all available daily diary entries of Nocturnal Stool for a 7-day period as described in Appendix 1.	Missing if fewer than 4 available measurements in the relevant 7 days.
Fatigue NRS	The Fatigue NRS is a single item that measures the "worst fatigue (weariness, tiredness) in the past 24 hours" using an 11-point NRS ranging from 0 (no fatigue) to 10 (fatigue as bad as can imagine).	Fatigue NRS Score	Calculated by averaging data from all available daily diary entries of Fatigue NRS for a 7-day period as described in Appendix 1.	Missing if fewer than 4 available measurements in the relevant 7 days.
Bristol Stool Scale	The Bristol Stool Scale is a single item that provides a pictorial and verbal description of stool consistency and form ranging from Type 1 (Hard Lumps) to Type 7 (Watery/liquid).	Bristol Stool Scale score	Calculated by using the worst value (i.e., largest number) from all available daily diary entries of Bristol Stool Scale for a 7-day period as described in Appendix 1.	Missing if fewer than 4 available measurements in the relevant 7 days.
		Loose Stool	Bristol Stool Scale score of 6 or 7.	Bristol Stool Scale score is missing

Measure	Description	Variable	Derivation / Comment	Definition of Missing
PGRC	Patient's Global Rating of Change (PGRC): The PGRC scale is a patient-rated instrument designed to assess the patients' rating of change in their symptom(s). Responses are graded on a 7-point Likert scale in which a score of 1 indicates that the subject's symptom(s) is "very much better," a score of 4 indicates that the subject's symptom has experienced "no change," and a score of 7 indicates that the subject's symptom(s) is "very much worse."	PGRC Score	Single Item.	Single item. Missing if missing.
IBDQ	Inflammatory Bowel Disease Questionnaire (IBDQ): A 32-item patient-completed questionnaire that measures 4 aspects of patients' lives: symptoms directly related to the primary bowel disturbance, systemic symptoms, emotional function, and social function (Guyatt et al. 1989; Irvine et al. 1994; Irvine et al. 1996). Responses are graded on a 7-point Likert scale in which 7 denotes "not a problem at all" and 1 denotes "a very severe problem."	Bowel symptoms subscore	Calculated as the sum of questions 1, 5, 9, 13, 17, 20, 22, 24, 26, 29.	If only one question is missing, imputed as the mean of the other items in the subscore. Missing if more than one item in the subscore is missing
		Systemic symptoms subscore	Calculated as the sum of questions 2, 6, 10, 14, 18.	
		Emotional function subscore	Calculated as the sum of questions 3, 7, 11, 15, 19, 21, 23, 25, 27, 30, 31, 32.	
		Social function subscore	Calculated as the sum of questions 4, 8, 12, 16, 28.	
		IBDQ score	Calculated as the sum of all questions. Scores range from 32 to 224; a higher score indicates a better quality of life.	If more than 4 questions are missing or more than 2 questions for any subscore are missing, then IBDQ Score is missing. Otherwise, missing questions imputed as the mean of the other items in each subscore.

Measure	Description	Variable	Derivation / Comment	Definition of Missing
		IBDQ response	≥16-point improvement from baseline in IBDQ score as described by Irvine et al. (1996).	If baseline IBDQ score or visit IBDQ score is missing, then IBDQ response is missing.
		IBDQ remission	IBDQ score ≥170 as described by Irvine (2008).	Missing if the IBDQ score is missing
EQ-5D-5L	The European Quality of Life-5 Dimensions-5 Level (EQ-5D-5L) is a standardized measure of health status used to provide a simple, generic measure of health for clinical and economic appraisal. The EQ-5D-5L consists of 2 components: a descriptive system of the respondent's health and a rating of his/her current health state using a 0- to 100-mm VAS. The descriptive system comprises the following 5 dimensions: Item 1: mobility Item 2: self-care Item 3: usual activities Item 4: pain/discomfort Item 5: anxiety/depression The respondent is asked to indicate his/her health state by ticking (or placing a cross) in the box associated with the most appropriate statement in each of the 5 dimensions.	EQ-5D-5L Items	Five health profile dimensions, each dimension has 5 levels: 1 = no problems 2 = slight problems 3 = moderate problems 4 = severe problems 5 = extreme problems It should be noted that the numerals 1 to 5 have no arithmetic properties and should not be used as a primary score.	Each dimension is a single item, missing if missing.
		EQ-5D-5L UK Population-based index score	Uses the concatenation of the value of each EQ-5D-5L dimension score in the order: Item 1, Item 2, Iitem3; Item 4; Item 5. Derive EQ-5D-5L UK Population-based index score according to the link by using the UK algorithm (Szende et al. 2006) to produce a patient-level index score between -0.59 and 1.0 (continuous variable): https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/crosswalk-index-value-calculator/	If any of the items is missing or equal to 9, the index score is missing
		EQ-5D VAS	Range from 0 = "worst imaginable health state" to 100 = "best imaginable health state". Note: higher value indicates better health state.	Single item, missing if missing

Measure	Description	Variable	Derivation / Comment	Definition of Missing
SF-36	The SF-36 Version 2 is a 36-item, patient-completed measure designed to be a short, multipurpose assessment of health (The SF Community – SF-36 Health Survey Update). The summary scores range from 0 to 100, with higher scores indicating better levels of function and/or better health. Items are answered on Likert scales of varying lengths. The SF-36 comprises 8 domain scores and 2 overarching component scores. SF-36 domain scores are: (1) Physical functioning, (2) Role-physical, (3) Role-emotional, (4) bodily pain, (5) vitality, (6) social functioning, (7) mental health and (8) general health. The component scores are: (1) the Physical Component Summary (PCS) and (2) Mental Component Summary (MCS).	SF-36 Domain scores and SF-36 Component Scores	Per copyright owner, the Quality Metric Health Outcomes™ Scoring Software will be used to derive SF-36 domain and component scores. After data quality-controls, the SF-36 software will re-calibrate the item-level responses for calculation of the domain and component scores. These raw scores will be transformed into the domain scores (t-scores) using the 4-week recall period. This entails exporting the patient data in a CSV or tab-delimited file for import, generation of the SF-36 scores and reports, and export of the calculated scores in a CSV or tab-delimited file for integration into SDTM/ADAM datasets.	Missing data handling offered by SF-36 software will be used. Maximum Data Recovery will be selected for Missing Score Estimator in the software.
		SF-36 PCS MCID Response	PCS component score increase (change from baseline) ≥5 as described by Coteur et al. (2009).	Missing if baseline or observed value is missing.
		SF-36 MCS MCID Response	MCS component score increase (change from baseline) ≥5 as described by Coteur et al. (2009).	Missing if baseline or observed value is missing.
WPAI:UC	The Work Productivity and Activity Impairment- (WPAI:UC) Questionnaire is a patient-reported instrument developed to measure the impact on work productivity and regular activities attributable to a specific health problem (Ulcerative Colitis). It contains 6 items that measure: 1) employment status, 2) hours missed from work due to the specific health problem, 3) hours missed from work for other reasons, 4) hours actually worked, 5) degree health affected productivity while working, and 6) degree health affected	Employment Status	Yes/No	Missing if question is missing
		Absenteeism Score (%)	$\frac{Q2}{(Q2+Q4)} \times 100$	Missing if Q2 or Q4 are missing. Also missing if Employment Status is No.
		Presenteeism Score (%)	$\frac{Q5}{10} \times 100$	Missing if Q5 is missing. Also missing if Employment Status is No.
		Work Productivity Loss Score	$\left[\frac{Q2}{Q2+Q4} + \left(1 - \frac{Q2}{Q2+Q4}\right)\frac{Q5}{10}\right] \times 100$	Missing if Q2, Q4 or Q5 is missing. Also missing if Employment

