Protocol B7981007

A PHASE 2A, DOUBLE-BLIND, RANDOMIZED, PLACEBO-CONTROLLED, PARALLEL GROUP STUDY TO EVALUATE THE EFFICACY AND SAFETY OF ORAL PF-06651600 AND PF-06700841 AS INDUCTION AND OPEN LABEL EXTENSION TREATMENT IN SUBJECTS WITH MODERATE TO SEVERE CROHN'S DISEASE

Statistical Analysis Plan
(SAP)

Version: Amendment 2

Date: 27-June-2022

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. VERSION HISTORY

This SAP amendment for study B7981007 is based on the protocol dated 27JUL2021.

Table 1. Summary of Major Changes in SAP Amendments

SAP Version	Change	Rationale
1	Not Applicable	Not Applicable
Amendment 1	Statistical analyses were aligned with the change of primary endpoint, additional sensitivity analyses were added. Due to strategic reasons, the main hypothesis of interest is to compare PF-06651600 and placebo. There will be no multiplicity adjustment.	Protocol Amendment 5
Amendment 2	For binary response defined on endpoints including CDAI, SF/AP, IBDQ and GHAS, and open label summary of all efficacy endpoints, missing values after study treatment discontinuation will be handled by non-responder imputation (NRI) method, others will be removed from the analysis (i.e., the missing visits will not contribute to the denominator in the calculation of the proportion).	Considerable amount of missing data before treatment discontinuation ends up in the study for exploratory data, and they are caused by various reasons such as site logistics in and poor participants diary compliance that are most likely not related to efficacy. For OLE period summary, imputation rule is relaxed as well.
	CDAI-100 and CDAI remission endpoint will be summarized only for subjects with baseline CDAI ≥220.	Endpoints are more clinically meaningful and traditionally defined for population with moderate and above disease population.
	If an AE started in induction and resolved in OLE, it will be counted as TEAE only in induction.	It is more clinically desired to present it once only.

2. INTRODUCTION

This SAP amendment provides the detailed methodology for summary and statistical analyses of the data collected in Study B7981007. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives

2.1.1. Induction Period

Study objectives and corresponding endpoints during 12-week (Weeks 1-12) induction period are provided in the Table 2 below.

Table 2. Study Objectives and Endpoints during Induction Period

Primary Objective(s):	**Primary Endpoint(s):**
• To evaluate the efficacy of PF-06651600 and PF-06700841 compared to placebo at Week 12 in subjects with moderate to severe Crohn's disease (CD).	• Proportion of subjects achieving SES--CD 50 (≥50% reduction in SESCD from baseline) at Week 12 as assessed by central reading.
Secondary Objective(s):	**Secondary Endpoint(s):**
• To evaluate the safety and tolerability of PF-06651600 and PF-06700841 compared to placebo in subjects with moderate to severe CD over 12 weeks.	• Incidence and severity of laboratory abnormalities, vital signs, 12-lead ECG, adverse events, serious adverse events and withdrawals due to adverse events. • Incidence of serious infections.
• To evaluate the efficacy of PF-06651600 and PF-06700841 compared to placebo during induction of additional endoscopic endpoints in subjects with moderate to severe CD.	• Proportion of subjects achieving clinically meaningful endoscopic improvement (reduction of ≥3 points from baseline in SES-CD score) at Week 12. • Change from baseline in SES-CD score at Week 12. • Proportion of subjects achieving SES-CD 25 and (≥25% reduction in SES-CD from baseline) at Week 12. • Proportion of subjects achieving endoscopic remission (SES-CD ≤2) at Week 12. • Proportion of subjects achieving mucosal healing (complete absence of ulcers) at Week 12.
Tertiary/Exploratory Objective(s):	**Tertiary/Exploratory Endpoint(s):**
• To evaluate the effect of PF-06651600 and PF-06700841 compared to placebo on outcomes based on additional clinical criteria.	• Proportion of subjects achieving clinical response and remission using stool frequency

	(SF) and abdominal pain (AP) measures at Weeks 2, 4, 6, 8, 10, and 12. • Proportion of subjects achieving deep remission (endoscopic remission by SES-CD and clinical remission by SF and AP) at Week 12. • Proportion of subjects with a Crohn's Disease Activity Index (CDAI)-100 response (defined by a decrease in CDAI score of at least 100 points from baseline) and proportion of subjects who are remitters (defined as CDAI <150) at Weeks 2, 4, 8, and 12. • The scores and change from baseline in Inflammatory Bowel Disease Questionnaire (IBDQ) total score and domains (Bowel Symptoms, Systemic Symptoms, Emotional Function and Social Function) at Weeks 4, 8, and 12. • The proportion of subjects with IBDQ total score ≥170 at Weeks 4, 8, and 12. • The proportion of subjects with ≥16 point increase in IBDQ total score from baseline at Weeks 4, 8, and 12. • Proportion of subjects achieving IBDQ bowel symptom domain response at Weeks 4, 8, and 12. • The scores and change from baseline in Euro Quality of Life Questionnaire 5 Dimensions 3 Levels (EQ-5D-3L) + Visual Analog Scale (VAS) at Weeks 4, 8, and 12. • The scores and change from baseline in Short Form-36, Version 2 Acute (SF-36 v2): physical and mental component summary scores (PCS & MCS), and 8 domain scores at Weeks 4, 8, and 12. • Change from baseline in Patient Global Impression of Severity (PGIS) score at Weeks 4, 8, and 12.
• To evaluate the effect of PF-06651600 and PF-06700841 compared to placebo on histopathology score.	• Change from baseline in Global Histologic Activity Score (GHAS) at Week 12. • Proportion of subjects achieving histologic remission at Week 12 (defined as GHAS ≤4).
• To assess the effect of PF-06651600 and PF-06700841 compared to placebo on disease and mechanistic biomarkers over time.	• Change from baseline in serum high sensitivity C-reactive protein (hsCRP) levels over time.

	• Change from baseline in fecal calprotectin over time. • Change from baseline in serum interferon gamma-induced protein 10 (IP-10) levels over time. • Change from baseline in (B-cell lymphoma 2) *BCL-2* gene expression. • Change from baseline in hematological values including reticulocytes, hemoglobin, neutrophils, platelets, and T, B and NK (TBNK) cells.
• To describe the pharmacokinetics (PK) of PF-06651600 and PF-06700841 compared to placebo in subjects with moderate to severe CD.	• PF-06651600 concentrations at Weeks 2, 4, 8, and 12. • PF-06700841 concentrations at Weeks 2, 4, 8, and 12.
• To collect non-banked samples (e.g., intestinal biopsies, stool for microbiome analysis, serum and plasma for analysis of proteins and a whole blood tube for RNA analysis) for exploratory research, unless prohibited by local regulations or ethics committee decision. • To collect banked biospecimens samples for exploratory research, unless prohibited by local regulations or ethics committee decision.	• Collection of non-banked exploratory samples unless prohibited by local regulations or ethics committee decision. • Collection of banked biospecimens unless prohibited by local regulations or ethics committee decision. Additional information on collection and potential use is provided in the Banked Biospecimens section of the protocol.

