Statistical Analysis Plan: I6T-MC-AMAM v2

A Phase 3, Multicenter, Randomized, Double-Blind, Placebo- and Active-Controlled, Treat-Through Study to Evaluate the Efficacy and Safety of Mirikizumab in Patients with Moderately to Severely Active Crohn's Disease

NCT03926130

Approval Date: 23 Aug 2023

Title Page

Protocol Title: A Phase 3, Multicenter, Randomized, Double-Blind, Placebo- and Active-Controlled, Treat-Through Study to Evaluate the Efficacy and Safety of Mirikizumab in Patients with Moderately to Severely Active Crohn's Disease

Protocol Number: I6T-MC-AMAM

Compound Number: LY3074828

Short Title: VIVID-1

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana USA 46285

Regulatory Agency Identifier Number(s)

Registry	ID
IND	30052
EudraCT	2018-004614-18

Approval Date: Statistical Analysis Plan Electronically Signed and Approved by Lilly on date provided below.

Table of Contents

Abbreviations and Definitions

Term	Definition
AC	adjudication committee
ADA	antidrug antibody
AE	adverse event
AESI	adverse event of special interest
ALT	alanine aminotransferase
ALP	alkaline phosphatase
ANCOVA	analysis of covariance
AP	abdominal pain
AST	aspartate transaminase
CD	Crohn's disease
CDAI	Crohn's Disease Activity Index
CDISC	Clinical Data Interchange Standards Consortium
CI	confidence interval
CMH	Cochran–Mantel–Haenszel
COVID-19	coronavirus disease 2019
CSR	clinical study report
CTR	Clinical Trial Registry
DMC	Data Monitoring Committee
ECG	Electrocardiogram
EIM	extraintestinal manifestation
eCRF	electronic case report form
EQ-5D-5L	European Quality of Life 5–Dimension 5 Level
ETV	early termination visit
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy–Fatigue
FWER	family-wise error rate

Term	Definition
HLT	High-Level Term
IBD-DI	Inflammatory Bowel Disease-Disability Index
IBDQ	Inflammatory Bowel Disease Questionnaire
ICE	intercurrent event
ISR	injection-site reaction
ITT	intent-to-treat
IV	intravenous
IWRS	interactive web-response system
LLT	Lowest Level Term
LS	least squares
mBOCF	modified baseline observation carried forward
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
mITT	modified intent-to-treat
MMRM	mixed effects model of repeated measures
mNRI	modified non-responder imputation
NAb	neutralizing anti-drug antibodies
NI	non-inferiority
NMSC	non-melanoma skin cancer
NR	non-responder
NRI	non-responder imputation
NRS	numeric rating scale
OI	opportunistic infections
PAS	Primary Analysis Set
PD	Pharmacodynamic

Term	Definition
PGIC	Patient Global Impression of Change
PGRS	Patient Global Rating of Severity
PhUSE	Pharmaceutical Users Software Exchange
PK	pharmacokinetic
PRO	patient-reported outcome
PT	Preferred Term
PY	participant years
Q4W	every 4 weeks
Q8W	every 8 weeks
QIDS-SR16	Quick Inventory of Depressive Symptomatology – Self Report
SAC	statistical analysis center
SAE	serious adverse event
SAP	Statistical Analysis Plan
SC	subcutaneous
SES-CD	Simple Endoscopic Score for Crohn's Disease
SF	stool frequency
SF-36	36-Item Short Form Health Survey
SMQ	Standardized MedDRA Query
SOC	System Organ Class
TBL	total bilirubin
TE	treatment-emergent
TE ADA	treatment-emergent anti-drug antibodies
TEAE	treatment-emergent adverse event
TFL	table, figure, and listing
ULN	upper limit of normal
WPAI-CD	Work Productivity and Activity Impairment Questionnaire Crohn's Disease

Version History

This SAP for Study I6T-MC-AMAM version 2 is based on the Protocol Amendment (e) dated 23 February 2022 and approved prior to unblinding.

SAP Version	Approval Date	Change	Rationale
1	26 Nov 2019	Not Applicable	Original version
2		Section 1.1 Objectives and Endpoints Updated and clarified the estimand language	Updated based on Protocol Amendment (e) on the changes of analysis population, primary and major secondary endpoints Clarified to be aligned with ICH guidance
		Sections 1.1 Objectives and Endpoints Updated objectives and endpoints	Updated based on Protocol Amendment (e) to address FDA feedback and to align with business needs
		Section 1.2.1 Study Conduct During Exceptional Circumstances Added this section.	Explained the changes in study conduct for COVID-19 pandemic or crisis caused by Russia-Ukraine war
		Section 2 Statistical Hypotheses Updated the co-primary and major secondary null hypotheses	Updated based on the co-primary and major secondary endpoints in Section 1.1
		Section 2.1 Multiplicity Adjustment Clarified Group 2 will proceed when the comparisons on both co-primary endpoints are met	Clarification
		Section 2.1 Multiplicity Adjustment	Updated based on the co-primary and major

SAP Version	Approval Date	Change	Rationale
		Updated the figure of graphical testing with the updated co-primary and major secondary endpoints	secondary endpoints in Section 1.1
		Section 3 Sample Size Determination Revised estimated power for the co-primary endpoints	Updated based on Protocol Amendment (e) to address FDA feedback
		Section 4 Analysis Sets Added the Primary Analysis Set for primary efficacy analysis Changed Modified Intent-to-Treat population for sensitivity analysis	Updated based on Protocol Amendment (e)
		Section 4 Analysis Sets Added sensitivity analysis population by excluding participants impacted by crisis	Address FDA feedback
		Section 5.1.1 Analysis Methods Selected stratification factors which will be in the analysis	To align with Protocol Amendment (e)
		Section 5.1.1 Analysis Methods Added the statement that the MMRM method will only be applied to variables with multiple measurements in each study period as the sensitivity analysis and the ANCOVA method will be used as the primary analysis method	Clarification The ANCOVA methodology was deemed to be better suited to the return to baseline approach which is incorporated into the estimand
		Section 5.1.2 Definition of Baseline Updated the baseline definition	Clarification
		Section 5.1.2 Definition of Study Intervention by Study Period Added Primary Analysis Set	Updated based on Protocol Amendment (e)

SAP Version	Approval Date	Change	Rationale
		Section 5.1.5.1 Non-Responder Imputation Added participants in placebo group begin mirikizumab at Week 12 as non-responders for all visits subsequent to Week 12	Clarification
		Section 5.1.5.1 Non-Responder Imputation Clarified the composite estimand strategy and hybrid estimand strategy	Clarification
		Section 5.1.5.2 Mixed-Effects Model for Repeated Measures Revised intercurrent events are in Section 5.1.5 Changed to the hybrid estimand strategy	Clarification and editorial consistency Updated based on estimands for continuous endpoints in Section 1.1
		Section 5.1.5.3 Modified Baseline Observation Carried Forward (mBOCF) Clarified the conditions Changed to the hybrid estimand strategy and modified the method of mBOCF	Clarification and editorial consistency Updated based on estimands for continuous endpoints in Section 1.1
		Section 5.1.5.4 Modified Nonresponder Imputation (mNRI) Added section to address intercurrent events due to COVID-19 pandemic, the crisis, sporadically missing and discontinuation not related to efficacy and safety	Added Sensitivity analysis
		Section 5.1.5.5 Tipping Point Analysis for Co-primary endpoints Clarified the missing data handling	Clarification

SAP Version	Approval Date	Change	Rationale
		Section 5.1.5.6 Tipping Point Analysis for selected endpoints at Week 52 for placebo comparison Added a tipping point analysis	Address FDA feedback
		Section 5.2 Participant Dispositions Added the analysis for mITT and PAS population	Updated based on Protocol Amendment (e)
		Section 5.3.1 Definition of Endpoints Revised co-primary endpoints	Updated based on Protocol Amendment (e) to address FDA feedback
		Section 5.3.1 Definition of Endpoints Added details of CDAI total score calculation	Clarification
		Section 5.3.2 Main analytical approach added the primary estimand	Address FDA feedback
		Section 5.3.2 Main Analytical Approach Revised the adjustment based on the selected stratification factors in the analysis	Clarification and editorial consistency
		Section 5.3.3 Sensitivity Analysis Revised the analysis populations Included tipping point analysis	Clarification and editorial consistency
		Section 5.3.4 Supplementary Analysis Section added for two supplementary analyses	Address FDA feedback
		Section 5.4.1.2 Main Analytical Approach and Sensitivity Analyses Added the estimands for major secondary endpoints in comparison to placebo	Clarifications Updated to align with the new graphical testing scheme

SAP Version	Approval Date	Change	Rationale
		Added the jump to reference sensitivity analysis method	
		Section 5.4.2 Other Secondary Endpoints and Exploratory Endpoints Added section for other exploratory endpoints Added supplementary analysis	Clarifications Provide extra information on endpoints of interest
		Section 5.5 Secondary Endpoints Analysis in Comparison to Ustekinumab Added the estimand	Clarifications
		Section 5.5.1.2 Additional Analysis on the Non-Inferiority Testing Added analyses to verify assumptions of ustekinumab clinical remission response rate Added to calculate the levels of 2-sided confidence intervals that exclude other non-inferiority margins.	Address FDA feedback
		Section 5.6 Safety Analyses Clarified the analysis for the study treatment period	Clarification
		Section 5.6.2 Adverse Events Added exposure adjusted incidence rates analysis Updated analysis table to add periods	Clarification and to align with compound level safety standards
		Section 5.6.7 Special Safety Topics Clarified the compound level safety standards will supersede the SAP	Clarification
		Section 5.6.8 Safety Subgroup Analysis Specified the safety subgroup analysis	Align with the compound level safety standards

SAP Version	Approval Date	Change	Rationale
		Section 5.7.3 Analysis for Japan Submission Added section for Japan submission planning	Clarification
		Section 5.7.4 Analysis for China Submission Added section for China submission planning	Clarification
		Section 5.7.5 Protocol Deviations Added section to clarify protocol deviation and important protocol deviation and analysis plan	Clarification
		Section 5.7.6 Trial Impact by COVID-19 Pandemic, and Russia/Ukraine Crisis Added specific listings for participants impacted by COVID-19 and crisis	New section to address exceptional circumstances
		Section 5.8.2 PK/PD Model Development Added section to clarify the unblinding data prior to the Week 52 primary endpoint database lock for PK/PD analysis as per the blinding and unblinding plan	Clarification
		Section 5.8.3 Week 52 Database Lock Revised language to clarify the Week 52 primary endpoint database lock	Clarification
		Appendix 1 (Section 6.1) Description and Derivation of Efficacy and Health Outcome Endpoints Modified derivation on the co-primary, major secondary, other secondary, exploratory endpoint	Updated based on Protocol Amendment (e) to address FDA feedback Clarification

SAP Version	Approval Date	Change	Rationale
		Appendix 2 (Section 6.2) Description of Analyses Updated the analysis methods on co-primary, major secondary, other secondary, exploratory endpoints per Sections 5.1.1, 5.3, 5.4, and 5.5	Updated based on Protocol Amendment (e) to address FDA feedback Made additional changes to major secondary and other, and sensitivity analyses
		Appendix 5 (Section 6.5) Study Intervention Compliance Clarified the derivation of the treatment compliance	Clarification
		Appendix 9 (Section 6.9) Study Visit or Week Definition for Daily Diary Added study visit definition for measurements collected from daily diary	Clarification
		Appendix 10 (Section 6.10) Specified Changes in Concomitant CD Medications (Intercurrent Event) Defined the specified changes in concomitant CD medications as intercurrent events	Clarification
		Appendix 11 (Section 6.11) IBD-DI Scoring Added the IBD-DI Scoring rule	Clarification
		Made other minor typographical corrections and clarifications without affecting content in the document	Clarification

1. Introduction

This SAP includes the analysis plan for efficacy, safety, biomarkers, and immunogenicity data.

Exploratory endpoints will be documented in supplemental SAPs.

The TFL specifications are contained in a separate document.

1.1. Objectives and Endpoints

Estimands for the co-primary and major secondary endpoints are defined as follows:

- Population: PAS (defined in Section 4)
- Strategies for ICE handling are specified below:
    - For binary endpoints, unless otherwise specified, a composite strategy is used where ICEs are included in the endpoint definition. Successful response only if:
        - response criteria from endpoint table met, and
        - no study intervention discontinuation prior to time point of interest, and
        - do not meet any of the specified changes in the concomitant CD medication prior to time point of interest (defined in Appendix 10 [Section 6.10]).
    - For the 2 major secondary binary endpoints, proportion of participants achieving endoscopic response at Week 52 and proportion of participants achieving clinical remission by CDAI at Week 52, a hybrid strategy is used to accommodate the additional ICE where participants in the placebo group switch to mirikizumab at Week 12. For this ICE a hypothetical scenario is envisaged in which these participants remained on placebo for the rest of the study and measurements after this ICE will be imputed (see details in Section 5.1.5.1 where non-responder imputation is described). For all other ICEs the composite strategy above will be used.
    - For continuous endpoints, a hybrid strategy is used. For ICEs of study intervention discontinuation and specified changes in the concomitant CD medication, the composite strategy will be used such that measurements after the ICEs will return to baseline. For the additional ICE where participants in the placebo group switch to mirikizumab at Week 12, a hypothetical scenario is envisaged in which these participants remained on placebo for the rest of the study and measurements after this ICE will be imputed (see details in Section 5.1.5.2).
- Population-level summary: The common risk difference will be used for binary endpoints and the LS mean difference will be presented for continuous endpoints.

Objectives	Endpoints
Co-primary	
To evaluate the efficacy of mirikizumab is superior to placebo as assessed by • clinical response by PRO at Week 12 and endoscopic response at Week 52 • clinical response by PRO at Week 12 and clinical remission by CDAI at Week 52	• Proportion of participants achieving clinical response by PRO[c] at Week 12 and endoscopic response[d] at Week 52 • Proportion of participants achieving clinical response by PRO[c] at Week 12 and clinical remission by CDAI[e] at Week 52
Major Secondary[a,b]	
To evaluate the efficacy of mirikizumab is superior to placebo at Week 52 as assessed by • endoscopic response • clinical remission by CDAI	• Proportion of participants achieving endoscopic response[d] at Week 52 • Proportion of participants achieving clinical remission by CDAI[e] at Week 52
To evaluate the efficacy of mirikizumab is superior to placebo at Week 12 as assessed by • endoscopic response • endoscopic remission • clinical response by PRO • clinical remission by CDAI • FACIT-Fatigue scores	• Proportion of participants achieving endoscopic response[d] at Week 12 • Proportion of participants achieving endoscopic remission SES-CD ≤ 4[i] at Week 12 • Proportion of participants achieving clinical response by PRO[c] at Week 12 • Proportion of participants achieving clinical remission by CDAI[e] at Week 12 • Change from baseline in FACIT-Fatigue scores at Week 12

Objectives	Endpoints
To evaluate the efficacy of mirikizumab is superior to placebo as assessed by both clinical response by PRO at Week 12 and each below, individually: • clinical remission by PRO at Week 52 • endoscopic remission at Week 52 • corticosteroid-free clinical remission by CDAI	Proportion of participants achieving clinical response by PRO[c] at Week 12 and each below, individually: • Clinical remission by PRO[f] at Week 52 • Endoscopic remission SES-CD ≤4[i] at Week 52 • Corticosteroid-free from Week 40 to Week 52 and clinical remission by CDAI[e] at Week 52
To evaluate the efficacy of mirikizumab in comparison to ustekinumab at Week 52 as assessed by • endoscopic response (superior) • clinical remission by CDAI (non-inferior)	Proportion of participants achieving • Endoscopic response[d] at Week 52 • Clinical remission by CDAI[e] at Week 52
Other Secondary	
To evaluate the efficacy of mirikizumab is superior to placebo at Week 12 as assessed by • clinical remission by PRO • clinical response by CDAI • endoscopic remission SES-CD Total Score 0-2 • endoscopic response and clinical response by CDAI • endoscopic response and clinical remission by CDAI • Urgency NRS less than or equal to 2 in participants with baseline Urgency NRS≥3	Proportion of participants achieving • Clinical remission by PRO[f] at Week 12 • Clinical response by CDAI[g] at Week 12 • Endoscopic remission[h] at Week 12 • Endoscopic response[d] and clinical response by CDAI[g] at Week 12 • Endoscopic response[d] and clinical remission by CDAI[e] at Week 12 • Urgency NRS less than or equal to 2 at Week 12 in participants with baseline Urgency NRS≥3
To evaluate the efficacy of mirikizumab is superior to placebo as assessed by both clinical response by PRO at Week 12 and each below, individually: • Clinical response by CDAI at Week 52 • Clinical response by PRO at Week 52	Proportion of participants achieving clinical response by PRO[c] at Week 12 and each below, individually: • Clinical response by CDAI[g] at Week 52 • Clinical response by PRO[c] at Week 52

Objectives	Endpoints
• Endoscopic remission SES-CD Total Score 0-2 at Week 52 • Stability of clinical remission by CDAI from Week 12 to Week 52 • Durability of endoscopic response at Week 12 and Week 52 • Durability of endoscopic remission at Week 12 and Week 52 • Endoscopic remission and clinical remission by CDAI at Week 52 • Endoscopic response and clinical remission by CDAI at Week 52 • Corticosteroid-free clinical remission by CDAI among participants who used corticosteroids at baseline • Urgency NRS less than or equal to 2 at Week 52 in participants with baseline Urgency NRS ≥3	• Endoscopic remission[h] at Week 52 • Stability of clinical remission by CDAI[e] from Week 12 to Week 52 • Durability of endoscopic response[d] at Week 12 and Week 52 • Durability of endoscopic remission[i] at Week 12 and Week 52 • Endoscopic remission[i] and clinical remission by CDAI[e] at Week 52 • Endoscopic response[d] and clinical remission by CDAI[e] at Week 52 • Corticosteroid-free from Week 40 to Week 52 and clinical remission by CDAI[e] at Week 52 in participants who used corticosteroids at baseline • Urgency NRS less than or equal to 2 at Week 52 in participants with baseline Urgency NRS ≥3.
To evaluate the efficacy of mirikizumab is superior to placebo at Week 52 as assessed by Urgency NRS	Change from baseline in Urgency NRS at Week 52
To evaluate the efficacy of mirikizumab is superior to placebo in clinical response by CDAI at Week 4	Proportion of participants achieving clinical response by CDAI[g] at Week 4
To evaluate the efficacy of mirikizumab is superior to placebo in mITT population as assessed by • clinical response by PRO at Week 12 and endoscopic response at Week 52 • clinical response by PRO at Week 12 and clinical remission by CDAI at Week 52	• Proportion of participants achieving clinical response by PRO[c] at Week 12 and endoscopic response[d] at Week 52 • Proportion of participants achieving clinical response by PRO[c] at Week 12 and clinical remission by CDAI[e] at Week 52

