PROTOCOL A3921288

A PHASE 3B/4, MULTI-CENTER, DOUBLE-BLIND, RANDOMIZED, PARALLEL GROUP STUDY OF TOFACITINIB (CP-690,550) IN SUBJECTS WITH ULCERATIVE COLITIS IN STABLE REMISSION

Statistical Analysis Plan
(SAP)

Version: 2.0

Date: 11 Mar 2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Table 1 Summary of Major Changes in Statistical Analysis Plan (SAP) Amendments

Version /Date	Associated Protocol Amendment	Rationale	Specific Changes
1.0 14 Sept 2017	Original 13 Jul 2017	Not Applicable	Not Applicable
2.0 11 Mar 2020	2 19 Jun 2019	Revision to align with protocol amendments 1 and 2.	1. Section 2.1 and Section 2.2 updated to align the study objectives and study design with the protocol 2. Section 3.2 updated to including detail of censoring rule for time to loss of remission; to move biomarker endpoint to this section to align with the protocol 3. Section 3.3.1 was removed as biomarker endpoints were listed as efficacy endpoints to align with the protocol 4. Section 3.5 definition for treatment emergent adverse event has been updated based on the new standard 5. Section 5.2 added details on reporting subgroups of subjects who are initially assigned to 5 mg BID and later are retreated with 10 mg BID 6. Section 5.2.1 added p-value, method for p-value, and sensitivity analyses 7. Section 5.2.2 modified the covariance matrix used in the linear mixed effects model 8. Section 5.3.1 modified the wording since the treatment duration has been changed to month 42. 9. Section 5.3.3 was added for method of handling data after dose escalation and dose reduction. 10. Section 6.2 moved Biomarker endpoints from section 6.3.1 11. Section 6.3.1 was removed. 12. Section 6.4 added P-value with method 13. Section 6.6.1 added details to compute incidence rates. 14. Section 7.1 added additional

			information about the unblinding plan. 15. Appendix 1 updated to include visits up to Month 42, and new windowing rule for retreatment subgroup.

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study A3921288.

2.1. Study Objectives

Primary Objective

- To evaluate the efficacy and safety of tofacitinib in subjects in stable remission on 10 mg BID who decrease the dose to and remain on 5 mg BID ("5 mg BID dose group") compared to subjects remaining on 10 mg BID.

Secondary Objectives

- To evaluate the efficacy and safety of tofacitinib in subjects in stable remission on 10 mg BID who decrease the dose to 5 mg BID with option for dose escalation for flare ("flexible dosing regimen") compared to subjects staying on 10 mg BID.

- To evaluate the efficacy and safety of the subset of subjects in stable remission on 10 mg BID who have flare after dose decrease to tofacitinib 5 mg BID and are re-treated with 10 mg BID.

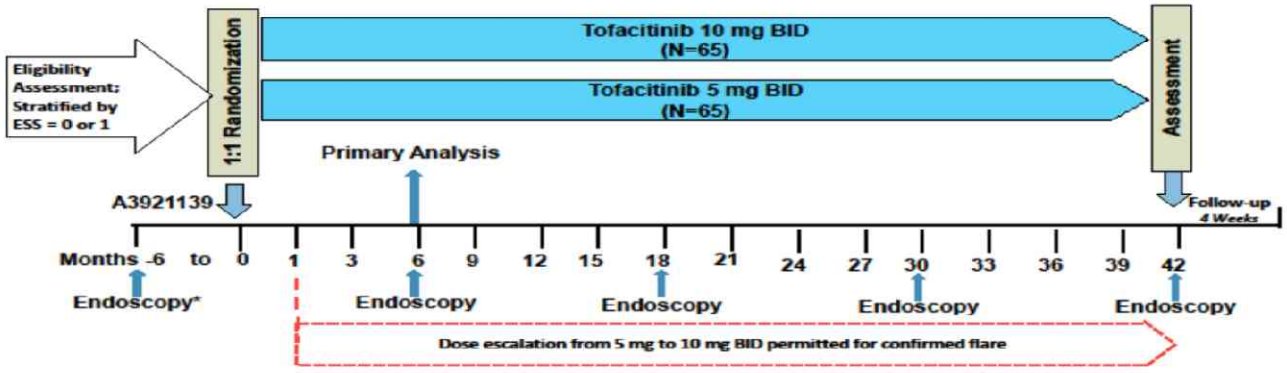
2.2. Study Design

This is a Phase 3b/4, multi center, randomized double blind study. This study will enroll subjects from currently ongoing Study A3921139 who are in stable remission on tofacitinib 10 mg BID for at least 6 months and who are not receiving any corticosteroids to treat their UC for at least 4 weeks prior to enrollment. Subjects must have been receiving tofacitinib 10 mg BID for a minimum of 2 consecutive years in Study A3921139 prior to baseline in order to be eligible for this study. Eligible subjects will be randomized in a 1:1 allocation ratio to receive either tofacitinib 5 mg BID or 10 mg BID at baseline of Study A3921288. Subjects will be stratified at baseline based on the endoscopic subscore (0 versus 1) of their most recent endoscopy.