Measure	Description	Variable	Derivation / Comment	Definition of Missing
	productivity in regular unpaid activities.	(%)		Status is No.
		Activity Impairment Score (%)	$\frac{Q6}{10} \times 100$	Missing if Q6 is missing. May still be present and nonmissing if patient is unemployed.
Histopathology	The histopathologic images will be read centrally in a blinded manner by a qualified pathologist and scoring performed using the Geboes Score (Geboes et al. 2000), Robarts Histopathology Index (RHI) (Mosli et al. 2015) and Nancy index (Marchal-Bressenot et al. 2017).	Geboes Grades	Geboes assigns values to each of 7 histological features: (0) structural [architectural change] (4 levels) (1) chronic inflammatory infiltrate (4 levels) (2a) lamina propria eosinophils (4 levels) (2b) lamina propria neutrophils (4 levels) (3) neutrophils in epithelium (4 levels) (4) crypt destruction (4 levels) (5) erosion or ulceration (5 levels)	Single items. Missing if missing.
		Geboes Score	The highest grade in which there is evidence of disease is assigned. For example, if <50% crypts involved is checked (i.e., Geboes Grade 3 is assigned a 2) and Crypt destruction is noted as 'none' (4.0) and Erosion or ulceration is 'No erosion, ulceration, or granulation tissue' (5.0), the subject will be assigned a score of 3.2.	Missing if any component of the definition is missing
		Robarts Histology Index (RHI)	The RHI score is based on the follow components of the Geboes: chronic inflammatory infiltrate, lamina propria neutrophils, neutrophils in epithelium, and erosion ulceration components (4 levels after combining Geboes 5.1 and 5.2). The RHI is calculated as: RHI = 1 × chronic inflammatory infiltrate level + 2 × lamina propria neutrophils level + 3 × neutrophils in epithelium level + 5 × erosion or ulceration level	Missing if any component of the definition is missing

Measure	Description	Variable	Derivation / Comment	Definition of Missing
		Nancy item scores	The Nancy item scores are: Ulceration, Acute inflammatory filtrate, and Chronic inflammatory filtrate.	Single items. Missing if missing.
		Nancy Index	Takes on possible grades of 0 to 4 based on the items according to the decision tree described by Marchal-Bressenot et al. (2017). A grade of 4 represents severely active disease while a grade of 0 represents no histological significant disease.	Missing if any component of the definition is missing.
		Primary Histologic Remission	Resolution of mucosal neutrophils, defined by Geboes histological subscore of 0 for grades: o 2b (lamina propria neutrophils), and o 3 (neutrophils in epithelium), and o 4 (crypt destruction), and o 5 (erosion or ulceration)	Missing if any component of the definition is missing
		RHI <3	RHI <3	Missing if any component of the definition is missing.
		Nancy Index <1	Nancy Index <1	Missing if any component of the definition is missing.
		Histologic Improvement	Geboes histological subscores of: o 0 (None) or 1 (<5% of crypts involved) for parameter 3 (neutrophils in epithelium), and o 0 (None) for parameter 4 (crypt destruction), and o 0 (None) for parameter 5 (erosion or ulceration)	Missing if any component of the definition is missing.
		Alternative Histologic Improvement	Geboes histological subscores of 0 for parameter: o 2B (neutrophils in lamina propria), and o 3 (neutrophils in epithelium), and o 5 (erosion or ulceration)	Missing if any component of the definition is missing.
Histo-Endo	Combined histology and endoscopic endpoints.	Histologic-Endoscopic Improvement	Histologic Improvement and endoscopic remission.	Missing if any component of the definition is missing.

Measure	Description	Variable	Derivation / Comment	Definition of Missing
		Histologic-Endoscopic Mucosal Remission	Primary histologic remission and endoscopic remission.	Missing if any component of the definition is missing.
CRP	C-reactive protein (CRP) is a biomarker of inflammation.	CRP	Lab value. May be transformed if needed.	Single lab value. Missing if missing.
Fecal calprotectin	Fecal calprotectin is used as a biomarker of intestinal inflammation in clinical practice.	Fecal calprotectin	Lab value. May be transformed if needed.	Single lab value. Missing if missing.
EIMs	Extraintestinal manifestations (EIMs) are collected using the medical history and adverse event eCRFs. Extraintestinal manifestations include, but are not limited to: uveitis, episcleritis, peripheral arthritis, dactylitis, enthesitis, sacroileitis, ankylosing spondylitis, erythema nodosum, pyoderma gangrenosum, primary sclerosing cholangitis, and oral aphthous ulcers.	EIM Subcategory	EIMs will also be categorized as: (1) Musculoskeletal; (2) Mucocutaneous; (3) Hepatic; (4) Occular.	No Imputation.
		Baseline EIMs	EIMs ongoing at first dose of study treatment.	No Imputation.
		Resolution of Baseline EIMs	Complete resolution of baseline EIMs at Week 12. If a patient has multiple baseline EIMs, then at least one EIM must have resolved.	No Imputation.
		Improvement or Resolution in Baseline EIMs	Reduction in the severity of baseline EIMs at Week 12, or complete resolution of baseline EIMs at Week 12. If a patient has multiple EIMs, then at least one EIM must have decreased in severity or resolved.	No Imputation.
		New EIMs	New EIMs at Week 12.	No Imputation.
		Worsening from baseline of an EIM	Increase in the severity of any baseline EIMs at Week 12.	No Imputation

Table AMAN.5.5. Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Time Point(s)[a]
Mayo Score and components	Clinical Remission (Primary Endpoint)	CMH analysis with NRI (Primary Analysis)	mITT (this is the primary analysis of the primary endpoint); PP	Week 12 Visit
		Common Risk Difference with mNRI	mITT; ITT	Week 12 Visit
		Logistic regression analysis with NRI	mITT	Week 12 Visit
	Clinical Remission – Alternative Estimand (Section 5.3.1)	CMH analysis with NRI and Alternative Estimand (Section 5.3.1)	mITT	Week 12 Visit
	Alternate Clinical Remission	CMH analysis with NRI	mITT; PP	Week 12 Visit
		Common Risk Difference analysis with mNRI	ITT	Week 12 Visit
	Alternate Clinical Remission 2	CMH analysis with NRI	mITT	Week 12 Visit
	Clinical Response	CMH analysis with NRI	mITT; PP mITT – In the biologic failed subpopulation; PP – In the biologic failed subpopulation;	Week 12 Visit
		Common Risk Difference analysis with mNRI	ITT; ITT – In the biologic failed subpopulation	
	Total Mayo Clinical Remission	CMH analysis with NRI	mITT	Week 12 Visit
	Total Mayo Clinical Response	CMH analysis with NRI	mITT	Week 12 Visit
	Endoscopic Remission	CMH analysis with NRI	mITT; PP	Week 12 Visit
		Common Risk Difference analysis with mNRI	ITT	
	Symptomatic Remission	CMH analysis with NRI	mITT; PP	Week 1 to 12
		Common Risk Difference Analysis with mNRI	ITT	
	Symptomatic Response	CMH analysis with NRI	mITT	Week 1 to 12
	SF component of clinical remission	CMH analysis with NRI	mITT	Week 1 to 12
	RB component of clinical remission	CMH analysis with NRI	mITT	Week 1 to 12

