Statistical Analysis Plan Version 4 I6T-MC-AMBG

A Phase 3, Multicenter, Randomized, Double-Blind, Parallel-Arm, Placebo-Controlled Maintenance Study of Mirikizumab in Patients with Moderately to Severely Active Ulcerative Colitis.

LUCENT 2

NCT03524092

Approval Date: 01-Oct-2021

Statistical Analysis Plan: I6T-MC-AMBG: A Phase 3, Multicenter, Randomized, Double-Blind, Parallel-Arm, Placebo-Controlled Maintenance Study of Mirikizumab in Patients with Moderately to Severely Active Ulcerative Colitis LUCENT 2

Mirikizumab (LY3074828) Ulcerative colitis

Study I6T-MC-AMBG (AMBG) is a Phase 3 clinical study that is designed to evaluate the safety and efficacy of mirikizumab in maintaining remission at Week 40 in patients who completed the 12-week induction study I6T-MC-AMAN (AMAN).

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I6T-MC-AMBG

Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 29-Oct-2018

Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 13-Dec-2019

Statistical Analysis Plan Version 3 electronically signed and approved by Lilly: 26-Feb-2021

Statistical Analysis Plan Version 4 electronically signed and approved by Lilly on date provided below.

Approval Date: 01-Oct-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 (SAP) Version 2 was approved on 12 December 2019 and was based on protocol amendment (a) approved 12 September 2019. The following updates were made in Version 2 after the first external safety Data Monitoring Committee (DMC) analysis but before the first unblinded analysis by the sponsor:

1. Made the following changes to the primary and major secondary endpoints:
    a. Moved "Clinical remission among induction responders" from major secondary endpoint to primary endpoint.
    b. Moved "Clinical remission among induction remitters" from primary endpoint to major secondary endpoint.
    c. Removed major secondary endpoints "clinical remission among biologic failed induction responders" and "endoscopic remission among biologic failed induction responders" from major secondary endpoints.
    d. Included "Stable maintenance of symptomatic remission" as a major secondary endpoint.
    e. Replaced "histological remission" with "mucosal healing" as a major secondary endpoint.
    f. Included "The change from baseline in the Urgency Numeric Rating Scale score" as a major secondary endpoint.
    g. Included "Alternate clinical remission" and "Alternate clinical remission among alternate clinical remitters" 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 (AMBG).
3. Clarified the study population and study cohort definition in Table AMBG.5.1 and Table AMBG.5.2. Clarified the description of study cohort throughout this SAP.
4. Clarified the study period definition in Table AMBG.5.3.
5. 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.
6. The graphical testing scheme (Section 5.7) was updated to reflect change for primary and major secondary endpoints.
7. Several modifications to the categories and calculations for baseline characteristics were made to Table AMBG.5.4.
8. Analysis of pre-existing conditions was added to Section 5.9.1.
9. Clarified the definition for treatment compliance.
10. The specific summary table for concomitant therapy was clarified (Section 5.11).
11. Made the following changes to Table AMBG.5.5 and Table AMBG.5.6:
    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. "Stable maintenance of symptomatic remission" endpoint was added.
e. "Alternative corticosteroid-free remission" endpoint was added.
f. "UCEIS endoscopic remission" endpoint was added.
g. EQ-5D-5L items and population-based index will no longer be in the primary SAP document.
h. "Urgency NRS >= 3 Point Improvement" endpoint was added.
i. Removed "QIDS SR16" from efficacy analysis.
j. 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.
k. The endpoints "Histologic-Endoscopic Improvement," "Mucosal Healing," and "Alternative Mucosal Healing" were added.
l. Analysis of extraintestinal manifestations was added.

12. Analysis of UC surgeries and hospitalization were removed from Table AMBG.5.5 and Table AMBG.5.6, and added to Section 5.13.
13. Removed by-patient listing of exposure in Section 5.17.1.
14. Clarified the analysis AESIs including the analysis of infections (Section 5.17.8.2), hypersensitivity (Section 5.17.8.3) and suicidal ideation/behavior and depression (Section 5.17.8.7).
15. Altered the calculation of visit date and clarified the calculations for daily diary in Appendix 1.
16. Added region definitions in Appendix 2.

Statistical analysis plan (SAP) Version 3 was approved on 24 February 2020 and was based on protocol amendment (a) approved 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 or edited the following sections based on an error found in some of the electronic clinical outcomes assessment (eCOA) devices:

    a. A description of the error was added in Section 4.2.
    b. The definition of the modified intent-to-treat (ITT) population found in Table AMBG.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 these 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 and Section 5.11, was added.

2. 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 4.3.
    b. A description of analysis considerations related to COVID-19 mitigations was added in Section 5.5, and sensitivity analysis added to Table AMBG.5.5.
    c. A clarification was added stating that even some mitigations approved under the COVID-19 addendum will cause patients to be excluded from the per-protocol population in Section 5.21.
3. Changed the definition of 'Primary histologic remission' and "Histologic-Endoscopic Mucosal Remission", renamed the 'Mucosal healing' endpoint to 'Histologic-Endoscopic Mucosal Remission' in Table AMBG.3.1 and Table AMBG.5.4.
4. Clarified the definition of *C. difficile* negative and remove 'stool toxin' in Section 4.1.
5. Clarified the definition of Cohort in Table AMBG.5.2.
6. Added analysis considerations to handle the Hungary Addendum (AMBG Protocol Addendum [11]) in Section 5.20 and the Hungarian Addendum extension period in Table AMBG.5.3.
7. Clarified in Section 5.1.4 that Fishers exact test would only be used as a supportive analysis for binary efficacy endpoints.
8. Clarified wording in Section 5.3.1 and 5.3.2 on estimands. Added details on how to analyze data for patients receiving protocol-defined rescue treatment with mirikizumab for mixed-effects model for repeated measures (MMRM) and modified baseline observation carried forward (mBOCF) analysis in Section 5.3.2 and 5.3.3.
9. Added a multiple imputation approach was as a sensitivity analysis in Section 5.3.4 and Table AMBG.5.5. The multiple imputation approach will be used for ITT analysis instead of the previously described nonresponder imputation (NRI) approach.
10. Added the tipping point analysis in Section 5.3.5.
11. Clarified the analysis considerations for patient disposition in Section 5.8 and removed irrelevant statement.
12. Removed and added categories in the Table AMBG.5.4 baseline characteristics table, changed the description of several variables, and edited the subgroup analysis.
13. Changed the definition of 'having had a study drug administration visit on a given date' in Section 5.10.
14. Clarified how to handle duplicate entries on the site-facing eCOA device in Section 5.12.
15. Added that Week 40 endoscopies performed outside of the window from Days 267 to 337 will be considered missing for analysis purposes in Section 5.5 and Table AMBG.5.5. This also impacted the diary window in Appendix 1.
16. Added endpoints and analysis for Alternate Clinical Remission 2 and Urgency Remission in Table AMBG.5.4 and Table AMBG.5.5.
17. Edited categories and analysis description for extraintestinal manifestations in Table AMBG.5.4 and Table AMBG.5.5.

18. Added analysis for nonresponder cohort for several endpoints in Table AMBG.5.5.
19. Edited a summary of sensitivity analysis in Sections 5.12.2 and 5.12.4.
20. Edited safety analyses in Table AMBG.5.7 and Table AMBG.5.8.
21. Added a section on subgroup analysis for the Urgency NRS endpoint in Section 5.18.
22. Added the section on the Japan subgroup analysis in Section 5.19.
23. Added additional details about the Week 40 database lock (DBL), final DBL, pharmacokinetics DBL, and maximized extended enrollment DBL in Section 5.20.
24. Made minor edits that did not change meaning throughout the document.

Statistical analysis plan (SAP) Version 4 was approved on 01-October-2021 and was based on protocol amendment (a) approved 12 September 2019. The following updates were made in Version 4 before the first unblinded analysis by the sponsor:

1. Changes to the primary and major secondary endpoints:
    a. Moved 'Alternate clinical remission at Week 40 among patients induced into alternate clinical remission' and 'Stable maintenance of symptomatic remission' from major secondary endpoints to other secondary endpoints.
    b. Changed the population of 'Corticosteroid-free remission' to be assessed in all patients instead of only among those receiving corticosteroid at baseline.
    c. Included "Urgency Numeric Rating Scale = 0 or 1" as a major secondary endpoint.
    d. Updated the graphical testing scheme in Figure AMBG.5.1.
2. Added 2 subgroups 'Number of failed biologics or tofacitinib (0,1, >=2) and 'Prior failure of conventional therapies but not biologics or tofacitinib (failed, not failed)' in Table AMBG.5.4.
3. Clarified 'Prior medications' are medications that start and stop prior to the date of first dose of AMAN study treatment.
4. Added endpoint 'Urgency NRS = 0 or 1 or 2' in Table AMBG.5.5 and Table AMBG.5.6.
5. Clarified that by-visit measurements will be summarized for binary endpoints derived from diary assessments through maintenance period, open-label extended induction and open-label extended maintenance periods, and weekly measurements will be summarized for binary endpoints through loss of response period.
6. Clarified that analyses will be done separately for visit-based records and diary-based records for Nocturnal Stool, Fatigue NRS, and Bristol Stool Scale in Table AMBG.5.5.
7. Added other secondary endpoints that combine Mayo Score and Urgency, or combine Mayo Score and Histology.
8. Clarified that only descriptive summaries will be provided for efficacy for patients in induction nonresponder and delayed responder cohorts.
9. Removed a few sensitivity analyses as a result of change in graphical scheme.
10. Changed the analysis method for corticosteroid dose change from baseline from MMRM to analysis of covariance (ANCOVA).

11. Added exploratory association analyses in Section 5.14 and anchor-based analyses in Section 5.15 to address FDA feedback.
12. Clarified in Section 5.17.7 that immunogenicity for patients treated in Study AMBG will be provided in the Integrated Summary of Immunogenicity.
13. Minor edits in language for clarification that did not change meaning throughout the document.

3. Study Objectives

3.1. Primary and Major Secondary Objectives

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

The estimand (ICH E9R1 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 AMBG.5.1 and Table AMBG.5.5.
- The endpoint/variables may be found in Table AMBG.3.1, Table AMBG.5.4, and Table AMBG.5.5.
- The handling of intercurrent events and missing data may be found in Section 5.3 and Table AMBG.5.5.
- Population summary measures are described in Section 5.1.4 and Table AMBG.5.6.