Although approximately 130 subjects are estimated to be enrolled into this study (based on availability of eligible subjects from Study A3921139 and based on the conservative assumptions of the attrition rate of 50% - 60% in Study A3921139), the final sample size may exceed 130 subjects. The Sponsor will continue to engage with investigative sites participating in Study A3921139 to assess the ability to enroll additional eligible subjects. This study will have a total of 42 months of treatment duration. The primary analysis will be conducted after the last subject enrolled reaches their Month 6 study visit. The study will remain double-blinded to the site and subject to the initial treatment assignment at baseline.

Study visits will occur at baseline (enrollment), and at Months 1, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39 and 42. All subjects, who withdraw early or who complete this study, will have a 4 week safety follow up evaluation after the last dose of investigational product. A schematic of the study design is shown in Figure 1.

Figure 1. Study Design



*All subjects must have an endoscopy performed in Study A3921139 ≤ 6 months prior to baseline of Study A3921288 with an endoscopic subscore of 0 or 1, in order to be eligible for enrollment. Local site read of endoscopy only.

During the course of the study, if a subject experiences an increase in clinical symptoms, such as an increase in rectal bleeding or an increase in stool frequency, then an endoscopy should be performed to assess if the subject is experiencing flare. Flare must be confirmed prior to performing any dose adjustments.

Flare is defined by meeting one of the following 4 criteria:

1. An increase in rectal bleeding subscore by at least 1 point and an increase in endoscopic subscore by at least 1 point; OR

2. An increase in rectal bleeding subscore by at least 2 points and an endoscopic subscore >0; OR

3. An increase in stool frequency subscore by at least 2 points and an increase in the endoscopic subscore by at least 1 point; OR

4. An increase in endoscopic subscore by at least 2 points.

Once the subject is confirmed to meet the definition for flare, documented by local read endoscopy, the investigator may adjust the dose through interactive response technology system. Subjects may have a dose increase to 10 mg BID or remain on 10 mg BID, depending on their initial treatment assignment which will remain blinded. Dose adjustments will not be permitted prior to the Month 1 study visit.

Per protocol Amendment 2, version 19 June 2019, the study investigator or designee will be required to ask each subject at each study visit if he/she has any newly-developed risk factors for pulmonary embolism (PE), and if one is identified, the subject will need to have their tofacitinib dose adjusted to open-label 5 mg BID if they are taking open label 10 mg BID or are receiving blinded study medication (5 mg BID or 10 mg BID). At the time of Amendment 2, 138 subjects have been enrolled in the study and have each completed various durations of participation. Therefore, this new procedure will only be collected at visits applicable to each subject's individual participation.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint

- Remission based on modified Mayo score at Month 6. Remission based on modified Mayo score is defined as an endoscopic subscore of 0 or 1, stool frequency subscore of 0 or 1, and rectal bleeding subscore of 0.

3.2. Secondary Endpoints

- Time to loss of remission (flare) based on modified Mayo score. Loss of remission (flare) is defined as meeting at least one of the 4 criteria for flare above in Section 2.2.

- Remission based on modified Mayo score at scheduled visits other than Month 6 for dose groups.

- Remission based on modified partial Mayo score. Remission based on modified partial Mayo score is defined as stool frequency subscore of 0 or 1, and a rectal bleeding subscore of 0.

- Remission based on total Mayo score. Remission based on total Mayo score is defined as total Mayo score of 2 points or lower, with no individual subscore exceeding 1 point and a rectal bleeding subscore of 0.

- Remission based on partial Mayo score. Remission based on partial Mayo score is defined as partial Mayo score of 2 points or lower, with no individual subscore exceeding 1 point and a rectal bleeding subscore of 0.
- Change from baseline (of Study A3921288) in modified Mayo score.
- Change from baseline (of Study A3921288) in modified partial Mayo score.
- Change from baseline (of Study A3921288) in total Mayo score.
- Change from baseline (of Study A3921288) in partial Mayo score.
- Mucosal healing. Mucosal healing is defined as an endoscopic subscore of 0 or 1.
- Clinical response based on total Mayo score. Clinical response based on total Mayo score is defined as decrease from baseline in total Mayo score of the Induction study, of at least 3 points and at least 30%, with an accompanying decrease in the subscore for rectal bleeding of at least 1 point, or absolute subscore for rectal bleeding of 0 or 1.
- Change from baseline in fecal calprotectin and change from baseline in hs-CRP levels.