Objectives	Endpoints
To evaluate the efficacy of mirikizumab is superior to placebo in not-biologic-failed and biologic-failed subgroups	Proportion of participants achieving • Endoscopic response[d] at Week 12 • Clinical remission by CDAI[e] at Week 12 Proportion of participants achieving clinical response by PRO[c] at Week 12 and each below at Week 52, individually: • Endoscopic response[d] • Endoscopic remission SES-CD ≤4[i] • Clinical remission by CDAI[e]
To evaluate the efficacy of mirikizumab in comparison to placebo in health outcomes and quality of life measures, symptomatic endpoints, inflammatory biomarkers	Proportion of participants achieving each below over time • Clinical remission by CDAI[e] • Clinical response by CDAI[g] • Clinical remission by PRO[f] • Clinical response by PRO[c] Change from baseline at Week 12 and Week 52 of each below: • C-reactive protein • Fecal calprotectin • FACIT-Fatigue scores (Week 52 only) • EQ-5D-5L index • WPAI-CD score • Medical Outcomes SF-36 Version 2 acute scores • IBDQ
To evaluate the efficacy of mirikizumab in comparison to placebo for other assessments	Proportion of participants • had no EIMs among those who had EIMs at baseline • Crohn's-related emergency room visits • Crohn's-related hospitalization • Crohn's-related surgeries Proportion of participants achieving clinical response by PRO[c] at Week 12

Objectives	Endpoints
	and each below evaluated at Week 24, and at Week 52, individually: • ≥50% reduction from baseline in the number of draining cutaneous fistulae • Closure of all draining cutaneous fistulae in patients who had any draining cutaneous fistulae at baseline
To evaluate the efficacy of mirikizumab in comparison to ustekinumab as assessed by • Endoscopic response at Week 12 (superior) • Endoscopic remission at Week 52 (superior) • Clinical remission by CDAI at Week 12 (non-inferior) • Clinical response by CDAI at Week 12 (non-inferior) • Clinical response by CDAI at Week 52 (non-inferior) • Corticosteroid-free clinical remission by CDAI at Week 52 (non-inferior) • Clinical response by PRO at Week 12 • Clinical response by PRO at Week 52 • Clinical remission by PRO at Week 12 • Clinical remission by PRO at Week 52	Proportion of participants achieving • Endoscopic response[d] at Week 12 • Endoscopic remission SES-CD ≤4[j] at Week 52 • Clinical remission by CDAI[e] at Week 12 • Clinical response by CDAI[g] at Week 12 • Clinical response by CDAI[g] at Week 52 • Corticosteroid-free clinical remission by CDAI[e] at Week 52 • Clinical response by PRO[c] at Week 12 • Clinical response by PRO[c] at Week 52 • Clinical remission by PRO[f] at Week 12 • Clinical remission by PRO[f] at Week 52
To evaluate the efficacy of mirikizumab in comparison to ustekinumab in not-biologic-failed and biologic-failed subgroups	Proportion of participants achieving • Endoscopic response[d] at Week 52 • Endoscopic remission SES-CD ≤4[i] at Week 52 • Clinical remission by CDAI[e] at Week 52 • Endoscopic response[d] at Week 12 • Clinical remission by CDAI[e] at Week 12

Objectives	Endpoints
To evaluate the pharmacokinetic and pharmacokinetic/pharmacodynamic relationships of mirikizumab	• Clearance and volume of distribution of mirikizumab • Relationship between mirikizumab exposure and efficacy
Tertiary/Exploratory	
CCI	

Abbreviations: AP = abdominal pain; CDAI=Crohn's Disease Activity Index; EIM = extraintestinal manifestation; EQ-5D-5L = European Quality of Life 5–Dimension 5 Level; FACIT-Fatigue = Functional Assessment of Chronic Illness Therapy-Fatigue; IBDQ = Inflammatory Bowel Disease Questionnaire; ITT = intent-to-treat; mITT = modified intent-to-treat; NRS = numeric rating scale; PAS = Primary Analysis Set; PRO = patient-reported outcome; SES-CD = Simple Endoscopic Score for Crohn's Disease; SF = stool frequency as per Bristol Stool Scale Category 6 or 7; SF-36 = Medical Outcomes Study 36-Item Short Form Health Survey; WPAI-CD = Work Productivity and Activity Impairment Questionnaire Crohn's Disease.

Note: mITT population is defined as all participants from ITT population who take at least one dose of study drug. PAS is defined as all participants from mITT population who have baseline SES-CD ≥7 (or ≥4 for isolated ileal disease)

[a] 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. A subset of these endpoints will be controlled at a 2-sided alpha level of 0.005 as described in Section 2.1. The graphical multiple testing procedure described in Bretz et al. (2009, 2011) will be used.

[b] The order of testing of the major secondary endpoints is determined from the result of the statistical simulation and is provided in Section 2.1.

[c] Clinical response by PRO is defined as at least a 30% decrease in SF and/or AP with neither score worse than baseline.

[d] Endoscopic response is defined as ≥50% reduction from baseline in SES-CD Total Score.

[e] Clinical remission by CDAI is defined as CDAI total score <150.

[f] Clinical remission by PRO is defined as SF≤3 and not worse than baseline (as per Bristol Stool Scale Category 6 or 7) and AP ≤1 and no worse than baseline.

[g] Clinical response by CDAI is defined as a decrease from baseline ≥100 and/or CDAI <150.

[h] Endoscopic remission is defined as SES-CD Total Score ≤2.

[i] Endoscopic remission SES-CD ≤4 is defined as SES-CD Total Score ≤4 and at least a 2-point reduction from baseline and no subscore >1.

1.2. Study Design

Study AMAM is a Phase 3, multicenter, randomized, double-blind, double-dummy, parallel group, active- and placebo-controlled, treat-through design (see schema below) clinical trial in participants with moderately-to-severely active CD.

Three intervention groups in the first period and 4 intervention groups in the second period will be studied in participants with moderate-to-severe CD:

- Mirikizumab 900 mg IV Q4W for 3 doses, then 300 mg SC Q4W
- Ustekinumab ~6 mg/kg IV for one dose, then 90 mg SC Q8W
- Placebo
    - When Period 1 concludes (Week 12), responders continue receiving placebo, and
    - Non-responders at Week 12 will receive mirikizumab as described above.

The total duration of study intervention with the investigational product is 52 weeks.

The maximum total duration of study participation for each participant, including screening and the post-treatment follow-up period, is 73 weeks. Participants who complete Study AMAM through Visit 17 will be given the option to enroll into the long-term extension study (I6T-MC-AMAX) if they are eligible. Participants who do not meet enrollment criteria for Study AMAX or who choose to not participate in Study AMAX will return for 2 post-treatment follow-up visits in Study AMAM (Visit 801 and Visit 802).

Participants who meet all criteria in Study AMAM for enrollment will be randomized at Visit 2 to receive either mirikizumab, ustekinumab, or placebo using a 6:3:2 randomization ratio. Assignment to study intervention groups will be determined by a computer-generated random sequence using an IWRS. To achieve between-group comparability, participants will be stratified to study intervention groups based upon these factors: a) biologic-failed status (yes/no), b) baseline corticosteroid use (yes/no), c) baseline SES-CD total score (<12, ≥12), d) region (North America/Europe/Other), and e) either baseline SF ≥7 and/or baseline AP ≥2.5 (yes/no).

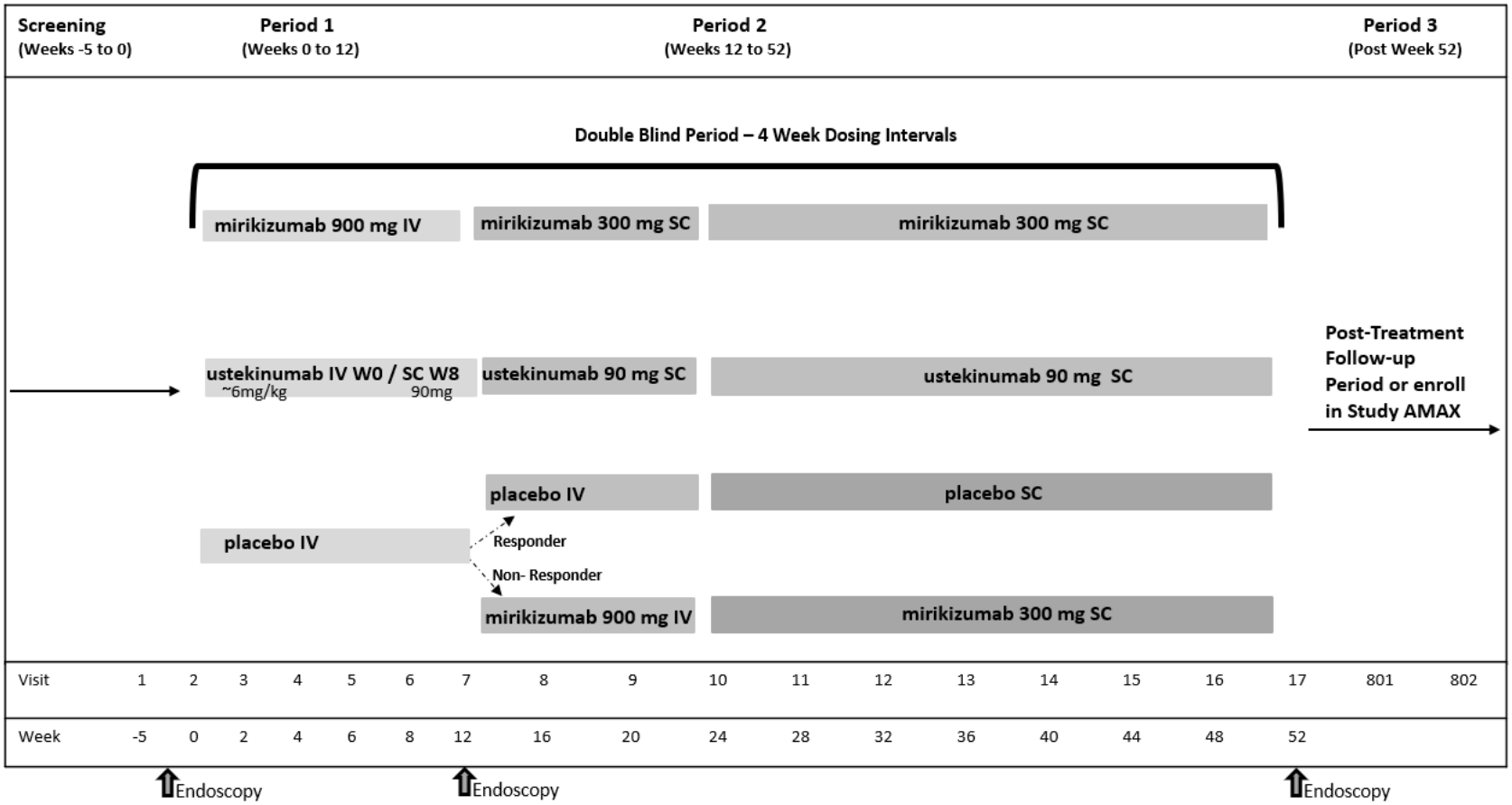
This stratification will be controlled by IWRS.

Participants in either active group will receive placebo to match the other active group using a double-dummy design. Participants in the placebo group receive both double-dummy placebo administrations.

Study intervention may be permanently discontinued or temporarily withheld during the study (see Sections 7.1.1 and 7.1.2 of the AMAM Protocol). Participants who permanently discontinue study drug early will undergo early termination procedures, which include an ETV and post-treatment follow-up visits (Visit 801 and Visit 802).

No rescue medication, other than moving placebo participants who do not have a clinical response by PRO at Week 12 to mirikizumab, is allowed during the study.

Participants who achieve clinical response by PRO and who are currently on corticosteroids will initiate corticosteroid tapering at or after Week 12, as described in the protocol (see Section 6.5.1 of the AMAM Protocol).



Abbreviations: IV = intravenous; SC = subcutaneous.

Note: From Week 8 through Week 20, all participants receive their assigned study intervention and matching placebo via both IV and SC administration.

1.2.1. Study Conduct During Exceptional Circumstances

Protocol Appendix 12: Provisions for Changes in Study Conduct During Exceptional Circumstances was added in Protocol Amendment (b). The changes to procedures are temporary measures intended to be used only during specific time periods as directed by the sponsor in partnership with the investigator. Exceptional circumstances are rare events that may cause disruptions to the conduct of the study. Examples include pandemics, natural disasters, or war. These disruptions may limit the ability of the investigators, participants, or both to attend on-site visits or to conduct planned study procedures.

Mitigations based on Protocol Appendix 12 are implemented for global COVID-19 pandemic and the crisis caused by the Russia and Ukraine war.

Study AMAM has been ongoing during the global COVID-19 pandemic, which resulted in some participants being unable or unwilling (e.g. fear of COVID-19 infection) to attend onsite clinical visits and have study procedures performed (see Section 5.1.5.4 for missing data handling). Addendum 8.2 has been approved by Eli Lilly and Company (Lilly) for China to allow local hematocrit test results from the hematology panel for analysis in the absence of central laboratory testing. Local regulatory review of the addendum is ongoing. Specific trial impacts by COVID-19 pandemic will be summarized (Section 5.7.6).

Study AMAM enrolled participants in Russia and Ukraine. During the crisis, the regular study conduct had been minimally impacted. Although all investigator sites are open and continue to conduct participant visits, it may be difficult or unsafe for participants to travel to those sites. Investigator sites in Russia and Ukraine that had not randomized participants have been closed. Study conduct, including data verification, is proceeding as usual for both countries.

Sites in Russia have had intermittent interruption of regular central laboratory testing. Addendum 21b has been approved to allow local hematocrit test results from the hematology panel for analysis in the absence of central laboratory testing.

Participants in Ukraine have reported delays and missed study visits. Central laboratory testing process has been impacted and local laboratory testing has been implemented. Addendum 20 has been approved to allow local hematocrit test results for analysis in absence of central laboratory testing.

In total, 174 participants from Russia and Ukraine randomized into this study. There were 70 participants (39 participants in Russia and 31 participants in Ukraine) that completed Week 52 or ETV before onset of the crisis (as of 24 Feb 2022) and with 100% data source verified. These 70 participants were determined to not be impacted, and therefore they are included in PAS population and relevant sensitivity analyses. Following FDA recommendation, all other participants from Russia and Ukraine who were ongoing in the study during the crisis (defined as randomized to or continuing in the study on or after 24 Feb 2022) are considered impacted by crisis and therefore they will be excluded in a sensitivity analysis (Section 4) and will be included in the PAS population (Section 4). Specific trial impacts by Russia/Ukraine crisis will be summarized (Section 5.7.6).

2. Statistical Hypotheses

The following is a list of primary and major secondary endpoints to be tested. The subscript for **H** denotes study intervention arms in the comparisons (m = mirikizumab, u = ustekinumab, and p = placebo), the numerical identifier of the endpoint within the comparison, and the type of hypothesis (0 for null, 1 for alternative), respectively.

Co-Primary Null Hypotheses:

- $\mathbf{H_{mp,1,0}}$: Proportion of mirikizumab participants achieving clinical response by PRO at Week 12 and endoscopic response at Week 52 is less than or equal to the proportion of placebo participants achieving clinical response by PRO at Week 12 and endoscopic response at Week 52
- $\mathbf{H_{mp,2,0}}$: Proportion of mirikizumab participants achieving clinical response by PRO at Week 12 and clinical remission by CDAI at Week 52 is less than or equal to the proportion of placebo participants achieving clinical response by PRO at Week 12 and clinical remission by CDAI at Week 52

Major Secondary Null Hypotheses:

- $\mathbf{H_{mp,3,0}}$: Proportion of mirikizumab participants achieving endoscopic response at Week 52 is less than or equal to the proportion of placebo participants achieving endoscopic response at Week 52
- $\mathbf{H_{mp,4,0}}$: Proportion of mirikizumab participants achieving clinical remission by CDAI at Week 52 is less than or equal to the proportion of placebo participants achieving clinical remission by CDAI at Week 52
- $\mathbf{H_{mp,5,0}}$: Proportion of mirikizumab participants achieving endoscopic response at Week 12 is less than or equal to the proportion of placebo participants achieving endoscopic response at Week 12
- $\mathbf{H_{mp,6,0}}$: Proportion of mirikizumab participants achieving endoscopic remission SES-CD $\leq$4 at Week 12 is less than or equal to the proportion of placebo participants achieving endoscopic remission SES-CD $\leq$4 at Week 12
- $\mathbf{H_{mp,7,0}}$: Proportion of mirikizumab participants achieving clinical remission by CDAI at Week 12 is less than or equal to the proportion of placebo participants achieving clinical remission by CDAI at Week 12
- $\mathbf{H_{mp,8,0}}$: The change from baseline to Week 12 in FACIT-Fatigue scores of mirikizumab participants is less than or equal to the change from baseline to Week 12 in FACIT-Fatigue scores of placebo participants
- $\mathbf{H_{mp,9,0}}$: Proportion of mirikizumab participants achieving clinical response by PRO at Week 12 and clinical remission by PRO at Week 52 is less than or equal to the proportion of placebo participants achieving clinical response by PRO at Week 12 and clinical remission by PRO at Week 52
- $\mathbf{H_{mp,10,0}}$: Proportion of mirikizumab participants achieving clinical response by PRO at Week 12 and endoscopic remission SES-CD $\leq$4 at Week 52 is less than or equal to the proportion of placebo participants achieving clinical response by PRO at Week 12 and endoscopic remission SES-CD $\leq$4 at Week 52
- $\mathbf{H_{mp,11,0}}$: Proportion of mirikizumab participants achieving clinical response by PRO at Week 12 and corticosteroid-free clinical remission by CDAI at Week 52 (for participants

who were steroid free from Week 40 to Week 52) is less than or equal to the proportion of placebo participants achieving clinical response by PRO at Week 12 and corticosteroid-free clinical remission by CDAI at Week 52 (for participants who were steroid free from Week 40 to Week 52)

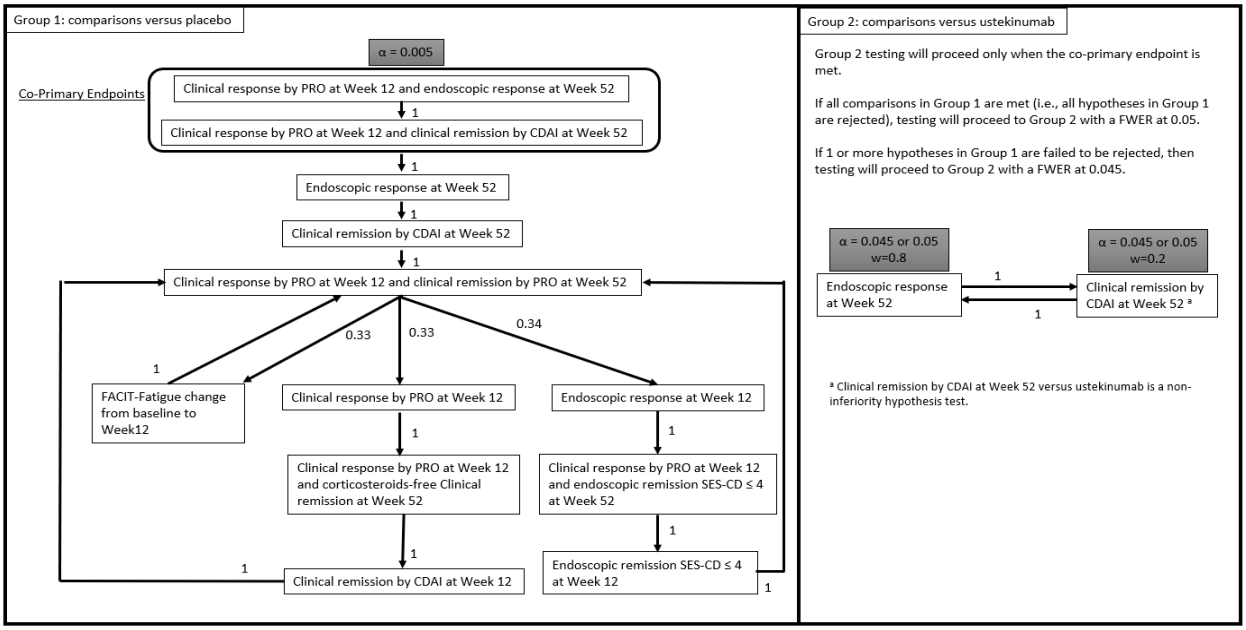
- $\mathbf{H_{mp,12,0}}$: Proportion of mirikizumab participants achieving clinical response by PRO at Week 12 is less than or equal to the proportion of placebo participants achieving clinical response by PRO at Week 12
- $\mathbf{H_{mu,1,0}}$: Proportion of mirikizumab participants achieving endoscopic response at Week 52 is less than or equal to the proportion of ustekinumab participants achieving endoscopic response at Week 52
- $\mathbf{H_{mu,2,0}}$: Proportion of mirikizumab participants achieving clinical remission by CDAI at Week 52 is less than or equal to the proportion of ustekinumab participants achieving clinical remission by CDAI at Week 52 by 10% or more

2.1. Multiplicity Adjustment

For testing the primary and major secondary hypotheses, a prespecified graphical scheme (Bretz et al. 2009, 2011) will be implemented to control the FWER at a 2-sided alpha level of 0.05 as described below:

Two groups including co-primary and major secondary hypotheses will be used. Group 1 will include the co-primary endpoints and all major secondary endpoints that involve comparisons versus placebo, and Group 2 will include all major secondary endpoints that involve comparisons versus ustekinumab. Within each group, the graphical scheme will control the FWER at a prespecified level. For Group 1, a FWER at 0.005 will be used. If all comparisons in Group 1 are met (i.e., all hypotheses in Group 1 are rejected), testing will proceed to Group 2 with a FWER at 0.05. If 1 or more hypotheses in Group 1 are failed to be rejected, while the comparisons on the co-primary endpoints must be met, then testing will proceed to Group 2 with a FWER at 0.045. More specifically, multiple testing adjusted p-values will be calculated using "Algorithm 2" described by Bretz and colleagues (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 is fully specified by the graph (including nodes, arrows and weights) along with the initial alpha allocation. The figure below describes the graphical scheme, and the initial alpha will be allocated to Group 1 and Group 2 as described above. Unless otherwise specified, there will be no adjustment for multiple comparisons for any other analyses outside the co-primary and major secondary endpoints. The testing scheme was finalized before the first unblinding of efficacy data.