Table AMBG.3.1. Primary Objective and Major Secondary Endpoints

Objective	Endpoint
Primary[a]	
• To test the hypothesis that mirikizumab is superior to placebo in achieving *clinical remission* at Week 40 among patients induced into **clinical response** with mirikizumab in Study AMAN	• The proportion of patients in *clinical remission* at Week 40, defined as: o Stool frequency (SF) subscore = 0, or SF = 1 with a ≥1-point decrease from induction 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 achieving *alternate clinical remission* at Week 40 among patients induced into **clinical response** with mirikizumab in Study AMAN[c]	• The proportion of patients in *alternate clinical remission* at Week 40, defined as[c]: o SF subscore = 0, or SF = 1, and o RB subscore = 0, and o ES = 0 or 1 (excluding friability).
• To evaluate the efficacy of mirikizumab compared to placebo in maintaining *clinical remission* at Week 40 (Week 52 of continuous therapy) among patients induced into **clinical remission** with mirikizumab	• The proportion of patients who were in *clinical remission* at Week 40 among patients in clinical remission at Week 12 in AMAN, with clinical remission defined as: o SF subscore = 0 or SF = 1 with a ≥1-point decrease from induction baseline, and o RB subscore = 0, and o ES = 0 or 1 (excluding friability)
• To evaluate the efficacy of mirikizumab compared to placebo on endoscopic remission at Week 40 among patients induced into **clinical response** with	• The proportion of patients in endoscopic remission at Week 40, defined as: • ES = 0 or 1 (excluding friability)

mirikizumab	
• To evaluate the efficacy of mirikizumab compared to placebo in achieving corticosteroid-free remission[d] without surgery among patients induced into **clinical response** with mirikizumab	• Corticosteroid-free remission without surgery at Week 40, defined as: o Clinical remission at Week 40, and o Symptomatic remission at Week 28, and o No corticosteroid use for ≥12 weeks prior to Week 40
• To evaluate the efficacy of mirikizumab compared to placebo in Histologic-Endoscopic Mucosal Remission at Week 40 among patients induced into **clinical response** with mirikizumab[e]	• The proportion of patients with Histologic-Endoscopic Mucosal Remission at Week 40, defined as achieving both[e]: • **Histologic remission** with resolution of mucosal neutrophils, defined using the Geboes scoring system with subscores 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) • **Endoscopic remission,** defined as ES = 0 or 1 (excluding friability)
• To evaluate the efficacy of mirikizumab compared to placebo on bowel movement urgency improvement at Week 40, among patients who were induced into **clinical response** with mirikizumab[f]	• The change from induction baseline in the Urgency Numeric Rating Scale score[g].
• To evaluate the efficacy of mirikizumab compared to placebo in achieving Urgency Remission at Week 40 among patients induced into clinical response with mirikizumab and had Urgency NRS ≥3 at induction baseline[h]	• The proportion of patients with Urgency Remission at Week 40, defined as: • Urgency Numeric Rating Scale = 0 or 1.

Abbreviations: MMS = modified Mayo Score; SAP = statistical analysis plan; UC = ulcerative colitis.

[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 0.05. 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] "Corticosteroid-free remission" will be assessed among all patients induced into clinical response with mirikizumab. This definition will supersede the protocol amendment (a), in which the endpoint was assessed only among patients with baseline corticosteroid use.

[e] Note that "Histologic-Endoscopic Mucosal Remission" was previously labeled as "Mucosal healing" and was listed as "Other Secondary" in the protocol amendment (a). The complete description of this endpoint is provided in this SAP. Also, "Histologic-Endoscopic Mucosal Remission" is now designated as a major secondary (i.e., multiplicity controlled) endpoint as a replacement of "Histologic remission". "Histologic remission" is listed as a major secondary endpoint in the protocol amendment (a). However, it will be designated as an "Other Secondary" endpoint (i.e., it will not be included in the multiplicity-controlled framework).

[f] The SAP language for the objective of "Bowel movement urgency improvement" supersedes the protocol language, which states "To evaluate the efficacy of mirikizumab compared to placebo on bowel movement urgency improvement at Week 40, among patients who: (1) had bowel urgency symptoms at induction baseline and (2) were induced into clinical response with mirikizumab".

[g] The SAP language for the endpoint related to "Bowel movement urgency improvement" supersedes the protocol language, which states "The proportion of patients with bowel movement urgency improvement at Week 40 as defined in the study SAP."

[h] Note that 'Urgency Remission' is added as a "major secondary" (i.e., multiplicity controlled) endpoint in the SAP and will supersede the protocol amendment (a).

Additional note: The endpoint 'Stable maintenance of symptomatic remission' is moved from "major secondary"(i.e., multiplicity controlled) endpoint to "other secondary" endpoint in the SAP and will supersede the protocol amendment (a).

3.2. Other Secondary and Exploratory Objectives

Other Secondary and Exploratory Endpoints can be found in Table AMBG.5.5 below.

4. Study Design

4.1. Summary of Study Design

Study I6T-MC-AMBG (AMBG) is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel-arm study evaluating the safety and efficacy of 200 mg mirikizumab Q4W subcutaneous (SC) in maintaining treatment response at Week 40 (i.e., 52 weeks of continuous therapy). The study also will evaluate the safety and efficacy of: (a) extended induction with 300 mg mirikizumab intravenous (IV) for patients who did not have a clinical response at Week 12 of Study I6T-MC-AMAN (AMAN), and (b) rescue induction for patients who achieved clinical response in Study AMAN and subsequently lose clinical response during Study AMBG.

Patient Population

Patients with moderately to severely active ulcerative colitis (UC) who completed Study AMAN and who meet eligibility requirements will enroll in this study. The study will enroll patients who achieve clinical response or clinical remission with blinded mirikizumab or placebo dosing in Study AMAN, as well as patients who do not achieve clinical response with blinded mirikizumab or placebo during Study AMAN.

Treatment Assignments

Maintenance study treatment assignment will be dependent on whether patients responded to study drug dosing in AMAN and whether they experience a loss of response (LOR) during this study as follows:

- **Mirikizumab Responders from Study AMAN**

  Patients who achieve clinical response with blinded mirikizumab in induction Study AMAN will be randomized to receive blinded 200 mg mirikizumab Q4W SC or blinded placebo SC Q4W (randomized withdrawal) in a 2:1 ratio. Randomization will be stratified to achieve between-group comparability, based on biologic-failed status (yes or no), induction remission status (yes or no), corticosteroid use (yes or no), and region (North America/Europe/Other).

  Patients will continue on the randomized treatment assignment for the remainder of Study AMBG unless they develop secondary LOR.

  Loss of Response is defined as:

  - ≥2-point increase from Study AMBG baseline in the combined stool frequency (SF) + rectal bleeding (RB) scores AND combined SF + RB score of ≥4, on 2 consecutive visits (≥7 days apart, and with confirmation of negative *C. difficile* testing),
  - AND
  - Confirmed by centrally read endoscopic subscore (ES) of 2 or 3

Subcutaneous dosing should be continued according to dosing schedule until endoscopy determines whether LOR is confirmed. If LOR is confirmed based on endoscopy at or after Week 12 of AMBG (and *C. difficile* testing is negative), patients will be rescued with open-label 300 mg mirikizumab Q4W IV for 3 doses. The first IV rescue dose may be administered as soon as LOR is confirmed by centrally read endoscopy, if ≥7 days from the most recent SC dose. Subsequent doses will be given every 4 weeks for total of 3 doses.

If endoscopy does not confirm secondary LOR, patients are encouraged to continue SC study drug dosing, maintaining the scheduled dosing interval. If study drug dosing is continued, an additional endoscopy would be performed at Week 40, early termination visit (ETV) or unscheduled visit (UV).

Patients who, in the opinion of the investigator, receive clinical benefit after completion of the LOR rescue therapy (12 weeks after first IV rescue dose) may be considered for enrollment into the long-term extension study I6T-MC-AMAP (AMAP) to receive further SC dosing. Once the LOR IV rescue therapy is initiated, no further SC dosing will be available in AMBG.

If study drug is discontinued, procedures for early termination from study drug will be performed and the patient should undergo posttreatment follow-up as described in the schedule of activities (AMBG Protocol Amendment [a] Section 2).

- **Placebo Responders from Study AMAN**

  Patients who achieve clinical response with blinded placebo in the induction study will continue to receive blinded placebo for the remainder of the maintenance study. Placebo SC injections will be administered Q4W to maintain study blind. If LOR is confirmed based on endoscopy at or after Week 12 (and *C. difficile* stool toxin testing is negative), patients will be rescued with open-label mirikizumab 300 mg Q4W IV for 3 doses. The same LOR assessments and procedures should be performed as described for the mirikizumab responders from Study AMAN.

- **Mirikizumab and Placebo Nonresponders from Study AMAN**

  Patients who do not achieve clinical response to blinded mirikizumab or blinded placebo in Study AMAN will receive open-label extended induction therapy with 300 mg mirikizumab IV at Weeks 0, 4, and 8, and undergo endoscopy at Week 12.

  Patients who achieve delayed clinical response (compared to induction study baseline) with extended mirikizumab induction therapy at Week 12 may subsequently receive open-label 200 mg mirikizumab Q4W SC starting at Week 12. Patients will continue on this dose regimen and undergo endoscopy at Week 40. Patients who, in the opinion of the investigator, receive clinical benefit may be considered for enrollment into the long-term extension Study AMAP to receive further SC dosing. Patients who discontinue study drug before Week 40 will undergo endoscopy and procedures for early termination

of the study drug, including posttreatment follow-up as described in the schedule of activities (AMBG Protocol Amendment [a] Section 2).

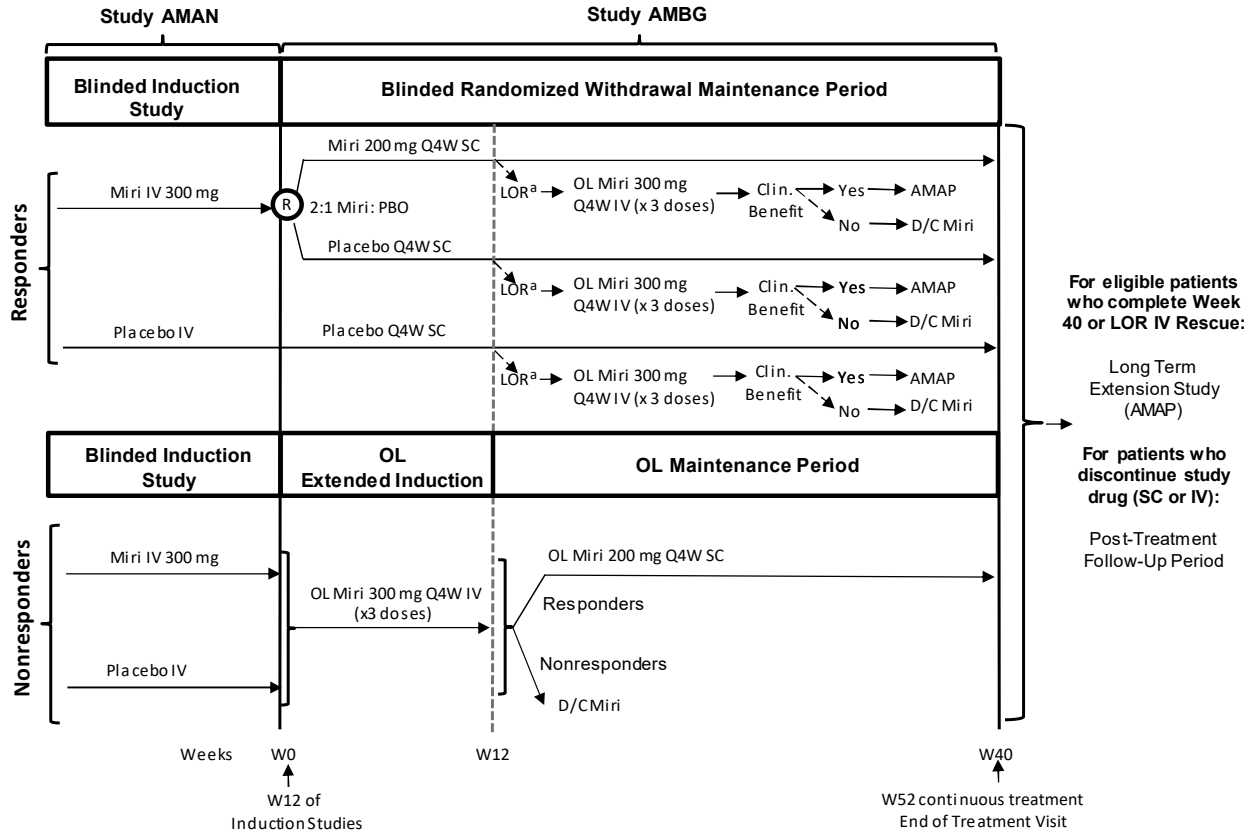
Patients who do not achieve clinical response to mirikizumab IV extended induction therapy at Week 12, compared to induction baseline, will discontinue study drug and undergo procedures for early termination of the study drug, including posttreatment follow-up as described in the schedule of activities (AMBG Protocol Amendment [a] Section 2).