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Time Point(s)[a]
	Endoscopic Response	CMH analysis with NRI	mITT	Week 12 Visit
	Endoscopic Normalization	CMH analysis with NRI	mITT	Week 12 Visit
	Time to first Symptomatic Remission	KM analysis (censoring described in Table AMAN.5.4)	mITT	During Induction Period
	Time to first Symptomatic Response	KM analysis (censoring described in Table AMAN.5.4)	mITT	During Induction Period
	Change from baseline in SF, RB and Total Symptomatic Score	MMRM; ANCOVA with mBOCF	mITT	Week 1 to 12
	Change from baseline in ES	ANCOVA with mBOCF	mITT	Week 12 Visit
UCEIS	Change from baseline in UCEIS and in individual components of the UCEIS	ANCOVA with mBOCF	mITT	Week 12 Visit
	UCEIS endoscopic remission	CMH analysis with NRI	mITT	Week 12 Visit
Urgency NRS	Change from baseline in Urgency NRS Score	MMRM;	mITT; PP; ITT	Week 1 to 12
		ANCOVA with mBOCF	mITT	Week 1 to 12
	Urgency NRS ≥3-Point Improvement	CMH analysis with NRI	mITT – In patients with an Urgency NRS ≥3 at baseline	Week 1 to 12
	Urgency Remission	CMH analysis with NRI	mITT – In patients with an Urgency NRS ≥3 at baseline	Week 1 to 12
Abdominal Pain NRS	≥30% improvement from baseline	CMH analysis with NRI	mITT – In patients with a NRS abdominal pain score ≥3 at baseline	Week 1 to 12
PGRS	Change from baseline in PGRS Score	MMRM; ANCOVA with mBOCF	mITT	Week 1 to 12
Nocturnal Stool	Change from baseline in Nocturnal Stool Score	MMRM; ANCOVA with mBOCF	mITT - In patients with a nocturnal stool score ≥1 at baseline	Week 1 to 12
Fatigue NRS	Change from baseline in Fatigue NRS Score	MMRM; ANCOVA with mBOCF	mITT	Week 1 to 12

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Time Point(s)[a]
Bristol Stool Scale	Loose Stool	CMH analysis with NRI (i.e., patients with missing data are assumed to have loose stool)	mITT – In patients with loose stool at baseline.	Week 1 to 12
PGRC	Mean PGRC Score	MMRM; ANCOVA as observed (Baseline will not be included as a covariate in model)	mITT	Week 4,8, and 12
IBDQ	Change from baseline in IBDQ Total Score and Subscores	ANCOVA with mBOCF	mITT	Week 12 Visit.
	IBDQ Response	CMH analysis with NRI	mITT	Week 12 Visit.
	IBDQ Remission	CMH analysis with NRI	mITT	Week 12 Visit.
EQ-5D-5L	Change from baseline of EQ-5D VAS score	ANCOVA with mBOCF	mITT	Week 12 Visit
SF-36	Change from baseline for Domain Scores and PCS and MCS Component Scores	ANCOVA with mBOCF	mITT	Week 12 Visit.
	SF-36 PCS MCID Response and SF-36 MCS MCID Response	CMH analysis with NRI	mITT	Week 1 to 12 Visit.
WPAI:UC	Change from baseline in WPAI:UC Scores (Absenteeism, Presenteeism, Work Productivity, Activity Impairment)	ANCOVA with mBOCF	mITT - In patients with baseline employment status of yes	Week 12 Visit.
Histopathology	Primary Histologic Remission	CMH analysis with NRI	mITT	Week 12 Visit
	RHI <3	CMH analysis with NRI	mITT	Week 12 Visit
	Nancy Index <1	CMH analysis with NRI	mITT	Week 12 Visit
	Histologic Improvement	CMH analysis with NRI	mITT	Week 12 Visit
	Alternative Histologic Improvement	CMH analysis with NRI	mITT	Week 12 Visit
Histologic-Endoscopic	Histologic-Endoscopic Improvement	CMH analysis with NRI	mITT	Week 12 Visit
	Histologic-Endoscopic Mucosal Remission	CMH analysis with NRI	mITT	Week 12 Visit

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Time Point(s)[a]
CRP	Change from baseline in CRP	ANCOVA with mBOCF	mITT	Week 12 Visit
Fecal calprotectin	Change from baseline in fecal calprotectin	MMRM; ANCOVA with mBOCF	mITT	Week 4 and 12 Visit
	Percent change from baseline in fecal calprotectin	MMRM; ANCOVA with mBOCF	mITT – In patients with fecal calprotectin >250 µg/g	Week 4 and 12 Visit
EIMs	"Resolution of Baseline EIMs," "Improvement or Resolution of Baseline EIMs" "New EIMs" "Worsening from baseline of an EIM"	Fisher's exact test for overall EIMs and each subcategory.	mITT – In patients with EIMs	Week 12 Visit

Abbreviations: ANCOVA = analysis of covariance; CMH = Cochran-Mantel-Haenzel; ITT = intent-to-treat; mBOCF = modified baseline observation carried forward; MCID = minimal clinically important difference; mITT = modified intent-to-treat; MMRM = mixed-effects model for repeated measures; mNRI = modified nonresponder imputation; NRI = nonresponder imputation; NRS = numeric rating scale; PP = per-protocol; Q = Question; SF-36 = 36-Item Short Form Survey; VAS = visual analog scale.

[a] MMRM analysis will be performed only for protocol-defined visits.

5.12.1. Primary Outcome and Methodology

Analysis of the primary endpoint (clinical remission) is described in Table AMAN.5.5. The primary endpoint analysis will utilize the CMH test (see Section 5.1.4) with NRI (see Section 5.3) for the mITT population (see Section 5.1.1).

5.12.2. Sensitivity Analyses of the Primary Outcome

The sensitivity analysis of the primary endpoint is described in Table AMAN.5.4 and Table AMAN.5.5. Sensitivity analysis for the primary endpoint includes the following:

- CMH test with NRI in the PP population
- Logistic regression analysis with NRI in the mITT population
- Common risk differences calculated with mNRI in the mITT population
- Common risk differences calculated with mNRI in the ITT population. Patient data impacted by the eCOA error will be imputed via multiple imputation in this approach
- CMH test with NRI in the ITT population with alternative estimand as described in Section 5.3.1

Depending upon the number of patients who are unable to meet the protocol-defined endoscopy window, we may provide the CMH test with NRI in the mITT population, removing patients who received their endoscopy outside of the 85 ±14 study day window.

5.12.3. Analyses of the Secondary Efficacy Outcomes

5.12.3.1. Major Secondary Efficacy Outcomes

The analysis of the major secondary endpoints is described in Table AMAN.5.4 and Table AMAN.5.5. The list of major secondary endpoints may be found Table AMAN.3.1. All of these endpoints except for the Urgency NRS endpoint are binary and the primary analysis of these endpoints will use the CMH test (see Section 5.1.4) with NRI (see Section 5.3) in the mITT population. The Urgency NRS endpoint will be analyzed using the MMRM analysis, including the planned study visits, as described in Section 5.1.4. A multiple testing procedure will be utilized to control the FWER at the 0.00125 significance level for the primary analysis of the primary endpoint and all major secondary endpoints (see Section 5.7).