Abbreviations: CDAI = Crohn's Disease Activity Index; FACIT-Fatigue = Functional Assessment of Chronic Illness Therapy-Fatigue; PRO = patient-reported outcome; SES-CD = Simple Endoscopic Score for Crohn's Disease.

3. Sample Size Determination

Approximately 3000 participants may be screened to achieve a total of approximately 1100 participants randomly assigned to study intervention. Based on a 6:3:2 randomization ratio, approximately 600 participants will be randomized to mirikizumab, 300 participants to ustekinumab, and 200 participants to placebo.

Approximately 90% of randomized participants are expected to meet the PAS definition as described in Section 4. A sample size of 990 participants (540 participants in mirikizumab and 180 participants in placebo) provides >90% power to demonstrate that mirikizumab is superior to placebo for the co-primary endpoints of: (1) clinical response by PRO at Week 12 and clinical remission by CDAI at Week 52, and (2) clinical response by PRO at Week 12 and endoscopic response at Week 52. This estimated power is based on a 2-sided chi-square test with alpha = 0.005 and assuming treatment response rates of the co-primary endpoints are 33% for mirikizumab and 10% for placebo.

The sample size based on the PAS also provides >90% power to demonstrate that mirikizumab is superior to ustekinumab for endoscopic response at Week 52. This calculation is based on a 2-sided chi-square test with alpha = 0.045 and assuming a difference of at least 16% between mirikizumab and ustekinumab in endoscopic response at Week 52.

Blinded sample-size re-estimation may be performed before the last participant has been enrolled in the study. For this re-estimation, response rates using blinded data will be evaluated and compared with the assumed response rates. The sample size from the study is estimated to be between a minimum sample size, 1100 participants, and a predefined, maximum sample size up to 1210 participants. If the re-estimated sample size is smaller than the planned minimum sample size, the study may enroll to the recalculated sample size. If the re-estimated sample size is larger than the planned minimum sample size, the team will decide whether to increase the sample size, up to a predefined maximum, or to accept the re-assessed reduced power.

4. Analysis Sets

For purposes of analysis, the following analysis sets are defined in the table below. Unless otherwise specified, all participants will be analyzed according to the study intervention to which they were randomized or assigned.

Population	Description
Screening Population	**Definition:** All participants who signed informed consent. **Purpose:** Used for disposition analysis.
PAS	**Definition:** All randomized participants who have baseline SES-CD ≥7 (or ≥4 for isolated ileal disease) and take at least 1 dose of study intervention, even if the participant does not take the assigned study intervention, does not receive the correct study intervention, or otherwise does not follow the protocol. **Purpose:** Used for efficacy, biomarkers, health outcomes, disposition and demographic.
PAS, exclude participants impacted by crisis	**Definition:** All participants in PAS, excluding all affected participants at affected sites by crisis (i.e., specific to Russia-Ukraine war). **Purpose:** Used for sensitivity analysis for the co-primary endpoints and major secondary efficacy endpoints.
PAS, Not-Biologic-Failed Population	**Definition:** All participants in PAS who have not failed any biologic medication regardless of prior biologic exposure. **Purpose:** Used for efficacy-related analysis.
PAS, Biologic-Failed Population	**Definition:** All participants in PAS who have failed at least 1 biologic medication. **Purpose:** Used for efficacy-related analysis.
mITT Population	**Definition:** All randomized participants who take at least 1 dose of study intervention, even if the participant does not take the assigned study intervention, does not receive the correct study intervention, or otherwise does not follow the protocol. **Purpose:** Used for sensitivity analysis for the co-primary endpoints, disposition and demographics.
ITT Population	**Definition:** All randomized participants, even if the participant does not take the assigned study intervention, does not receive the correct study intervention, or otherwise does not follow the protocol. **Purpose:** Used for disposition, demographics.
Safety Population	**Definition:** Same as mITT Population. **Purpose:** Safety analysis for the Period 1 and for the Study Treatment Period will be conducted on this population.
All Active Treatment Safety Population	**Definition:** All randomized participants who received at least 1 dose of mirikizumab or ustekinumab. **Purpose:** Safety analysis for any active study intervention will be conducted in this population

Abbreviations: GCP = Global Clinical Practice; ITT = intent-to-treat; mITT = modified intent-to-treat; PAS = Primary Analysis Set; SES-CD = Simple Endoscopic Score for Crohn's Disease;

5. Statistical Analyses

5.1. General Considerations

Statistical analysis of this study will be the responsibility of Lilly or its designee.

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

Unless otherwise specified, efficacy analyses will be conducted on the PAS, and safety analyses will be conducted on the safety populations as described in Section 4.

When reported, descriptive statistics will include the number of participants; mean, standard deviation, median, minimum, and maximum for continuous measures; and frequency counts and percentages for categorical measures.

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

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

5.1.1. Analysis Methods

All tests for the co-primary and major secondary endpoints will be conducted under the multiplicity-controlled framework described in Section 2.1. For the analyses of hypothesis under Group 1 with a FWER at 0.005, a 2-sided 99.5% CI will be provided along with the p-value. For analyses of hypothesis under Group 2, if all hypotheses in Group 1 are rejected, a FWER at 0.05 will be used. In this case, a 2-sided 95% CI with the p-value will be provided for the analyses of hypothesis under Group 2. If 1 or more hypotheses in Group 1 are failed to be rejected, while the comparisons on the co-primary endpoints must be met, then testing will proceed to Group 2 with a FWER at 0.045. In this case, a 2-sided 95.5% CI with the p-value will be provided. For other secondary endpoints without multiplicity control, superiority comparisons will be performed at a 0.05 2-sided alpha level. The corresponding p-value along with the 2-sided 95% CI will be provided.

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

- Unadjusted proportions for each study intervention 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) is the difference in proportions adjusted for the selected stratification factors: biologic-failed status (yes/no), baseline SES-CD total score (<12, ≥12), and either baseline SF ≥7 and/or baseline AP ≥2.5 (yes or unknown/no). SAS PROC FREQ will be used for the estimates and CIs, where the CIs are calculated by using Mantel-Haenszel-Sato method (Sato 1989).

- Cochran–Mantel–Haenszel (CMH) test will be used to compare the study intervention groups while adjusting for the selected stratification factors. The CMH p-value will be reported, and the CMH adjusted odds ratio along with the 2-sided asymptotic (i.e., not continuity corrected) CIs at the levels specified above. The CMH test will adjust for biologic-failed status (yes/no), baseline SES-CD total score (<12, ≥12), and either baseline SF ≥7 and/or baseline AP ≥2.5 (yes or unknown/no). The CMH chi-square p-value and the relative risk along with its 2-sided CI will be provided. In addition, the absolute study intervention difference in proportions will be provided along with the 2-sided CI estimate. If deemed necessary, 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 if necessary.

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

For continuous efficacy and health outcome variables with multiple measurements in each study period, a restricted maximum likelihood-based MMRM will be used as a sensitivity analysis. The model will include study intervention, biologic-failed status (yes/no), baseline SES-CD total score (<12, ≥12), either baseline SF ≥7 and/or baseline AP ≥2.5 (yes or unknown/no), visit, and study intervention by visit interaction as fixed categorical effects and baseline score and baseline score by visit interaction as fixed continuous effects. An unstructured covariance structure will be used to model the between- and within-subject errors. If this analysis fails to converge, the heterogeneous Toeplitz covariance structure, followed by the heterogeneous autoregressive covariance structure will be used. The Kenward–Roger method will be used to estimate the degrees of freedom. Type III tests for the LS means will be used for the statistical comparison; a 2-sided CI will also be reported. Contrasts will be set up within the model to test study intervention groups at specific time points of interest.

Study intervention comparisons of continuous efficacy and health outcome variables will be made using ANCOVA as a primary analysis method with study intervention, biologic-failed status (yes/no), baseline SES-CD total score (<12, ≥12), either baseline SF ≥7 and/or baseline AP ≥2.5 (yes or unknown/no), and baseline score in the model. Type III sums of squares for LS means will be used for statistical comparison between study intervention groups. The LS mean difference, standard error, p-value, and a 2-sided CI, unless otherwise specified, will also be reported. Missing data imputation method for the ANCOVA model is specified in Section 5.1.5. To handle longitudinal repeated data, the ANCOVA model will be applied to analyze selected time points one at a time.

For variables that are not collected at each postbaseline visit, data may exist at visits where the variable was not scheduled to be collected. In these situations, data from the early discontinuation visit that do not correspond to the planned collection schedule will be excluded from the MMRM analysis (Andersen and Millen 2013). Also, for by-visit summaries/displays such as boxplots, the weeks when data was not scheduled to be collected may not be displayed.

However, unscheduled assessments within any defined study period will still be used in the shift analyses, and for imputing values for the change from baseline to mBOCF endpoint analyses.

Fisher's exact test will be used for categorical safety data (example, AEs) for between study intervention group comparisons. Continuous safety variables (example, laboratory variables) will be analyzed by ANCOVA with study intervention and baseline value in the model.

5.1.2. Definition of Baseline

Visit 2 (Week 0) is the baseline randomization and the first dose of study intervention visit. The centrally read baseline SES-CD score from the screening endoscopy is considered the baseline for endoscopic response and endoscopic remission endpoints. Daily diary entries obtained prior to first dose of study intervention are also considered baseline for clinical remission by PRO, clinical response by PRO, and CDAI clinical response. Baseline score for daily diary entries will be calculated by averaging the most recent 7 days (possibly nonconsecutive) in the 12 days prior to the day of Visit 2, after removing the day(s) of the endoscopy prep, the day of endoscopy procedure, and the 2 days following the endoscopy procedure. If less than 4 days of data are available, the baseline score will be set to missing. 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 dose of study intervention (Period 1 start date), unless otherwise specified.

Baseline for safety analysis is described in Section 5.6.

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.

5.1.3. Definition of Study Period Time Interval

The table below displays a list of study periods along with the definition of which participants 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 the table below should be understood to mean "the day before" while the words "after" should be understood to mean "the day after." For the purpose of determining whether a date/time lies within an interval these words are intended to convey whether the start or end of the period is inclusive of the specified date.

Study Period	Interval Start Definition	Interval End Definition
Screening	Informed consent date	Prior to the start of Period 1.
Period 1	At the date/time[a] of first dose of study intervention. For participants who are randomized but not dosed, Period 1 starts on the date of randomization.	Prior to the start of Period 2. For participants who discontinue before or on the Week 12 visit, Period 1 ends at the latest date of study intervention discontinued date or last study intervention visit date.
Period 2	At the Week 12 dosing date/time[a]. If the participant is unable to be dosed at the Week 12 visit, Period 2 starts at the Week 12 visit. If the participant misses the Week 12 visit, the Period 2 starts at Day 91.	After the Week 52 visit date. For participants who discontinue prior to Week 52, Period 2 ends at the latest date of study intervention disposition date or last study intervention visit date.
Study Treatment Period	Same as Period 1 interval start	The latest of the following dates: (1) after the end of Period 1, (2) after the end of Period 2.
Follow-up Period	All participants who had Visit 801 or Visit 802 are considered to have entered the Follow-up Period. The latest of Period 1or Period 2 interval end date.	The last date of the last study visit and study disposition date.
All Active Treatment Period	At the date/time[a] of first dose of study intervention with mirikizumab or ustekinumab (that is dosing with placebo does not start the period).	The latest of the following dates: (1) the end of Period 1, (2) the end of Period 2.
All Active Treatment + Follow-up Period	At the date/time[a] of first dose of study intervention with mirikizumab or ustekinumab (that is dosing with placebo does not start the period).	The last date of the last study visit and study disposition date.

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

5.1.4. Definition of Study Intervention by Study Period

The table below provides the study intervention groups to be displayed for each analysis population and analysis period.

Analysis Population	Analysis Period	Study Intervention Groups:	Study Intervention-Group Comparison, Unless Otherwise Specified
PAS[a], mITT,	Study Treatment Period	900 mg miri IV Q4W / 300 mg miri SC Q4W (miri) 6 mg/kg uste IV W0 / 90 mg uste SC Q8W (uste) Placebo IV Q4W/ SC Q4W (pbo) Total	For efficacy: miri vs pbo miri vs uste
Safety Population	Period 1	Placebo IV Q4W (pbo) 900 mg IV Q4W miri (miri) 6 mg/kg IV W0 / 90 mg SC QW8 uste (uste) Total	For safety: miri vs pbo miri vs uste
Safety Population	Study Treatment Period	900 mg miri IV Q4W / 300 mg miri SC Q4W (miri) 6 mg/kg uste IV W0 / 90 uste SC Q8W (uste) Placebo IV Q4W/ SC Q4W (pbo)[b]	For safety: miri vs uste
All Active Treatment Safety Population	All Active Treatment Period, All Active Treatment + Follow-up Period	All miri (includes any time when a participant was in miri) 6 mg/kg uste IV W0 / 90 uste SC Q8W (uste)	For safety: All miri vs uste

Abbreviation: IV = intravenous; miri = mirikizumab; mITT = modified intent-to-treat; PAS = Primary Analysis Set; pbo = placebo; Q4W = every 4 weeks; Q8W = every 8 weeks; SC = subcutaneous; uste = ustekinumab.

[a] include the additional analysis populations based on Primary Analysis Set in Section 4.

[b] While on placebo: patients randomized to placebo who do not meet response criteria at the 12-week assessment will be censored at the time they begin mirikizumab treatment (to be added to footnote for corresponding output).

5.1.5. Missing Data Imputation

The Schedule of Activities outlined in the protocol specifies the allowable windows for assessments. Assessments performed outside these windows will not be excluded from the analysis (unless otherwise specified) but will be reported as a protocol deviation.

ICEs (FDA 2017) are events which occur after the study intervention initiation and make it impossible to measure a variable or influence how it should be interpreted. Section 1.1 includes the following ICEs, which may lead to missing endpoint data depending on the estimand of interest:

1. Discontinuation of study intervention prior to time point of interest. Note: participants who take prohibited medications are required to discontinue the study treatment.
2. Specified changes in concomitant CD medications (Appendix 10 [Section 6.10]) prior to time point of interest.
3. Non-responders at Week 12 in Placebo arm switch to mirikizumab

Participants may also have sporadically missing data due to reasons other than ICEs (e.g. failure to fill out a daily diary or attend an office visit).

The methods described below will be used to handle missing data.

5.1.5.1. Non-Responder Imputation (NRI)

As described in Section 1.1, the composite strategy will be used to handle binary endpoints; patients who discontinue study treatment or have specified changes in concomitant CD medications are categorized as treatment failures. As such, these patients are not considered missing from the perspective of the estimand of interest. A small number of patients who completed study treatment up to the time point of interest but are sporadically missing the binary endpoint data will still require imputation. These patients will be imputed using NRI.

Additionally, the NRI method is used for all visits subsequent to Week 12 when the estimand of interest uses the hypothetical strategy for handling the additional ICE of participants in the placebo group beginning study intervention with mirikizumab at Week 12. The assumption behind this imputation method is that, had the participant stayed on placebo, a non-response would have been observed for the endpoint of interest.

5.1.5.2. Mixed-Effects Model for Repeated Measures (MMRM)

As a sensitivity analysis for continuous variables with multiple postbaseline measurements in a study period, the MMRM approach will be used with the missing at random assumption for handling missing data. This analysis takes into account both missingness of data and the correlation of the repeated measurements.

For continuous endpoints, a hybrid estimand strategy is used. For ICEs of study intervention discontinuation and specified changes in the concomitant CD medication, the composite strategy will be used such that measurements after the ICEs will return to baseline. As such, these patients are not considered missing from the perspective of the estimand of interest.

For the additional ICE where participants in the placebo group switch to mirikizumab at Week 12, a hypothetical scenario is envisaged in which these patients remained on placebo for the rest of the study, leading to a missing data problem. Because we do not think measurements after this ICE can be considered as missing at random, we use the baseline observation carry forward method to impute data after this ICE before we apply the MMRM approach.

The MMRM approach will be used to handle the remaining sporadic missingness.

5.1.5.3. Modified Baseline Observation Carried Forward (mBOCF)

As a primary analysis for continuous variables, the ANCOVA with mBOCF approach will be used for handling missing data.

For continuous endpoints, a hybrid estimand strategy is used. For ICEs of study intervention discontinuation and specified changes in the concomitant CD medication, the composite strategy will be used such that measurements after the ICEs will return to baseline. As such, these patients are not considered missing from the perspective of the estimand of interest.

For the additional ICE where participants in the placebo group switch to mirikizumab at Week 12, a hypothetical scenario is envisaged in which these patients remained on placebo for the rest of the study. We use the baseline observation carry forward method to impute data after this ICE.

For all participants with sporadically missing observations prior to any ICEs, the last non-missing observation before the sporadically missing data will be carried forward to the corresponding visit.