Posttreatment Follow-up Period

Patients will undergo a maximum 16-week posttreatment follow-up period:

- Patients who discontinue study drug with **last dose administered IV** will return for posttreatment follow-up visits (Visit 801 and 802) at 4 and 16 weeks after the end-of-treatment visit.
- Patients who discontinue study drug with **last dose administered SC** will return for posttreatment follow-up visits (Visit 801 and 802) at 4 and 12 weeks after the end-of-treatment visit.
- Patients who subsequently enter the long-term extension Study AMAP do not need to complete the posttreatment follow-up period.

Figure AMBG.4.1 illustrates the AMBG study design.



Abbreviations: D/C = discontinue; IV = intravenous; LOR = loss of response; Miri = mirikizumab; OL = open-label; PBO = placebo; Q4W = every 4 weeks; R = randomization; RB = rectal bleeding; SC = subcutaneous; SF = stool frequency; W = week.

Note: LOR is defined as ≥2-point increase from maintenance baseline in the combined SF + RB scores, AND combined SF + RB score of ≥4 on 2 consecutive visits, AND confirmed by an endoscopic subscore of 2 or 3.

a = Loss of response at or after Week 12 and up to and including Week 28.

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

4.2. Electronic Clinical Outcomes Assessment Transcription Error Protocol Addenda

Questions to assess patient-reported outcome (PRO) measures of SF and RB are recorded using electronic clinical outcomes 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. 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. Two addenda were created in Poland and Turkey separately, to allow patients impacted by the eCOA transcription errors, who do not receive clinical benefit in the opinion of

the investigator, to enroll into the extended induction period of Study AMBG. Some patients had already enrolled into Study AMBG as Study AMAN induction responders at the time when the transcription error was discovered, and these patients will continue on the treatment assignment for the remainder of study unless they develop secondary LOR. For the impacted patients from Poland and Turkey, LOR is defined as:

- combined SF + RB score of ≥4, on 2 consecutive visits (≥7 days apart), and with confirmation of negative *C. difficile* testing (first assessment can start as early as Week 8/Visit 3), AND
- confirmed by centrally read ES of 2 or 3 not sooner than Week 12/Visit 4.

Patients who were not Study AMAN responders but experience clinical benefit from extended induction dosing, in the opinion of investigator, at Study AMBG Week 12 will continue to receive extended maintenance SC treatment.

Additional details of the issue and planned analysis can be found in the addenda and Section 5.4 below.

4.3. COVID-19 Addendum

Study AMBG 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 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 40 endoscopy assessment
- allowing a patient missing the Study AMAN Week 12 endoscopy, despite the window extension, to continue into the maintenance study. *Note although allowed by the Protocol Addendum I6T-MC-AMBG (16), no patients rolled over into Study AMBG without the Study AMAN Week 12 endoscopy.*
- allowing a patient missing the AMBG Week 12 endoscopy (nonresponder path), despite the window extension, to continue in the maintenance study
- allowing a patient missing the LOR endoscopy (responder path) to receive rescue IV dosing with mirikizumab
- allowing a patient missing the Week 40 endoscopy, despite the window extension, to continue into the long-term extension study, AMAP.
- 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)
- home administration of SC study drug

Additional details of the mitigations and planned analysis are provided in the Protocol Addendum I6T-MC-AMBG (16) and in Section 5.5 below.

4.4. Determination of Sample Size

Assuming that 90% of patients complete Study AMAN (which is expected to randomize approximately 1160 patients), Eli Lilly and Company (Lilly) anticipates approximately 1044 patients will enroll in Study AMBG. It is expected that approximately 470 of these patients will enter Study AMBG as clinical responders to mirikizumab and then will be randomized 2:1 to 200 mg mirikizumab SC (313 patients) and placebo (157 patients). Among the approximately 470 mirikizumab clinical responders, approximately 180 mirikizumab clinical remitters will be randomized to 200 mg mirikizumab SC (120 patients) and placebo (60 patients). This assumes that:

- The induction study (AMAN, which has a mixed population with approximately 50% biologic-failed patients) is expected to have an overall clinical remission rate of 23% and response rate of 60% with mirikizumab.
- 75% of induction patients receive treatment with mirikizumab, based on a 3:1 randomization ratio for the induction study.
- 10% dropout rate from induction to maintenance.

The primary endpoint, clinical remission at Week 40, will be assessed on patients who achieved clinical response to mirikizumab induction treatment. Assuming mirikizumab and placebo clinical remission rates of 47% and 27%, respectively, this study based on the 470 mirikizumab induction responders is expected to have >95% power to demonstrate that mirikizumab is superior to placebo by using a chi-square test with a 2-sided significance level of 0.05. In addition, the sample size is expected to provide adequate power (>80%) to demonstrate that mirikizumab is superior to placebo for endoscopic remission, and corticosteroid-free remission at Week 40, among responders to mirikizumab induction treatment by using a chi-square test with a 2-sided significance level of 0.05.

4.5. Method of Assignment to Treatment

Assignment to treatment groups for patients entering Study AMBG as clinical responders will be determined by a computer-generated random sequence using an interactive web-response system (IWRS) in a double-blind manner. To achieve between-group comparability, patients will be stratified to these arms based upon biologic-failed status (yes or no), corticosteroid use (yes or no), region (North America/Europe/Other), and induction clinical remission status (yes or no). This stratification will be controlled by IWRS.

4.5.1. Selection and Timing of Doses

The doses will be administered at each scheduled visit as described in the Schedule of Activities (AMBG Protocol Amendment [a] Section 2). The actual time of all dose administrations will be recorded in the patient's electronic case report form (eCRF). Doses, other than rescue mirikizumab IV doses, should not be administered during unscheduled visits, unless the

unscheduled visit occurs during the next scheduled dosing window. Rescue mirikizumab IV doses may be administered if LOR has been confirmed and if it is ≥7 days from the most recent SC dose. Subsequent rescue mirikizumab IV doses will be given every 4 weeks (±3 days) for total of 3 doses.

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 CSR.

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

5.1.1. Populations for Analyses

Patient populations are defined in Table AMBG.5.1 along with the analysis that the patient population will be used for. The treatment groups and inferential comparisons described in Table AMBG.5.2 will be used unless otherwise specified. Unless otherwise specified, for all populations/analysis, patients will be analyzed according to the treatment to which they were assigned.

Table AMBG.5.1. Analysis Populations in Study AMBG

Population	Description
All entered patients	**Definition:** All patients who signed informed consent. **Purpose:** Used for disposition analysis.
Modified Intent-to-treat (mITT) Population	**Definition:** All assigned/randomized patients who received any amount of study treatment excluding patients impacted by the eCOA transcription error in Poland and Turkey (regardless of if the patient does not receive the correct treatment, or otherwise does not follow the protocol). **Purpose:** Used for efficacy and health outcomes analysis.
Safety Population	**Definition:** All assigned/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, including patients impacted by the eCOA error from Poland and Turkey). **Purpose:** Used for safety-related analysis.
Intent-to-Treat (ITT) Population	**Definition:** All assigned/randomized patients. Patients will be analyzed according to the treatment to which they were assigned. **Purpose:** Used for sensitivity analysis of the primary and major secondary efficacy

	endpoints.
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 5.21). **Purpose:** Used as a sensitivity analysis for the primary and major secondary efficacy endpoints.
Pharmacokinetic Evaluable	**Definition**: All patients who received at least 1 dose of investigational product and have sufficient blood sampling to allow for pharmacokinetic evaluation **Purpose:** Used for Pharmacokinetic and Pharmacodynamic analyses.

Abbreviations: eCOA = electronic clinical outcomes assessment; GCP = good clinical practice.
Analysis cohorts are defined in Table AMBG.5.2.

Table AMBG.5.2. Analysis Cohorts in Study AMBG

Cohort	Description	Treatment Groups	Inferential Comparisons
Mirikizumab Induction responder (Randomized maintenance cohort/Main cohort)	Patients who responded, as per the IWRS, to mirikizumab induction dosing at Week 12 of AMAN and then are re-randomized to 200 mg mirikizumab SC or placebo	placebo SC 200 miri SC	placebo SC vs 200 miri SC
Mirikizumab Induction remitter	Patients classified as clinical remitters at Week 12 of AMAN and are re-randomized, per the IWRS, to 200 mg mirikizumab SC or placebo.	placebo SC 200 miri SC	placebo SC vs 200 miri SC
Induction responder (Cohort 1 + Cohort 3 in the protocol)	Patients enrolled into AMBG who were classified as clinical responders, as per the IWRS, at Week 12 of AMAN.	placebo SC (PR) placebo SC (MR) 200 miri SC (MR)	placebo SC (MR) vs 200 miri SC (MR)
Loss of Response (LOR) cohort	Patients who responded, as per the IWRS, to induction dosing at Week 12 of AMAN, lost response, and received a dose of open label IV mirikizumab rescue therapy.	placebo SC (PR) placebo SC (MR) 200 miri SC (MR)	No comparison
Induction nonresponder (Cohort 2 + Cohort 4 in the protocol)	Patients enrolled into AMBG who were classified as clinical nonresponders, as per the IWRS, at Week 12 on AMAN.	300 miri IV (MN) 300 miri IV (PN)	No comparison
Delayed clinical responder	Subset of induction nonresponders, as per the IWRS, who achieved delayed clinical response, as per the IWRS, entered the open label maintenance period, and received SC mirikizumab dosing	200 miri SC (MN) 200 miri SC (PN)	No comparison

Abbreviations: IWRS = interactive web-response system; MN = Mirikizumab induction nonresponder; MR = Mirikizumab induction responder; PN = Placebo induction nonresponder; PR = Placebo induction responder.