5.12.3.2. Sensitivity Analysis of the Major Secondary Efficacy Outcomes

As described Table AMAN.5.4 and Table AMAN.5.5, the following analyses will be performed as sensitivity analysis for selected major secondary binary outcomes:

- Common risk differences calculated with mNRI in the ITT population. Patient data impacted by the eCOA error will be imputed via multiple imputation in this approach
- CMH test with NRI in the PP population

For the Urgency NRS endpoint:

- MMRM analysis including the planned study visits in the ITT population

- MMRM analysis including the planned study visits in the PP population
- ANCOVA analysis with mBOCF in the mITT population

Additional sensitivity analysis for selected endpoints, such as using the mNRI multiple imputation approach, may also be performed.

5.12.3.3. Other Secondary Efficacy Outcomes

The analysis of the other secondary endpoints is described in Table AMAN.5.4 and Table AMAN.5.5.

5.12.3.4. Exploratory Efficacy Endpoints

The analysis of exploratory efficacy endpoints is described in Table AMAN.5.4 and Table AMAN.5.5.

5.13. Health Outcomes/Quality-of-Life Analyses

5.13.1. Health Care Utilization

Hospitalization is recorded in the hospitalization events eCRF, which is triggered by the AE eCRF. Categories of hospitalization include: Emergency Ward, General Ward, Hospital, Intensive Care Unit, and Other Care Facility. UC-related hospitalizations may be determined from the related adverse event eCRF. Summary statistics will be reported for the number and percentage of patients with UC-related hospitalization overall and within each category by treatment group. Also, for the hospitalization combined category, we will report: the exposure-adjusted incidence rates (number of patients with the event / total person years*100) by treatment, the relative risk, and p-value. Both the relative risk and p-value will be derived from a Poisson regression model with treatment as explanatory variables. The p-value will be based on the likelihood ratio test. This analysis will be conducted for the induction period.

UC-related surgery is recorded in the Surgical Procedures eCRF, which is triggered by the AE eCRF. Types of surgery include proctocolectomy, total colectomy, partial colectomy, and other. As with hospitalizations, summary statistics will be reported for the number and percentage of patients with any surgery, a colectomy surgery (i.e., proctocolectomy, total colectomy, partial colectomy) and within each surgery category by treatment group. Also, for the colectomy surgery category, analysis of the exposure-adjusted incidence rates may be performed similar to the analysis above for hospitalization if a sufficient number of surgeries are performed to justify the analysis.

5.13.2. Additional Health Outcomes/Quality-of-Life Analyses

Details of the additional Health Outcome/Quality-of-Life analyses, including the psychometric analysis for the Urgency NRS, will be provided in supplemental SAP documents.

5.14. Pharmacokinetic/Pharmacodynamic Analyses

The pharmacokinetic/pharmacodynamics (PK/PD) analyses will be conducted by the PK/PD and Pharmacometrics group at Eli Lilly. The PK of mirikizumab will be characterized using

graphical evaluations and mixed-effect (population PK) modeling approaches. Various structural and error models will be evaluated during development of the mixed-effect model. Intrinsic factors (e.g., age, body weight, gender, anti-drug antibodies [ADAs], etc.) and extrinsic factors (e.g., co-medications) will be investigated to assess their influence on model parameters. Model evaluation will include a visual predictive check. Estimates of PK model parameters and covariate effects and the corresponding 90% CIs will be reported.

Analyses of exposure-response relationships will be conducted using both exploratory graphical approaches and model-based approaches. Exploratory graphical analysis approaches may consist of graphs showing the percentage of patients who achieve clinical response, (alternate) clinical remission, and endoscopic remission at different percentiles (e.g., quartiles) of exposure of mirikizumab at Week 12. Measures of exposure may include population PK estimated average concentrations (C_{avg}) between Week 0 and Week 12, or estimated or observed trough concentrations at Week 12. Model based analyses will utilize population exposure-response logistic regression models, where maximum effect (E_{max}) or other model structures may be used to relate exposure to the probability of achieving clinical response, clinical remission, and endoscopic remission. These models may be used to evaluate patient factors that may impact the relationship between exposure and the probability of achieving the endpoint. Longitudinal exposure-response models for SF and RB subscores may be developed, which relate the time course and magnitude of mirikizumab exposure to the time course of these subscores.

Additional analyses may be conducted if they are deemed appropriate. Data from this study may be combined with other study data, if appropriate. Further details on PK and PK/PD analyses will be provided separately in the PK/PD analysis plan.

5.15. Safety Analyses

The planned analyses of safety data will be performed with an intent to maintain consistency with compound level safety standards. These standards are based on internal standards which were informed by Clinical Data Interchange Standards Consortium (CDISC) standards, regulatory guidance (e.g., FDA Clinical Review Template), and cross-industry standardization efforts (e.g., Pharmaceutical Users Software Exchange [PhUSE] white papers from the Standard Analyses and Code Sharing Working Group provided in the PhUSE Computational Science Deliverables Catalog [WWW]).

As detailed in Table AMAN.5.1, the safety analysis population is defined as all randomized patients who received any amount of study treatment (regardless if the patient does not receive the correct treatment, or otherwise does not follow the protocol). Inferential comparisons will be 300 mg mirikizumab Q4W IV vs. placebo Q4W IV. The analysis period (see Table AMAN.5.2) of interest will be the Induction period. Analysis of the safety data collected for the Follow Up period will be performed in the integrated safety analysis.

Unless otherwise noted, Fisher's exact test will primarily be used to compare percentages, and odds ratios will be provided. Odds ratios will be created with mirikizumab treatment as the numerator and placebo as the denominator.

Treatment differences in mean change for continuous measurements will be assessed using an ANCOVA model containing terms for treatment and the continuous covariate of baseline measurement. Type 3 sums of squares will be used. The significance of within-treatment group changes from baseline will be evaluated by testing whether the treatment group LS mean changes from baseline are different from zero using a t-statistic.

5.15.1. Extent of Exposure

Duration of exposure to study treatment will be summarized by treatment group for the safety population. For the treatment period of interest associated with the safety analysis population, exposure will be calculated as the time period length in years (see Section 5.1.2) with start and end dates described in Table AMAN.5.2. Exposure will be calculated for the Induction Period.

Total patient-years (PY) of exposure will be reported by treatment. Descriptive statistics (n, mean, SD, minimum, first quartile, median, third quartile, and maximum) will be provided for patient-weeks of exposure and the frequency of patients falling into different exposure ranges will be summarized. Exposure ranges will generally be reported in weeks using the following as a guide:

- >0, ≥4 weeks, ≥8 weeks, ≥12 weeks.
- >0 to <4 weeks, ≥4 weeks to <8 weeks, ≥8 weeks to < 12 weeks, ≥ 12 weeks.

Additional exposure ranges may be considered if necessary. No p-values will be reported in these tables as they are intended to describe the study populations, rather than test hypotheses about them.

5.15.2. Adverse Events

A TEAE is defined as an event that first occurred or worsened in severity after baseline. The MedDRA Lowest Level Term (LLT) will be used in the treatment-emergent computation. The maximum severity for each LLT during the baseline period will be used as baseline. The treatment period will be included as post-baseline for the analysis. For events with a missing severity during the baseline period, it will be treated as 'mild' in severity for determining treatment-emergence. Events with a missing severity during the postbaseline period will be treated as 'severe' and treatment-emergence will be determined by comparing to baseline severity. For events occurring on the day of first taking study medication, the start times of the study treatment and AE will be used to determine whether the event was pre- versus post-treatment. If start time for the AE is missing, it will be assumed to have started in the postbaseline period.