5.1.5.4. Modified Nonresponder Imputation (mNRI)

For the co-primary endpoints, missing data will be imputed using hybrid imputation as a sensitivity analysis. Missing data for reasons including the COVID-19 pandemic, Russia/Ukraine crisis and treatment discontinuation due to lost to follow-up and pregnancy will be imputed by multiple imputation (MI), while missing data due to treatment discontinuation for other reasons such as AE and lack of efficacy will be imputed by NRI. Measurements after ICEs of specified changes in the concomitant CD medication and treatment switch to mirikizumab also will be handled by NRI. Sporadically missing data (i.e., when a patient was still in the treatment period but data was not collected) will be imputed by MI.

The hybrid imputation method will be implemented as follows:

1. Identify all missing data and what caused the missingness. The reasons are categorized as: (1) the COVID-19 pandemic, Russia/Ukraine crisis, treatment discontinuation due to lost to follow-up, pregnancy or sporadically missing; (2) other reasons. Data points that occur following ICEs in Section 1.1 will be set to "missing" prior to Step 2.
2. For each treatment arm and each longitudinal variable, missing data will be imputed under multivariate normal assumption. The imputation model will adjust for prespecified baseline variables. A total of 100 imputed datasets will be created.
3. For each of these imputed complete datasets from Step 2, the imputed data for missing due to other reasons (as identified in Step 1) and for measurements after ICEs of specified changes in the concomitant CD medication and treatment switch to mirikizumab will be set to missing, and the NRI method will be implemented. All other data, including imputed or observed data, will be used for the analysis or to derive the binary outcomes.
4. The Mantel-Haenszel 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 pooling the CMH test statistic using the Wilson-Hilferty transformation (Wilson and Hilferty 1931).

5.1.5.5. Tipping Point Analysis for Co-primary endpoints

Tipping point analysis will be conducted as sensitivity analysis for co-primary endpoints. Within each analysis, the most extreme case will be considered, in which all sporadically missing data (i.e. missing not due to ICEs) for participants randomized to mirikizumab will be imputed using the worst possible outcomes and all sporadically missing data for participants randomized to placebo and continued with placebo will be imputed with the best possible outcomes.

- Missing responses in the mirikizumab group will be imputed with a range of response probability, including probabilities of 0, 0.2, 0.4, 0.6, 0.8, and 1.0.
- For missing responses in placebo continued with placebo group, a range of response probabilities, including probabilities of 0, 0.2, 0.4, 0.6, 0.8, and 1.0 will be used to impute the missing values. Multiple imputed datasets (m=100) will be generated for each response probability.

Study intervention differences between mirikizumab and placebo will be analyzed for each imputed data set using CMH test (Section 5.1.1). Results across the imputed data sets will be

aggregated using SAS® Proc MIANALYZE in order to compute a p-value or a 99.5% CI for the study intervention comparisons for the given response probability. If the probability values do not allow for any variation between the multiple imputed datasets (for example, all missing responses in the placebo continued with placebo and mirikizumab groups are imputed as responders and NRs respectively), then the p-value from the single imputed dataset will be used.

5.1.5.6. Tipping Point Analysis for selected endpoints at Week 52 for placebo comparison

For the 2 major secondary endpoints, endoscopic response at Week 52 and clinical remission by CDAI at Week 52, an additional sensitivity analysis will be performed. As previously described, a hybrid strategy is used to accommodate the ICE where participants in the placebo group switch to mirikizumab at Week 12. For this ICE a hypothetical scenario is envisaged in which these patients remained on placebo for the rest of the study and for all other ICEs the composite strategy above will be used (see details in Section 1.1 and Section 5.1.5.1). In our primary analysis of these 2 endpoints, we impute participants who switch to mirikizumab as non-responders. In this sensitivity analysis, we use alternative assumptions to perform the imputation. Within each analysis, participants in the placebo non-responder group who switch from placebo to mirikizumab at Week 12 will be imputed with a range of response probabilities from 0 and up to the observed response rate at Week 52 among placebo non-responders by PRO at Week 12. This upper bound is an extremely conservative assumption which corresponds approximately to estimating the Week 52 placebo rate ignoring the fact that placebo non-responders by PRO at Week 12 had received 40 Weeks of active mirikizumab therapy by Week 52. For sporadically missing data, NRI will be used. Multiple imputed datasets (m=100) will be generated for each response probability where the multiple imputation will only be applied to participants who switch to mirikizumab at Week 12.

Study intervention differences between mirikizumab and placebo will be analyzed for each imputed data set using CMH test (Section 5.1.1). Results across the imputed data sets will be aggregated using SAS® Proc MIANALYZE in order to compute a p-value or a 99.5% CI for the study intervention comparisons for the given response probability. If the probability values do not allow for any variation between the multiple imputed datasets (for example, all missing responses in the placebo continued with placebo and mirikizumab groups are imputed as responders and NRs respectively), then the p-value from the single imputed dataset will be used.

5.2. Participant Dispositions

Screen failures and reason for screen failure will be summarized. The treatment disposition and study disposition will be summarized by study intervention group for the mITT, ITT and PAS population. Frequency counts and percentages of all participants who are randomized and complete the study intervention, who complete study, who discontinue the study intervention early, and who discontinue the study early will be presented overall at Week 12 and at Week 52. Reasons for early discontinuation of the study intervention and the study will be summarized.

All participants who are randomized (i.e., the ITT population) and discontinued from study intervention or study 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.

5.3. Primary Endpoint(s) Analysis

5.3.1. Definition of Endpoint(s)

The co-primary endpoint is comprised of 2 separate endpoints:

- the proportion of participants achieving clinical response by PRO at Week 12 and endoscopic response at Week 52
- the proportion of participants achieving clinical response by PRO at Week 12 and clinical remission by CDAI at Week 52.

Endoscopies performed for Week 52 that occur up to a maximum of 14 days after the Week 52 visit date, but before any additional dosing, will be used for the analysis of Week 52.

For CDAI total score, if central lab data is not available, hematocrit data from local lab will be used. If scheduled hematocrit data at a given visit is not available, the hematocrit value obtained the closest and within ±7 days of the date of the visit which is also within study period will be used. If hematocrit data is still not available, the closest hematocrit result from the preceding visit will be used. If scheduled weight at a given visit is not available, the closest measured weight from the preceding visit up to 7 days after the date of visit will be used, which must be measured within study period. CDAI score is missing if any of the 3 patient-reported items or 3 physician-reported items is missing.

Full descriptions and derivations of these endpoints are shown in Appendix 1 (Section 6.1).

5.3.2. Main Analytical Approach

The primary estimand represents the primary clinical question of interest: what is the difference between mirikizumab and placebo with successful responses to the co-primary endpoints, separately, after adhering to the 52 weeks study intervention, without any specified changes in the concomitant CD medication, in the PAS. Section 1.1 describes the primary estimand that will be used to assess the co-primary objective of this study.

The primary hypothesis that will be tested in this study is that mirikizumab is superior to placebo with regards to the co-primary endpoint. The missing values will be imputed using NRI (Section 5.1.5.1).

The primary analysis will use the CMH chi-square test to compare mirikizumab to placebo adjusting for the selected stratification factors with NRI (Section 5.1.5.1). The common risk difference and the odds ratio adjusted for the selected stratification factors with the 2-sided 99.5% CI will be presented (see Section 5.1.1). Additional details are described in Appendix 2 (Section 6.2).

5.3.3. Sensitivity Analysis

The co-primary endpoints will also be analyzed using the same approach described in Section 5.3.2 in the following analysis populations (definition can be found in Section 4):

- the mITT population
- PAS, exclude participants impacted by crisis

The logistic regression analysis (see Section 5.1.1) with the same imputation method described in Section 5.3.2 will be used to analyze the 2 co-primary endpoints for the PAS as a sensitivity analysis.

Tipping point analysis will be used as another sensitivity analysis for the 2 co-primary endpoints in the PAS (see Section 5.1.5.5).

Additional details of the sensitivity analyses are described in Appendix 2 (Section 6.2).

5.3.4. Supplementary Analysis

There is also an interest to further evaluate the impact of the additional condition of clinical response by PRO at Week 12 on clinical remission by CDAI and endoscopic response at Week 52. In addition to the analyses of the co-primary endpoints described in Section 5.3 and of these Week 52 major secondary endpoints as described in Section 5.4, two supplementary analyses are also proposed:

1) For these 2 major secondary endpoints, an additional ICE for participants in the placebo group is switching from placebo to mirikizumab at Week 12. This leads to a missing data problem because the estimand strategy for this ICE is the hypothetical strategy (see Section 1.1). For the primary missing data imputation method, the participants in the placebo group who switch to mirikizumab are imputed using NRI. To further explore different missing data imputation strategies, a tipping point analysis is included for participants in the placebo group who switch to mirikizumab at Week 12 (SAP Section 5.1.5.6) to evaluate the impact on these two major secondary endpoints if they had not switched to mirikizumab.
2) It is also of interest to compare the treatment effect of mirikizumab and placebo at Week 52 among clinical responders at Week 12. However, the causal effect of mirikizumab and placebo cannot be estimated in this subpopulation without adjustment because clinical response by PRO is a post-intervention outcome. Therefore, randomization is not guaranteed and participants who respond to mirikizumab may be different from patients who respond to placebo at Week 12. To address this issue, the principal stratum estimand strategy will be used to estimate the treatment effect among the principal stratum of always clinical responders at Week 12, which is defined under the potential outcome framework as participants who would achieve clinical response by PRO at Week 12 under both mirikizumab and placebo. To identify participants belonging to this principal stratum, there is a missing data problem for the unobserved clinical response by PRO status at Week 12 had the patient hypothetically received the alternative treatment. We will consider the following imputation strategies for clinical response by PRO at Week 12.
    a. An imputation model will be used to determine for patients who are observed to achieve clinical response by PRO at Week 12 under mirikizumab or placebo, whether they would also achieve clinical response by PRO under the alternative treatment.
    b. An imputation model will only be used to determine whether participants who are observed to achieve clinical response by PRO at Week 12 under mirikizumab

treatment would also achieve clinical response under placebo. Monotonicity will be assumed for placebo patients, such that patients who achieve clinical response by PRO at Week 12 under placebo will be assumed to also achieve clinical response had they received mirikizumab.

A total of 100 imputations will be taken, the set of always clinical responders will be defined as participants who are observed/imputed to achieve clinical response by PRO at Week 12 for both placebo and mirikizumab. After identifying the principal stratum of always clinical responders within each imputation, the Mantel-Haenszel common risk difference will be used to estimate the Week 52 treatment effect by comparing the observed outcomes between mirikizumab and placebo patients. The point estimate will be obtained as the mean risk difference across imputations, and bootstrap will be used to estimate confidence intervals for the treatment effect (Bartlett and Hughes 2020; Lipkovich 2022).

5.4. Secondary Endpoint(s) Analysis in Comparison to Placebo

5.4.1. Major Secondary Endpoint(s)

5.4.1.1. Definition of Endpoint(s)

Major secondary endpoints are listed in Section 1.1.

Descriptions and derivations of these endpoints are shown in Appendix 1 (Section 6.1).

5.4.1.2. Main Analytical Approach and Sensitivity Analyses

The estimand for the major secondary binary endpoints is the same as the primary estimand for co-primary endpoints. The estimand to assess the clinical question of interest: what is the difference between mirikizumab and placebo with successful responses to the major secondary binary endpoints, individually, after adhering to the relevant duration of study intervention, without any specified changes in the concomitant CD medication, in the PAS.

The estimand for the major secondary continuous endpoint is to assess the mean difference between mirikizumab and placebo prior to discontinuation of study drug, any specified changes in the concomitant CD medications in the PAS.

Statistical hypotheses for major secondary endpoints are stated in Appendix 2 (Section 6.2).

The analyses for major secondary binary endpoints will utilize the CMH test similar to the primary analysis (Section 5.3.2). The logistic regression analysis in Section 5.3.3 for the co-primary endpoints will be also used for major secondary binary endpoints as sensitivity analyses. The analysis for major secondary continuous endpoint will be using ANCOVA (see Section 5.1.1). A sensitivity analysis for the major secondary continuous endpoint will be performed. The jump to reference imputation will be applied with ANCOVA and the placebo arm will be the reference. In this sensitivity analysis, all missing data in the placebo arm will be imputed under the assumption of missing at random (MAR). For the mirikizumab arm, the sporadically missing data will be imputed assuming MAR and the measurements after ICEs will be set to missing and be imputed assuming behavior like the placebo arm.

Additional details of analyses for major secondary endpoints are described in Appendix 2 (Section 6.2).

5.4.2. Other Secondary Endpoint(s) and Exploratory Endpoints

Other secondary and exploratory objectives are listed in Section 1.1. Descriptions and derivations of these endpoints are shown in Appendix 1 (Section 6.1).

The main analytical approach for other secondary binary endpoints of mirikizumab in comparisons to placebo is similar as for the co-primary endpoints (see Section 5.3.2). For other secondary continuous endpoints over time, the main analytical approach is MMRM (see Section 5.1.1). For other secondary continuous endpoints with a single postbaseline time point in each study period, the analysis will be made using ANCOVA (see Section 5.1.1).

As a supplementary analysis, a principal stratum analysis similar to that for the primary and major secondary endpoint (see Section 5.3.4) will be conducted to evaluate the proportion of participants achieve urgency NRS ≤ 2 among participants with baseline Urgency NRS ≥ 3 at Week 52 within the principal stratum of always clinical responders at Week 12. The Mantel-Haenszel common risk difference will be used to estimate the Week 52 treatment effect by comparing the observed outcomes between mirikizumab and placebo patients at Week 52 within this principal stratum.

Additional details for other and exploratory secondary endpoints are described in Appendix 2 (Section 6.2) or will be provided in supplemental SAP documents.

5.5. Secondary Endpoint(s) Analysis in Comparison to Ustekinumab

Major secondary hypotheses are that mirikizumab is superior to ustekinumab in below 2 endpoints:

- endoscopic response at Week 52
- endoscopic remission SES-CD ≤ 4 at Week 52

Another major secondary hypothesis is that mirikizumab is non-inferior to ustekinumab at Week 52 in clinical remission by CDAI at Week 52.

Other secondary objectives and endpoints in the comparison of mirikizumab and ustekinumab are in Section 1.1.

The estimand to assess the clinical question of interest: what is the difference (or is there any clinical meaningful difference) between mirikizumab and ustekinumab with successful responses to the major secondary binary endpoints, individually, after adhering to the 52 weeks of study intervention, without any of specified changes in the concomitant CD medication in the PAS. The missing values will be handled using NRI (Section 5.1.5.1).

The analyses for these major secondary endpoints will utilize the CMH test similar to the primary analysis (Section 5.3.2). The logistic regression analysis in Section 5.3.3 for the co-primary endpoints will be also used for these major secondary binary endpoints as sensitivity analyses.

5.5.1. Noninferiority Analysis

To assess noninferiority of mirikizumab to ustekinumab, the lower bound of the 2-sided 95% or 95.5% CI of the estimated common risk difference (see Section 5.1.1) in proportions between mirikizumab and ustekinumab response will be compared to the noninferiority margin. To establish that mirikizumab is noninferior to ustekinumab, the upper bound for the ustekinumab minus mirikizumab proportions must be less than the prespecified noninferiority margin. That is, the hypothesis test is significant if $UL < M$. Equivalently, a p-value for the non-inferiority hypothesis test will be calculated using the following formula:

$$p - value = minimum\left[1, 2 \times \left(1 - \Phi\left(\frac{\hat{\delta} + M}{SE_{\hat{\delta}}}\right)\right)\right]$$

Here M is the non-inferiority margin, $\hat{\delta}$ is the common risk difference, $SE_{\hat{\delta}}$ is the standard error of $\hat{\delta}$, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Note that the multiplication factor of "2" in the formula accounts for the fact that in the CI approach to non-inferiority testing, a 2-sided CI is used. As previously mentioned (see Section 5.1.1), the common risk difference and its standard error will be calculated using the Mantel-Haenszel-Sato method (Sato 1989) as implemented in SAS PROC FREQ to adjust for the selected stratification factors.

5.5.1.1. Justification of Noninferiority Margin

There is no universally accepted value for what is considered to be a clinically unimportant difference between 2 treatments based on CDAI remission. Global regulatory guidance (EMA 2005; FDA 2016) indicate that selection of the noninferiority margin is based upon a combination of statistical and clinical grounds.

The data from 2 induction studies (UNITI1 and UNITI2) followed by a single maintenance study (IM-UNITI) were used to demonstrate the efficacy of ustekinumab versus placebo in CDAI remission. Patients who were randomized to ustekinumab in UNITI1 or UNITI2 and who achieved CDAI response at Week 8 were eligible to enroll in the randomized placebo-withdrawal portion of IM-UNITI as the primary population. All placebo patients and patients who did not achieve clinical response at Week 8 after being dosed with ustekinumab were eligible to enroll in the nonrandomized portion of IM-UNITI. These patients received ustekinumab at Week 8 and were assessed for CDAI response at Week 16. Those who achieved CDAI response at Week 16 were eligible to continue dosing with ustekinumab while the patients who did not achieve clinical response were discontinued (Feagan et al. 2016).

For Study AMAM, the CDAI remission rate for ustekinumab 90 mg Q8W at Week 52 will be estimated as below:

> Prob (CDAI Remission at Week 52) = Prob (Week 52 CDAI Remission |Week 8 CDAI Response) * Prob (Week 8 CDAI Response) + Prob (Week 52 CDAI Remission |^CDAI Response at Week 8) * Prob (^Week 8 CDAI Response).

The following table shows the CDAI response rates at Week 8 for the 6-mg/kg ustekinumab group:

Induction Study	CDAI Response at Week 8	CDAI Nonresponse at Week 8
UNITI 1	94/249 (37.8%)	155/249 (62.2%)
UNITI 2	121/209 (57.9%)	88/209 (42.1%)
Overall	215/458 (46.9%)	243/458 (53.1%)

Abbreviation: CDAI = Crohn's Disease Activity Index.
Note: Values in the table are n/N (%) where: n = number of patients achieving a response; N = total number of patients in category. Data based on non-responder imputation.
Source: Feagan et al. 2016.

CDAI remission rates for Week 52 are estimated as follows:

- Probability of CDAI remission at Week 52 for the 90-mg Q8W treatment group for patients achieving CDAI response at Week 8 was 68/128 (53.1%).
- Probability of CDAI remission at Week 52 for the 90-mg Q8W treatment group for patients not achieving CDAI response at Week 8 is estimated as 126/467 (27.0%). This probability assumes that patients who received ustekinumab at both Weeks 0 and 8 and did not achieve CDAI response at Week 8 and at Week 16 did not achieve CDAI remission at Week 52.

Based on the above, the expected CDAI remission rate at Week 52 for the 90-mg Q8W treatment group is 39.2% with 95% CI (34.6, 43.9).

CDAI remission rates for patients who were randomized to placebo and continued in placebo during maintenance are not available at Week 52 in IM-UNITI. However, CDAI response and CDAI remission rates from Week 2 through Week 26 were shown to be stable with a slow decrease (approximately 5%) from Week 8 to Week 26 in PRECISE 1 (Sandborn et al. 2007). Assuming that CDAI remission rates in placebo patients have a similar decrease of approximately 5% between Week 8 and Week 52, CDAI remission at Week 52 for the placebo group will be 12.3% (6.8%, 17.9%) based on the summary below.

The following table shows the CDAI remission rates for the placebo group at Week 8:

Induction Study	CDAI Remission at Week 8
UNITI-1	18/247 (7.3%)
UNITI-2	41/209 (19.6%)
Overall	59/456 (12.9%)

Abbreviation: CDAI = Crohn's Disease Activity Index.
Note: Values in the table are n/N (%) where: n = number of patients achieving a response; N = total number of patients in category. Data based on non-responder imputation.
Source: Feagan et al. 2016.