2.1.2. Open Label Extension Period

Study objectives and corresponding endpoints during the 52-week (Weeks 12-64) open label extension (OLE) period are provided in the Table 3 below.

Table 3. Study Objectives and Endpoints during the OLE Period

Primary Objective(s):	**Primary Endpoint(s):**
• To assess the safety and tolerability of PF-06651600 and PF-06700841 therapy during open label extension period for subjects with moderate to severe CD.	• Incidence and severity of laboratory abnormalities, vital signs, 12-lead ECG, adverse events, serious adverse events and withdrawals due to adverse events.
Exploratory Objective(s):	**Exploratory Endpoint(s):**
• To evaluate the efficacy of PF-06651600 and PF-06700841 therapy during open label extension period for subjects with moderate to severe CD.	• Proportion of subjects achieving clinically meaningful endoscopic improvement (reduction of ≥3 points from baseline in SES-CD) at Week 64. • Proportion of subjects achieving SES-CD 25 and SES-CD 50 at Week 64.

	• Proportion of subjects achieving endoscopic remission at Week 64. • Proportion of subjects achieving mucosal healing (complete absence of ulcers) at Week 64.
• To evaluate the effect of PF-06651600 and PF-06700841 on outcomes based on additional clinical criteria.	• Proportion of subjects achieving clinical response and remission as defined by SF and AP endpoints at Weeks 16, 20, 24, 32, 40, 48, 56, and 64. • Proportion of subjects achieving deep remission (endoscopic remission by SES-CD and clinical remission by SF, AP) at Week 64. • Proportion of subjects with a CDAI-100 response or CDAI remission at Weeks 16, 32, and 64. • The scores and change from baseline in IBDQ Total score and domains (Bowel Symptoms, Systemic Symptoms, Emotional Function and Social Function) at Weeks 16, 32, and 64. • The proportion of subjects with IBDQ total score ≥170 at Weeks 16, 32, and 64. • The proportion of subjects with ≥16 point increase in IBDQ total score from baseline at Weeks 16, 32, and 64. • Proportion of subjects achieving IBDQ bowel symptom domain response at Weeks 16, 32, and 64. • The scores and change from baseline in EQ-5D-3L + VAS at Weeks 16, 32, and 64. • The scores and change from baseline in SF-36 v2: PCS & MCS, and 8 domain scores at Weeks 16, 32, and 64. • Change from baseline in PGIS score at Weeks 16, 32, and 64.
• To evaluate the effect of PF-06651600 and PF-06700841 compared to placebo on histopathology score.	• Change from baseline in GHAS at Week 64. • Proportion of subjects achieving histologic remission at Week 64.
• To describe the PK of PF-06651600 and PF-06700841 in subjects with moderate to severe CD.	• PF-06651600 concentrations at Weeks 16, 20, 32, 56, and 64. • PF-06700841 concentrations at Weeks 16, 20, 32, 56, and 64.

• To explore the relationship between PK, PD, and clinical endpoints.	• Change from baseline in serum hsCRP levels over time. • Change from baseline in fecal calprotectin. • Change from baseline in serum IP-10 levels over time. • Change in baseline in *BCL-2* gene expression. • Change from baseline in hematological values including reticulocytes, hemoglobin, neutrophils, platelets, and TBNK cells.
• To collect non-banked samples (e.g., intestinal biopsies, stool for microbiome analysis, serum and plasma for analysis of proteins and a whole blood tube for RNA analysis) for exploratory research, unless prohibited by local regulations or ethics committee decision. • To collect banked biospecimens samples for exploratory research, unless prohibited by local regulations or ethics committee decision.	• Collection of non-banked exploratory samples unless prohibited by local regulations or ethics committee decision. • Collection of banked biospecimens unless prohibited by local regulations or ethics committee decision. Additional information on collection and potential use is provided in the Banked Biospecimens section of the protocol.

2.2. Study Design

This is a Phase 2a, randomized, double-blind, placebo-controlled, parallel group, multicenter study in subjects with moderate to severe active CD. The entire study consists of 1) a screening period of up to 6-weeks, 2) a 12-week induction period, 3) a 52-week open label extension period, and 4) a 4-week follow up period. Approximately 230~250 subjects in total will be randomized into the study.

The 12-week induction period will be placebo-controlled and double-blind within each investigational product. Subjects who meet the eligibility criteria at the baseline visit will be randomly assigned to receive either active or placebo treatments. In the induction period, 200 mg QD for 8 weeks followed by 50 mg QD for 4 weeks of PF-06651600 and matching placebo in a 2:1 ratio and 60 mg QD for 12 weeks of PF-06700841 and matching placebo in a 2:1 ratio will be investigated. The hybrid dosing regimen for PF-06651600 during the 12-week induction period is a consequence of available nonclinical long-term toxicity data supporting 200 mg treatment for only up to 8 weeks. For analysis, placebo groups will be combined.

After the completion of the induction period, subjects will enter the 52 weeks OLE period. There will be no rerandomization at the beginning of the OLE period. Subjects will receive the same study drug that they were randomized to receive during the induction period, and there will be no placebo arms. Placebo subjects from the induction period will also receive active drug in the OLE period. The matching placebo subjects from the double-blind PF06651600 treatment/placebo induction period will receive 50 mg of PF-06651600, while

the corresponding placebos from the double-blind PF-06700841/placebo induction period will receive 30 mg of PF-06700841 for 52 weeks.

After completion of the OLE period, subjects will enter the 4-week follow-up period. Any subjects, who discontinues early from the double-blind period prior to the Week 12 visit, should undergo the procedures for an Early Termination (Induction) visit on the last day the subject takes the investigational product or as soon as possible thereafter. For subjects who discontinue early from OLE period (after the Week 12 visit, but prior to the Week 52 visit), the procedures scheduled for an Early Termination (ET) visit will be performed on the last day the subject takes the investigational product or as soon as possible thereafter. These early withdrawal subjects, along with subjects who complete the induction period but are not willing to participate in the OLE period, will be asked to complete the Follow-up visit approximately 4 weeks after the last dose of study drug.

The original objectives of this study were to evaluate the efficacy [based on clinically meaningful endoscopic improvement (reduction of ≥3 points from baseline in SES-CD score) at Week 12 as assessed by central reading], safety, tolerability, PK, and PD of 200 mg for 8 weeks followed by 50 mg for 4 weeks of PF-06651600 (ritlecitinib) dosed once daily and 60 mg of PF-06700841 (brepocitinib) dosed once daily during an induction period of 12 weeks, followed by an open label extension period at doses of 50 mg and 30 mg of PF-06651600 (ritlecitinib) and PF-06700841 (brepocitinib), respectively, for 52 weeks.