CCI

[redacted]

3.4. Baseline Variables

Unless specifically stated otherwise, the baseline value is defined as the last non-missing measurement collected prior to the first administration of study medication in this study, A3921288, at Day 1. Baseline Mayo scores will be derived based on the baseline individual subscores.

The endoscopic subscore (0 versus 1) at baseline is a stratification factor in randomization and will be used as a covariate or stratification factor in analyses. The endoscopic subscore based on the case report form will be used for the statistical analyses.

3.5. Safety Endpoints

The safety endpoints in this study include:

- Incidence and severity of adverse events (AEs).
- Incidence and severity of clinical laboratory abnormalities, and change from baseline in clinical laboratory values.
- Incidence of vital sign abnormalities and change from baseline in vital signs.

- Incidence of clinically significant changes in physical examinations from baseline.
- Safety events of interest including adjudicated safety events (eg, opportunistic infections, malignancy, gastrointestinal perforation, hepatic events, and cardiovascular events).

An adverse event is considered treatment emergent relative to a given treatment if the event starts after the first dose. The first dose is defined as the first dose in Study A3921288.

A 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into 1 of 3 tiers. Different analyses will be performed for different tiers (See Section 6.6.1).

Tier-1 events: These are pre-specified events of clinical importance and are maintained in a list in the product's Safety Review Plan.

Tier-2 events: These are events that are not tier-1 but are "common". A MedDRA Preferred Term is defined as a tier-2 event if there are at least 4 in any treatment group.

Tier-3 events: These are events that are neither tier-1 nor tier-2 events

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 releasing the database and classifications will be documented per standard operating procedures.

4.1. Full Analysis Set

The primary analysis population will be the Full Analysis Set (FAS) defined as all subjects who are randomized into the study and receive at least 1 dose of investigational product.

4.2. Per Protocol Analysis Set

No per protocol analysis set will be defined.

4.3. Safety Analysis Set

The safety analysis set will be defined as all subjects who receive at least 1 dose of investigational product.

A randomized but not treated subject will be excluded from the safety analysis set. A treated but not randomized subject will be included in the safety analyses.

4.4. Other Analysis Sets

No other analysis sets will be used.

5. GENERAL METHODOLOGY AND CONVENTIONS

The primary analysis will be performed to assess efficacy and safety when all subjects enrolled complete their Month 6 study visit (or drop out before Month 6 visit) and data are cleaned and locked. However, the final analyses will be performed after all subjects complete the study.

5.1. Hypotheses and Decision Rules

No formal hypotheses are to be tested.

5.2. General Methods

For the primary objective, analyses will be performed for comparisons of the following two dose groups:

- Tofacitinib 10 mg BID dose group: subjects who are initially assigned to tofacitinib 10 mg BID at randomization
- Tofacitinib 5 mg BID dose group: subjects who are initially assigned to tofacitinib 5 mg BID at randomization, regardless of whether the dose is escalated back to tofacitinib 10 mg BID or not (Data collected after the dose escalation from tofacitinib 5 mg BID to 10 mg BID will not be included in the analyses.)

For the first secondary objective, analyses will be performed for comparisons of the following two regimen groups:

- Tofacitinib 10 mg BID regimen group: subjects who are initially assigned to tofacitinib 10 mg BID at randomization.
- Tofacitinib 5 mg BID regimen group: subjects who are initially assigned to tofacitinib 5 mg BID at randomization, staying on tofacitinib 5 mg BID or having dose escalation from tofacitinib 5 mg BID to 10 mg BID. Data collected after the dose escalation will be included in the analyses.

For the second secondary objective, summary descriptive statistics will be reported for the subgroup of subjects who are initially assigned to tofacitinib 5 mg BID and later are re-treated with tofacitinib 10 mg BID based on the observed case data using the original windowing rules. In addition, the efficacy data after the dose escalation will be summarized using the new windowing rule by resetting the baseline (Day 0) at the dose escalation time.

5.2.1. Analyses for Binary Efficacy Data

The number and proportion of responders will be presented by dose group, regimen group and the subgroup of subjects who are initially assigned to tofacitinib 5 mg BID and later are retreated with tofacitinib 10 mg BID as described in Section 5.2.

For dose comparison between tofacitinib 10 mg BID and 5 mg BID, the stratified estimation of the treatment difference between tofacitinib 10 mg BID and 5 mg BID in the proportion of subjects will be presented along with its 95% confidence interval (CI) with the endoscopic

subscore at baseline (0 versus 1) as the stratification factor. The Cochran Mantel Haenszel (CMH) weight method will be used for the stratified estimation of the treatment difference. The stratified CI will be constructed using the NewCombe method by Yan and Su (2010).[1] In addition, p-value based on CMH Chi-square test stratified by endoscopic subscore at baseline (0 versus 1) will be presented. The same analysis method will be used for regimen groups.