Based on these rates, the expected treatment effect for ustekinumab 6 mg/kg followed by 90 mg Q8W versus placebo at Week 52 in a treat-through study is 26.9% with 95% CI (19.5, 34.6). Confidence intervals were estimated based on bootstrap re-sampling.

Using the fixed 95%/95% margin method, a 10% NI margin represents clinical judgement about the amount of the active control effect that must be retained. Assuming that Study AMAM will have similar proportions of biologic-failed and not-biologic-failed patients as those observed in the UNITI program and that the constancy assumption holds, the proposed NI margin is expected to preserve 50% of the expected ustekinumab effect in CDAI remission at Week 52 in a treat-through study. If the lower 97.5% lower bound of the CI for the difference between mirikizumab and ustekinumab is greater than -10%, it would rule out a loss of more than half of the benefit expected for ustekinumab for CDAI remission at Week 52.

5.5.1.2. Additional Analysis on the Noninferiority Testing

Since a placebo group will be included in Study AMAM, assay sensitivity as well as the constancy assumption will be checked by (1) comparing ustekinumab to placebo for CDAI response at Week 8, (2) the CDAI remission rate at Week 52 among the CDAI Week 8 responders in ustekinumab.

In addition, the level of a 2-sided CI that excludes 8% NI margin and 5% NI margin will also be provided. To assess noninferiority of mirikizumab to ustekinumab, the lower bound of the 2-sided 95% or 95.5% CI of the estimated common risk difference (see Section 5.1.1) in proportions between mirikizumab and ustekinumab response will be compared to the noninferiority margin. Additional details are described in Appendix 2 (Section 6.2).

5.6. Safety Analyses

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

In general, safety evaluations will be based on the following safety analysis populations with their associated study periods (Section 5.1.3):

- Safety Population
- All Active Treatment Safety Population

Fisher exact test will be used to compare percentages, and odds ratios will be provided. Odds ratios will be created with mirikizumab treatment as the numerator, and placebo or ustekinumab as the denominator. For the study treatment period, the formal statistical comparison of mirikizumab and placebo will not be performed.

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

For document writing purposes, tests with 2-sided p-values less than 0.05 will be referred to as “having strong evidence for a treatment difference,” unless otherwise noted. However, p-values should not be over-interpreted for these safety analyses.

5.6.1. Extent of Exposure

Duration of exposure to study intervention will be summarized by study intervention group for the safety analysis populations. For the treatment period of interest associated with each safety analysis population, exposure will be calculated as (Date of end date in the Interval – Date of start date in the Interval + 1 day) described in Section 5.1.3. The following periods will be used for calculations:

- For the Safety Population, we will use:
  - Period 1
  - Study Treatment Period where exposures for patients who are randomized to placebo and switch to mirikizumab will not be counted .
- For the All Active Treatment Safety Population, the All Active Treatment Period will be used.

Total patients-years (PY) of exposure will be reported for above safety analysis populations by study intervention group in Section 5.1.4. Descriptive statistics will be provided for participants-weeks of exposure and the frequency of participants falling into different exposure ranges will be summarized.

- ≥0, ≥4 weeks, ≥8 weeks, ≥12 weeks, ≥16 weeks, ≥24 weeks, ≥32 weeks, ≥40 weeks, ≥48 weeks.
- >0 to < 4 weeks, ≥4 weeks to <8 weeks, ≥8 weeks to <12 weeks, ≥12 weeks to 16 weeks, …, ≥48 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.

Reasons for not taking study intervention and reasons for not taking the planned amount of study intervention will be reviewed.

5.6.2. Adverse Events

A TEAE is defined as an AE that first occurred or worsened in severity after baseline. The MedDRA 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 the start time for the AE is missing, it will be assumed to have started in the later period.

For the safety populations, the baseline period and postbaseline will be defined as follows:

- Safety Population: The baseline period is the Screening Period. The postbaseline period will be the Period 1 and Study Treatment Period.

- All Active Treatment Safety Population. The baseline period for participants randomized to mirikizumab and ustekinumab during Period 1 is the Screening Period. For participants randomized to placebo and switched to mirikizumab, the baseline events are those events which are ongoing at the time of the first injection with mirikizumab (that is, the baseline period is a moment in time). Two different postbaseline periods will be used:
  - For "all miri" and "uste," the all Active Treatment Period will be used.
  - For "all miri+follow-up" and "uste+follow-up," the All Active Treatment Period + Follow-Up Period will be used.

The summary analyses will be presented for the safety populations corresponding to the periods as described in the table below. Summary tables will include the number and percentage of participants reporting an event. For events that are sex-specific (as defined by MedDRA), the number of participants at risk will include only participants from the given sex. Comparisons will be performed using Fisher's exact test. In addition, exposure adjusted incidence rates will be performed as described in the compound-level safety standards.

Analysis	Population/Period[a]
Overview of AEs	S/P1;S/TP; A/AP
Summary of TEAE PTs by decreasing frequency	S/P1; S/TP; A/AP
Summary of TEAE PTs occurring in ≥1% of participants by decreasing frequency	S/P1; S/TP; A/AP
Summary of TEAE PTs by decreasing frequency within SOC	S/P1; S/TP; A/AP
Summary of TEAE PTs by maximum severity by decreasing frequency within SOC	S/P1; S/TP; A/AP
Summary of SAE PTs by decreasing frequency within SOC	S/P1; S/TP; A/AP
Summary of AEs leading to study treatment discontinuation by decreasing frequency	S/P1; S/TP; A/AP
Listing of SAEs	S
Listing of Deaths	All Entered participants

Abbreviations: AE = adverse event; PT = Preferred Term; SAE = serious adverse event; SOC = System Organ Class; TEAE = treatment-emergent adverse event.

a Populations are abbreviated as follows: S = Safety Population; A = All Active Treatment Safety Population. Periods are abbreviated as follows: P1 = Period 1; TP = Study Treatment Period; AP = All Active Treatment Period.

5.6.2.1. Common Adverse Events

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

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

The number and percentage of participants reported with an SAE, including those resulting in death during the treatment period, will be summarized by treatment using MedDRA PT nested within SOC. A listing of SAEs will be provided. A listing of all deaths from screening to end of study participation will be provided for all entered participants.

The number and percentage of participants 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.

5.6.3. Clinical Laboratory Evaluations

As described fully in compound-level safety standards and in the laboratory-related PhUSE white papers (PhUSE 2013, 2015), the clinical laboratory evaluations will be summarized with the following displays described in the table below.

Analysis	Population[a]
Box plots of observed values (and change from baseline values) by visit. 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 5.6.	S
Treatment-emergent abnormal high lab values (that is, participants shifting from a normal/low maximum baseline value to a high maximum postbaseline value) or low lab values (that is, participants shifting from normal/high minimum baseline value to a low minimum postbaseline value)	S, A
Scatter plot of maximum (minimum) postbaseline value versus maximum (minimum) baseline value	S, A
Shift tables showing the number of participants 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 cut-offs defined in the compound-level safety standards.	S, A
Listing of abnormal findings	S (all study periods)

Abbreviation: ANCOVA = analysis of covariance.

[a] Populations are abbreviated as follows: S = Safety Population; A = All Active Treatment Safety Population.

The baseline is the last non-missing assessment in the baseline period. The postbaseline periods will be identical to those described in Section 5.6.2. Postbaseline measurement of continuous analysis (for example, boxplots) will include only scheduled measurements, while postbaseline categorical analysis (for example, shifts) will include both scheduled and unscheduled measurements.

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

- Safety Population:
  - For analyses of continuous measurements: the last scheduled or unscheduled non-missing measurement recorded during the Screening Period.
  - For analyses of categorical measurements: all scheduled or unscheduled non-missing measurements recorded during the Screening Period.
- All Active Treatment Safety Population:
  - For analyses of categorical measurements: (1) all scheduled or unscheduled non-missing measurements recorded during the Screening Period for the participants randomized to mirikizumab and ustekinumab, (2) the last scheduled or unscheduled non-missing measurement recorded before first mirikizumab intervention for participants randomized to placebo and switched to mirikizumab.

For any lab given on the day of first taking study medication at the start of the postbaseline period, the start time of the study intervention 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

(that is, 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.

5.6.4. Vital Signs and Other Physical Findings

As described fully in compound-level safety standards and in the vital signs-related PhUSE white papers (PhUSE 2013, 2015), vital signs and weight will be summarized similarly to the clinical laboratory evaluation (see Section 5.6.3) as defined in the compound-level safety standards. 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.

5.6.5. Electrocardiograms

ECGs will be read locally. Complete ECG data will not be part of the clinical database. Any clinically significant findings from ECGs that result in a diagnosis and that occur after the participant receives the first dose of study drug will be reported as an AE via eCRF.

5.6.6. Immunogenicity

An individual sample is potentially examined multiple times in a hierarchical procedure to produce a sample ADA assay result and potentially a sample NAb assay result. A participant has TE ADA when ADAs are induced or boosted by exposure to study drug (i.e., when at least 1 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 greater than the minimum required dilution of 1:10 (if no ADAs were present at baseline).

Compound-level safety standards will be followed in the analyses of immunogenicity. Listings of immunogenicity assessments will be provided along with the summary of specified TEAEs by TE ADA status. The summary of TE ADA and NAb status will be produced for the Safety Population and All Active Treatments Safety Population, where the postbaseline period for reporting is the same as described for AEs in Section 5.6.2. Additional assessment of the relationship between immunogenicity and efficacy may be performed.

5.6.7. Special Safety Topics

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 AESIs relevant to these special safety topics will be identified by one or more SMQs, by a Lilly-defined MedDRA PT listing based upon the review of the most current version of MedDRA, 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 populations during their associated study periods using the baseline and postbaseline definitions described in Sections 5.6.2 and 5.6.3.

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. In the event that the listing of terms or analysis changes for a special safety topic it will be documented in the compound level safety standards which will supersede this document; it will not warrant an amendment to the individual study SAP.

5.6.7.1. Hepatic Safety

Analyses for laboratory analyte measurements are described in Section 5.6.3. This section describes additional analyses for the topic.

Hepatic labs include ALT and AST, TBL and serum 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 participants with a measurement greater than or equal to 3 times (3X), 5 times (5X), and 10 times (10X) the performing lab ULN during the treatment period for all participants with a postbaseline value and for subsets based on various levels of baseline value.
- TBL and ALP: The percentages of participants with a measurement greater than or equal to 2 times (2X) the performing lab ULN during the treatment period will be summarized for all participants with a postbaseline value and for subsets based on various levels of baseline value.
- Plots of maximum postbaseline ALT versus maximum postbaseline TBL (entire safety population), maximum postbaseline AST versus maximum postbaseline TBL, maximum postbaseline ALP versus maximum postbaseline TBL.
- A listing of the information collected on the Hepatic-Safety eCRF.

5.6.7.2. Infections, Including Opportunistic Infections and Serious Infections

Infections will be defined using the PTs from the MedDRA Infections and Infestations SOC. TE infections will be analyzed for: all infections (by maximum severity), serious infections and OI. The MedDRA terms used to identify infections considered to be OI in participants with immune-mediated inflammatory conditions treated with immunomodulatory drugs are based on Winthrop and colleagues (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 OIs. Analyses will include:

- Infections/Serious Infections: TE Infections by PT.
- OIs: TE OI by narrow terms and broad terms separately.

5.6.7.3. Hypersensitivity Reactions

Hypersensitivity reactions is used as an overarching term to describe events that are systemic or localized reactions that likely have an allergic/hypersensitivity etiology. Participants will be evaluated by the investigator for signs and symptoms suggestive of hypersensitivity, and investigators will complete a follow-up eCRF designed to record additional information.

Potential hypersensitivity reactions will be determined using the following SMQs: anaphylactic reaction (SMQ 20000021), hypersensitivity (SMQ 20000214), and angioedema (SMQ 20000024). Potential hypersensitivity will be categorized as immediate (i.e., occurring within 24 hours from end of the study drug administration) and non-immediate (i.e., occurring after the day of study drug administration but prior to subsequent drug administration), based on the timing of the reaction.

Analyses will include:

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

5.6.7.4. Infusion/Injection Site Reactions

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

ISRs will be defined using the following MedDRA HLT:

- ISR, excluding certain PTs (e.g., those PTs related to joint), and
- ISR, excluding certain PTs (e.g., those PTs related to joint).

Analyses will include:

- TE ISRs by overall, HLT and PT.
- The additional data collected on the ISR follow-up form will be summarized in 2 distinct ways: at the participant level and at the event level. A by-participant listing of these data will be provided.

5.6.7.5. Cerebro-Cardiovascular Events

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

Analyses will include:

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

5.6.7.6. Malignancies

Malignancies will be defined using PTs from the Malignant tumors SMQ. Malignant tumor events will be summarized separately for the categories: NMSC and Malignancies, excluding NMSC.

Analyses will include:

- TE malignancy by category and PT, and
- by-participant listing for all participants having a TEAE of malignancy at any time

5.6.7.7. Depression

During the study, depression will be assessed prospectively by the investigator via signs and symptoms and the QIDS-SR16.

For QIDS-SR16, the shift tables will be provided showing the number and percentage of participants within each baseline category (maximum value) versus each postbaseline category (maximum value) by study intervention. Additionally, outcomes such as any increase in depression will be compared between study interventions (further described in the compound-level safety standards).

5.6.8. Safety Subgroup Analysis

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

5.7. Other Analyses

5.7.1. Health Outcomes/Quality of Life

The health outcome and quality of life measures including Urgency NRS, FACIT-Fatigue, EQ-5D-5L, WPAI-CD, SF-36, and IBDQ will be analyzed using methods described for measurements as described for efficacy measures in Section 1.1.

5.7.2. Efficacy Subgroup Analyses

Subgroup analyses will be conducted for all primary and major secondary endpoints in the PAS. The subgroups to be analyzed are listed in Appendix 4 (Section 6.4) along with the demographic characteristics. Additional subgroup analysis which are not based on baseline/demographic characteristics in Appendix 4 (Section 6.4) include TE anti-mirikizumab antibody status. Some additional subgroup analyses may be performed to meet regulatory requirements in specific countries. The analysis of additional subgroups will not require an amendment to the SAP.

Within each subgroup category, the proportion of responders by study intervention, study intervention differences, and 95% CIs will be displayed. Also, p-values using Fisher's exact test for study intervention comparison will be provided. Forest plots may be generated to display the odds ratios and 95% CIs for selected efficacy subgroup analyses.

A logistic regression model with study intervention, subgroup, and the interaction of subgroup-by-study intervention, and the covariates described in Section 1.2. The subgroup-by-study intervention interaction will be tested using the Firth correction (Firth 1993) at the significance level of 0.05. If any group within the subgroup is less than 10% of the total population, only summaries of the efficacy data will be provided (that is, no inferential testing). Changes to the variables used for the subgroup analysis will not require the SAP to be updated.

5.7.3. Analysis for Japan Submission

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

5.7.4. Analysis for China Submission

A subset of the planned efficacy, health outcomes, and safety analyses will be also reproduced based on participants from China sites, in support of the regulatory submission in China. The list of tables, listings, and figures for the participants from China sites will be in a separate document.

5.7.5. Protocol Deviations

Protocol deviations will be identified throughout the study. Important protocol deviations are defined as those deviations from the protocol likely to have a significant impact on the completeness, accuracy, and/or reliability of the study data or that may significantly affect a participant's rights, safety, or well-being.

A separate document known as the "The AMAM Trial Issues Management Plan" describes the categories and subcategories of important protocol deviations and the source of the deviation identified.

The number and percentage of participants having important protocol deviation(s) will be summarized within category and subcategory of deviations by study intervention. A by-participant listing of important protocol deviations will be provided.

5.7.6. Trial Impact by COVID-19 Pandemic, and Russia/Ukraine Crisis

Participants who experience impact by COVID-19 pandemic or Russia/Ukraine crisis will be summarized by the type of impact. Specific impacts may include protocol deviations, which contains out-of-window visits, treatment interruptions, treatment and/or study discontinuations, and missed visits. The summary will be provided for the overall mITT populations.

- listing of all randomized participants who discontinue study treatment due to COVID-19 pandemic
- listing of all study disruptions related to COVID-19 pandemic
- listing of AEs or deaths related to COVID-19 pandemic
- listing of important protocol deviations due to COVID-19 pandemic
- listing of all randomized participants who discontinue study treatment due to the Russia-Ukraine war

- listing of all study disruptions related to the Russia-Ukraine crisis
- listing of AEs or deaths related to the Russia-Ukraine crisis, and
- listing of important protocol deviations due to the Russia-Ukraine crisis.

5.8. Interim Analyses

5.8.1. Data Monitoring Committee (DMC)

A DMC consisting of members external to Lilly is established for interim safety monitoring across Studies AMAM and AMAX in participants with CD. This committee consists of 5 voting members, including a designated chairperson, 3 additional physicians with gastroenterology and/or clinical trial expertise, and a statistician.

No DMC member may have contact with study sites. A SAC is external to the mirikizumab team that may be Lilly employees or from third-party organization designated by Lilly. No member of the SAC will have contact with study sites. Access to the unblinding safety data will be limited to the DMC and the SAC or their designees. The study team will not have access to the unblinded data. The DMC will advise Lilly regarding continuing participant safety; however, the DMC may request key efficacy data to put safety observations into context and to assess a reasonable benefit/risk profile for ongoing participants in the studies. Study AMAM will not be stopped for positive efficacy. There will be no alpha adjustment for these interim assessments. Study sites will receive information about interim assessment ONLY if they need to know for the safety of their participants.

Details of the planned safety data analyses, the roles and responsibilities, and the data review process are included in the DMC Charter. Unblinding details are specified in a separate unblinding plan.

5.8.2. PK/PD Model Development

A limited number of preidentified Lilly employees or their designees who are not in direct contact with clinical sites may gain access to unblinded PK data, as specified in the unblinding plan prior to the Week 52 database lock in order to initiate the PK/PD model development process for the final analysis. Information that may unblind the study during the analyses will not be shared with study sites or the blinded study team until the primary database lock has occurred. Unblinding details can be found in the unblinding plan. PK/PD analysis details can be found in the Population PK/PD Analysis Plan.

5.8.3. Week 52 Database Lock

A primary database lock is planned after all participants have completed the Week 52 visit or the ETV. This is the final analysis for the primary efficacy objective of the study. However, the study may be ongoing for the posttreatment follow-up period for patients remaining in the study at the time of this database lock. All study site personnel (except where access to unblinded data is allowed by the protocol) and patients will remain blinded until the final database is complete and locked.

A final database lock will occur after all participants complete the study, including the Safety Posttreatment Follow-Up Period. This will be used for updating analysis with data from follow-up period.