Amendment 5 of Protocol B7981007 revises the original design to eliminate the PF-06700841 (brepocitinib) and placebo arms. Therefore, upon approval of Amendment 5 by regional Regulatory Authorities and Ethics Committees, Protocol B7981007 will be conducted as a Phase 2a, randomized, double-blind, placebo-controlled, parallel group study focused on the evaluation of the efficacy and safety profile of the PF-06651600 (ritlecitinib) in subjects with moderate to severe active CD. Participants that have been randomized under previous amendments will continue to receive PF-06700841 (brepocitinib) through completion. This study modification is solely a strategic decision on the part of the sponsor to prioritize future development of PF-06651600 (ritlecitinib), a JAK3/TEC inhibitor currently in development for a variety of other inflammatory/autoimmune indications and to enable efficient comparison with contemporary and emerging trials in Crohn's disease. The decision to eliminate the PF-06700841 (brepocitinib) cohort of the study is not due to any specific safety or efficacy concerns that would negatively affect the overall benefit/risk for participants in this trial or in other trials.

Figure 1. Original Study Design Schematic

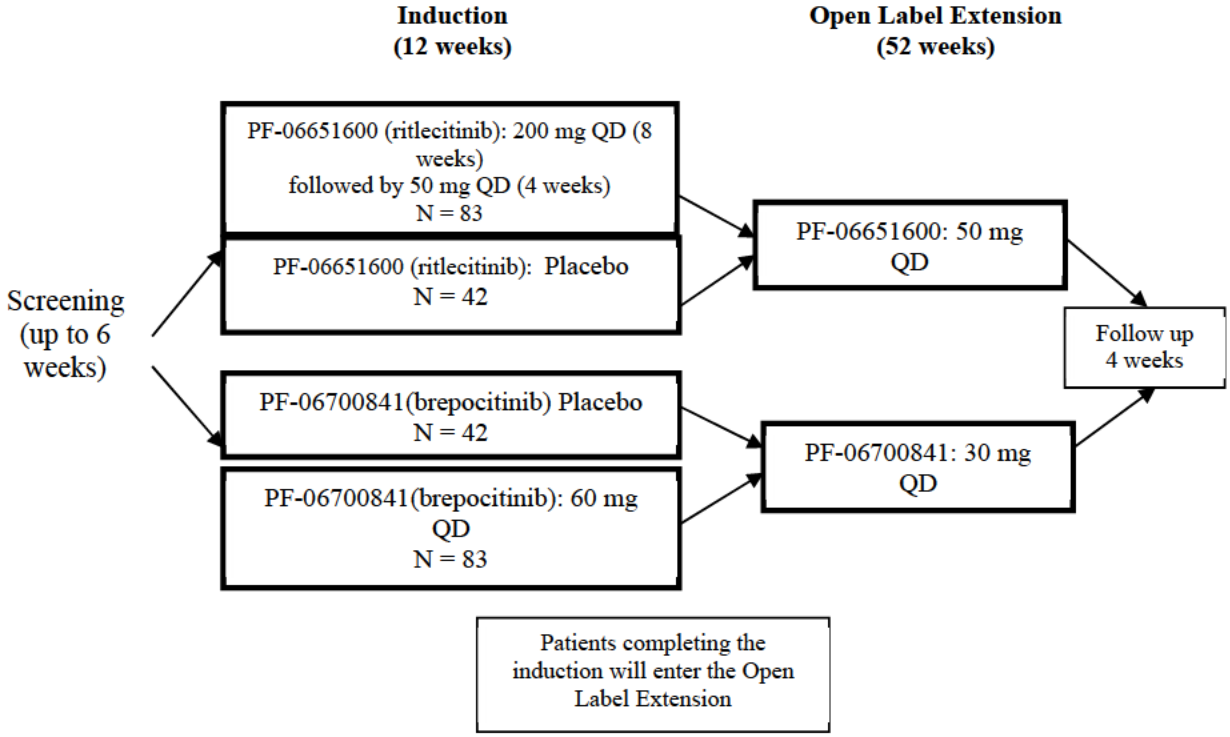
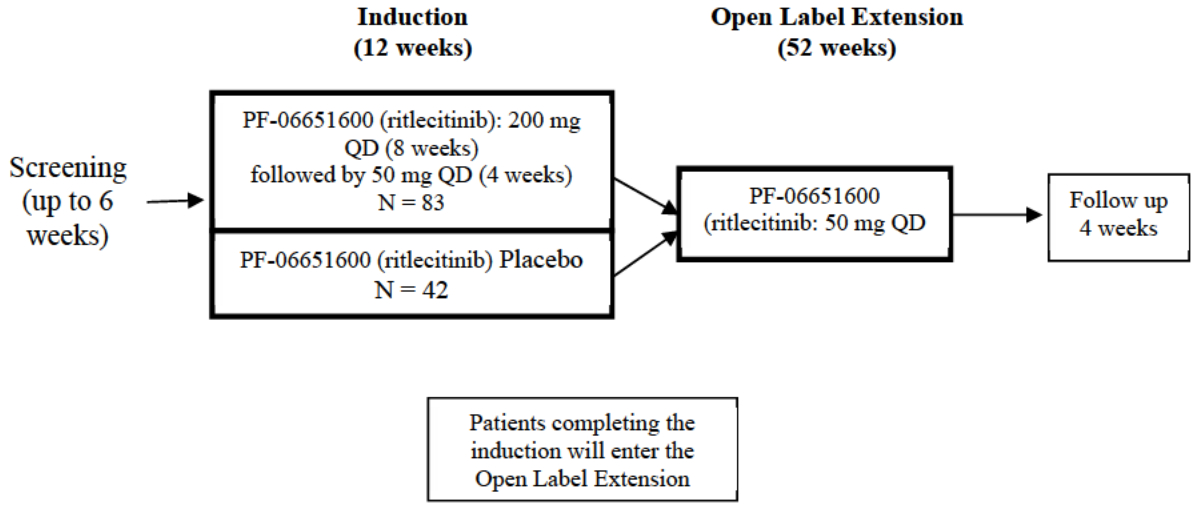


Figure 2. Study Schematic Post Implementation of Protocol Amendment 5



3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Induction Period

3.1.1. Primary Endpoint(s)

The primary efficacy endpoint during the induction period is:

- Proportion of subjects achieving SES-CD 50 (≥ 50% reduction in SES-CD from baseline) at Week 12.