5.1.2. Study Time Intervals

Table AMBG.5.3 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 AMBG.5.3 should be understood to mean "the day/time before" while the words "after" should be understood to mean "the day/time 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 AMBG.5.3. Definition of Study Period Time Intervals

Study Period Definition	Interval Start Definition	Interval End Definition
Maintenance Period (M): All patients who responded to AMAN treatment, as per the IWRS, and are randomized/assigned to the study are considered as entering the Maintenance Period.	At the date/time[a] of first AMBG study drug administration following randomization/assignment. For patients who are randomized/assigned but not dosed, the Maintenance Period starts on the date of randomization/assignment.	The maximum of treatment discontinued date and last treatment visit date. For patients who enter LOR period, the Maintenance Period ends prior to LOR period start. For Hungarian patients who enter the blinded Hungarian Addendum extension period (HE), the Maintenance Period ends prior to Hungarian Addendum Extension Period start.
Loss of Response Period (LOR): All patients who are eligible for and satisfy LOR criteria and receive LOR dosing are considered as entering the LOR period.	At the date/time[a] of first dose of AMBG rescue therapy.	The maximum of treatment discontinued date and last treatment visit date.
Open Label Extended Induction Period (EI) All patients who did not achieve clinical response to AMAN treatment, as per the IWRS, and are assigned to the study are considered as entering the Open Label Extended Induction Period.	At the date/time[a] of first AMBG study drug administration in the OL Extended IV Induction period. For patients who are assigned but not dosed, this period starts on the date of AMBG treatment assignment.	The maximum of treatment discontinued date and last treatment visit date. For patients who enter Open Label Maintenance period, the Open Label extended Induction period ends prior to Open Label Maintenance period start.

Open Label Maintenance Period (EM) All patients who did not achieve clinical response to AMAN treatment, achieved delayed response at AMBG Week 12, as per the IWRS, and received Open Label Maintenance dosing are considered as entering the Open Label Maintenance Period.	At the first date/time[a] of AMBG study drug administration in the OL Maintenance Period. If a patient is unable to be dosed at AMBG Week 12, the EM period starts at the Week 12 visit.	The maximum of treatment discontinued date and last treatment visit date.
Posttreatment follow-up period (FUP) All patients who had Visit 801 or 802 are considered to have entered the follow-up period	After last date of any study treatment period defined above	The maximum of the last study visit and study disposition date.
Hungarian Addendum Extension Period (HE)	At the date/time of Week 40 study drug administration.	The maximum of the last study visit and study disposition date.

Abbreviation: IWRS = interactive web-response system.

[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

For efficacy and health outcomes analysis, unless otherwise specified, baseline refers to the values or observations collected before the initiation of study treatment in Study AMAN.

The baseline for variables collected as part of the daily diary (including the PRO components of the Modified Mayo Score, SF and RB subscores) will be calculated from valid daily diary entries obtained prior to baseline endoscopy preparation (see Appendix 1). The baseline endoscopy component of the Mayo Score will use the endoscopic appearance of the mucosa at the AMAN screening endoscopy. For other efficacy and health outcome 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) in Study AMAN.

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 0.05 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 0.05, a 2-sided 95% 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 as mentioned in Section 5.2. 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 as mentioned in Section 5.2. The CMH p-value will be reported.
- The odds ratio and the corresponding CIs, adjusting for the stratification factors, as mentioned in Section 5.2, will be provided.
- As a secondary measurement of efficacy, the relative risk along with its 2-sided CI will be provided, adjusting for the stratification factors in Section 5.2 using the Mantel-Haenszel estimator. If deemed necessary as a supportive analysis, additional analyses of categorical efficacy variables may be conducted to address sparse data or small sample sizes. A Fisher's exact test may be utilized. When specified as a sensitivity analysis for binary endpoints, logistic regression with a Firth penalized likelihood (Firth 1993) will be used. The model will include the treatment groups and the covariates described in Section 5.2. Firth correction can be implemented in PROC Logistic by including *'firth'* as an option in the model statement.

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) corticosteroid use (yes/no), (d) AMAN Clinical Remission status (yes or no), (e) region (North America/Europe/Other), (f) baseline value, (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 postbaseline 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) AMAN Clinical Remission status (yes or no), (e) region (North America/Europe/Other), and (f) baseline value in the model. Type III sums of squares for 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. Unless otherwise specified, for ANCOVA, reported data from only planned visits will be used as the primary analysis.

5.2. Adjustments for Covariates

The randomization is stratified by (a) biologic failed patient (yes/no), (b) corticosteroid use (yes/no) at AMAN baseline, (c) region (North America/Europe/Other) and (d) AMAN Clinical Remission status (yes or no). These factors will be adjusted for as described in Section 5.1.4.

5.3. Handling of Dropouts or Missing Data

Intercurrent events (ICH E9R1 2017) 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. Assessments performed outside these windows will not be excluded from any analysis (unless otherwise specified), but will be reported as a protocol deviation (see Section 5.21).

5.3.1. Nonresponder Imputation (NRI)

For analysis of binary efficacy and health outcomes variables, missing data will be imputed using a nonresponder imputation (NRI) method. Patients will be considered nonresponders for the NRI analysis if they do not meet the categorical efficacy criteria, have missing clinical efficacy data at a time point of interest, or take rescue medication prior to the time point of interest.

The NRI method can be justified based on the composite strategy (ICH E9R1 2017) for handling intercurrent events. In this strategy, patients are defined as responders only if they meet the clinical efficacy criteria at the predefined time, complete the study treatment period without missing relevant data, and do not receive rescue dosing of IV mirikizumab. Patients failing any of the above criteria will be categorized as nonresponders.

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

For continuous variables, the primary analysis will be MMRM under the missing at random (MAR) assumption for 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. Data from planned visits prior to treatment discontinuation will be used in the analysis regardless of whether the patient took prohibited concomitant rescue medication, was not compliant with treatment, or otherwise violated the protocol during their time in the treatment period. Patients who receive protocol-defined rescue treatment with mirikizumab will be treated as missing after receiving rescue for the analysis of the randomized maintenance cohort. Therefore, our analysis under the missing at random assumption may be justified as consistent in intent with the treatment policy estimand with regard to use of prohibited concomitant medication and adherence to treatment. With regard to the specific intercurrent event of receiving protocol-defined rescue therapy after LOR, the estimand is intended to reflect the hypothetical scenario where the patient continued on assigned treatment.

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 the analysis of the randomized maintenance cohort, patients who receive protocol-defined rescue treatment with mirikizumab will be treated as missing after receiving rescue; therefore, their last observation will be carried forward. For all patients with intermittent 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., rescue treatment and discontinuation due to reasons other than an AE) or intermittent 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 at Week 40 for patients impacted by the eCOA transcription error, missing data will be imputed using modified nonresponder imputation (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 Week 40 ES for patients who received an endoscopy outside of the Days 267 to 337 window will be imputed.

The multiple imputation method will be implemented as follows:

- The modified Mayo subscores for all scheduled visits will be jointly 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 40 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 40 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 Mantel-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 the 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 Electronic Clinical Outcomes Assessment Error Patients in Poland and Turkey

Data from patients impacted by the Poland and Turkey eCOA transcription error addendum (as described in Section 4.2) 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 analyses for efficacy will be performed on primary and selected key secondary endpoints by including impacted patients using the intent-to-treat (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 5.17).

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 submission. The planned safety analyses will include the following summary tables for the re-randomized maintenance cohort: 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 (TE) abnormal lab results, TE 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 with a COVID-19-related AE, discontinuations from study treatment due to COVID-19-related issues (e.g., due to COVID-19-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 40 endoscopies which occurred more than 2 weeks early or more than 8 weeks late (i.e., Days 267 to 337) will be considered missing for analysis purposes (e.g., will be considered nonresponders for NRI analysis), and
  - 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 at Week 40 by treating patients who received their endoscopy outside of the 281 ± 14-day window as nonresponders.

- missing endoscopies:
  - Patients who miss the Week 12 endoscopy (nonresponder path) will be treated as nonresponders for all Study AMBG Week 12 endpoints that require an endoscopy. Patients who transition from the extended induction to extended maintenance period without an endoscopy at Study AMBG Week 12, per the COVID-19 addendum, will be included in the analysis for the extended maintenance period. Depending on the number of patients who rolled over, additional sensitivity analyses may be performed for selected efficacy endpoints at the end of extended maintenance period excluding those patients.
  - Patients who are rescued with mirikizumab IV treatment without an endoscopy, per the COVID-19 addendum, will be censored (e.g., subsequently treated as nonresponders) for the analysis of the maintenance treatment period. For the primary analysis of the LOR rescue period, these patients will be included in the analysis.
  - Patients who miss the Week 40 endoscopy will be treated as nonresponders for all endpoints that require an endoscopy. Multiple imputation of the endoscopy score will be performed as additional sensitivity analysis.
- other mitigations:
  - The primary analysis of all endpoints will include patients in the analysis populations as described in Table AMBG.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 5.19. No additional sensitivity analyses are planned.

Safety analyses will be performed as planned in Section 5.17, and described more fully in compound level safety standards, 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 crossindustry 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 0.1, 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 0.05, 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 0.05 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 will be fully specified by the graph (including nodes, arrows and weights) along with the initial alpha allocation.

Figure AMBG.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 improvement at Week 40) are binary and will be analyzed using the CMH test with NRI imputation. Bowel Movement Urgency improvement at Week 40 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. Revisions to the following graphical scheme, if any, will be based on blinded assessment of AMBG data.

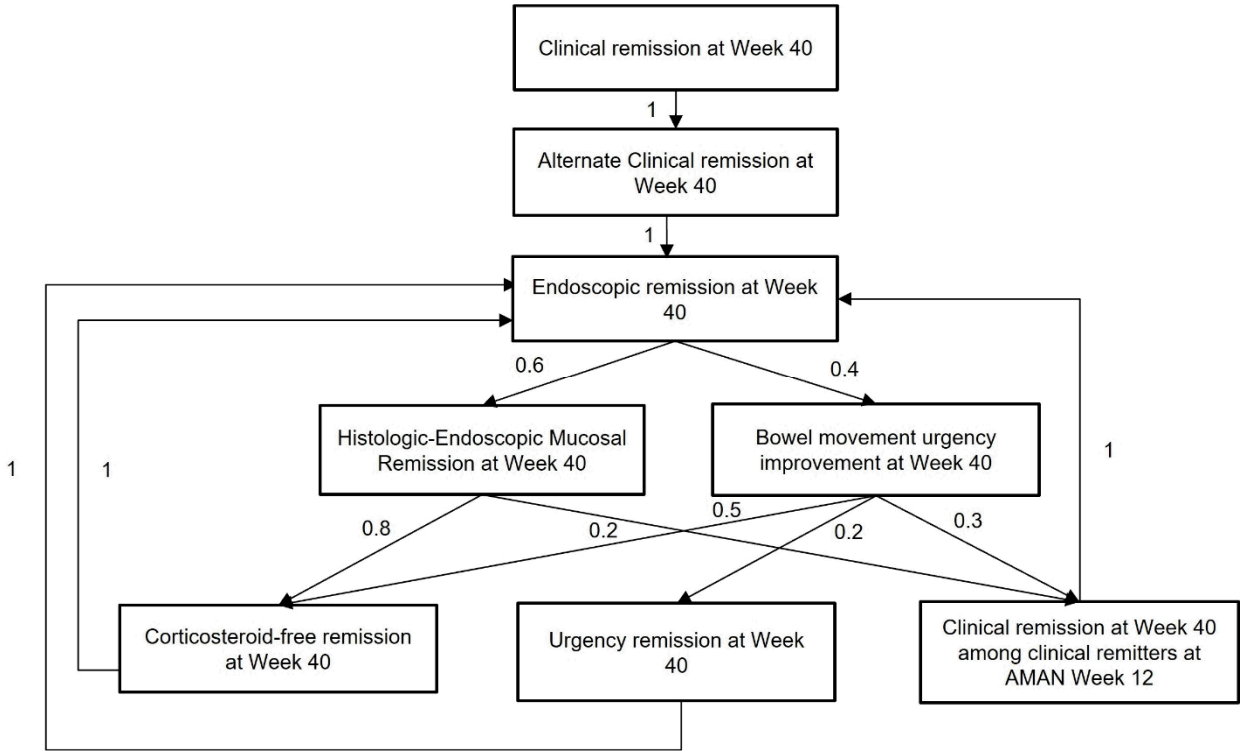


Figure AMBG.5.1. Graphical testing procedure for primary and major secondary endpoints.