6. Supporting Documentation

6.1. Appendix 1: Description and Derivation of Efficacy and Health Outcome Endpoints

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
SES-CD	The SES-CD is an endoscopic scoring system for CD based on 4 endoscopic variables (presence and size of ulcers, proportion of surface covered by ulcers, proportion of surface affected by disease, and presence and severity of stenosis), which are assessed in 5 ileocolonic bowel segments (ileum; right, transverse, and left colon; and rectum). Each of the 4 endoscopic variables is scored from 0 to 3: presence and size of ulcers (none = score 0; diameter 0.1 cm to 0.5 cm = score 1; 0.5 cm to 2 cm = score 2; >2 cm = score 3); extent of ulcerated surface (none = 0; <10% = 1; 10% to 30% = 2; >30% = 3); extent of affected surface (none = 0; <50% = 1; 50% to 75% = 2; >75% = 3); and presence and type of narrowing (none = 0; single, can be passed = 1; multiple, can be passed = 2; cannot be passed = 3). The grand total is obtained as the sum of all endoscopic scores across all bowel segments.	SES-CD total score	SES-CD total score is calculated as average of total scores from all readers.	Missing if endoscopy was not done, if 2 or more endoscopies were deemed unreadable, was not done within study period for Week 12 or was not done within 14 days after Week 52 visit	**NA**

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
	The endoscopic scores for each bowel segment are called subscores. Total scores range from 0 to 56, with higher scores indicating more severe disease.	Change from baseline in SES-CD total score	Change from baseline in SES-CD total score is calculated as SES-CD total score – baseline SES-CD total score	Missing if baseline or observed value is missing.	**NA**
		Endoscopic Response	Endoscopic response is defined as ≥50% improvement from baseline in SES-CD total score. If [100* (SES-CD total score – baseline SES-CD total score) / baseline SES-CD total score] ≤-50, then endoscopic response is achieved.	Missing if baseline or observed value is missing.	Y
					N

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		Endoscopic Response-sensitivity	Endoscopic response is defined as >50% improvement from baseline in SES-CD total score. If [100* (SES-CD total score – baseline SES-CD total score) / baseline SES-CD total score] <-50, then endoscopic response is achieved.	Missing if baseline or observed value is missing.	Y N
		Endoscopic remission SES-CD ≤4	Endoscopic remission SES-CD ≤4 is defined as SES-CD Total Score ≤4 and at least a 2-point reduction from baseline and no subscore >1. If SES-CD total score ≤4, SES-CD total score – baseline SES-CD total score ≤-2, and each SES-CD subscore ≤1, then endoscopic remission SES-CD ≤4 is achieved. SES-CD subscore is calculated as the average of observed subscores from all readers.	Missing if baseline or observed value is missing.	Y N

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		Durability of endoscopic response	This endpoint is achieved when endoscopic response is achieved at both Week 12 and Week 52.	Missing if baseline, or any postbaseline observed value is missing.	Y
		Durability of endoscopic remission	This endpoint is achieved when endoscopic remission is achieved at both Week 12 and Week 52.	Missing if baseline or any postbaseline observed value is missing.	Y
		Endoscopic remission SES-CD 0-2	Endoscopic remission SES-CD 0-2 is defined as a SES-CD Total Score of 0-2.	Missing if endoscopy was not done or if 2 or more endoscopies were deemed unreadable.	Y

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
CDAI	CDAI is an 8-item disease activity measure comprised of a composite of 3 patient-reported and 5 physician-reported/laboratory items (physical signs and a laboratory parameter [hematocrit]). Participant responses are summed over a 7-day period and all items are subsequently weighted.	CDAI total score	CDAI total score is based on the CDAI questionnaire in Appendix 7 (Section 6.7). It also utilizes the standard weights table in that section. For the patient-reported items, the most recent 7 days are included (possibly nonconsecutive) out of the 12 days prior to the day of the visit (see Appendix 9 [Section 6.9] for details), after removing the day(s) of the endoscopy prep, the day of endoscopy procedure, and the 2 days following the endoscopy procedure.	CDAI total score is missing if any of below is missing: Patient-reported items are missing if less than 4 days of data are available. Physician-reported questionnaire is not answered, or section 5/6 is missing. Note that if none of the options in Section 4 of the CDAI questionnaire is checked and questionnaire is answered, it will be assumed that no extra-intestinal manifestations were present.	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
				Hematocrit is missing if no central or local lab is collected from the preceding visit up to 7 days post visit date and within the study period. Weight is missing if no measure is done from the preceding visit up to 7 days post visit date and within the study period.	
		Change from baseline in CDAI total score	Change from baseline in CDAI score is calculated as CDAI score – baseline CDAI score.	Missing if CDAI total score is missing at either baseline or time point of interest.	NA
		Clinical remission by CDAI	Clinical remission by CDAI is defined as a CDAI total score <150.	Missing if CDAI total score is missing at time point of interest and the score of available items <150.	Y
					N
		Clinical response by CDAI	Clinical response by CDAI is defined as a decrease from baseline	Missing if CDAI total score is missing at baseline or at time	Y

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
			in the CDAI total score ≥ 100 and/or a CDAI total score <150.	point of interest, while the CDAI total score at time point of interest is >150.	N
	Number of liquid or very soft stools per Bristol Stool Scale Category 6 or 7 is the first item reported by patient in CDAI (Appendix 7 [Section 6.7])	SF average	SF average will be calculated by averaging the most recent 7 days (possibly nonconsecutive) out of the 12 days prior to the day of the visit (see Appendix 9 [Section 6.9] for details), after removing the day(s) of the endoscopy prep, the day of endoscopy procedure, and the 2 days following the endoscopy procedure.	Missing if less than 4 days of data are available	NA
		Change from baseline in SF average	Calculated as observed SF average minus baseline SF average	Missing if baseline or time point of interest is missing	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
	AP is one of the patient-reported items in CDAI (Appendix 7 [Section 6.7]) AP score is classified as 0 = none, 1 = mild, 2 = moderate, 3 = severe.	AP average	AP average will be calculated by averaging the most recent 7 days (possibly nonconsecutive) out of the 12 days prior to the day of the visit (see Appendix 9 [Section 6.9] for details), after removing the day(s) of the endoscopy prep, the day of endoscopy procedure, and the 2 days following the endoscopy procedure.	Missing if less than 4 days of data are available	NA
		Change from baseline in AP average	Calculated as observed AP average minus baseline AP average	Missing if baseline or time point of interest is missing	NA
	PRO comprises of stool frequency (SF) per Bristol Stool Scale Category 6 or 7, and abdominal pain (AP) based on the 0-3 scale of CDAI (Appendix 7 [Section 6.7])	Clinical remission by PRO	Clinical remission by PRO is defined as a SF average ≤3 and AP average ≤1 with both values no worse than baseline.	Missing if SF average or AP average is missing at either baseline or time point of interest and available SF average and AP average at time point of interest are not worse than baseline	Y
					N

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		Clinical response by PRO	Clinical response by PRO is defined as at least a 30% decrease in SF or AP, and neither worse than baseline. If [100 * (SF average – baseline SF average) / baseline SF average] ≤ -30 or [100*(AP average – baseline AP average) / baseline AP average] ≤ -30, and SF average ≤ baseline SF average and AP average ≤ baseline AP average, then clinical response by PRO is achieved.	Missing if SF average or AP average is missing at either baseline or time point of interest and available SF average and AP average at time point of interest are not worse than baseline.	Y
					N
		Stability of clinical remission by CDAI	Stability of clinical remission by CDAI is defined as achieving clinical remission by CDAI in at least 80% of the visits from Week 12 to Week 52 (at least 9 of 11 visits)	Missing if clinical remission by CDAI is missing for more than 2 visits from Week 12 to Week 52.	Y

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		Corticosteroid-free clinical remission by CDAI	Corticosteroid-free clinical remission by CDAI is defined as achieving clinical remission by CDAI and being corticosteroid-free from Week 40 to Week 52.	Missing if clinical remission by CDAI is missing at timepoint of interest	Y
					N
Composite SES-CD and CDAI endpoints	See SES-CD and CDAI sections above.	Endoscopic response + clinical response by CDAI	This endpoint is achieved when both endoscopic response and clinical response by CDAI are achieved.	Missing if either endoscopic response or clinical response by CDAI are missing at time point of interest	Y
		Endoscopic remission SES-CD ≤4 + clinical remission by CDAI	This endpoint is achieved when both endoscopic remission SES-CD ≤4 and clinical remission by CDAI are achieved.	Missing if either endoscopic remission SES-CD ≤4 or clinical remission by CDAI are missing at time point of interest	Y
		Endoscopic response + clinical remission by CDAI	This endpoint is achieved when both endoscopic response and clinical remission by CDAI are achieved.	Missing if either endoscopic response or clinical remission by CDAI are missing at time point of interest.	Y

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
Urgency NRS	The Urgency NRS is a single patient-reported item that measures the severity for the urgency (sudden or immediate need) to have a bowel movement in the past 24 hours using an 11-point NRS ranging from 0 (no urgency) to 10 (worst possible urgency).	Urgency NRS Score	Urgency NRS score is calculated by averaging the most recent 7 days (possibly nonconsecutive) out of the 12 days prior to the day of the visit (see Appendix 9 [Section 6.9] for details), after removing the day(s) of the endoscopy prep, the day of endoscopy procedure, and the 2 days following the endoscopy procedure.	Missing if less than 4 days of data are available	NA
		Change from baseline in urgency NRS score	Calculated as observed Urgency NRS minus baseline Urgency NRS	Missing if baseline or time point of interest is missing	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		Urgency NRS clinically meaningful within-participant Improvement	Urgency NRS clinically meaningful within-participant improvement ≥3-point improvement is defined as a decrease from baseline in the urgency NRS score of ≥3 points. If urgency NRS score – baseline urgency NRS score ≤-3, then urgency NRS ≥3-point improvement is achieved.	Missing if baseline or observed value is missing	Y
		Urgency NRS ≤2 point	This endpoint is achieved when Urgency NRS is ≤2 point	Missing if Urgency NRS score is missing at time point of interest	Y
FACIT-Fatigue	The FACIT-Fatigue is a 13-item instrument developed to measure fatigue in chronic illness participants. It has been validated for use in IBD participants. Total score ranges from 0 to 52 based on a rating of 4-point Likert scale. Higher scores are better.	FACIT-Fatigue Score	All responses in the questionnaire are added with equal weight to obtain the total score. For additional details see Appendix 8 (Section 6.8).	Missing if more than 50% of items (i.e. 8 or more out of 13) are missing. If 7 or less items are missing, the total score is prorated from the score of the answered items.	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		Change from baseline in FACIT-Fatigue score	Change from baseline in FACIT-Fatigue score is calculated as FACIT-Fatigue score minus baseline FACIT-Fatigue score.	Missing if baseline or observed value is missing.	NA
		FACIT-Fatigue thresholds of clinically meaningful within-participant improvement	Change from baseline of FACIT-Fatigue ≥6; ≥7; ≥8; ≥9 points (4 thresholds, respectively). Endpoints are achieved when FACIT-Fatigue score – baseline FACIT-Fatigue score is equal to or greater than the candidate thresholds. .	Missing if FACIT-Fatigue score is missing at time point of interest or at baseline.	Y
IBDQ	IBDQ is 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."	IBDQ score	IBDQ total score is 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.	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		Change from baseline in IBDQ total score	Calculated as IBDQ total score minus baseline IBDQ total score	Missing if baseline or time point of interest is missing	NA
		Change from baseline in IBDQ subscore	Calculated as IBDQ subscore minus baseline IBDQ subscore	Missing if baseline or time point of interest is missing	NA
		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.	NA
		Systemic symptoms subscore	Calculated as the sum of Questions 2, 6, 10, 14, 18.		NA
		Emotional function subscore	Calculated as the sum of Questions 3, 7, 11, 15, 19, 21, 23, 25, 27, 30, 31, 32.		NA
		Social function subscore	Calculated as the sum of Questions 4, 8, 12, 16, 28.		NA
		IBDQ response	≥16 point improvement from baseline in IBDQ score as described by Irvine et al. (1996).	Missing if either baseline or observed value is missing.	Y
		IBDQ remission	IBDQ score ≥170 as described by Irvine (2008).	Missing if the IBDQ score is missing	Y

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
PGRS	PGRS is a 1-item patient-rated questionnaire designed to assess the participants' 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 participant has no symptoms (that is, "none") and a score of 6 indicates that the participant's symptom are "very severe."	PGRS score	The average will be calculated using daily diary data from the most recent 7 days (possibly nonconsecutive) out of the 12 days prior to the day of the visit (see Appendix 9 [Section 6.9] for details), after removing the day(s) of the endoscopy prep, the day of endoscopy procedure, and the 2 days following the endoscopy procedure.	Missing if less than 4 days of data are available	NA
		Change from baseline in PGRS	Calculated as Observed PGRS – baseline PGRS	Missing if baseline or observed value is missing	NA
PGIC	PGIC is a patient-rated instrument designed to assess the participants' 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 participant's symptom is "very much better," a score of 4 indicates that the participant's symptom has experienced "no change," and a score of 7 indicates that the participant's symptom is "very much worse."	PGIC score	Observed score is used. No additional derivation	Missing if score is missing	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
EQ-5D-5L	The EQ-5D-5L is a standardized measure of health status used to provide a simple, generic measure of health for clinical and economic appraisal. The EQ-5D-5L consists of 2 components: a descriptive system of the respondent's health and a rating of his/her current health state using a 0- to 100-mm VAS. The descriptive system comprises the following 5 dimensions: Item 1: mobility Item 2: self-care Item 3: usual activities Item 4: pain/discomfort Item 5: anxiety/depression The respondent is asked to indicate his/her health state by ticking (or placing a cross) in the box associated with the most appropriate statement in each of the 5 dimensions.	EQ-5D-5L index score	For each of the 5 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. The index uses the concatenation of the value of each EQ-5D-5L dimension score in the order: Item 1, Item 2, Item3; Item 4; Item 5.	If any of the items is missing or equal to 9, the index score is missing	NA
			Index score is calculated based on the responses to the 5 dimensions, providing a single value on a scale from less than 0 (where zero is a health state equivalent to death; negative values are valued as worse than dead) to 1 (perfect health). The UK		

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
			algorithm is used (Szende et al. 2007).		
		Change from baseline of EQ-5D-5L index scores	Calculated as Observed score – baseline score	Missing if baseline or observed value is missing	NA
		EQ-5D VAS	Range from 0 = "worst imaginable health state" to 100 = "best imaginable health state"	Single item, missing if missing	NA
		Change from baseline of EQ-5D VAS	Calculated as Observed score – baseline score	Missing if baseline or observed value is missing	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
SF-36	The SF-36 Version 2 is a 36-item, patient-completed measure designed to be a short, multipurpose assessment of health (The SF Community – SF-36 Health Survey Update). The summary scores range from 0 to 100, with higher scores indicating better levels of function and/or better health. Items are answered on Likert scales of varying lengths. The SF-36 comprises 8 domain scores and 2 overarching component scores. SF-36 domain scores are: (1) Physical functioning, (2) Role-physical, (3) Role-emotional, (4) bodily pain, (5) vitality, (6) social functioning, (7) mental health and (8) general health. The component scores are: (1) the PCS and (2) MCS.	SF-36 Domain scores and SF-36 Component Scores	Per copyright owner, the Quality Metric Health Outcomes™ Scoring Software will be used to derive SF-36 domain and component scores. After data quality-controls, the SF-36 software will re-calibrate the item-level responses for calculation of the domain and component scores. These raw scores will be transformed into the domain scores (t-scores) using the 4-week recall period. This entails exporting the participant 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.	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		SF-36 change from baseline for domain and component score.	Calculated as observed SF-36 score – baseline SF-36 score.	Missing if baseline or observed value is missing.	NA
		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.	Y
		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.	Y
WPAI-CD	WPAI-CD is a patient-reported instrument developed to measure the impact on work productivity and regular activities attributable to a specific health problem (WPAI-CD). 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. Four scores are calculated from the responses to	Employment Status	Yes/No	Missing if question is missing	NA
		Absenteeism Score (%)	$\frac{Q2}{(Q2+Q4)} \times 100$	Missing if Q2 or Q4 are missing. Also missing if Employment Status is No.	NA
		Presenteeism Score (%)	$\frac{Q5}{10} \times 100$	Missing if Q5 is missing. Also missing if Employment Status is No.	NA
		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.	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
	these 6 items: absenteeism, presenteeism, work productivity loss, and activity impairment (Reilly Associates [WWW]). Scores are calculated as impairment percentages (Reilly et al. 1993), with higher numbers indicating greater impairment and less productivity (Reilly Associates [WWW]); that is, worse outcomes. Patients will record their responses to the WPAI-CD electronically as source data in the tablet device at appropriate visits.	Activity Impairment Score (%)	$\frac{Q6}{10} \times 100$	Missing if Q6 is missing. May still be present and nonmissing if patient is unemployed.	NA
EIMs	EIMs will be collected in the eCRF and include: a. arthritis, arthralgia; b. iritis, uveitis; c. erythema nodosum, pyoderma gangrenosum, aphthous, stomatitis.	EIMs Count	EIMs count will be derived by summing the number of EIMs defined by subcategories a-c.	NA	NA
		Resolution of Baseline EIMs	No EIMs at Week 12 and Week 52	NA	Y
					N
		New EIMs	New EIMs at Week 12 and Week 52	NA	Y
					N

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
Fistulae	Draining cutaneous and draining rectal/vaginal fistulae will be collected in the eCRF	Number of draining cutaneous fistulae	Draining fistulae count will be calculated by adding the number of draining cutaneous fistulae.	NA	NA
		At least 50% reduction in draining cutaneous fistulae	At least 50% reduction in draining cutaneous fistulae is defined as percent change from baseline ≤ -50	NA	Y
		Draining rectal/vaginal fistulae	Yes/No	NA	Y
		Closure of draining cutaneous fistulae	The number of participants that have a count of zero for draining cutaneous fistulae at endpoint will be considered as achieving closure of draining cutaneous fistulae	NA	Y
Medical resource utilization and health economics	Crohn's related ER visits, hospitalizations, surgeries related to Crohn's disease will be collected in the eCRF.	Crohn's-related ER visits count	The number of Crohn's-related ER visits will be calculated up to Week 12 and Week 52.	No imputation	NA
		Crohn's-related ER visits	Participants that have Crohn's-related ER visits up to Week 12 and Week 52	NA	Y

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		Crohn's related hospitalizations count	The number of hospitalizations will be calculated up to Week 12 and Week 52.	No imputation	NA
		Crohn's related hospitalizations	Participants that have Crohn's-related hospitalizations up to Week 12 and Week 52.	NA	Y
		Crohn's related surgeries count	The number of Crohn's-related surgeries will be calculated up to Week 12 and Week 52.	No imputation	NA
		Crohn's related surgeries	Participants that have Crohn's-related surgeries up to Week 12 and Week 52	NA	Y

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
Biomarkers	CRP is a biomarker of inflammation.	CRP	Lab value. May be transformed if needed	Single lab value. Missing if missing.	NA
	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.	NA
[redacted]					NA
					NA
					NA
					NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
	[redacted]				NA
					NA
					NA
					NA
					NA
					NA
					NA
					Y

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
		CCI			N
					Y
					N
DSI-CD (collect at baseline only)	DSI-CD is clinician's reported 16-item measurement scored by reviewing patient symptoms, physical assessment, labs, medications, physical activity, and pain. Scores range from 0 to 100, with a higher score indicating worse disease severity.	DSI-CD score	DSI-CD score is calculated as the sum of all questions.	Missing if score is missing	NA

Measure	Description	Variable	Derivation / Comment	Definition of Missing	Composite Endpoint from Week 16 to Week 52 with Clinical Response by PRO at Week 12
IBD-DI (collect at baseline only)	IBD-DI is a 15-item patient reported questionnaire developed to measure IBD-related disability by reviewing the patient's symptoms, mood and overall well-being over the past 7 days. Total scores range from 0 to 100, with higher scores representing greater levels of IBD-related disability. Patients will record their responses to the IBD-DI electronically as source data in the tablet device at the appropriate visit.	IBD-DI score	IBD-DI score is calculated as S*100/n*4, where n = number of questions which have been answered and S = sum of the n questions score. Questions 2 and 3 are combined as 1 question: "Yes" if either Question 2 or Question 3 is "Yes"; "No" if both Question 2 and 3 are "No"; or else is missing. (Appendix 11 [Section 6.11])	Missing if (14-n)/14 ≥20%	NA

Abbreviations: ADaM = Analysis Data Model; AP = abdominal pain; CDAI = Crohn's Disease Activity Index; CD = Crohn's disease; eCRF = electronic case report form; CRP = C-reactive protein; CSV = Comma Separated Values; DSI-CD = Disease Severity Index-Crohn's Disease; EIM = extraintestinal manifestation; EQ-5D-5L = European Quality of Life 5–Dimension 5 Level; ER = emergency room; FACIT-Fatigue = Functional Assessment of Chronic Illness Therapy–Fatigue; CCI IBD = inflammatory bowel disease; IBD-DI = Inflammatory Bowel Disease-Disability Index; IBDQ = Inflammatory Bowel Disease Questionnaire; MCID = minimal clinical important difference; MCS = mental component summary; NRS = numeric rating scale; PCS = physical component summary; PGRS = Patient Global Rating of Severity; PGIC = Patient Global Impression of Change; PRO = patient-reported outcomes; Q = Question; CCI ; SDTM = Study Data Tabulation Model; SES-CD = Simple Endoscopic Score for Crohn's Disease; SF = stool frequency; SF-36 = 36-Item Short Form Health Survey; VAS = Visual Analog Scale; WPAI-CD = Work Productivity and Activity Impairment Questionnaire Crohn's Disease.