3.1.2. Secondary Endpoint(s)

Secondary endpoints during the induction period are:

- Incidence and severity of laboratory abnormalities, vital signs, 12-lead ECG, adverse events, serious adverse events, and withdrawals due to adverse events.
- Incidence of serious infections.
- Change from baseline in SES-CD score at Week 12.
- Proportion of subjects achieving SESCD 25 (≥ 25% reduction in SES-CD from baseline) and CMEI (Clinically Meaningful Endoscopic Improvement defined as a reduction of ≥3 points from baseline) at Week 12.
- Proportion of subjects achieving endoscopic remission (SES-CD ≤2) at Week 12.
- Proportion of subjects achieving mucosal healing (complete absence of ulcers) at Week 12.

3.1.3. Tertiary/Exploratory Endpoint(s)

Below are tertiary/exploratory endpoints during the induction period. CDAI-100 and CDAI remission will only be summarized for subjects with baseline CDAI≥220.

- Proportion of subjects achieving clinical response and remission using stool frequency (SF) and abdominal pain (AP) measures at Weeks 2, 4, 6, 8, 10, and 12.
- Proportion of subjects achieving deep remission (endoscopic remission by SES-CD and clinical remission by SF and AP) at Week 12.
- Proportion of subjects with a CDAI-100 response at Weeks 2, 4, 8, and 12.
- Proportion of subjects with CDAI remission (CDAI <150) at Weeks 2, 4, 8, and 12.
- The scores and change from baseline in IBDQ total score and domains (Bowel Symptoms, Systemic Symptoms, Emotional Function and Social Function) at Weeks 4, 8, and 12.

- Proportion of subjects with IBDQ total score ≥170 at Weeks 4, 8, and 12.
- Proportion of subjects with ≥16 point increase in IBDQ total score from baseline at Weeks 4, 8, and 12.
- Proportion of subjects achieving IBDQ bowel symptom domain response at Weeks 4, 8, and 12.
- The scores and change from baseline in EQ-5D-3L + VAS at Weeks 4, 8, and 12.
- The scores and change from baseline in Short Form-36, Version 2 Acute (SF-36 v2): physical and mental component summary scores (PCS & MCS), and 8 domain scores at Weeks 4, 8, and 12.
- Change from baseline in PGIS score at Weeks 4, 8, and 12.
- Change from baseline in GHAS at Week 12.
- Proportion of subjects achieving histologic remission (GHAS ≤4) at Week 12.
- Change from baseline in serum hsCRP levels over time.
- Change from baseline in fecal calprotectin over time.
- Change from baseline in serum IP-10 levels over time.
- Change from baseline in *BCL-2* gene expression.
- Change from baseline in hematological values including reticulocytes, hemoglobin, neutrophils, platelets, and T, B and NK (TBNK) cells.
- PF-06651600 concentrations at Weeks 2, 4, 8, and 12.
- PF-06700841 concentrations at Weeks 2, 4, 8, and 12.

3.2. Open Label Extension Period

3.2.1. Primary Endpoint(s)

The primary endpoint during the OLE period is:

- Incidence and severity of laboratory abnormalities, vital signs, 12-lead ECG, adverse events, serious adverse events, and withdrawals due to adverse events.

3.2.2. Exploratory Endpoint(s)

Exploratory endpoints during the OLE period are:

- Proportion of subjects achieving clinically meaningful endoscopic improvement (reduction of ≥3 points from baseline in SES-CD) at Week 64.
- Proportion of subjects achieving SES-CD 25 and SES-CD 50 at Week 64.

- Proportion of subjects achieving endoscopic remission at Week 64.
- Proportion of subjects achieving mucosal healing (complete absence of ulcers) at Week 64.
- Proportion of subjects achieving clinical response and remission as defined by SF and AP endpoints at Weeks 16, 20, 24, 32, 40, 48, 56, and 64.
- Proportion of subjects achieving deep remission (endoscopic remission by SES-CD and clinical remission by SF, AP) at Week 64.
- Proportion of subjects with a CDAI-100 response or CDAI remission at Weeks 16, 32, and 64.
- The scores and change from baseline in IBDQ Total score and domains (Bowel Symptoms, Systemic Symptoms, Emotional Function and Social Function) at Weeks 16, 32, and 64.
- The proportion of subjects with IBDQ total score ≥170 at Weeks 16, 32, and 64.
- The proportion of subjects with ≥16 point increase in IBDQ total score from baseline at Weeks 16, 32, and 64.
- Proportion of subjects achieving IBDQ bowel symptom domain response at Weeks 16, 32, and 64.
- The scores and change from baseline in EQ5D3L + VAS at Weeks 16, 32, and 64.
- The scores and change from baseline in SF36 v2: PCS & MCS, and 8 domain scores at Weeks 16, 32, and 64.
- Change from baseline in PGIS score at Weeks 16, 32, and 64.
- Change from baseline in GHAS at Week 64.
- Proportion of subjects achieving histologic remission at Week 64.
- PF-06651600 concentrations at Weeks 16, 20, 32, 56, and 64.
- PF-06700841 concentrations at Weeks 16, 20, 32, 56, and 64.
- Change from baseline in serum hsCRP levels over time.
- Change from baseline in fecal calprotectin.
- Change from baseline in serum IP-10 levels over time.

- Change in baseline in *BCL-2* gene expression.
- Change from baseline in hematological values including reticulocytes, hemoglobin, neutrophils, platelets, and TBNK cells.

3.3. Baseline Variables

Baseline data collected/measured at Screening visit are:

- Medical history, history of alcohol, smoking and drug abuse, and height
- HBsAg, HBcAb, HCV Ab, HCV RNA PCR if HCV Ab positive
- Fecal calprotectin and SES-CD

Baseline data collected/measured at Baseline (Day 1) are listed below:

- Weight, complete physical examination, and vital signs & temperature
- hsCRP, SF, AP, CDAI, PGIS, IBDQ, EQ-5D-3L + VAS, and SF-36
- Hematology (including coagulation panel) and serum chemistry (including fasting lipid panel)

3.4. Safety Endpoints

Safety will be assessed by the spontaneous reporting of AEs, physical examinations, and clinical laboratory results in all subjects who receive at least one dose of the investigational product. Unscheduled safety assessments may be performed at any time during the study to assess any perceived safety concerns.

3.4.1. Adverse Events

An adverse event is considered treatment-emergent adverse event (TEAE) to a given treatment if the event started during the effective duration of treatment regardless of whether a similar event of equal or greater severity existed in the baseline period.

Safety endpoints will be assessed by the spontaneous reporting of:

- Incidence of TEAEs and SAEs
- Incidence of withdrawals due to AEs
- Incidence of serious infections

3.4.2. Laboratory Data

Below is a list of hematology and serum chemistry test parameters.