Summary tables of AEs to include the following:

- Overview of AEs
- Summary of TEAE PTs by decreasing frequency
- Summary of TEAE PTs occurring in >=1% of patients by decreasing frequency
- Summary of TEAE PTs by decreasing frequency within SOC

- Summary of TEAE PTs by maximum severity by decreasing frequency within SOC
- Summary of SAE PTs by decreasing frequency
- Summary of AEs leading to study discontinuation
- Listing of SAEs

Summary tables will include the number and percentage of patients reporting an event. For events that are gender-specific (as defined by MedDRA), the number of participants at risk will include only patients from the given gender. Comparisons will be performed using Fisher's exact test. P-values should be interpreted cautiously due to the fact that multiplicity is not controlled.

The baseline period and post-baseline periods (see Table AMAN.5.2) will be defined as follows:

- The baseline period is the Screening period.
- The postbaseline period will be the Induction period interval.

5.15.2.1. Common Adverse Events

The percentages of patients with TEAEs will be summarized by treatment using MedDRA PT for the common TEAEs (occurred in ≥1% before rounding of mirikizumab-treated patients). Events will be ordered by decreasing frequency in mirikizumab.

5.15.3. Deaths, Other Serious Adverse Events, and Other Notable Adverse Events

The number and percentage of patients who reported a SAE (including those resulting in death) during the treatment period will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the mirikizumab group within SOC. This analysis will be conducted for the Induction period. A listing of SAEs will be provided.

The number and percentage of patients who permanently discontinued from study treatment due to an AE (including AEs that led to death) during the treatment period will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the mirikizumab group within SOC. This summary will be conducted for the induction period.

5.15.4. Clinical Laboratory Evaluations

As described more fully in compound level safety standards and in the laboratory-related PhUSE white papers [PhUSE 2013; PhUSE 2015], the clinical laboratory evaluations will be summarized with the following displays:

- Box plots of observed values (and change from baseline values) by visit.
- Change from baseline to last observation will be summarized within the box plot of changes (rightmost column), and descriptive summary statistics will be included in a table below the box plot along with a p-value using the ANCOVA model described in Section 5.1.4

- Treatment emergent abnormal high lab values (i.e. patients shifting from a normal/low maximum baseline value to a high maximum post-baseline value) or low lab values (i.e. patients shifting from normal/high minimum baseline value to a low minimum post-baseline value)
- Scatter plot of maximum (minimum) post-baseline value vs. maximum (minimum) baseline value
- Shift tables showing the number of patients who shift from each category of maximum (minimum) baseline observation to each category of maximum (minimum) post-baseline observation. Here categories may be low, normal or high with cut-offs defined in the compound level safety standards.

For these displays, the post-baseline period will be the induction period. Post-baseline measurement for continuous analysis (e.g., boxplots) will include *only* scheduled measurements, while post-baseline categorical analysis (e.g., shifts) will include *both* scheduled and unscheduled measurements.

Measurements are defined to be in the baseline periods as follows:

- For analyses of continuous measurements: the last scheduled or unscheduled nonmissing measurement recorded during the Screening Period.
- For analyses of categorical measurements: all scheduled or unscheduled nonmissing measurements recorded during the Screening Period.

For any lab performed on the day of first taking study medication at the start of the post-baseline period, the start time of the study treatment will be used to determine whether the lab was pre- versus post-baseline. If time for the lab is missing, it will be assumed to be in the baseline period (i.e. we assume the protocol defined order of procedures was followed). Following the compound level safety standards, for some labs a safety concern may exist for only high (or only low) values. For these labs, displays with only maximum (or minimum) values will be used and shift tables will be presented accordingly.

5.15.5. Vital Signs and Other Physical Findings

As described more fully in compound level safety standards and in the vital signs-related PhUSE white papers [PhUSE 2013; PhUSE 2015], vital signs will be summarized similarly to the clinical laboratory evaluation (see Section 5.15.4). For vital signs, the low and high limits are based on a combination of a specified value and a change or percentage change. In this case, the PhUSE white paper recommends providing scatter plots and shifts to low/high. Boxplots will also be presented.

5.15.6. Electrocardiograms

Complete electrocardiogram (ECG) data will not be part of the clinical database for the individual studies. Any clinically significant findings from ECGs that result in a diagnosis and that occur after the patient receives the first dose of the investigational treatment will be reported

to Lilly or its designee as an AE via eCRF. Aside from standard AE summary tables no additional analysis of ECG data will be performed.

5.15.7. Immunogenicity

An individual sample is potentially examined multiple times in a hierarchical procedure to produce a sample ADA assay result and potentially a sample neutralizing anti-drug antibodies (NAb) assay result. A patient has treatment-emergent anti-drug antibodies (TE ADA) when ADAs are boosted or induced by exposure to study drug. That is, when at least one postbaseline ADAs sample has a 4-fold increase in titers compared to baseline (if ADA were present at baseline) or has a titer 2-fold greater than the minimum required dilution of 1:10 (if no ADAs were present at baseline). Compound level safety standards will be followed in the analyses of immunogenicity. Listings of immunogenicity assessments will be provided along with the summary of specified TEAEs by TE ADA status. The summary of TE ADA and NAb status will be produced, where the postbaseline period for reporting is the induction period. Analyses of the relationship between immunogenicity and PK will be conducted as part of the PK/PD analyses as described in Section 5.14. Additional assessments of the relationship between immunogenicity and efficacy will be performed as deemed appropriate.

5.15.8. Special Safety Topics including Adverse Events of Special Interest

This section includes areas of interest whether due to observed safety findings, potential findings based on drug class, or safety topics anticipated to be requested by a regulatory agency for any reason. In general, potential adverse events of special interest (AESI) relevant to these special safety topics will be identified by one or more standardized MedDRA query(ies) (SMQs), by a Lilly defined MedDRA PT listing based upon the review of the most current MedDRA Version, or by treatment emergent relevant laboratory changes, as described below. Additional special safety topics may be added as warranted.

Unless otherwise specified, the AESIs will be summarized for the safety population during the induction period using the baseline and postbaseline definitions described in Sections 5.15.2.

Full details of the search terms and rules for deriving AESIs in each of the sections below are described in the compound level safety standards along with information about the types of summaries and listings to be provided.

5.15.8.1. Hepatic Safety

Hepatic labs include alanine aminotransferase (ALT) and aspartate transaminase (AST), total bilirubin (TBL) and serum alkaline phosphatase (ALP). When criteria are met for hepatic evaluations, investigators will complete a follow-up hepatic safety eCRF.

Analyses will include:

- ALT and AST: The percentages of patients with a measurement greater than or equal to 3 times (3X), 5 times (5X), and 10 times (10X) the performing lab upper limit of normal (ULN) during the treatment period for all patients with a post-baseline value and for subsets based on various levels of baseline value.
- TBL and ALP: The percentages of patients with a measurement greater than or equal to 2 times (2X) the performing lab ULN during the treatment period will be summarized for all patients with a post-baseline value and for subsets based on various levels of baseline value.
- Plot of maximum post-baseline ALT vs. maximum post-baseline total bilirubin (entire safety population).
- A listing of the information collected on the hepatic-safety CRF.