6.2. Appendix 2: Description of Analyses

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
Composite CDAI & SES-CD	Clinical response by PRO at Week 12 and endoscopic response (Week 52)	Primary Efficacy Analysis Set (PAS) Miri vs. pbo	CMH analysis with NRI	Co-primary / MCP
		PAS Miri vs. pbo	Logistic regression analysis with NRI	Sensitivity
		PAS Miri vs. pbo	CMH analysis with mNRI	Sensitivity
		mITT, , PAS – exclude participants impacted by crisis, PAS –Not-Biologic-Failed, PAS – Biologic Failed	CMH analysis with NRI	Sensitivity, other secondary
		PAS Miri vs. pbo	CMH analysis with tipping point analysis	Sensitivity
	Clinical response by PRO at Week 12 and endoscopic response-sensitivity (>50% improvement SES-CD) (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Sensitivity
	Clinical response by PRO at Week 12 and endoscopic remission ≤4 (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Major secondary / MCP in Group 1
		PAS Miri vs. pbo	Logistic regression analysis with NRI	Sensitivity
		PAS Miri vs. pbo	CMH analysis with mNRI	Sensitivity
		PAS – exclude participants impacted by crisis, PAS – Not-Biologic-Failed, PAS – Biologic Failed Miri vs. pbo	CMH analysis with NRI	Sensitivity, Other secondary
	Clinical response by PRO at Week 12 and endoscopic remission 0-2 (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
	Endoscopic response and clinical response by CDAI (Week 12)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
	Endoscopic response and clinical remission by CDAI (Week 12)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12, Endoscopic remission and clinical remission by CDAI (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12, endoscopic response and clinical remission by CDAI (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12, durability of endoscopic response (Week 12 & Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12, durability of endoscopic remission ≤ 4 (Week 12 & Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
CDAI	Clinical response by PRO at Week 12 and clinical remission by CDAI (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Co-primary / MCP
		PAS Miri vs. pbo	Logistic regression analysis with NRI	sensitivity
		mITT, PAS –exclude participants impacted by crisis, PAS – Not-Biologic-Failed, PAS – Biologic Failed Miri vs. pbo	CMH analysis with NRI	Sensitivity, Other secondary

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
		PAS Miri vs. pbo	CMH analysis with tipping point analysis	sensitivity
	Clinical remission by CDAI (Week 12) (Week 52)	PAS Miri vs. pbo Miri vs. uste	CMH analysis with NRI	Major secondary/ MCP in Group 1, Major secondary/ MCP in Group 2
			Logistic regression analysis with NRI	Sensitivity
		PAS Miri vs. pbo	CMH analysis with NRI (alternative estimand for placebo NRs at Week 12)	Sensitivity
		PAS – exclude participants impacted by crisis Miri vs. pbo	CMH analysis with NRI	Sensitivity
		PAS – Not-Biologic-Failed, PAS – Biologic Failed Miri vs. pbo Miri vs. uste	CMH analysis with NRI	Other secondary
		PAS Miri vs. pbo	CMH analysis with tipping point analysis at Week 52	supplementary
		Always clinical responders Miri vs. pbo	Principal stratum analysis for always clinical responder treatment effect	supplementary
	Clinical response by CDAI (Week 4) (Week 12)	PAS Miri vs. pbo Miri vs. uste	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12 and clinical remission by PRO (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Major secondary / MCP in Group 1
		PAS – exclude participants impacted by crisis	CMH analysis with NRI	Sensitivity

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
		Miri vs. pbo		
	Clinical response by PRO (Week 12) (Week 52)	PAS Miri vs. pbo Miri vs. uste	CMH analysis with NRI	Major secondary / MCP in Group 1 Other secondary
	Clinical remission by PRO (Week 12) (Week 52)	PAS Miri vs. pbo Miri vs. uste	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12 and clinical response by CDAI (Week 52)	PAS Miri vs pbo	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12 and clinical response by PRO (Week 52)	PAS Miri vs pbo	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12 and stability of clinical remission by CDAI	PAS Miri vs pbo	CMH analysis with NRI	Other secondary
	CDAI total score, change from baseline in CDAI total score	PAS Miri vs. pbo Miri vs. uste	MMRM; ANCOVA with mBOCF	supplementary
	AP average, change from baseline in AP average	PAS Miri vs. pbo Miri vs. uste	MMRM; ANCOVA with mBOCF	supplementary
	SF average, change from baseline in SF average	PAS Miri vs. pbo Miri vs. uste	MMRM; ANCOVA with mBOCF	supplementary
	Clinical response by PRO at Week 12 and corticosteroid-free clinical remission by CDAI (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Major secondary / MCP in Group 1

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
		PAS - participants with baseline corticosteroid use Miri vs. pbo	CMH analysis with NRI	Other secondary
		PAS – exclude participants impacted by crisis Miri vs. pbo	CMH analysis with NRI	Sensitivity
	Corticosteroid-free clinical remission by CDAI (Week 52)	PAS Miri vs. uste	CMH analysis with NRI	Other secondary
SES-CD	Endoscopic response (Week 12) (Week 52)	PAS Miri vs. pbo Miri vs. uste	CMH analysis with NRI	Major Secondary / MCP in Group 1 Major Secondary / MCP in Group 2 Other secondary
			Logistic regression analysis with NRI	sensitivity
		PAS Miri vs. pbo	CMH analysis with NRI (alternative estimand for placebo NRs at Week 12)	Sensitivity
		PAS – exclude participants impacted by crisis (Week 12)	CMH analysis with NRI	Sensitivity
		PAS – Not-Biologic-Failed, PAS – Biologic Failed Miri vs. pbo Miri vs. uste	CMH analysis with NRI	Other secondary
		PAS Miri vs. pbo	CMH analysis with tipping point analysis at Week 52	supplementary

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
		Always clinical responders Miri vs. pbo	Principal stratum analysis for always clinical responder treatment effect	supplementary
	Endoscopic response-sensitivity (>50% improvement SES-CD) (Week 12) (Week 52)	PAS Miri vs. pbo Miri vs. uste	CMH analysis with NRI	Sensitivity
	Endoscopic remission SES-CD ≤4 (Week 12) (Week 52)	PAS Miri vs. Pbo Miri vs. Uste	CMH analysis with NRI	Major secondary/ MCP in Group 1 Other secondary
			Logistic regression analysis with NRI	sensitivity
		PAS – exclude participants impacted by crisis Miri vs. pbo (Week 12) Miri vs. uste (Week 52)	CMH analysis with NRI	Sensitivity
		PAS – Not-Biologic-Failed, PAS – Biologic Failed Miri vs. pbo Miri vs. uste	CMH analysis with NRI	Other secondary
	Endoscopic remission SES-CD 0-2 (Week 12)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
	SES-CD total score, change from baseline in SES-CD total score	PAS Miri vs. pbo Miri vs. uste	ANCOVA with mBOCF	supplementary
Urgency NRS	Change from baseline in Urgency NRS (Week 12)	PAS Miri vs. pbo	MMRM ANCOVA with mBOCF	Other secondary

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
		PAS – exclude participants impacted by crisis Miri vs. pbo	MMRM	Sensitivity
	Change from baseline in Urgency NRS (Week 52)	PAS Miri vs. pbo	MMRM; ANCOVA with mBOCF	Other secondary
	Change from baseline in Urgency NRS (Week 52)	Always clinical responders Miri vs pbo	Principal stratum analysis for always clinical responder treatment effect	supplementary
	Urgency NRS	PAS Miri vs. pbo	MMRM; ANCOVA with mBOCF	supplementary
	Clinical response by PRO at Week 12 and Urgency NRS ≤2 (Week 12) (Week 52)	PAS – participants with baseline Urgency NRS ≥ 3) Miri vs. pbo	CMH analysis with NRI	Other secondary
		PAS – exclude participants impacted by crisis Miri vs. pbo	CMH analysis with NRI	Sensitivity
	Urgency NRS clinically meaningful within-participant improvement	PAS Miri vs. pbo	CMH analysis with NRI	supplementary
FACIT- Fatigue	Change from baseline in FACIT-Fatigue score (Week 12) (Week 52)	PAS Miri vs. pbo	ANCOVA with mBOCF	Major secondary/ MCP in Group 1 Other secondary
	FACIT-Fatigue score	PAS Miri vs. pbo	ANCOVA with mBOCF	Supplementary
	FACIT-Fatigue clinically meaningful within-participant improvement	PAS Miri vs. pbo	CMH analysis with NRI	supplementary
IBDQ	IBDQ total score	PAS Miri vs. pbo	ANCOVA with mBOCF	Other secondary

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
	Change from baseline in IBDQ total score (Week 12) (Week 52)			
	IBDQ bowel symptoms subscore Change from baseline in IBDQ bowel symptoms subscore IBDQ systemic symptoms subscore Change from baseline in IBDQ systemic symptoms subscore IBDQ emotional function subscore Change from baseline in IBDQ emotional function subscore IBDQ social function subscore Change from baseline in IBDQ social function subscore	PAS Miri vs. pbo	ANCOVA with mBOCF	Supplementary
	IBDQ response (Week 12)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12 and IBDQ response (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
	IBDQ remission (Week 12)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
	Clinical response by PRO at Week 12 and IBDQ remission (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
EQ-5D-5L	EQ-5D-5L index score Change from baseline in EQ-5D-5L index score (Week 12) (Week 52)	PAS Miri vs. pbo	ANCOVA with mBOCF	Other secondary
SF-36	SF-36 component scores Change from baseline in SF-36 component scores SF-36 domain scores Change from baseline in SF-36 domain scores	PAS Miri vs. pbo	ANCOVA with mBOCF	Other secondary
	SF-36 PCS MCID response SF-36 MCS MCID response (Week 12)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
	Clinical response by PRO at Week 12 and SF-36 PCS MCID response Clinical response by PRO at Week 12 and SF-36 MCS MCID response (Week 52)	PAS Miri vs. pbo	CMH analysis with NRI	Other secondary
WPAI-CD	WPAI-CD Scores Change from baseline in WPAI-CD score (Week 12) (Week 52)	PAS – participant with baseline Employment Status of Yes Miri vs. pbo	ANCOVA with mBOCF	Other secondary
Biomarkers	CRP (log-transformed) Change from baseline in CRP (log-transformed)	PAS Miri vs. pbo	MMRM; ANCOVA with mBOCF	Other secondary

Measure	Endpoint / Variable (Time Point of Interest)	Population Group Comparison(s)	Analysis Method (Section 5.1.1)	Type of Analysis
	Fecal calprotectin (log-transformed) Change from baseline in fecal calprotectin (log-transformed)	PAS Miri vs. pbo	MMRM; ANCOVA with mBOCF	Other secondary
EIMs	Proportion of participants with EIMs resolution	PAS – in participants with EIMs at baseline Miri vs pbo	CMH analysis	Other secondary
	Proportion of participants with new EIMs	PAS – in participants without EIMs at baseline Miri vs pbo	CMH analysis	Other secondary
Fistula	Proportion of participants with at least 50% reduction in draining cutaneous fistulae	PAS – in participants with draining cutaneous fistulae at baseline Miri vs pbo	CMH analysis with NRI	Other secondary
	Proportion of participants with closure of draining cutaneous fistulae	PAS – in participants with draining cutaneous fistulae at baseline Miri vs pbo	CMH analysis with NRI	Other secondary
Others	Proportion of participants had Crohn's-related emergency room visit	PAS Miri vs pbo	CMH analysis	Other secondary
	Proportion of participants had Crohn's-related hospitalization	PAS Miri vs pbo	CMH analysis	Other secondary
	Proportion of participants had Crohn's-related surgeries	PAS Miri vs pbo	CMH analysis	Other secondary

CCI

Abbreviations: ANCOVA = analysis of covariance; AP = abdominal pain; CDAI = Crohn's Disease Activity Index; CMH = Cochran–Mantel–Haenszel; CRP = C-reactive protein; EIM = extraintestinal manifestation; EQ-5D-5L = European Quality of Life 5–Dimension 5 Level; FACIT-Fatigue = Functional Assessment of Chronic Illness Therapy-Fatigue; IBDQ = Inflammatory Bowel Disease Questionnaire; mBOCF = modified baseline observation carried forward; NR = non-responder; NRI = non-responder imputation; NRS = numeric rating scale; MCID = minimal clinical important difference; MCP = multiple comparisons procedure; MCS = mental component summary; miri = mirikizumab; mITT = modified intent-to-treat population; mNRI = modified non-responder imputation; MMRM = mixed effects model of repeated measures; PAS = Primary Analysis Set; pbo = placebo; PCS = physical component summary; PRO = patient-reported outcome; SES-CD = Simple Endoscopic Score for Crohn's Disease; SF = stool frequency; SF-36 = 36-Item Short Form Health Survey; WPAI-CD = Work Productivity and Activity Impairment Questionnaire Crohn's Disease; uste = ustekinumab.

6.3. Appendix 3: Changes to Protocol-Planned Analyses

As compared with protocol amendment (e), the SAP Version 2 content differs in several respects. The analysis and endpoints describe in the SAP will supersede the language in the protocol.

- The definition of which endpoints will be designated as a "major secondary" vs. an "other secondary" endpoint has changed due to business considerations. The SAP Version 2 with these changes was approved prior to the primary database lock.
    - The comparisons of mirikizumab and placebo on change from baseline in Urgency NRS at Week 12 and Week 52 are changed from major secondary objectives to other secondary objectives.
    - The comparison of mirikizumab and placebo on change from baseline in FACIT-Fatigue scores at Week 12 is changed from another secondary objective to a major secondary objective.
    - The comparison of mirikizumab and placebo on "Clinical Response by PRO at Week 12 and either Clinical Remission by CDAI at Week 52 or endoscopic remission SES-CD ≤4 at Week 52 and also corticosteroid-free from Week 40 to Week 52" is replaced by "clinical response by PRO at Week 12 and clinical remission by CDAI at Week 52 and also corticosteroid-free from Week 40 to Week 52" as a major secondary objective.
    - The comparison of mirikizumab and placebo on Clinical Response by PRO at Week 12 is changed from another secondary objective to a major secondary objective.
    - The comparison of mirikizumab and ustekinumab on endoscopic remission SES-CD ≤4 at Week 52 is changed from a major secondary objective to another secondary objective.
- The primary analysis for continuous endpoints was changed from MMRM analysis to ANCOVA. This change was introduced based on the more detailed discussion of estimands described in the Section 1.1. It was deemed that the ANCOVA methodology would be better suited to the return to baseline approach which is incorporated into the estimand.

6.4. Appendix 4: Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized descriptively by study intervention group for ITT, mITT and PAS; no testing will be performed for baseline characteristics. For continuous measures, summary statistics will include sample size, mean, standard deviation, median, minimum, and maximum. For categorical measures, summary statistics will include sample size, frequency, and percentages.

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)	
	No	Asia, North America, Central America/South America, 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/m^2), Normal (≥18.5 and <25 kg/m^2), Overweight (≥25 and <30 kg/m^2), Obese (≥30 and <40 kg/m^2), Extreme obese (≥40 kg/m^2)	X
Tobacco use	No	Never, Current, Former	X
Prior CD Therapy			
Prior biologic exposure	No	Ever used, Never used	X
Prior biologic failure[d]	No	Ever, Never	X
Inadequate response or loss of response to a biologic	No	Ever, Never	X
Inadequate response to a biologic	No	Ever, Never	
Loss of response to a biologic	No	Ever, Never	
Intolerance to a biologic	No	Ever, Never	
Number of prior biologics used	No	0, 1, 2, >2	
Number of failed[d] biologics	No	0, 1, 2, >2	
Prior biologic failure[d] and prior biologic exposures	No	Not exposed, Exposed but not failed, Exposed and failed at least one	
Prior anti-TNF failure[d, e]	No	Ever, Never	X

Variable	Continuous Summary	Categorical Summary	Subgroup Analysis[a]
Number of failed (unique) prior anti-TNFs[d, e]	No	0,1, 2, >2	
Prior anti-integrin failure[d, f]	No	Ever, Never	X
Prior corticosteroid failure	No	Ever, Never	
Prior immunomodulator failure[d, g]	No	Ever, Never	
Baseline CD Therapies			
Baseline corticosteroid use	No	Yes, No	X
Baseline prednisone equivalent dose	Yes	None	
Budesonide	No	Yes, No	
Baseline immunomodulator use[g]	No	Yes, No	X
Baseline corticosteroid and immunomodulator use[g]	No	Corticosteroid only, Immunomodulator, Neither, Both	
Baseline use of oral aminosalicylates	No	Yes, No	
Baseline use of methotrexate	No	Yes, No	
Baseline use of thiopurine	No	Yes, No	
Baseline Disease Characteristics			
Duration of CD[h]	Yes	<1 year, ≥1 to <5 years, ≥5 year	X
Age at Diagnosis of CD[i]	Yes	<10 year, ≥10 to <17 years, ≥17 year to <40 years, ≥40 years	
Baseline Disease Location[j]	No	Ileal, Colonic, Ileal-colonic	X
History of Surgical Bowel Resection	No	Yes, No	
Number of Surgical Bowel Resection	Yes	None	
Baseline Fecal Calprotectin	Yes	≤250 µg/g, >250 µg/g	X
Baseline CRP	Yes	≤10 mg/L, >10 mg/L	X
Baseline SES-CD total score	Yes	SES-CD (< 12, ≥ 12)	X
Baseline AP average	Yes	AP average (< 2, ≥ 2)	X
Baseline SF average	Yes	SF average (< 7, ≥ 7)	X
Baseline CDAI	Yes	CDAI total score (<300, ≥ 300)	X
		CDAI total score (<150, ≥150 to 220, ≥220 to ≤450, >450)	X
Baseline IBDQ Total Score and Domain Scores	Yes	None	
Baseline Urgency NRS	Yes	Urgency NRS(<3, ≥3)	
Baseline FACIT-Fatigue	Yes		

Variable	Continuous Summary	Categorical Summary	Subgroup Analysis[a]
Other Baseline Patient-Reported Outcomes			
Baseline SF-36 PCS, MCS and Domain scores	Yes	None	
Baseline WPAI-CD employment status	No	Yes, No	
Baseline WPAI-CD score	Yes	None	
EQ-5D-5L VAS and index scores	Yes	None	

Abbreviations: AP = abdominal pain; BMI = body mass index; CD = Crohn's disease; CDAI = Crohn's Disease Activity Index; CRP = C-reactive protein; eCRF = electronic case report form; EQ-5D-5L = European Quality of Life 5–Dimension 5 Level; IBDQ = Inflammatory Bowel Disease Questionnaire; MCS = mental component summary; NRS = numeric rating scale; PCS = physical component summary; SES-CD = Simple Endoscopic Score for Crohn's Disease; SF = stool frequency; SF-36 = 36-Item Short Form Health Survey; VAS = Visual Analog Scale; WPAI-CD = Work Productivity and Activity Impairment Questionnaire Crohn's Disease.