- **Hematology**: hemoglobin, hematocrit, red blood cell count, platelet count, white blood cell count, total neutrophils (Abs), eosinophils (Abs), basophils (Abs), lymphocytes (Abs), monocytes (Abs), reticulocyte count, and coagulation panel

- **Serum chemistry**: blood urea nitrogen, creatinine, cystatin C, glucose (fasting), calcium, sodium, potassium chloride, ttotal CO_2 (bicarbonate), aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin, direct bilirubin, alkaline phosphatase, uric acid, albumin, total protein, creatine kinase (CK), CK fractionation, total cholesterol, triglycerides, high-density lipoproteins (HDL), and low-density lipoprotein (LDL)

3.4.3. Vital Signs, including Height and Weight

Vital sign measurements are pulse rate, blood pressures, respiratory rate, oral or tympanic temperature, respiratory rate, pulse rate, and blood pressure.

Height, weight and Body Surface Area (BSA) are collected at pre- and post-treatment.

3.4.4. Physical Examinations

Complete physical examinations consist of assessments of general appearance; skin; head, eyes, ears, nose and throat; mouth, heart; lungs; breast (optional); abdomen; external genitalia (optional); extremities; neurologic function; and lymph nodes.

4. ANALYSIS SETS

Data for all subjects will be assessed to determine if subjects meet the criteria for inclusion in each analysis population prior to unblinding and release of the database and classifications will be documented per standard operating procedures. For OLE period summary, only subjects who received at least one OLE treatment will be included correspondingly for full analysis set, safety and PK analysis sets.

4.1. Full Analysis Set

The full analysis set analysis set will include all randomized subjects who received at least one dose of the randomized investigational drug (PF-06651600, PF-06700841, or placebo).

4.2. Pharmacokinetics Set

The PK concentration population is defined as all enrolled subjects who received at least one dose of PF-06651600 or PF-06700841 and in whom at least one concentration value is reported.

4.3. Safety Analysis Set

The safety (SAF) analysis set is defined as those subjects who received at least one dose of the investigational drug.

5. GENERAL METHODOLOGY AND CONVENTIONS

Final analyses will occur after database lock after Last Subject Last Visit (LSLV).

An interim analysis may be performed when Week 12 data has been accumulated for approximately 50% of the subjects initially randomized.

After the induction treatment period (Week 0-12) is finished for all the subjects, some members of the study team, who are otherwise not responsible for continued study conduct until the database is locked, will be unblinded so that a report for the induction period data can be generated. The subjects, investigators, and site personnel will continue to be blinded to randomized study treatments throughout the remaining period of the study (including the extension period). The database will be officially released after last subject last visit occurs. The final analysis will be then conducted, and the CSR will be issued. The decision rules for the final analyses are also described in the next section.

5.1. Hypotheses and Decision Rules

5.1.1. Induction Period

Null hypothesis is the proportion of subjects achieving Week 12 centrally read SES-CD 50 is the same for the active (PF-06651600 or PF-06700841) and placebo arms. Using a between-group comparison at Week 12, for PF-06651600 vs placebo and PF-06700841 vs placebo, the treatment will be considered superior to placebo if the lower bound of the 90% confidence interval is greater than 0. There will be no multiplicity adjustment across the compounds. Due to strategic reasons, the main hypothesis of interest is to compare PF-06651600 and placebo.

5.1.2. Open Label Extension Period

There will be no hypothesis testing for the two active treatment (PF-06651600 and PF-06700841) arms in the OLE period.

5.2. General Methods

In general, number and percent will be presented for binary and categorical variables. Number, mean, standard deviation (or standard error of the mean), median, minimum, and maximum will be presented for continuous variables. In addition, graphics may be used to present the data – specific details will be outlined in the study List of Table (LOT).

For analyses purposes, the two placebo arms from the induction period will be combined. There will be no comparisons between the two active treatments PF-06651600 and PF-06700841.

5.2.1. Analyses for Binary Data

The binary data comparing active treatment group and placebo group will be analyzed using the Cochran Mantel Hanzel (CMH) test adjusting the baseline disease stratification factor below:

- Baseline disease activity/extent
    - no isolated ileal disease and baseline SES-CD >15
    - no isolated ileal disease and baseline SES-CD ≤15
    - isolated ileal disease

Rate (%) differences and the corresponding 2-sided 90% confidence intervals adjusted for stratification factors will be computed using the methods proposed by Cochran with the minimum risk weights proposed by Mehrotra and Railkar. It is anticipated to have small numbers in each investigation site, so the investigation site effect will not be an adjustment factor.

For analyses without adjustment by stratification such as subset analyses, the estimated risk difference and the associated confidence interval of the risk differences will be presented. The unconditional exact method as described by Chan and Zhang (1999) will be used to compute the confidence intervals.

5.2.2. Analyses for Longitudinal Continuous Data

Mixed Model for Repeated Measurements (MMRM): longitudinal continuous data will be analyzed by MMRM including the fixed effects of treatment, visit, and treatment-by-visit interaction, along with baseline value as a covariate when available. Unstructured covariance matrix will be assumed when possible. If there are a sufficient number of subjects in each stratum, the stratification factors will be used as covariates in the model.

At each visit, estimates of mean values and the mean differences between the active treated group and the placebo group will be derived from the model, together with the corresponding p-values, standard errors and 90% confidence intervals.

5.2.3. Analyses for Non-longitudinal Continuous Data

Analysis of Covariance (ANCOVA): non-longitudinal continuous data will be analyzed by ANCOVA with treatment as the factor and baseline value as a covariate. If there are a

sufficient number of subjects in each stratum, the stratification factors will be used as covariates in the model.

Estimates of mean values and the mean differences between the active treated group and the placebo group will be derived from the model, together with the corresponding p-values, standard errors and 90% confidence intervals.

5.2.1. Analyses for Categorical Data

None.

5.2.2. Analyses for Time to Event Data

None.

5.3. Methods to Manage Missing Data

Observed data will be used for descriptive statistics, i.e., missing data will not be imputed. For PK concentration, zero will be imputed for values below lower limits of detection (LLOD) or quantification (LLOQ). For biomarkers and other lab parameters, ½*LLOD or ½*LLOQ will be imputed.

5.3.1. Binary Endpoint

For statistical analysis of binary response defined on SES-CD endpoints, subjects with missing values will be handled by non-responder imputation (NRI) method, i.e., setting any missing values to be non-responsive (0), except that the visits missing due to COVID-19 will be removed from the analysis (i.e., the missing visits will not contribute to the denominator in the calculation of the proportion).

For statistical analysis of binary response defined on all other endpoints including CDAI, SF/AP, IBDQ and GHAS, missing values after study treatment discontinuation will be handled by non-responder imputation (NRI) method, others will be removed from the analysis (i.e., the missing visits will not contribute to the denominator in the calculation of the proportion).

5.3.2. Continuous Endpoints

For non-patient reported outcome variables, the missing continuous data will be used as is (Observed Case – OC).