5.15.8.2. Infections, Including Opportunistic Infections and Serious Infections

Infections will be defined using the PTs from the MedDRA Infections and Infestations SOC. Treatment-emergent infections will be analyzed for: all infections (by maximum severity), serious infections and opportunistic infections (OI). The MedDRA terms used to identify infections considered to be OI in patients with immune mediated inflammatory conditions treated with immunomodulatory drugs are based on Winthrop et al. (2015) and are listed in the compound level safety standards. The list contains narrow (more specific) and broad (less specific) PTs with respect to these prospectively defined opportunistic infections.

Analyses will include:

- Treatment-emergent (TE) Infections by PT;
- Serious Infections by PT;
- Opportunistic Infections: TE OI by narrow terms and broad terms separately.

5.15.8.3. Hypersensitivity

Hypersensitivity reactions is used as an overarching term to describe events that are systemic or localized reactions that likely have an allergic/hypersensitivity etiology. The evaluation of study drug-related systemic hypersensitivity reactions will be through the unsolicited reporting of TEAEs and through the use of the Hypersensitivity, Anaphylactic, and Infusion-Related Reaction Follow-up Forms completed by the investigator.

Potential hypersensitivity reaction AEs will be determined using the following SMQs: anaphylactic reaction, hypersensitivity, and angioedema. Potential hypersensitivity AEs will be categorized as Immediate (i.e., on day of study drug administration) and non-immediate (i.e., occurring after the day of study drug administration) based on the timing of the reaction.

Analyses will include:

- For Immediate Hypersensitivity: (1) combined narrow/algorithmic search (i.e., any narrow term from any one of the SMQs, or anaphylaxis algorithm), (2) narrow/algorithmic search (i.e., any narrow/algorithmic term) by SMQ, (3) broad search (i.e., any narrow or broad term) by SMQ, and (4) TEAEs (occurring on the day of study drug administration) by PT not in any of the 3 SMQs.
- For Non-Immediate Hypersensitivity: (1) combined narrow search (i.e., any narrow term from any one of the SMQs), (2) narrow search (i.e., any narrow term) by SMQ, and (3) broad search (i.e., any narrow or broad term) by SMQ.

5.15.8.4. Infusion Site Reactions (ISR)

Infusion site reactions are AEs localized to the immediate site of the administration of a drug. The evaluation of study drug related ISRs will be through the unsolicited reporting of ISR TEAEs and through the use of an Infusion Site Reaction Follow-up Form completed by the investigator for each ISR reported.

Infusion site reactions will be defined using the following MedDRA High Level Term (HLT): Infusion site reaction, excluding certain PTs (e.g., those PTs related to joint).

Analyses will include:

- TE ISRs, HLT, and PT.
- The additional data collected on the ISR follow-up form will be summarized in two distinct ways: at the patient level and at the event level. A by-patient listing of these data will be provided.

5.15.8.5. Cerebro-Cardiovascular Events

The cerebro-cardiovascular events reported in the study will be adjudicated by an independent, external adjudication committee (AC). All confirmed events after adjudication will be used for the analysis of cerebro-cardiovascular events. Categories of events include: Cardiovascular, Cerebrovascular and Peripheral Vascular Events. As detailed in the compound level safety standards, the categories are further categorized into subcategories.

Analyses will include:

- TE cerebro cardiovascular confirmed events by category, subcategory and PT.
- By-patient listing for all patients having a TEAE of cerebro-cardiovascular (confirmed event, no event, or insufficient documentation for event determination) at any time.

5.15.8.6. Malignancies

Malignancies will be defined using PTs from the Malignant tumors SMQ. Malignant tumor events will be summarized separately for the categories: Non-Melanoma skin cancer (NMSC) and Malignancies excluding NMSC.

Analyses will include:

- TE malignancy by category and PT.
- By-patient listing for all patients having a TEAE of malignancy at any time.

5.15.8.7. Suicidal Ideation/Behavior and Depression

During the study, suicidal ideation and behavior, and depression will be assessed prospectively by the investigator via signs and symptoms and through the use of the Columbia-Suicide Severity Rating Scale (C-SSRS) [Screening] and the Quick Inventory of Depressive Symptomatology Self Report (QIDS-SR16) [Weeks 0, 12, ETV].

Analyses will include:

- C-SSRS: Only a listing of the C-SSRS data will be provided.
- QIDS-SR16: Shift tables will be provided showing the number and percentage of patients within each baseline category (maximum value) versus each post-baseline category (maximum value) by treatment. Additionally, outcomes such as any increase in depression will be compared between treatments (further described in the compound level safety standards).

5.16. Subgroup Analyses

5.16.1. Efficacy Subgroup Analysis

Subgroup analyses will be conducted for all primary and key secondary endpoints (excluding the key secondary endpoint which is in the subgroup of biologic-failed patients) in the mITT Population. The subgroups to be analyzed are listed Table AMAN.5.3 along with the demographic characteristics. Additional subgroup analysis which are not based on baseline/demographic characteristics in Table AMAN.5.3 include treatment-emergent anti-mirikizumab antibody status. Some additional subgroup analyses may be performed to meet regulatory requirements in specific countries. Changes to subgroup analysis will not require an amendment to the SAP.

For binary endpoints, a logistic regression model with treatment, subgroup, and the interaction of subgroup-by-treatment, and the covariates mentioned in Section 5.2. The subgroup-by-treatment interaction will be tested using the Firth correction (Firth 1993) at the significance level of 0.05. Within each subgroup category the proportion of responders by treatment, treatment differences and 95% CIs will be displayed. Also, p-values using Fisher's exact test for treatment comparison will be provided.

For the Week 12 Urgency NRS endpoint, MMRM analysis will be performed for each subgroup for select subgroups. Within each subgroup, LS means by treatment, LS mean differences, and 95% CIs will be displayed. To test for interaction, an MMRM model with a subgroup-by-treatment interaction term for each visit will be fit.

Forest plots may be generated to display the treatment difference and 95% CIs for selected efficacy subgroup analyses. If the number of patients in any subgroup category is <10% of the total population, only summaries of the efficacy data will be provided (that is, no inferential testing for that subgroup).

5.16.2. Safety Subgroup Analysis

A summary of TEAEs will be produced for the biologic-failed subgroup. Additional safety subgroup analyses may be performed if there is a potentially relevant finding during the periodic study safety reviews. Also, subgroup analysis for safety related endpoints will be performed within the context of the integrated safety analysis.

5.16.3. Analysis for Japan Submission

A subset of the planned analyses (e.g., patient disposition, demographic and baseline characteristics, efficacy, health outcomes, and safety analyses) will be reproduced based on patients from Japan sites, in support of the regulatory submission in Japan. The list of tables, listings, and figures for the patients from Japan sites (Japanese population) will be in a separate document.

5.17. Protocol Violations

Protocol deviations will be identified throughout the study. Important protocol deviations (IPDs) are defined as those deviations from the protocol that would potentially compromise patients' safety, data integrity, or study outcome.

The important protocol deviations excluded from per-protocol analysis (IPDPPs), which are a subset of the important protocol deviations, are the IPDs that might have significant impact on the primary efficacy results. The impact of IPDPPs on the efficacy results will be assessed by assessing the robustness of the study results and conclusions to the choice of analysis population, both by including and excluding patients with IPDPPs. As specified in Table AMAN.5.1, the Induction PPS population is defined as all randomized patients who do not have any IPDPPs. Mitigations approved under the COVID-19 addendum which would otherwise have been classified IPDPPs (had they not been approved under the addendum) will still result in patients being excluded from PP analysis.

A separate document known as the "The AMAN Trial Issues Management Plan (TIMP)" describes the categories and subcategories of important protocol deviations, whether or not these deviations are IPDPPs, and how the IPDs would be identified. The TIMP will be finalized before the Week 12 database lock (DBL).