5.8. Patient Disposition

The treatment disposition and study disposition will be summarized by treatment group for the mITT and ITT population. Summaries will also include reason for discontinuation from the study tabulated by treatment group. The study disposition of all patients who are randomized will be summarized along with the reason for discontinuation. The completers will be categorized into those completers who entered AMAP and those who did not. Summary will be by the groups: (1) treatment disposition completers, and (2) treatment disposition noncompleters.

All ITT patients who 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 mITT Population in the Maintenance Period, 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

Patient demographic variables and baseline characteristics will be summarized by dose and overall for the mITT and ITT population. The baseline characteristics of patients entering

additional study periods may also be summarized if necessary. 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 AMBG.5.4 describes the specific variables and how they will be summarized. The final column specifies variables used for the efficacy subgroup analysis described in Section 5.18. Changes to the Table AMBG.5.4, including the summary of additional patient characteristics and subgroup analysis will not require an amendment to the SAP.

Table AMBG.5.4. 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	X
	No	By Country (listed in other documents)	X
	No	Asia, North America, Central America/South America, East Europe, West Europe and ROW (rest of the world)	X
Height (cm)	Yes	None	
Weight (kg)	Yes	<80 kg, ≥80 kg	
		<100 kg, ≥100 kg	X
BMI[c]	Yes	Underweight (<18.5 kg/m2), Normal (≥18.5 and <25 kg/m2), Overweight (≥25 and <30 kg/m2), Obese (≥30 and <40 kg/m2), Extreme obese (≥40 kg/m2)	X
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	

Variable	Continuous Summary	Categorical Summary	Subgroup Analysis[a]
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[d] or tofacitinib used	No	0, 1, 2, >2	
Number of failed[e] biologics[d] or tofacitinib	No	0, 1, 2, >2	
Number of failed[e] biologics[d] or tofacitinib	No	0, 1, ≥2	X
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	
Prior failure of conventional therapies but not biologics or tofacitinib	No	Failed, Not failed	X
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 beclometasone[k]	No	Yes, No	
Baseline immunomodulator use[i]	No	Yes, No	X
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	

Variable	Continuous Summary	Categorical Summary	Subgroup Analysis[a]
		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 Subcore	No	Possible values of 4-point scale in Table AMBG.5.5.	
Baseline Stool Frequency Mayo Subscore	No	Possible values of 4-point scale in Table AMBG.5.5.	
Baseline Rectal Bleeding Mayo Subscore	No	Possible values of 4-point scale in Table AMBG.5.5.	
Baseline PGA Mayo Subscore	No	Possible values of 4-point scale in Table AMBG.5.5.	
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 scores	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 5.18 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 AMAN 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 AMAN prior med eCRF for Corticosteroid include: prednisone and other corticosteroid. Note that this is not exactly the same as the inclusion criteria defined in the protocol.

[h] Options on the AMAN 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 the 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 the 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.

5.9.1. Preexisting Conditions

Preexisting condition is defined as the condition/event recorded on the Pre-Existing Conditions and Medical History eCRF page with a start date prior to the date of AMBG first dosing, and no end date (i.e., the event is ongoing) or an end date on or after the date of first dosing. In addition, the AEs starting prior to first dose are also included. Notice if a preexisting condition worsens in severity on or after the date of first dosing, it will be recorded as an AE on Adverse Event eCRF page with the date of worsening as the start date.

The number and percentage of patients with pre-existing 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 Maintenance Period. The treatment compliance of patients while being treated during other study periods may also be summarized if necessary. Treatment compliance for each patient will be calculated as:

$$\textit{Treatment compliance}\ (\%) = 100 \times \frac{\textit{Total number of study drug administration visits}}{\textit{Total number of study drug administration visits planned per protocol}}$$

The "Total number of study drug administration visits planned per protocol" is based on the number of planned administration visits before the patient discontinued study drug. Each patient will be defined as having had a study drug administration visit on a given date if:

- For visits where the patient is to receive an IV infusion, they received the planned dose (i.e., a partial dose does not count) as derived from the Exposure eCRF page
- For visits where the patient is to receive SC injections, they received the planned number of injection doses (i.e., 2 doses) as derived from the Exposure eCRF page

"Overall compliance" with therapy is defined as having at least 80% treatment compliance. Proportions of patients who satisfy the definition of overall compliance during the Maintenance Period will be compared between treatment groups using Fisher's exact test. Overall Compliance may be summarized for other study periods if deemed necessary. Patient treatment compliance will be summarized for the mITT population.

5.11. 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 AMAN study treatment. *Concomitant medications* are those medications that start before, on or after the first day of AMBG 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 PTs 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 and ITT population.

Summary tables include the following:

- Summary of prior medications for mITT population for induction responders
- Summary of concomitant medication for mITT and ITT population by cohort and period (induction responder at maintenance period and LOR period, induction nonresponder at open label extended induction period and extended maintenance period)

- Summary of concomitant medication for mITT and ITT population within class of interest (corticosteroid therapy, immunomodulatory therapy) by cohort and period (induction responder at maintenance period and LOR period, induction nonresponder at open label extended induction period and extended maintenance period). Definition of these two classes of interest will be based on compound level safety standards.

Summaries for other study periods may be generated if deemed necessary.

5.12. Efficacy Analyses

Table AMBG.5.5 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 on these devices with different responses on the same visit, the first nonmissing response will be used.

Table AMBG.5.6 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 AMBG.5.5. 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 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 of the AMAN trial. • 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.	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. Additionally, for patients with the Week 40 endoscopy outside of the Days 267 to 337 (inclusive) window, the Week 40 endoscopy score will be treated as missing for analysis.
		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	Missing if SF, RB

Measure	Description	Variable	Derivation / Comment	Definition of Missing
	Each subscore is on a 4-point scale, ranging from 0 to 3.		decrease from baseline • RB subscore = 0, and • ES subscore = 0 or 1 (excluding friability)	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.
		Alternate Clinical Remission 3	• SF subscore = 0, or SF = 1 with a ≥1-point decrease from baseline • RB subscore = 0, and • ES subscore = 0	Missing if SF, RB or ES subscores are 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.
		Modified Mayo Score (MMS)	Calculated as: SF + RB + ES.	Missing if SF, RB or ES 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,	Missing if total Mayo Score is

Measure	Description	Variable	Derivation / Comment	Definition of Missing
			and • A decrease of >=1 point in the RB subscore from baseline or a RB score of 0 or 1	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.
		Stable maintenance of symptomatic remission	• The proportion of patients in symptomatic remission for at least 7 out of 9 visits from Week 4 to Week 36 and in symptomatic remission at Week 40 among patients in symptomatic remission (and clinical response) at Week 12 of AMAN. • Symptomatic remission is defined as: o SF = 0, or SF = 1 with a ≥1-point decrease from induction baseline o RB = 0	Missing if Symptomatic remission is missing at Week 0, Week 40, or more than 2 Weeks between Week 4 and Week 36.
		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
		Corticosteroid-free Remission	Corticosteroid-free remission without surgery at Week 40, defined as: • Clinical remission at Week 40 • Symptomatic remission at Week 28, and • No corticosteroid use for ≥12 weeks prior to Week 40	Missing if Clinical Remission at wk40 or Symptomatic Remission at w28 Missing
		Alternative Corticosteroid-free Remission	Corticosteroid-free remission without surgery at Week 40, defined as: • Clinical remission at Week 40	Missing if Clinical Remission at wk40 or Symptomatic

Measure	Description	Variable	Derivation / Comment	Definition of Missing
			• Symptomatic remission at Week 28, and • No corticosteroid use for ≥24 weeks prior to Week 40	Remission at w28 Missing
		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 (cit Travis). 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 numeric rating scale (NRS) is a single item that measures the severity for the urgency (sudden or immediate	Urgency NRS	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.

Measure	Description	Variable	Derivation / Comment	Definition of Missing
	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 >=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.
		Urgency Remission	Urgency NRS = 0 or 1	Missing if urgency NRS score is missing.
		Urgency NRS = 0 or 1 or 2	Urgency NRS = 0 or 1 or 2	Missing if urgency NRS score is missing.
Abdominal Pain NRS	The Abdominal Pain NRS is single 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 (i.e., “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 NRS for a 7-day period as described in Appendix 1.	Missing if fewer than 4 available measurements in the relevant 7 days.

Measure	Description	Variable	Derivation / Comment	Definition of Missing
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 waken from sleep.	Nocturnal Stool Score	For visit-based records collected at site, use the single item collected. For diary-based records, calculated by averaging data from all available daily diary entries of Nocturnal Stool NRS for a 7-day period as described in Appendix 1.	For visit-based records, missing if missing. For diary-based records, 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 you can imagine").	Fatigue NRS Score	For visit-based records collected at site, use the single item collected. For diary-based records, 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	For visit-based records collected at site, use the single item collected. For diary-based records, 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.	Missing if 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, impute 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 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

Measure	Description	Variable	Derivation / Comment	Definition of Missing
		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.
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	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, Item 3, Item 4, and Item5. Derive EQ-5D-5L UK Population-based index score using the UK algorithm (Szende et al. 2006) to produce a patient-level index score between -0.59 and 1.0 (continuous variable).	If any of the items is missing or equal to 9, the index score is missing

Measure	Description	Variable	Derivation / Comment	Definition of Missing
	associated with the most appropriate statement in each of the 5 dimensions.	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
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,	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.