For PRO endpoints, rules suggested by the developers of these PROs will be followed in calculating the values of a given component at a scheduled assessment. If these rules are not enough for imputing a value, then the missing values will be handled in the same way as non-patient reported outcome variables.

6. ANALYSES AND SUMMARIES

6.1. Induction Study Period

Analysis will be done pairwise between an active (PF-06651600 or PF-06700841) treated group and placebo. All efficacy analyses will be based on full analysis set.

6.1.1. Primary Endpoint(s)

- Endpoint: Proportion of subjects achieving SES-CD 50 at Week 12
- Statistical Method: Risk Difference using Mehrotra and Railkar method in Section 5.2.1
- Missing Data: Missing data, except that the visits missing due to COVID-19, will be handled by NRI.

Primary Analysis (NRI) - Setting any missing values to be non-responsive (0) for those treatment failure subjects withdraw from the study prior to Week 12

The reason of change of primary endpoint was to enable efficient comparison with contemporary and emerging trials for Crohn's disease, and same measurement SES-CD is still used for the primary endpoint, thus concern for inflation of type 1-error is minimized. For final analysis in the end of study, all study data will be used cumulatively, and hypothesis testing will be performed at alpha=0.05 for each compound. Considering the alpha spent at the past interim for CMEI is 0.03, the overall type 1 error for each compound is controlled at 0.08.

Sensitivity Analysis

In order to address the concern for potential inflation of type 1-error due to the change of primary endpoint, we consider a hypothetic adaptive design where both unblinded analyses of CMEI and SES-CD 50 were performed at the past interim when approximately 60% of subjects completed week 12 visit, and SES-CD 50 was then decided to be the primary endpoint going forward, the following analysis that still preserves type 1-error will be performed.

Inverse normal method is a way to preserve type-1 error for an adaptive design combining result from different stages.

$Z = w\Phi^{-1}(p_1) + \sqrt{1-w^2}\Phi^{-1}(p_2), w = \sqrt{0.6}$, since interim was at 60% subjects

p_1 and p_2 are p-value for stage 1 (before interim) and stage 2 (after interim).

Close testing procedure is a way to control type-1 error for multiple hypothesis tests. Define below two hypothesis tests:

H_{10}: $p_{1s} = p_{0s}$ where p_{1s} and p_{0s} are the SES-CD 50 rate for PF-06700841 and combined placebo respectively.
H_{20}: $p_{1c} = p_{0c}$, where p_{1c} and p_{0c} are the CMEI rate for PF-06700841 and combined placebo respectively.

To control the family wise error rate in strong sense under this hypothetical adaptive design, close testing procedure requires that the adaptively chosen hypothesis (H_{10}) will only be rejected at a type 1 error alpha level if combined p-value (by inverse normal method) for the intersection test ($H_{10} \cap H_{20}$)<alpha and combined p-value for H_{10}<alpha

For intersection test $H_{10} \cap H_{20}$:

- P-value for stage 1 is defined as min $\{2p_{(1)}, p_{(2)}\}$, where $p_{(1)} \leq p_{(2)}$ are the smaller and bigger p-value of the two hypothesis tests with only stage 1 data from subjects who completed week 12 or discontinued early from induction phase before interim.

- P-value for stage 2 is the p-value for SES-CD 50 with stage 2 data from subjects who completed week 12 or discontinued early from induction phase after interim.

For H_{10}, the p-values at each stage are based on independent SES-CD 50 data from each stage.

6.1.2. Secondary Endpoint(s)

The endpoint below will be analyzed using ANCOVA method (see Section 5.2.3) with no imputations for any missing data (OC).

- Change from baseline in SES-CD score at Week 12

The secondary binary efficacy endpoints below will be analyzed using CMH method with NRI for any missing data.

- Proportion of subjects achieving SES-CD 25 and CMEI at Week 12.
- Proportion of subjects achieving endoscopic remission (SES-CD ≤2) at Week 12.
- Proportion of subjects achieving mucosal healing (complete absence of ulcers) at Week 12.

6.1.3. Tertiary/Exploratory Endpoint(s)

The tertiary/exploratory non-longitudinal endpoint below will be analyzed using ANCOVA method with no imputations for any missing data (OC).

- Change from baseline in GHAS at Week 12.

The tertiary/exploratory longitudinal efficacy endpoints below will be analyzed using MMRM method (see Section 5.2.2) with no imputations for any missing data (OC), i.e. subjects with partial data will be included in MMRM analysis.

- The change from baseline in IBDQ total score and domains (Bowel Symptoms, Systemic Symptoms, Emotional Function and Social Function) at Weeks 4, 8, and 12.
- The change from baseline in EQ-5D-3L + VAS at Weeks 4, 8, and 12.
- The change from baseline in Short Form-36, Version 2 Acute (SF-36 v2): physical and mental component summary scores (PCS & MCS), and 8 domain scores at Weeks 4, 8, and 12.
- Change from baseline in PGIS score at Weeks 4, 8, and 12.
- Change from baseline in serum hsCRP levels over time.
- Change from baseline in fecal calprotectin over time.
- Change from baseline in serum IP10 levels over time.
- Change from baseline in *BCL-2* gene expression.
- Change from baseline in hematological values including reticulocytes, hemoglobin, neutrophils, platelets, and T, B and NK (TBNK) cells.

The tertiary/exploratory binary efficacy endpoints below will be analyzed using CMH method. Exact methods will be used for the all the timepoints. CDAI-100 and CDAI remission will only be summarized for subjects with baseline CDAI ≥220. Missing values after study treatment discontinuation will be handled by non-responder imputation (NRI) method, others will be removed from the analysis (i.e., the missing visits will not contribute to the denominator in the calculation of the proportion).

- Proportion of subjects achieving clinical response (≥30% reduction from baseline in AP or SF with neither worse than baseline) and remission (defined as SF ≤1.5 and AP ≤1, and both not worse than baseline) using stool frequency (SF) and abdominal pain (AP) measures at Weeks 2, 4, 6, 8, 10, and 12.
- Proportion of subjects achieving deep remission (endoscopic remission by SES-CD and clinical remission by SF and AP) at Week 12.
- Proportion of subjects with a CDAI-100 response at Weeks 2, 4, 8, and 12.
- Proportion of subjects with CDAI remission (CDAI <150) at Weeks 2, 4, 8, and 12.
- Proportion of subjects with IBDQ total score ≥170 at Weeks 4, 8, and 12.

- Proportion of subjects with ≥16 point increase in IBDQ total score from baseline at Weeks 4, 8, and 12.

- Proportion of subjects achieving IBDQ bowel symptom domain response at Weeks 4, 8, and 12. The improvement is defined as an increase of at least 1.2 points from baseline in average score among IBDQ bowel symptom domain (items 1, 5, 9, 13, 17, 20, 22, 24, 26, 29).