[a] These subgroup analyses will be used for efficacy endpoints.

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

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

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

[e] Anti-TNF alpha biologics include: infliximab, infliximab biosimilar, adalimumab, adalimumab biosimilar, and golimumab, and certolizumab pegol.

[f] Anti-integrin biologics include: natalizumab, and vedolizumab

[g] Prior immunomodulators include: 6-mercaptopurine, azathioprine. other thiopurines and >/=25mg weekly intramuscular/subcutaneous methotrexate. Note that this is not exactly the same as the inclusion criteria defined in the protocol.

[h] Length of the interval from the date of CD diagnosis to the date of informed consent.

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

[j] Ileal is defined as SES-CD scores of 0 for all colonic segments at screening. Ileal-colonic is defined as any non-zero SES-CD score in any colonic segment and non-zero SES-CD score in ileal segment at screening. Colonic is defined as SES-CD scores of 0 in ileal segment at screening. The majority rule applies to scores from multiple central readers.

6.5. Appendix 5: Study Intervention Compliance

Study intervention compliance with investigational product will be summarized for the mITT population. Study intervention compliance for each participant will be calculated as:

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

Here the planned drug administrations per protocol is based on the number of visits before the participant discontinued study drug. Each participant will be defined as having received a dose on a given date if the dose is administered as derived from the Exposure eCRF page. "Overall compliance" with therapy is defined as having at least 80% treatment compliance . Proportions of participants who meet the definition of *overall compliance* during the Induction Period will be compared between study intervention groups using Fisher's exact test.

6.6. Appendix 6: Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the 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 study intervention 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 study intervention 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 participants in every study intervention group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures (e.g., the CSR, manuscripts, and so forth.

6.7. Appendix 7: CDAI Questionnaire

The CDAI score is calculated for each visit using the algorithm below (Best et al. 1976). The standard weights can be determined using the Standard Weight table on the following page.

FOR REVIEW PURPOSES ONLY

Questionnaire obtained by: Lilly	Study ID	Subject Number	Visit/Cycle Number	Signature of Individual Completing Form
	Investigator Number	Page 1 of 1		Date Signed by Individual Completing Form

Patient reported outcomes in Crohn's disease

(a) Crohn's Disease Activity Index (CDAI)

VARIABLE	DAY 1	2	3	4	5	6	7	7 Day Total	Weighting Factor	Total
1. Number of liquid or very soft stools									x 2 =	
2. Abdominal pain 0=none, 1=mild, 2=moderate, 3=severe									x 5 =	
3. General well-being 0=generally well, 1=slightly under par, 2=poor, 3=very poor, 4=terrible									x 7 =	
4. Extra-intestinal manifestations, Current									Check all that apply	
a. Arthritis/arthralgia										
b. Iritis/uveitis										
c. Erythema nodosum, pyoderma gangrenosum,aphthous stomatitis										
d. Anal fissure, fistula, or abscess										
e. Other fistula										
f. Fever over 37.8C (100F) during past 7 days										
Total number of checked boxes=										
x 20 =										
5. Lomitil, Imodium, Opiates for diarrhea in the last 7 days									No = 0, Yes = 1	
x 30 =										
6. Abdominal mass									None = 0, Questionable = 2, Definite = 5	
x 10 =										
7. Local Haematocrit (%, rounded to whole)									If Male, 47-_____= If Female,42-_____= If negative, enter 0	
x 6 =										
8. Body weight calculation									Percentage deviation from standard weight x 1 =	
CDAI TOTAL=										

Standard Weight Table Based on Height and Sex

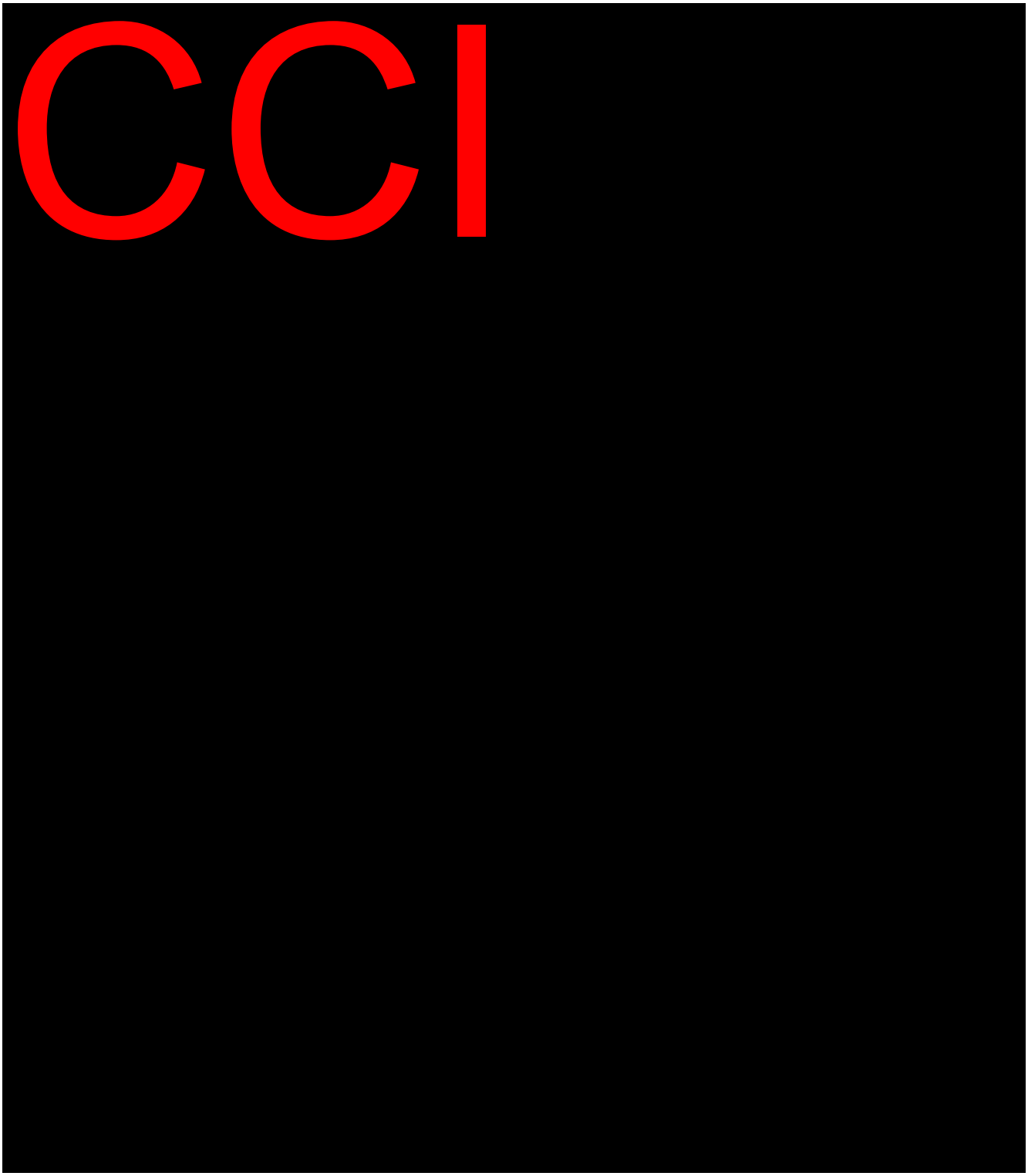
WOMEN	
Height in cm *without shoes*	**Standard Weight in Kg**
148	53.1
149	53.6
150	54.1
151	54.5
152	55.0
153	55.4
154	55.9
155	56.4
156	57.0
157	57.5
158	58.1
159	58.6
160	59.1
161	59.6
162	60.2
163	60.7
164	61.3
165	61.9
166	62.4
167	62.9
168	63.4
169	63.9
170	64.5
171	65.0
172	65.5
173	66.0
174	66.6
175	67.2
176	67.7
177	68.3
178	68.8
179	69.3
180	69.8
181	70.3
182	70.9
183	71.5
184	72.1
185	72.7
186	73.4

MEN	
Height in cm *without shoes*	**Standard Weight in Kg**
158	62.6
159	62.9
160	63.3
161	63.7
162	64.1
163	64.6
164	65.0
165	65.5
166	66.0
167	66.6
168	67.1
169	67.6
170	68.1
171	68.7
172	69.2
173	69.7
174	70.3
175	70.8
176	71.3
177	71.9
178	72.4
179	73.0
180	73.6
181	74.3
182	74.8
183	75.5
184	76.2
185	76.9
186	77.6
187	78.2
188	78.8
189	79.6
190	80.4
191	81.0
192	81.6
193	82.2
194	82.8
195	83.4
196	84.0

Modified for height without shoes from the 1983 Metropolitan Life Insurance Ideal Weights for Height tables.

NOTE: if height is outside of the range in the table, the closest height in the table will be used.


CCI


CCI

6.9. Appendix 9: Study Visit or Week Definition for Daily Diary

CDAI-SF, CDAI-AP, CDAI Well-Being, Urgency NRS, PGRS and additional measures are collected using Patient Daily Diary, entries will be mapped to study week by the following:

Visit Number / Week Number	Start Day	End Day
Visit 2 / Baseline	Date of First Injection -12 days	Date of First Injection – 1 day
Visit 3 / Week 2	Max (Date of First Injection, Week 2 Assessment Date – 12 days)	Week 2 Assessment Date – 1 day
Visit 4 / Week 4	Max (Week 2 Assessment Date, Week 4 Assessment Date – 12 days)	Week 4 Assessment Date – 1 day
Visit 5 / Week 6	Max (Week 4 Assessment Date, Week 6 Assessment Date – 12 days)	Week 6 Assessment Date – 1 day
Visit 6 / Week 8	Max (Week 6 Assessment Date, Week 8 Assessment Date – 12 days)	Week 8 Assessment Date – 1 day
Visit 7 / Week 12	Max (Week 8 Assessment Date, Min (Week 12 Assessment Date, Start date of Period 2) – 12 days)	Min (Week 12 Assessment Date, Start date of Period 2)– 1 day
Visit 8 / Week 16	Max (Week 12 Assessment Date (i.e. Date of First Injection of Period 2), Week 16 Assessment Date – 12 days)	Week 16 Assessment Date – 1 day
Visit 9 / Week 20	Max (Week 16 Assessment Date, Week 20 Assessment Date – 12 days)	Week 20 Assessment Date – 1 day
Visit 10 / Week 24	Max (Week 20 Assessment Date, Week 24 Assessment Date – 12 days)	Week 24 Assessment Date – 1 day
Visit 11 / Week 28	Max (Week 24 Assessment Date, Week 28 Assessment Date – 12 days)	Week 28 Assessment Date – 1 day
Visit 12 / Week 32	Max (Week 28 Assessment Date, Week 32 Assessment Date – 12 days)	Week 32 Assessment Date – 1 day
Visit 13 / Week 36	Max (Week 32 Assessment Date, Week 36 Assessment Date – 12 days)	Week 36 Assessment Date – 1 day
Visit 14 / Week 40	Max (Week 36 Assessment Date, Week 40 Assessment Date – 12 days)	Week 40 Assessment Date – 1 day
Visit 15 / Week 44	Max (Week 40 Assessment Date, Week 44 Assessment Date – 12 days)	Week 44 Assessment Date – 1 day
Visit 16 / Week 48	Max (Week 44 Assessment Date, Week 48 Assessment Date – 12 days)	Week 48 Assessment Date – 1 day
Visit 17 / Week 52	Max (Week 48 Assessment Date, Week 52 Assessment Date – 12 days)	Week 52 Assessment Date – 1 day

6.10. Appendix10: Specified Changes in Concomitant CD Medications (Intercurrent Event)

Corticosteroids in Period 1

1. Initiation of oral corticosteroids or oral budesonide due to worsening Crohn's disease.
2. Initiation of CCI
3. CCI
4. CCI
5. CCI due to worsening Crohn's disease.

Corticosteroids in Period 2

1. Initiation of CCI
2. CCI
3. CCI
4. CCI
5. CCI due to worsening Crohn's disease.

Immunomodulator in Period 1 and/or Period 2

1. Initiation of CCI .
2. CCI
3. CCI
4. CCI

6.11. Appendix 11: IBD-DI Scoring

Scoring:

ANSWERS:

- Question 1:
  - 0 = Very good; 1 = Good; 2 = Moderate; 3 = Bad; 4 = Very Bad
- Question 2&3:
  - 0 = No; 4 = Yes or uncertain
- Question 4 Stool frequency:
  - 0 = 0; 1 = 1-7; 2 = 8-18; 3 = 19-29; 4 = >29
- Table questions:
  - 0 = None; 1 = Mild; 2 = Moderate; 3 = Severe; 4 = Extreme or cannot do

Total score = S*100/(n*4) n = number of questions which have been answered S = sum of the n questions score S is possible if (14-n)/14<20%	Total score
	Ranging from 0 (no disability) to 100 (highest disability level)

7. References

Agresti A. *Categorical Data Analysis.* 3rd Edition. Hoboken, NJ: John Wiley & Sons Inc., 2013.

Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med.* 2014;33(4):693-713. doi: 10.1002/sim.5974

Andersen SW, Millen BA. On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharm Stat*. 2013;12(1):7-16. doi: 10.1002/pst.1548

Bartlett JW, Hughes RA. Bootstrap inference multiple imputation under uncongeniality and misspecification. *Stat Methods Med Res*. 2020;29(12):3533-3546. doi: 10.1177/0962280220932189

Best WR, Becktel JM, Singleton JW, Kern F Jr. Development of a Crohn's disease activity index: National Cooperative Crohn's Disease Study. *Gastroenterology.* 1976;70(3):439-444.

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med.* 2009;28(4):586-604. doi: 10.1002/sim.3495

Bretz F, Posch M, Glimm E, et al. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J.* 2011;53(6):894-913. doi: 10.1002/bimj.201000239

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(9):1032-1041. doi: 10.1111/j.1365-2036.2009.03966.x

D'Haens AGR, Geboe K, Peeters M, et al. Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded ileum. *Gastroenterology*. 1998;114(2):262-267. doi: 10.1016/s0016-5085(98)70476-7

[EMA] European Medicines Agency. Guideline on the choice of non-inferiority margin. EMEA/CPMP/EWP/2158/99. Published 27 July 2005. Accessed March 08, 2019. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf

[FDA] Food and Drug Administration. E9(R1): statistical principles for clinical trials: addendum: estimands and sensitivity analyses in clinical trials: guidance for industry. Draft Guidance. Published 16 June 2017. Accessed 23 November 2019. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM582738.pdf

[FDA] United States Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness: guidance for industry. Published November 2016. Accessed November 23, 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/non-inferiority-clinical-trials

Feagan BG, Sandborn WJ, Gasink C, et al. UNTI-IM-UNTI Study Group. Ustekinumab as induction and maintenance therapy for Crohn's disease. *N Engl J Med*. 2016;375(20):1946-1960. doi: 10.1056/NEJMoa1602773

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

Guyatt G, Mitchell A, Irvine EJ, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology.* 1989;96(30):804-810.

Irvine EJ, Feagan B, Rochon J, et al. Quality of Life: a valid and reliable measure of therapeutic efficacy in the treatment of inflammatory bowel disease. Canadian Crohn's Relapse Prevention Trial Study Group. *Gastroenterology*. 1994;106(2):287-296. doi: 10.1016/0016-5085(94)90585-1

Irvine EJ, Zhou Q, Thompson AK. The short inflammatory bowel disease questionnaire: a quality of life instrument for community physicians managing inflammatory bowel disease. CCRPT Investigators. Canadian Crohn's Relapse Prevention Trial. *Am J Gastroenterol.* 1996;91(8)1571-1578.

Irvine EJ. Quality of life of patients with ulcerative colitis: past, present, and future. *Inflamm Bowel Dis*. 2008;14(4):554-565. doi: 10.1002/ibd.20301

Lipkovich I, Ratitch B, Qu Y, et al. Using principal stratification in analysis of clinical trials. *Stat Med.* 2022;41(19):3837-3877. doi: 10.1002/sim.9439. Epub 2022 Jun 13

Mosli HM, Feagan BG, Zou G, et al. Development and validation of a histological index for UC. *Gut*. 2017;66(1):50-58. doi: 10.1136/gutjnl-2015-310393

[PhUSE] Pharmaceutical Users Software Exchange. Computational Science Symposium Development of Standard Scripts for Analysis and Programming Working Group. Analyses and displays associated with measures of central tendency – focus on vital sign, electrocardiogram, and laboratory analyte measurements in phase 2-4 clinical trials and integrated submission documents. Published 10 October 2013. Accessed November 23, 2019. http://www.phusewiki.org/docs/CSS%20White%20Papers%202016/CSS_WhitePaper_CentralTendency_v1.0.pdf

[PhUSE] Pharmaceutical Users Software Exchange. Computational Science Symposium Development of Standard Scripts for Analysis and Programming Working Group. Analysis and Display White Papers Project Team. Analyses and displays associated with outliers or shifts from normal to abnormal: focus on vital signs, electrocardiogram, and laboratory analyte measurements in phase 2-4 clinical trials and integrated summary documents. Published 10 September 2015. Accessed November 23, 2019. http://www.phusewiki.org/docs/CSS%20White%20Papers%202016/CS_WhitePaper_OutliersShifts_v1.0.pdf

Reilly Associates resources pages. WPAI Scoring. Accessed November 25, 2019. http://www.reillyassociates.net/WPAI_Scoring.html

Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics*. 1993;4(5):353-365. doi: 10.2165/00019053-199304050-00006

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

Sandborn WJ, Feagan BG, Stoinov S, et al. PRECISE 1 Study Investigators. Certolizumab pegol for the treatment of Crohn's disease. *N Engl J Med*. 2007;357:228-238.

Sato T. On the variance estimator for the Mantel-Haenszel risk difference. *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.

Wilson EB, Hilferty MM. The distribution of chi-square. *Proc Natl Acad Sci U S A*. 1931;17(12):684-688.

Winthrop KL, Novosad SA, Baddley JW, et al. Opportunistic infections and biologic therapies in immune-mediated inflammatory diseases: consensus recommendations for infection reporting during clinical trials and postmarketing surveillance. *Ann Rheum Dis.* 2015;74(12):2107-2116. doi: 10.1136/annrheumdis-2015-207841

Signature Page for VV-CLIN-028402 v1.0

Approval	PPD 23-Aug-2023 17:03:05 GMT+0000

Signature Page for VV-CLIN-028402 v1.0