Measure	Description	Variable	Derivation / Comment	Definition of Missing
	(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 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-UC (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 productivity in regular unpaid activities.	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 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

Measure	Description	Variable	Derivation / Comment	Definition of Missing
				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 2000), Robarts Histopathology Index (RHI) (Mosli 2015) and Nancy index (Marchal-Bressenot 2017).	Geboes Grades	Geboes assigns values to each of seven 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 following 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
		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	Missing if any component of the

Measure	Description	Variable	Derivation / Comment	Definition of Missing
			by Marchal-Bressenot (2017). A grade of 4 represents severely active disease while a grade of 0 represents no histological significant disease.	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.
		Primary Histologic Remission	Resolution of mucosal neutrophils, defined by Geboes histological subscore of 0 for grades: • 2b (lamina propria neutrophils) • 3 (neutrophils in epithelium) • 4 (crypt destruction), and • 5 (erosion or ulceration)	Missing if any component of the definition is missing.
		Histologic Improvement	Geboes histological subscores of: • 0 (None) or 1 (<5% of crypts involved) for parameter 3 (neutrophils in epithelium) • 0 (None) for parameter 4 (crypt destruction), and • 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: • 2B (neutrophils in lamina propria), and • 3 (neutrophils in epithelium), and 5 (erosion or ulceration)	Missing if any component of the definition is missing.
Histo-endo	Combined histology and endoscopic endpoints	Histologic-Endoscopic Mucosal Improvement	Histologic improvement and endoscopic remission	Missing if histologic remission is missing or endoscopic remission is missing

Measure	Description	Variable	Derivation / Comment	Definition of Missing
		Histologic-Endoscopic Mucosal Remission	Primary histologic remission and endoscopic remission.	Missing if histologic remission is missing or endoscopic remission is missing
		Alternative Histologic-Endoscopic Mucosal Remission	Primary histologic remission and endoscopic mayo subscore = 0	Missing if histologic remission is missing or endoscopic remission is missing
Enhanced Efficacy	Combined Mayo SF + RB + ES + Urgency endpoints	Clinical Response and Urgency NRS >=3 Point Improvement	Clinical Response and Urgency NRS >=3 Point Improvement	Missing if baseline MMS, Week 40 MMS, or Urgency NRS is missing.
		Clinical Remission and Urgency Remission	Clinical Remission and Urgency NRS = 0 or 1	Missing if SF, RB, or ES subscores or urgency NRS at baseline or Week 40 are missing.
		Clinical Remission and Urgency NRS = 0 or 1 or 2	Clinical Remission and Urgency NRS = 0 or 1 or 2	Missing if SF, RB, or ES subscores or urgency NRS at baseline or Week 40 are missing.
	Combined Mayo SF + RB + ES + Histology endpoints	Symptomatic Remission and Histologic-Endoscopic Mucosal Remission	Symptomatic Remission and Histologic-Endoscopic Mucosal Remission	Missing if SF, RB, or ES subscores or Geboes score at baseline or Week 40 are missing.
		Symptomatic Remission and Alternative Histologic-Endoscopic Mucosal Remission	Symptomatic Remission and Alternative Histologic-Endoscopic Mucosal Remission	Missing if SF, RB, or ES subscores or Geboes score at baseline or Week 40 are missing.

Measure	Description	Variable	Derivation / Comment	Definition of 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.
Corticosteroid dose	The ATC code for corticosteroid is listed in the compound level safety standards	Corticosteroid dose	Concomitant medication. Will transform to prednisone equivalent for analysis.	No Imputation.
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 AMAN study treatment.	No Imputation.
		Resolution of Baseline EIMs	Complete resolution of baseline EIMs at Week 40. If a patient has multiple baseline EIMs, then at least one of the EIMs must have resolved.	No Imputation.
		Improvement or Resolution in Baseline EIMs	Reduction in the severity of baseline EIMs at Week 40, or complete resolution of baseline EIMs at Week 40. If a patient has multiple EIMs, then at least one of the EIMs must have decreased in severity or resolved.	No Imputation.
		New EIMs	New EIMs at Week 40	No imputation
		Worsening from baseline of an EIM	Increase in the severity of any baseline EIMs at Week 40	No imputation

Table AMBG.5.6. Description of Efficacy/Health Outcomes Analyses

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Cohort, Time Point(s)[b]
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	Induction responder at Week 40
		Logistic regression analysis with NRI	mITT	
		Common risk difference with mNRI	mITT, ITT	Mirikizumab Induction responder at Week 40
		Tipping Point analysis	mITT	
		CMH analysis with NRI	mITT; PP	Induction clinical remitter at Week 40
		Common risk difference with mNRI	ITT	Mirikizumab Induction clinical remitter at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	Alternate Clinical Remission	CMH analysis with NRI	mITT	Induction responder at Week 40
			PP	
		Common Risk Difference with mNRI	ITT	Mirikizumab Induction responder at Week 40
		CMH analysis with NRI	mITT—In patients achieving Alternate Clinical Remission at AMAN Week 12	Mirikizumab Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	Alternate Clinical Remission 2	CMH analysis with NRI	mITT	Induction responder at Week 40
	Alternate Clinical Remission 3	CMH analysis with NRI	mITT	Induction responder at Week 40
	Clinical Response	CMH analysis with NRI	mITT	Induction responder at Week 40

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Cohort, Time Point(s)[b]
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	Endoscopic Remission	CMH analysis with NRI	mITT; PP	Induction responder at Week 40
		Common Risk Difference with mNRI	ITT	Mirikizumab Induction responder at Week 40
		CMH analysis with NRI	mITT—In patients achieving endoscopic remission at AMAN Week 12	Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	Symptomatic Remission	CMH analysis with NRI	mITT	Induction responder, by visit through Week 40
		Descriptive summary	mITT	Induction nonresponder, by visit through Week 12 (and Week 40[a]) Loss of response cohort, weekly through LOR rescue therapy period
	Stable maintenance of symptomatic remission	CMH analysis with NRI	mITT; PP. Both in patients achieving Symptomatic Remission at AMAN Week 12	Induction responder at Week 40
		Common Risk Difference with mNRI	ITT—In patients achieving Symptomatic Remission at AMAN Week 12	Mirikizumab Induction responder at Week 40
	Symptomatic Response	CMH analysis with NRI	mITT	Induction responder, by visit through Week 40
		Descriptive summary	mITT	Induction nonresponder, by visit through Week 12 (and Week 40[a]) Loss of response cohort, weekly through LOR rescue therapy period

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Cohort, Time Point(s)[b]
	Alternative Corticosteroid-free Remission	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
	Corticosteroid-free Remission	CMH analysis with NRI	mITT—In patients receiving corticosteroids at AMAN baseline	Induction responder at Week 40
			mITT; PP	Induction responder at Week 40
		Common Risk Difference with mNRI	ITT	Mirikizumab Induction responder at Week 40
		CMH analysis with NRI	mITT	Induction Clinical remitter at Week 40
		Descriptive summary	mITT	Induction nonresponder [a] hat Week 40
	SF component of clinical remission	CMH analysis with NRI	mITT	Induction responder, by visit through Week 40
		Descriptive summary	mITT	Induction nonresponder, by visit through Week 12 (and Week 40[a])
	RB component of clinical remission	CMH analysis with NRI	mITT	Induction responder, by visit through Week 40
		Descriptive summary	mITT	Induction nonresponder, by visit through Week 12 (and Week 40[a])
	Total Mayo Clinical Remission	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
	Total Mayo Clinical Response	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
	Endoscopic Response	CMH analysis with NRI	mITT	Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	Endoscopic Normalization	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
	Change from baseline in SF, RB and Total Symptomatic Score	MMRM; ANCOVA with mBOCF	mITT	Mirikizumab induction responder through Week 40
		Descriptive summary	mITT	Induction nonresponder through Week 12 (and Week 40[a])
	Change from baseline in ES	ANCOVA with mBOCF	mITT	Mirikizumab induction responder at Week 40

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Cohort, Time Point(s)[b]
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
UCEIS	Change from baseline in UCEIS and in individual components of the UCEIS	ANCOVA with mBOCF	mITT	Mirikizumab induction responder at Week 40
	UCEIS endoscopic remission	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
Urgency NRS	Change from baseline in Urgency NRS Score	MMRM	mITT; ITT; PP	Mirikizumab Induction responder through Week 40
		ANCOVA with mBOCF	mITT	Mirikizumab Induction responder through Week 40
		Descriptive summary	mITT	Induction nonresponder through Week 12 (and Week 40[a])
	Urgency NRS >= 3 Point Improvement	CMH analysis with NRI	mITT—In patients with Urgency NRS >= 3 at AMAN baseline	Mirikizumab induction responder, by visit through Week 40
		Descriptive summary	mITT—In patients with Urgency NRS >= 3 at AMAN baseline	Induction nonresponder, by visit through Week 12 (and Week 40[a])
	Urgency Remission	CMH analysis with NRI	mITT—In patients with Urgency NRS ≥3 at AMAN baseline	Induction responder, by visit through Week 40
			mITT	
		Descriptive summary	mITT—In patients with Urgency NRS ≥3 at AMAN baseline	Induction nonresponder, by visit through Week 12 (and Week 40[a])
	Urgency NRS = 0 or 1 or 2	CMH analysis with NRI	mITT—In patients with Urgency NRS ≥3 at AMAN baseline	Induction responder, by visit through Week 40

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Cohort, Time Point(s)[b]
Enhanced Efficacy	Clinical Response and Urgency NRS >=3 Point Improvement	CMH analysis with NRI	mITT—In patients with Urgency NRS ≥3 at AMAN baseline	Induction responder at Week 40
	Clinical Remission and Urgency Remission	CMH analysis with NRI	mITT—In patients with Urgency NRS ≥3 at AMAN baseline	Induction responder at Week 40
	Clinical Remission and Urgency NRS = 0 or 1 or 2	CMH analysis with NRI	mITT—In patients with Urgency NRS ≥3 at AMAN baseline	Induction responder at Week 40
	Symptomatic Remission and Histologic-Endoscopic Mucosal Remission	CMH analysis with NRI	mITT	Induction responder at Week 40
	Symptomatic Remission and Alternative Histologic-Endoscopic Mucosal Remission	CMH analysis with NRI	mITT	Induction responder at Week 40
Abdominal Pain NRS	>= 30% improvement from baseline	CMH analysis with NRI	mITT—In patients with an abdominal pain NRS score ≥3 at baseline	Mirikizumab induction responder, by visit through Week 40
PGRS	Change from baseline in PGRS Score	MMRM; ANCOVA with mBOCF	mITT	Mirikizumab induction responder through Week 40
Nocturnal Stool	Change from baseline in Nocturnal Stool Score	MMRM; ANCOVA with mBOCF	mITT—In patients with a nocturnal stool score ≥1 at baseline	Mirikizumab induction responder, by visit through Week 40
Fatigue NRS	Change from baseline in Fatigue NRS Score	MMRM; ANCOVA with mBOCF	mITT	Mirikizumab induction responder, by visit through Week 40
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.	Mirikizumab induction responder, by visit through Week 40

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Cohort, Time Point(s)[b]
PGRC	Mean PGRC Score	ANCOVA as observed (Baseline will not be included as a covariate in model)	mITT	Mirikizumab Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
IBDQ	Change from baseline in IBDQ Total Score and Subscores	ANCOVA with mBOCF	mITT	Mirikizumab Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	IBDQ Response	CMH analysis with NRI	mITT	Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	IBDQ Remission	CMH analysis with NRI	mITT	Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
EQ-5D-5L	Change from baseline of EQ-5D VAS	ANCOVA with mBOCF	mITT	Mirikizumab Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
SF-36	Change from baseline for Domain Scores and PCS and MCS Component Scores	ANCOVA with mBOCF	mITT	Mirikizumab Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 40[a]
	SF-36 PCS MCID Response and SF-36 MCS MCID Response	CMH analysis with NRI	mITT	Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 40[a]
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	Mirikizumab Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 40[a]
Histopathology	Primary Histologic Remission;	CMH analysis with NRI	mITT	Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Cohort, Time Point(s)[b]
	RHI <3	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
	Nancy Index <1	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
	Histologic Improvement	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
	Alternative Histologic Improvement	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
Histologic-Endoscopic	Histologic-Endoscopic Mucosal Improvement	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	Histologic-Endoscopic Mucosal Remission	CMH analysis with NRI	mITT; ITT; PP	Induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	Alternative Histologic-Endoscopic Mucosal Remission	CMH analysis with NRI	mITT	Mirikizumab induction responder at Week 40
CRP	Change from baseline in CRP	ANCOVA with mBOCF	mITT	Mirikizumab induction responder at Week 40
Fecal calprotectin	Change from baseline in fecal calprotectin	ANCOVA with mBOCF	mITT	Mirikizumab induction responder at Week 40
		Descriptive summary	mITT	Induction nonresponder at Week 12 (and Week 40[a])
	Percent change from baseline in fecal calprotectin	ANCOVA with mBOCF	mITT—In patients with fecal calprotectin > 250 µg/g	Mirikizumab induction responder at Week 40
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	Mirikizumab induction responder at Week 40