- Proportion of subjects achieving histologic remission (GHAS ≤4) at Week 12.

6.1.4. Pharmacokinetic Endpoint(s)

The PK endpoints are:

- PF-06651600 concentrations at Weeks 2, 4, 8, and 12.

- PF-06700841 concentrations at Weeks 2, 4, 8, and 12.

The concentrations will be summarized and presented with summary statistics. A population model may be developed for the purpose of estimating PK parameters. Any population PK model developed to characterize the PK data will be reported separately.

6.2. Open Label Extension Period

Summary statistics for any endpoints collected/measured during the OLE period (see Section 3.2) will be provided. There will be no direct comparisons (i.e. testing or p-values generated) between the two active treatment arms in the OLE period. All open label efficacy data, missing values after study treatment discontinuation will be handled by non-responder imputation (NRI) method, others will be removed from the analysis (i.e., the missing visits will not contribute to the denominator in the calculation of the proportion). For SES-CD 50 and CDAI remission at week 12 and week 64, within subject comparison will be made by McNemar's test for active treatment sequences.

6.3. Subgroup Analyses

Proportion of subjects achieving SES-CD 50 at Week 12 may be summarized for the subsets below (tables will be generated if we have sufficient number of subjects in each stratum):

- Anti-TNF experience (yes or no).

- Steroid use at baseline (yes or no).

- Baseline disease activity/extent

    - no isolated ileal disease and baseline SES-CD >15

- no isolated ileal disease and baseline SES-CD ≤15
- isolated ileal disease

6.4. Baseline and Other Summaries and Analyses

6.4.1. Baseline Summaries

Demographics and medical history including variables defined in Section 3.3 will be summarized by treatment group according to Pfizer standards.

6.4.2. Study Conduct and Subject Disposition

Subjects evaluation, disposition, discontinuation will be summarized, separately for induction and OLE periods, according to Pfizer standards.

6.4.3. Study Treatment Exposure

A summary of compliance and the number of doses received as well as the median total dose by visit and treatment group will be provided for each period.

The exposure to study drug will be summarized by total number of applications, the total number of days of dosing, and number and percentage of subjects who are compliant with the dosing regimen.

6.4.4. Concomitant Medications and Non-Drug Treatments

Prior drug and non-drug treatment, concomitant drug and non-drug treatment will be summarized according to Pfizer standards.

6.5. Safety Summaries and Analyses

Safety analysis will be based on the SAF analysis set.

All clinical AEs, SAEs, treatment-emergent signs and symptoms (TEAEs), withdrawal due to AEs, ECGs, vital signs, and safety laboratory data will be reviewed and summarized on an ongoing basis during the study to evaluate the safety of subjects.

Safety data will be presented in tabular and/or graphical format and summarized descriptively, where appropriate. All safety endpoints will be listed and summarized in accordance with Pfizer Standards. Categorical outcomes (e.g., AEs) will be summarized by subject counts and percentage. Continuous outcome (e.g., blood pressure, pulse rate, etc.) will be summarized using N, mean, median, standard deviation, etc. Change from baseline in laboratory data, ECGs and vital signs will also be summarized. Subject listings will be produced for these safety endpoints accordingly.

6.5.1. Adverse Events

The safety data will be summarized in accordance with Pfizer Data Standards. An adverse event is considered treatment-emergent adverse event (TEAE) to a given treatment if the event started during the effective duration of treatment regardless of whether a similar event of equal or greater severity existed in the baseline period. TEAE will be summarized separately for the induction period and open label extension period. If an AE started in induction and resolved in OLE, it will be counted as TEAE only in induction. However if such cross-period AE caused subject discontinuation during OLE period, that cross-period AE will be reported under OLE treatment period for summary of subject discontinuation from study drug or study.

All safety data will be summarized descriptively through appropriate data tabulations, descriptive statistics, categorical summaries, and graphical presentations. Safety endpoints for the study include:

- Treatment-emergent AEs and SAEs;
- Withdrawals due to AEs;
- Serious infections, defined as any infection (for e.g., viral, bacterial, and fungal) requiring hospitalization or parenteral antimicrobials. This definition programmatically should be consistent with 7.1.7 in the protocol.

As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. There will be no adjustment for multiple comparisons or stratification factors in the analyses unless specified.

6.5.2. Laboratory Data

Laboratory data will be listed and summarized in accordance with the Pfizer reporting standards. The incidence of clinical laboratory abnormalities and changes in lipid profile and creatinine will be summarized for each treatment group.

In addition, the absolute value, change from baseline, and percent change from baseline will be computed by visit and treatment group for hemoglobin, hematocrit, platelets, neutrophils, lymphocytes, serum creatinine, serum cystatin, creatinine based eGFR (estimated glomerular filtration rate), and cystatin based eGFR, liver function tests (ALT, AST, bilirubin, creatinine kinase), total cholesterol, triglycerides, HDL, LDL, and HDL to LDL ratio. Correspondingly, plots over time of these laboratory data will generated.

Laboratory test values meeting retest criteria and discontinuation criteria in protocol will be summarized, respectively.

6.5.3. Vital Signs, including Height and Weight

Vital signs including weight will be summarized at:

- Baseline, Weeks 2, 4, 8, 12, and Early Termination of induction period
- Weeks 16, 20, 24, 32, 40, 48, 56, 64, and 68/Early Termination of OLE period

6.5.4. Electrocardiogram

ECG parameters, if applicable, will be summarized at:

- Baseline, Week 12, and Early Termination of induction period
- Weeks 32, 48, and 64/Early Termination of OLE period

6.5.5. Physical Examination

Physical examinations will be summarized at:

- Baseline, Week 12, and Early Termination of induction period
- Week 64 and Early Termination of OLE period

7. INTERIM ANALYSES

7.1. Introduction

There will be one interim analysis plan (IAP) for this study.

Details regarding the analysis procedures used for the past interim analysis based on CMEI endpoint when approximately 60% of subjects completed Week 12 visit was provided in the interim analysis plan.

7.2. Interim Analyses and Summaries

The interim analysis results will be used to facilitate internal decision-making. The results will only be distributed to a select list of individuals involved in the internal decision-making process in order to protect the integrity of the study. This list of individuals will be provided in the interim analysis plan. The results of the interim analysis will not enable individuals directly involved in running the study (such as investigators) to identify treatment assignments for individual subjects still in the study. There are no prospective plans to stop the study early for success as a result of the interim analyses. During the interim analysis, some members of the study team may be unblinded and replaced with blinded colleagues. The subjects, investigators, and individuals from the sponsor (or designee) who interact with the investigators and monitor safety will continue to be blinded to individual study treatments throughout the follow up period of the study. Additional interim analyses may be performed based on emerging data.

7.3. Data Monitoring Committee

This study uses an External Data Monitoring Committee (E-DMC).