The number and percentage of patients having IPDs will be summarized within category and subcategory of deviations by dosing regimen for mITT population.

A by-patient listing of IPDs will be provided.

5.18. Interim Analysis and Data Monitoring

Data Monitoring Committee: One Data Monitoring Committee (DMC) consisting of members external to Lilly will be established for periodic monitoring of clinical trial data across all Phase 3 trials for the UC adult program. This committee will consist of a minimum of 3 members, including a physician with expertise in gastroenterology and a statistician.

No member of the DMC may have contact with study sites. A statistical analysis Center (SAC) will prepare and provide unblinded data to the DMC. The SAC members may be Lilly employees or from third-party organizations designated by Lilly. However, they will be external to the study team and will have no contact with sites and no privileges to influence changes to the ongoing studies. The timing and frequency of the periodic clinical trial data review by the DMC will be detailed in the DMC charter for the UC adult program.

The DMC is authorized to evaluate unblinded interim efficacy and safety analyses. The DMC will make recommendation to the Lilly Research Laboratories Senior Management Designee, who may order the immediate implementation of the DMC recommendation, or may convene an internal review committee (IRC), which is independent from the study team, to review the recommendation according to standard Lilly policy. Study sites will receive information about interim results ONLY if it is required for the safety of their patients.

Week 12 DBL: An unblinded analysis will be performed after all patients have completed the Week 12 Visit or discontinued study treatment. This DBL will include all data collected by the cutoff date, including follow-up data from patients that have begun the posttreatment follow-up period. This is the final analysis for the efficacy endpoints up to Week 12. However, the study may be ongoing for the posttreatment follow-up period for patients remaining in the study at the time of this DBL.

Final DBL: A final DBL will occur if needed after the posttreatment follow-up period from all active patients is completed.

Pharmacokinetics Analysis: In addition, a limited number of preidentified, internal Lilly personnel that are not in contact with clinical sites may gain access to unblinded data, including PK, as specified in the unblinding plan. The unblinded data will be restricted and will NOT be shared with anyone outside this preidentified group until after the Week 12 DBL. Unblinding details will be provided in the unblinding plan.

5.19. Annual Report Analyses

Based on regulatory requirements for the Development Safety Update Report (DSUR), reports will be produced (if not already available from the study CSR) for the reporting period covered by the DSUR

5.20. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

- Summary of AEs, provided as a dataset which will be converted to an XML file. Both SAEs and 'Other' AEs are summarized: by treatment group, by MedDRA PT.
- An AE is considered 'Serious' whether or not it is a TEAE.

- An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term
    - the number of events experienced.
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures for example, the CSR, manuscripts, and so forth.

6. Unblinding Plan

Details will be provided in a separate unblinding plan document.

7. References

Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med.* 2014;33(4):693-713.

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

Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes or parametric tests. *Biom J.* 2011;53(6):894-913.

Coteur G, Feagan B, Keininger DL, Kosinski M. Evaluation of the meaningfulness of health-related quality of life improvements as assessed by the SF-36 and the EQ-5D VAS in patients with active Crohn's disease. *Aliment Pharmacol Ther*. 2009;29:1032-1041.

Firth D. Bias reduction of maximum likelihood estimates. *Biometrika.* 1993;80 (1):27-38.

Geboes K, Riddell R, Ost A, Jensfelt B, Persson T, Löfberg R. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. *Gut.* 2000;47(3):404-409.

International Conference on Harmonization. ICH E9 (R1): Addendum to statistical principles for clinical trials on choosing appropriate estimands and defining sensitivity analyses in clinical trials. 2017. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM582738.pdf.

Irvine EJ. Quality of life of patients with ulcerative colitis: past, present, and future. *Inflamm Bowel Dis*. 2008;14: 554-565.

Irvine EJ. Development and subsequent refinement of the inflammatory bowel disease questionnaire: a quality-of-life instrument for adult patients with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 1999;28: S23-S27.

Jin M, Liu G. Estimand framework: delineating what to be estimated with clinical questions of interest in clinical trials. *Contemp Clin Trials*. 2020;96:106093.

Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, Bastien C, Cahn V, Cadiot G, Diebold MD, Danese S, Reinisch W, Schreiber S, Travis S, Peyrin-Biroulet L. Development and validation of the Nancy histological index for UC. *Gut.* 2017;66(1):43-49.

Metha CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stats Med.* 2011;30(28):3267-3284.

Mosli M, Feagan BG, Zou G, Sandborn WJ, D'Haens G, Khanna R, Shackelton LM, Walker CW, Nelson S, Vandervoort MK, Frisbie V, Samaan MA, Jairath V, Driman DK, Geboes K, Valasek MA, Pai RK, Lauwers GY, Riddell R, Stitt LW, Levesque BG. Development and validation of a histological index for UC. *Gut*. 2015;0:1-9.

Nilsson M, Crowe B, Anglin G, Ball G, Munsaka M, Shahin S, Wang W. Clinical trial drug safety assessment for studies and submissions impacted by COVID-19 [published online August 05, 2020]. *Stat Biopharm Res*.

Pharmaceutical Users Software Exchange [PhUSE] resources page. PhUSE web site. Available at https://www.phuse.eu/css-deliverables. Accessed September 18, 2017.

Pharmaceutical Users Software Exchange [PhUSE resources page. Analyses & Displays Associated with Demographics, Disposition, & Medications in Phase 2-4 Clinical Trials & Integrated Summary Documents. 2018. Available at: https://www.phuse.eu/documents//working-groups/deliverables/analyses-displays-associated-with-demographics-disposition-medications-in-phase-2-4-clinical-trials-version-20-02-mar-18-11808.pdf. Accessed October 23, 2018.

Pharmaceutical Users Software Exchange [PhUSE] resources page. Analyses and Displays Associated with Measures of Central Tendency – Focus on Vital Sign, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Submission Documents. 2013. Available at: https://www.phusewiki.org/docs/CSS%20White%20Papers%202016/CSS_WhitePaper_CentralTendency_v1.0.pdf. Accessed October 23, 2018.

Pharmaceutical Users Software Exchange [PhUSE] resources page. Analyses and Displays Associated with Outliers or Shifts from Normal to Abnormal: Focus on Vital Signs, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Summary Documents. 2015. Available at https://www.phusewiki.org/docs/CSS%20White%20Papers%202016/CS_WhitePaper_OutliersShifts_v1.0.pdf. Accessed October 23, 2018.

Pharmaceutical Users Software Exchange [PhUSE] resources page. Analysis and Displays Associated with Adverse Events: Focus on Adverse Events in Phase 2-4 Clinical Trials and Integrated Summary Document. 2017. Available at https://www.phuse.eu/documents//working-groups/cs-whitepaper-adverseevents-v10-4442.pdf. Accessed October 23, 2018.

Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc*. 1996;91(434):473-489.

Szende A, Oppe M, Devlin N, eds. EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Dordrecht, Netherlands: Springer; 2007.

Sato, T. On the variance estimator of the Mantel-Haenszel risk difference [editorial]. *Biometrics*. 1989;45(4):1323-1324.

Travis SP, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel JF, Feagan BG, Hanauer SB, Lémann M, Lichtenstein GR, Marteau PR, Reinisch W, Sands BE, Yacyshyn BR, Bernhardt CA, Mary JY, Sandborn WJ. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). *Gut*. 2012;61(4):535-542.