Measure	Variable	Analysis Method (Section 5.1.4)	Population (Section 5.1.1)	Cohort, Time Point(s)[b]
Corticosteroid dose (prednisone equivalent)	Change from induction baseline	ANCOVA with mBOCF	mITT—In patients taking corticosteroid with prednisone equivalent at induction baseline	Mirikizumab induction responder, weekly through Week 40

Abbreviations: ANCOVA = analysis of covariance; ATC = Anatomical Therapeutic Chemical; CMH = Cochran-Mantel-Haenszel; EIM = extraintestinal manifestations; EQ-5D-5L = European Quality of Life Questionnaire; ES = endoscopic subscore; IBDQ = Inflammatory Bowel Disease Questionnaire; ITT = intent-to-treat; mBOCF = modified baseline observation carried forward; MCID = minimal clinically important difference; MCS = Mental Component score; mITT = modified intent-to-treat; MMRM = mixed-effects model for repeated measures; MMS = modified Mayo Score; mNRI = modified nonresponder imputation; NRI = nonresponder imputation; NRS = Numeric Rating Scale; PGRC = Patient's Global Rating of Change; PGRS = Patient's Global Rating of Severity; PSC = Physical Component score; PP = per-protocol; RB = rectal bleeding; RHI = Robarts Histology index; SF = stool frequency; SF-36 = Short Form-36; UCEIS = Ulcerative Colitis Endoscopic Index of Severity; VAS = visual analog scale WPAI UC = Work Productivity and Activity Impairment Questionnaire Ulcerative Colitis.

[a] Among patients who are dosed in the extension maintenance study period.

[b] Unless otherwise specified, MMRM and ANCOVA analysis will be performed only for protocol-defined visits. Descriptive summary will include observed and NRI-imputed values for binary endpoints, and observed and mBOCF cases for continuous endpoints.

5.12.1. Primary Outcome and Methodology

Analysis of the primary endpoint (clinical remission) is described in Table AMBG.5.5. The primary endpoint analysis will utilize the CMH test (see Section 5.1.4) with NRI (see Section 5.3.1) for the mITT population (see Section 5.1.4).

5.12.2. Sensitivity Analyses of the Primary Outcome

As described in Table AMBG.5.5 and Table AMBG.5.6, the following analysis will be performed as sensitivity analysis for the primary endpoint:

- CMH test with NRI in the PP population
- common risk difference calculated with mNRI in the mITT population
- common risk difference calculated with mNRI in the ITT population; patient data impacted by the eCOA error will be imputed via multiple imputation in this approach, and
- logistic regression analysis with NRI in the mITT population.

5.12.3. Analyses of the Major Secondary Efficacy Outcomes

The analysis of the major secondary endpoints is described in Table AMBG.5.5 and Table AMBG.5.6. All of these endpoints except for the Urgency NRS endpoint are binary/categorical and the primary analysis of these endpoints will use the CMH test (see Section 5.1.4) with NRI (see Section 5.3.1) in the mITT population. The Urgency NRS endpoint will be analyzed using the MMRM analysis including the planned study visits in the mITT population as describe in Section 5.1.4). A multiple testing procedure will be utilized to control the FWER at the 0.05 significance level for the primary analysis of the primary endpoint and all major secondary endpoints.

5.12.4. Sensitivity Analysis of the Major Secondary Efficacy Outcomes

The following analysis will be performed as sensitivity analysis for the major secondary outcomes excluding the Urgency NRS endpoint:

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

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, and
- ANCOVA analysis with mBOCF in the mITT population.

5.12.5. Analyses of the Other Secondary Efficacy Outcomes

The analysis of the other secondary endpoints is described in Table AMBG.5.5 and Table AMBG.5.6.

5.12.6. Exploratory Efficacy Endpoints

The analysis of exploratory efficacy endpoints is described in Table AMBG.5.5 and Table AMBG.5.6.

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. Ulcerative colitis related hospitalizations may be determined from the related AE eCRF. Summary statistics will be reported for the number and percentage of patients by UC related hospitalizations 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.

Ulcerative colitis 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. Exploratory Association Analyses

Additional analyses will be conducted in mITT patients to assess the association of histologic-endoscopic efficacy endpoints and longer-term UC clinical outcomes at Week 40, such as Inflammatory Bowel Disease Questionnaire (IBDQ), Short Form-36 (SF-36), corticosteroid-free remission, and health care utilization.

5.15. Anchor-Based Analyses

Anchor-based analyses will be conducted to verify the threshold for the components of symptomatic remission. The Patient's Global Rating of Severity (PGRS) and Patient's Global Rating of Change (PGRC) will be used as anchor variables. Sensitivity, specificity, positive predictive value, negative predictive value, and Youden's index (Sensitivity + Specificity − 1) will be computed for a sequence of thresholds for the stool frequency and rectal bleeding subscores against the binary anchor variables at Week 40.

5.16. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

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 (such as, age, body weight, gender, anti-drug antibodies [ADAs], etc.) and extrinsic factors (such as, 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 40. Measures of exposure may include population PK estimated average concentrations (C_{avg}) between Week 0 and Week 40, or estimated or observed trough concentrations at Week 40. 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.

5.17. Safety Analyses

The planned analyses of safety data will be performed with an intent to maintain consistency with compound level safety standards (Program Safety Analysis Plan: Mirikizumab [LY3074828] or PSAP). 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 AMBG.5.1, the safety analysis population is defined as all randomized /assigned patients who received any amount of study treatment (regardless of if the patient does not receive the correct treatment, or otherwise does not follow the protocol). The cohorts and study periods of interest, are described in Table AMBG.5.2 and Table AMBG.5.3 respectively. Inferential comparisons will only be performed in the Maintenance Period on the Mirikizumab induction responder cohort (double blind 200mg miri Q4W SC vs. placebo Q4W SC treatment groups). 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.

Unless otherwise stated, summaries of safety collected during the follow-up periods will be analyzed in the Integrated Safety Analysis.

5.17.1. Extent of Exposure

Duration of exposure to study treatment will be summarized by treatment group for the safety population. For the study periods of interest, exposure will be calculated as the time period length in years (see Section 5.1.2) with start and end dates described in Table AMBG.5.3. Exposure will be calculated for the Maintenance Period (Induction responder), LOR Period (induction responder), OL Extended Induction Period (Induction nonresponder), and OL Maintenance Period (Induction nonresponder).

Total patient-years (PY) of exposure will be reported by treatment. Descriptive statistics 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 to <4 weeks, ≥4 weeks to <8 weeks, ≥8 weeks to < 12 weeks, ≥ 12 weeks to < 16 weeks, ≥ 16 weeks to < 24 weeks, ≥ 24 weeks to < 32 weeks, ≥ 32 weeks to 40 weeks.
- >0, ≥4 weeks, ≥8 weeks, ≥12 weeks, ≥16 weeks, ≥24 weeks, ≥32 weeks, ≥40 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.

5.17.2. 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 TE computation. The maximum

severity for each LLT during the baseline period will be used as baseline. The treatment period will be included as postbaseline 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 posttreatment. If start time for the AE is missing, it will be assumed to have started in the postbaseline period.

Summary tables as described in Table AMBG.5.7 will be presented. 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.

Postbaseline study periods are defined in Table AMBG.5.3. The baseline period will be defined as follows:

- For all events in induction responder in the M and induction nonresponder in EI study periods (and excluding those occurring after satisfying LOR criteria): The baseline events are those events which are ongoing at the start of the AMBG study (i.e., the baseline period is a moment in time).
- For all events in induction nonresponder patients in the EM study periods: The baseline events are those events which are ongoing at the start of the EM period (i.e., the baseline period is a moment in time).
- For events occurring in LOR cohort patients after having satisfied LOR criteria: the baseline events are those events which are ongoing at the time of first dosing with mirikizumab LOR IV treatment.

Adverse events that occur during follow-up will be reported in a listing.

Table AMBG.5.7. Summary Tables Related to Adverse Events

Analysis	Period[a] (Cohort)
Overview of AEs	M (Induction responder), EI/EM [b] (Induction nonresponder), LOR (Loss of response cohort)
Summary of TEAE PTs by decreasing frequency	M (Induction responder) EI/EM [b] (Induction nonresponder), LOR (Loss of response cohort)
Summary of TEAE PTs occurring in >=1% of patients by decreasing frequency	M (Induction responder)
Summary of TEAE PTs by decreasing frequency within SOC	M (Induction responder) , EI/EM [b] (Induction nonresponder), LOR (Loss of response cohort)
Summary of TEAE PTs by maximum severity by decreasing frequency within SOC	M (Induction responder)
Summary of SAE PTs by decreasing frequency within SOC	M (Induction responder), EI/EM [b] (Induction nonresponder),

	LOR (Loss of response cohort)
Summary of AEs leading to study treatment discontinuation	M (Induction responder), EI/EM [b] (Induction nonresponder), LOR (Loss of response cohort)
Listing of SAEs	All periods including follow-up, Enrolled Patients
Listing of AEs	All periods including follow-up, Enrolled Patients

Abbreviations: AE = Adverse Event; TEAE = Treatment-Emergent Adverse Event; SAE = Serious Adverse Event; PT = Preferred Term; SOC = System Organ Class.

[a] Population/Period are abbreviated as follows: M = Safety during Maintenance Period; EI = Safety during OL Extended Induction; EM = Safety during OL Maintenance; LOR = Loss of Response "-" = Not Applicable.

[b] among patients who are dosed in the extension maintenance study period.

5.17.2.1. 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.

5.17.3. 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 on the populations/periods defined in Table AMBG.5.7. 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 on the populations/periods defined in Table AMBG.5.7.

5.17.4. 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 described in Table AMBG.5.8:

Table AMBG.5.8. Summaries/Displays/Analysis for Clinical Laboratory Evaluations

Analysis	**Period[a] (Cohort(s))**
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.	M (Induction responder) EI/EM [b] (Induction nonresponder), LOR (Loss of response cohort)
Treatment-emergent abnormal high lab values (i.e., patients shifting from a normal/low maximum baseline value to a high maximum postbaseline	M (Induction responder), EI/EM [b] (Induction nonresponder),

value) or low lab values (i.e., patients shifting from normal/high minimum baseline value to a low minimum postbaseline value)	LOR (Loss of response cohort)
Scatter plot of maximum (minimum) postbaseline value vs. maximum (minimum) baseline value	M (Induction responder)
Shift tables showing the number of patients who shift from each category of maximum (minimum) baseline observation to each category of maximum (minimum) postbaseline observation. Here categories may be low, normal or high with cutoffs defined in the compound level safety standards.	M (Induction responder)

[a] Population/Period are abbreviated as follows: M = Safety during Maintenance Period; EI = Safety during OL Extended Induction; EM = Safety during OL Maintenance; LOR = Loss of Response (cohort 3 only) "-" = Not Applicable.