The E-DMC will be responsible for ongoing monitoring of the efficacy and safety of subjects in the study according to the charter. The E-DMC will review accumulating renal safety data and propose changes to the protocol as needed to ensure subject safety. The recommendations made by the E-DMC to alter the conduct of the study will be forwarded to Pfizer for final decision. Pfizer will forward such decisions, which may include summaries of aggregate analyses of endpoint events and of safety data that are not endpoints, to regulatory authorities, as appropriate. Additional information can be obtained in the E-DMC charter.

8. REFERENCES

1. Siddiqui O, MMRM vs. LOCF: A Comprehensive comparison based on simulation study and 25 NDA Datasets, Journal of Biopharmaceutical Statistics 19: 227-246, 2009.

2. Lanchin JM, A review of methods for futility stopping based on conditional power, Statistics in Medicine, 2005; 24: 2747-2764.

3. Mehrotra, D., Raikar, R. “Minimum Risk Weights for Comparing Treatments in Stratified Binomial Trials”. Statistics in Medicine, 2000, 19, pp. 811-825.

9. APPENDICES

Appendix 1. Summary of Efficacy Analyses

Efficacy analyses comparing active and placebo arms in the induction period will be based on the full analysis set.

Efficacy Endpoints	Analysis Method	Missing Data
Proportion of subjects achieving SES-CD 50 at Week 12	CMH	NRI
Proportion of subjects achieving SES-CD 50 at Week 12	CMH	OC
Change from baseline in SES-CD score at Week 12	ANCOVA	OC
Proportion of subjects achieving SES-CD 25 and CEMI at Week 12	CMH	NRI
Proportion of subjects achieving endoscopic remission (SES-CD ≤2) at Week 12	CMH	NRI
Proportion of subjects achieving mucosal healing (complete absence of ulcers) at Week 12	CMH	NRI
Change from baseline in GHAS at Week 12	ANCOVA	OC
The scores and change from baseline in IBDQ total score and domains (Bowel Symptoms, Systemic Symptoms, Emotional Function and Social Function) at Weeks 4, 8, and 12	MMRM	OC

Efficacy Endpoints	Analysis Method	Missing Data
The scores and change from baseline in EQ-5D-3L + VAS at Weeks 4, 8, and 12	MMRM	OC
The scores and change from baseline in Short Form-36, Version 2 Acute (SF-36 v2): physical and mental component summary scores (PCS & MCS), and 8 domain scores at Weeks 4, 8, and 12	MMRM	OC
Change from baseline in PGIS score at Weeks 4, 8, and 12	MMRM	OC
Change from baseline in serum hsCRP levels over time	MMRM	OC
Change from baseline in fecal calprotectin over time	MMRM	OC
Change from baseline in serum IP-10 levels over time	MMRM	OC
Change from baseline in *BCL-2* gene expression	MMRM	OC
Change from baseline in hematological values including reticulocytes, hemoglobin, neutrophils, platelets, and T, B and NK (TBNK) cells	MMRM	OC
Proportion of subjects achieving clinical response and remission using stool frequency (SF) and abdominal pain (AP) measures at Weeks 2, 4, 6, 8, 10, and 12	CMH	NRI
Proportion of subjects achieving deep remission (endoscopic remission by SES-CD and clinical remission by SF and AP) at Week 12	CMH	NRI
Proportion of subjects with a CDAI-100 response at Weeks 2, 4, 8, and 12	CMH	NRI
Proportion of subjects with CDAI remission (CDAI <150) at Weeks 2, 4, 8, and 12	CMH	NRI
Proportion of subjects with IBDQ total score ≥170 at Weeks 4, 8, and 12	CMH	NRI
Proportion of subjects with ≥16 point increase in IBDQ total score from baseline at Weeks 4, 8, and 12	CMH	NRI
Proportion of subjects achieving IBDQ bowel symptom domain response at Weeks 4, 8, and 12	CMH	NRI
Proportion of subjects achieving histologic remission (GHAS ≤4) at Week 12	CMH	NRI

Appendix 2. Definition and Use of Visit Windows in Reporting

Visit windows will be used for efficacy variables, and for any safety data that display/summarize by study visit. For other endpoints (e.g., ECG, vital signs), visit windows will be applied for summary statistics by study visits if required. For induction phase completers, any SES-CD results collected within the first 10 days and other efficacy variables within the first 5 days in OLE will be analyzed under induction treatment.

Visit Label	Target Day	Definition [Day window]
Screening		Days -28 to Day 0
Induction Period Weeks 0-12		
Week 0	Day 1, Baseline	Day 1
Week 2	15	Day 2 to 22
Week 4	29	Day 23 to 43
Week 8	57	Day 44 to 71
Week 12	85	Day 72 to 99
OLE Period Weeks 12-64		
Week 16	113	Days 100 to 127
Week 20	141	Days 128 to 155
Week 24	169	Days 156 to 197
Week 32	225	Days 198 to 253
Week 40	281	Days 254 to 309
Week 48	337	Days 310 to 365
Week 56	393	Days 366 to 421
Week 64	449	Days 422 to 463
Follow Up/End of Study		
Week 68	-	Days 464 to -

For the lab values, if the calculated study day for the labelled baseline visit is not study Day 1, but falls within 28 days before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

For the other values, if the calculated study day for the labelled baseline visit is not study Day 1, but falls before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

If two or more visits fall into the same window, keep the one closest to the Target Day. If two visits are equaled distant from the Target Day in absolute value, the later visit should be used.

Safety analysis will follow Pfizer standards.

Appendix 3. Calculation of IBDQ Scores

The IBDQ is a questionnaire consisting of 32 items. Each item is scored from 1 to 7 with 1 being the worst possible response and 7 being the best possible response.

Bowel Function Domain Score is calculated as the sum of the scores for questions 1, 5, 9, 13, 17, 20, 22, 24, 26 and 29.

Emotional Status Doman Score is calculated as the sum of the score for questions 3, 7, 11, 15, 19, 21, 23, 25, 27, 28, 30 and 31.

Systematic Symptoms Domain Score is calculated as the sum of the scores for questions 2, 6, 10, 14 and 18.

Social Function Domain score is calculated as the sum of the scores for questions 4, 8, 12, 16 and 32.

The Total IBDQ Score is calculated as the sum of the four domain scores above.

For the four domains, the score will be calculated as long as at least 50% of items within the domain are non-missing. Missing scores will be replaced using the mean of the non-missing scores within the domain. This leads to each domain score being calculated as:

$$\text{Domain score} = \frac{\text{Sum of non-missing scores} \times \text{Number of items in the domain}}{\text{Number of non-missing items in domain}}$$

If any of the four domain scores cannot be calculated due to more than 50% of the items having missing scores, then the total score will not be calculated.