Yucel RM, He Y, Zaslavsky AM. Gaussian-based routines to impute categorical variables in health surveys. Stat Med. 2011;30(29):3447-3460.

Appendix 1. Daily Diary Calculations

Weekly summary measures of daily diary data will be created for each patient. The 7-day period associated with each week will be defined using a visit centric approach. The table below displays the interval for each week,

Week (Visit)	Start Day[a]	End Day[a]
Screening	Max(Informed Consent Date, Week 0 Visit Date - 14)	Week 0 Visit Date - 8
Baseline	Week 0 Visit Date - 7	Week 0 Visit Date - 1
Week 1	Max(Week 0 Visit Date, Week 2 Visit Date - 14)	Week 2 Visit Date - 8
Week 2 (V2)	Max(Week 0 Visit Date, Week 2 Visit Date - 7)	Week 2 Visit Date - 1
Week 3	Max(Week 2 Visit Date, Week 4 Visit Date - 14)	Week 4 Visit Date - 8
Week 4 (V3)	Max(Week 2 Visit Date, Week 4 Visit Date - 7)	Week 4 Visit Date - 1
Week 5	Max(Week 4 Visit Date, Week 8 Visit Date - 28)	Week 8 Visit Date - 22
Week 6	Max(Week 4 Visit Date, Week 8 Visit Date - 21)	Week 8 Visit Date -15
Week 7	Max(Week 4 Visit Date, Week 8 Visit Date - 14)	Week 8 Visit Date - 8
Week 8 (V4)	Max(Week 4 Visit Date, Week 8 Visit Date - 7)	Week 8 Visit Date - 1
Week 9	Max(Week 8 Visit Date, Week 12 Visit Date - 28)	Week 12 Visit Date - 22
Week 10	Max(Week 8 Visit Date, Week 12 Visit Date - 21)	Week 12 Visit Date - 15
Week 11	Max(Week 8 Visit Date, Week 12 Visit Date - 14)	Week 12 Visit Date - 8
Week 12 (V5)	Max(Week 8 Visit Date, Week 12 Visit Date - 7)	Week 12 Visit Date - 1

Abbreviation: V = Visit.

[a] If End Day < Start Day, do not assign specified visit week. Visit date will be calculated by selecting the first date from the following list (i.e., first in list order): (1) date of earliest bowel preparation if bowel preparation date is available, (2) date of endoscopy if endoscopy was performed, (3) date of treatment if treatment was given, or (4) office visit date if available, or (5) date of visit center of the protocol-defined window for that visit (i.e. study day 85). For patients who received their endoscopy outside of the window from study days 71 to 113, the visit date will be calculated as study day 85. The screening endoscopy is assumed to be associated with the Week 0 visit.

For the Mayo SF and RB subscores, the most recent 3 nonmissing days of the 7-day period in the table above will be averaged and rounded to the nearest integer to calculate the weekly score for each patient. Patients with less than 3 measurements in the 7 day period will be considered missing.

For the Bristol Stool Scale the worst (i.e. maximum) of the available measures during the 7 period in the table above will be used to calculate a weekly score for each patient. If fewer than

4 days are available (i.e., not missing), the patient will be considered to be missing data for that week.

For all other daily diary measures, all available days of the 7 days will be averaged and rounded to the nearest integer to calculate the weekly score for each patient. If fewer than 4 days are available (i.e., not missing), the patient will be considered to be missing data for that week.

If multiple diary assessments on a single day are present, use the earliest nonmissing assessment. Data from the following days will be considered missing: (i) days when patients receive bowel preparation, (ii) the day of an endoscopy, and (iii) the day after an endoscopy.

If the baseline assessment is missing per the above rules, the first available postbaseline assessment starting with Week 1 will be used to impute the baseline so that the patient can be included in the analysis.

Appendix 2. Countries and Regions

Country	Region 1	Region 2
AUSTRIA	Europe	Western Europe
BELGIUM	Europe	Western Europe
CROATIA	Europe	Eastern Europe
CZECH REPUBLIC	Europe	Eastern Europe
DENMARK	Europe	Western Europe
FRANCE	Europe	Western Europe
GERMANY	Europe	Western Europe
HUNGARY	Europe	Eastern Europe
IRELAND	Europe	Western Europe
ITALY	Europe	Western Europe
LATVIA	Europe	Eastern Europe
LITHUANIA	Europe	Eastern Europe
NETHERLANDS	Europe	Western Europe
POLAND	Europe	Eastern Europe
ROMANIA	Europe	Eastern Europe
SLOVAKIA	Europe	Eastern Europe
SPAIN	Europe	Western Europe
SWITZERLAND	Europe	Western Europe
UNITED KINGDOM	Europe	Western Europe
CANADA	North America	North America
UNITED STATES	North America	North America

ARGENTINA	Other	Central America/S America
AUSTRALIA	Other	ROW
BRAZIL	Other	Central America/S America
CHINA	Other	Asia
INDIA	Other	Asia
ISRAEL	Other	ROW
JAPAN	Other	Asia
KOREA, SOUTH	Other	Asia
MALAYSIA	Other	Asia
MEXICO	Other	Central America/S America
RUSSIAN FEDERATION	Other	ROW
SAUDI ARABIA	Other	ROW
SERBIA	Other	ROW
TAIWAN	Other	Asia
TURKEY	Other	ROW
UKRAINE	Other	ROW

Abbreviation: ROW = Rest of World.

Appendix 3. Summary of eCOA Transcription Errors

Rectal Bleeding (RB) electronic clinical outcomes assessment (eCOA) Error in Poland

Rectal bleeding is a patient-reported outcome (PRO) measure and a component of the Mayo score. The score normally requires patients to indicate the degree of blood seen with bowel movements and includes the following options:

(0) No blood seen
(1) Streaks of blood with stool less than half of the time
(2) Obvious blood (more than just streaks) or streaks of blood with stool most of the time
(3) Blood alone passed

The transcription error on the eCOA device resulted in option (3) having a label of "No blood seen." Patients in Poland saw the options:

(0) No blood seen
(1) Streaks of blood with stool less than half of the time
(2) Obvious blood (more than just streaks) or streaks of blood with stool most of the time
(3) No blood seen

Thus, patients with no RB may have selected the option normally assigned to the worst grade (3) or may have selected the appropriate grade (0). Patients who had severe bleeding would not have been able to select "Blood alone passed" and could have selected a less-severe RB option.

This error in the electronic daily diary impacted all **112** patients from Poland enrolled in the I6T-MC-AMAN (AMAN) (induction) study at the time the error was discovered. Of the 112 impacted patients, **98** had already completed Study AMAN and entered Study I6T-MC-AMBG (AMBG) (maintenance).

Stool Frequency eCOA Error in Turkey

Stool frequency is a PRO measure and a component of the Mayo score. In completing the daily diary, the patient should have seen the question:

> **"How many stools did you have in the past 24 hours?"**

Instead, the patient saw a duplicate of the Nocturnal Stool question, as follows:

> **"How many stools did you have during the night causing you to waken from sleep?"**

This error in the electronic daily diary impacted all **6** patients from Turkey enrolled in AMAN (induction) study at the time the error was discovered. Of the 6 impacted patients, **3** had already completed Study AMAN and entered Study AMBG (maintenance).

Leo Document ID = 9fb256cc-09d9-442a-bc5c-ff9c561b1e80

Approver: PPD

Approval Date & Time: 14-Jan-2021 22:54:06 GMT

Signature meaning: Approved