[b] among patients who are dosed in the extension maintenance study period.

For these displays, the postbaseline periods will be identical to those described in Table AMBG.5.3. Postbaseline measurement for continuous analysis (e.g., boxplots) will include *only* scheduled measurements, while postbaseline categorical analysis (e.g., shifts) will include *both* scheduled and unscheduled measurements.

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

- For induction responder in the M and induction nonresponder in EI study periods (and excluding those occurring after satisfying LOR criteria):
    - For analyses of continuous and categorical measurements: the last scheduled or unscheduled nonmissing measurement recorded during the AMAN trial (i.e., the baseline period only includes a single assessment).
- For induction nonresponders in the EM study periods:
    - For analyses of continuous and categorical measurements: the last scheduled or unscheduled nonmissing measurement recorded during the EI study period (i.e., the baseline period only includes a single assessment).
- For LOR cohort after having satisfied LOR criteria:
    - For analyses of continuous and categorical measurements: the last scheduled or unscheduled nonmissing measurement recorded prior to LOR dosing (i.e., the baseline period only includes a single assessment).

For any lab performed on the day of first taking study medication at the start of the postbaseline period, the start time of the study treatment will be used to determine whether the lab was pre- versus postbaseline. 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.

5.17.5. 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 5.17.4). 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.

5.17.6. Electrocardiograms

Complete electrocardiogram (ECG) data will not be part of the clinical database for this study. 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.

5.17.7. 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 ADA sample has a ≥4-fold increase in titers compared to baseline (if ADA were present at baseline) or has a titer ≥2-fold increase than the minimum required dilution of 1:10 (if no ADAs were present at baseline).

Due to the nature of how immunogenicity is determined and calculated, and because patients enrolled into Study AMBG from a prior trial, anti-drug antibodies cannot be characterized strictly within the context of Study AMBG. Immunogenicity for patients enrolled in Study AMBG is better characterized within the holistic context of all exposures across all trials. Therefore, immunogenicity for patients treated in Study AMBG will be provided in the Integrated Summary of Immunogenicity.

5.17.8. 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 TE 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 for the Maintenance (induction responder), LOR (induction responder), OL Extended Induction

(induction nonresponder), and OL Maintenance Periods (induction nonresponder) using the baseline and postbaseline definitions described in Section 5.17.2 and Section 5.17.4.

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.

5.17.8.1. 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 postbaseline 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 postbaseline value and for subsets based on various levels of baseline value.
- Plot of maximum postbaseline ALT vs. maximum postbaseline total bilirubin (entire safety population).
- A listing of the information collected on the hepatic-safety CRF.

5.17.8.2. 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:

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

5.17.8.3. 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., occurring on day of study drug administration) and Nonimmediate (i.e., occurring after the day of study drug administration) based on the timing of the reaction. More details will be included in the compound level safety standards.

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 Nonimmediate 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.

5.17.8.4. Infusion/Injection Site Reactions (ISR)

The evaluation of study drug related ISRs will be through the unsolicited reporting of ISR TEAEs and through the use of an Infusion or Injection Site Reaction Follow-up Form completed by the investigator for each ISR reported.

Infusion/Injection site reactions will be defined using the following MedDRA High Level Terms (HLT):

- Infusion site reaction, excluding certain PTs (e.g., those PTs related to joint).
- Injection site reaction, excluding certain PTs (e.g., those PTs related to joint).

Analyses will include:

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

5.17.8.5. 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 including major adverse 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.

5.17.8.6. 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.

5.17.8.7. 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) [Week 0] and the Quick Inventory of Depressive Symptomatology Self Report (QIDS-SR16) [Weeks 12, 24, 40, ETV].
Analyses will include:

- C-SSRS: Only a listing of the C-SSRS 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 postbaseline 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).

5.18. Subgroup Analyses

Subgroup analyses will be conducted for all primary and major secondary endpoints in the mITT Population. The subgroups to be analyzed are listed Table AMBG.5.4 along with the demographic characteristics. Additional subgroup analyses may include TE anti-mirikizumab antibody status. Some additional subgroup analyses may be performed to meet regulatory requirements in specific countries. Changes to subgroups analyses 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 described in Section 5.2. The subgroup-by-treatment interaction will be tested using the Firth correction (Firth 1993) at the significance level of 0.1.

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 40 Urgency NRS endpoint, MMRM analysis will be performed for selected 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 differences 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 (i.e., no inferential testing for that subgroup).

5.18.1. Safety Subgroup Analysis

Subgroup analysis for safety related endpoints will be performed within the context of the integrated safety analysis. No safety subgroup analysis will be performed specifically for this study unless there is a potentially relevant finding during the periodic study safety reviews.

5.19. 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.

5.20. Analysis for Hungary Addendum

Per the Study AMBG Protocol Addendum (11), Hungary patients who received blinded study treatment at Study AMBG Week 40 will have an extension period to receive blinded mirikizumab 200 mg SC or placebo until the Study AMBG Week 40 database lock (DBL). A by-patient listing of AEs, SAEs, and discontinuations due to AEs will be provided for events that occur in this period.

5.21. 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 PP 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 AMBG.3.1, the PP population is defined as all randomized patients who do not have IPDPPs. Mitigations approved under the COVID-19 addendum that 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 "AMBG 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 40 DBL.

The number and percentage of patients having IPDs will be summarized within category and subcategory of deviations by dosing regimen for the mITT population during the Maintenance, OL Extended Induction, and OL Maintenance Periods.

A by-patient listing of IPDs will be provided.

5.22. 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 40 DBL: An unblinded analysis will be performed after all patients have completed the Week 40 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 40. However, the study may be ongoing for the posttreatment follow-up period at the time of this DBL.

Final DBL: A final DBL will occur after the posttreatment follow-up period 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 40 DBL. Unblinding details will be provided in the unblinding plan.

Maximized extended enrollment (ME2) DBL: A ME2 DBL is planned in China after all patients in the ME2 cohort have completed all study procedures, including the posttreatment follow-up period, or have discontinued the trial. Details of this cohort and analysis considerations will be provided in the unblinding plan and China SAP Addendum.

5.23. 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

5.24. 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 et al. 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, Öst 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.

Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, Bastien C, Cahn V, Cadiot G, Diebold MD, et al. 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 et al. 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. *Stat Biopharm Res*. 2020;4;498-505.

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

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

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

Travis et al. 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.

8. Appendices

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]
Induction Baseline	AMAN Baseline diary data to be used	
AMAN Week 12	AMAN Visit 5 (Week 12) diary data to be used	
Week 1	Max(Week 0 Visit Date, Week 4 Visit Date – 28)	Week 4 Visit Date - 22
Week 2	Max(Week 0 Visit Date, Week 4 Visit Date – 21)	Week 4 Visit Date - 15
Week 3	Max(Week 0 Visit Date, Week 4 Visit Date – 14)	Week 4 Visit Date -8
Week 4 (V2)	Max(Week 0 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 (V3)	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 (V4)	Max(Week 8 Visit Date, Week 12 Visit Date – 7)	Week 12 Visit Date – 1
Week 13	Max(Week 12 Visit Date, Week 16 Visit Date – 28)	Week 16 Visit Date - 22
Week 14	Max(Week 12 Visit Date, Week 16 Visit Date – 21)	Week 16 Visit Date - 15
Week 15	Max(Week 12 Visit Date, Week 16 Visit Date – 14)	Week 16 Visit Date -8
Week 16 (V5)	Max(Week 12 Visit Date, Week 16 Visit Date – 7)	Week 16 Visit Date – 1
Week 17	Max(Week 16 Visit Date, Week 20 Visit Date – 28)	Week 20 Visit Date -22
Week 18	Max(Week 16 Visit Date, Week 20 Visit Date – 21)	Week 20 Visit Date -15
Week 19	Max(Week 16 Visit Date, Week 20 Visit Date – 14)	Week 20 Visit Date -8
Week 20 (V6)	Max(Week 16 Visit Date, Week 20 Visit Date – 7)	Week 20 Visit Date – 1
Week 21	Max(Week 20 Visit Date, Week 24 Visit Date – 28)	Week 24 Visit Date -22
Week 22	Max(Week 20 Visit Date, Week 24 Visit Date – 21)	Week 24 Visit Date -15
Week 23	Max(Week 20 Visit Date, Week 24 Visit Date – 14)	Week 24 Visit Date -8
Week 24 (V7)	Max(Week 20 Visit Date, Week 24 Visit Date – 7)	Week 24 Visit Date – 1
Week 25	Max(Week 24 Visit Date, Week 28 Visit Date – 28)	Week 28 Visit Date - 22

Week 26	Max(Week 24 Visit Date, Week 28 Visit Date – 21)	Week 28 Visit Date - 15
Week 27	Max(Week 24 Visit Date, Week 28 Visit Date – 14)	Week 28 Visit Date -8
Week 28 (V8)	Max(Week 24 Visit Date, Week 28 Visit Date – 7)	Week 28 Visit Date – 1
Week 29	Max(Week 28 Visit Date, Week 32 Visit Date – 28)	Week 32 Visit Date -22
Week 30	Max(Week 28 Visit Date, Week 32 Visit Date – 21)	Week 32 Visit Date -15
Week 31	Max(Week 28 Visit Date, Week 32 Visit Date – 14)	Week 32 Visit Date -8
Week 32 (V9)	Max(Week 28 Visit Date, Week 32 Visit Date – 7)	Week 32 Visit Date – 1
Week 33	Max(Week 32 Visit Date, Week 36 Visit Date – 28)	Week 36 Visit Date -22
Week 34	Max(Week 32 Visit Date, Week 36 Visit Date – 21)	Week 36 Visit Date -15
Week 35	Max(Week 32 Visit Date, Week 36 Visit Date – 14)	Week 36 Visit Date -8
Week 36 (V10)	Max(Week 32 Visit Date, Week 36 Visit Date – 7)	Week 36 Visit Date – 1
Week 37	Max(Week 36 Visit Date, Week 40 Visit Date – 28)	Week 40 Visit Date -22
Week 38	Max(Week 36 Visit Date, Week 40 Visit Date – 21)	Week 40 Visit Date -15
Week 39	Max(Week 36 Visit Date, Week 40 Visit Date – 14)	Week 40 Visit Date -8
Week 40 (V11)	Max(Week 36 Visit Date, Week 40 Visit Date – 7)	Week 40 Visit Date – 1

[a] If End Day < Start Day, do not assign specified visit week. Visit date will be calculated by selecting the first available date from the following list (i.e., first in list order): (1)date of earliest bowel preparation if bowel prep date is available, (2) date of endoscopy if endoscopy was performed, (3) date of treatment if treatment was given, (4) office visit date if available, or (5) date of visit center of the protocol-defined window for that visit (e.g., study Day 281 for V11). For patients who received their Week 40 endoscopy outside of the window from study Days 267 to 337 , the visit date will be calculated as study Day 281.

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-day 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.

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

Leo Document ID = f7573c31-ec70-4b9c-abca-bce89ec7e268

Approver: PPD

Approval Date & Time: 01-Oct-2021 16:45:31 GMT

Signature meaning: Approved