For the second secondary objective, in the subgroup of subjects who are initially assigned to tofacitinib 5 mg BID and later are retreated with tofacitinib 10 mg BID, the number and proportion of responders will be also presented overall and by the endoscopic subscore at baseline (0 versus 1).

As sensitivity analyses for handling of missing data, remission based on modified Mayo score and remission based on total Mayo score will also be analyzed using a generalized linear mixed-effects model with a logit link and with dose group, visit, dose group by visit interaction, endoscopic subscore at baseline (0 versus 1) all as fixed effects, and subject as a random effect based on data collected at Month 6, Month 18, Month 30 and Month 42 if sufficient data are available at Month 42. The estimated mean, mean difference, with p-values and 95% confidence interval based on the model will be presented at each time point for dose group and regimen group.

5.2.2. Analyses for Continuous Efficacy Data

The descriptive statistics will be used to summarize continuous data. The continuous data and change from baseline will be summarized descriptively by visit and dose group, regimen group and the subgroup of subjects who are initially assigned to tofacitinib 5 mg BID and later are retreated with tofacitinib 10 mg BID.

For continuous efficacy endpoints based on total Mayo score evaluated for primary objective at Month 6, the change from baseline will be analyzed using an analysis of covariance (ANCOVA) model with dose group, the endoscopic subscore at baseline (0 versus 1) as factors and baseline score as a covariate. Otherwise, for continuous efficacy endpoints, the change from baseline will be analyzed using linear mixed effects model with baseline value, dose group or regimen group, the endoscopic subscore at baseline (0 versus 1), visit, and dose group or regimen group by visit interaction all as fixed effects, and subject as a random effect. The adjusted estimations and associated 95% CI for the overall difference between the two dose groups or two regimen groups will be computed at each visit.

5.2.3. Analyses for Categorical Efficacy Data

For categorical variables, the descriptive statistics such as proportions in each category will be reported.

5.2.4. Analyses for Time to Event Data

Kaplan-Meier estimates at the scheduled visits will be computed for time-to-event endpoint.

Kaplan-Meier plots for the time-to-event variables will be produced for dose comparison and regimen comparison.

The median time-to-event will be reported for each dose group and each regimen group. Time-to-event endpoints will be tested between dose groups or regimen groups, using a log-rank test stratified by endoscopic subscore at baseline (0 versus 1). In addition, cumulative event rates and associated CIs will be estimated from the Kaplan-Meier curves for each dose group and each regimen group. The dose or regimen comparison for the proportions of the event will be made using Wald test statistics.

5.2.5. Analysis of Safety Data

Safety data will be summarized based on observed-case data by dose group, regimen group and the subgroup of subjects who are initially assigned to tofacitinib 5 mg BID and later are retreated with tofacitinib 10 mg BID. Missing data will not be imputed.

For dose groups, data collected after the dose escalation from tofacitinib 5 mg BID to 10 mg BID will not be included in the analyses.

For regimen group, data from subjects with dose escalation from tofacitinib 5 mg BID to 10 mg BID will be treated as it is.

For tier-1 adverse events, an unconditional exact method for risk difference proposed by Chan and Zhang (1999) [2] will be used to compare tofacitinib 10 mg BID to 5 mg BID (dose group), and tofacitinib fixed 10 mg BID to flexible 5 mg/10 mg BID (regimen group). P-values and 95% confidence intervals will be formed. For tier-2 events, the risk difference will be presented along with its 95% confidence interval using the normal approximation for the difference in binomial proportions. For tier-3 events, the risk difference will be reported.

5.3. Methods to Manage Missing Data

5.3.1. Efficacy Endpoints

For binary efficacy endpoints derived from total Mayo score or partial Mayo score, subjects with missing scores in an analysis window will be considered as non-responders.

For continuous endpoints that are measured repeatedly over time, the missing values will be handled in a linear mixed effects model where missing values are assumed to be missing at random.

For continuous endpoints based on total Mayo score and modified Mayo score, data collected after baseline to Month 6 will be considered as Month 6 if a subject has missing data at Month 6; data collected after Month 6 to Month 18 will be considered as Month 18 if a subject has missing data at Month 18; data collected after Month 18 to Month 30 will be considered as Month 30 if a subject has missing data at Month 30; data collected after Month 30 to Month 42 will be considered as Month 42 if a subject has missing data at Month 42. The description of the handling of missing data within individual components of the Mayo score is described in an Appendix 2.1.

Subjects who drop out or complete the study without meeting the flare criteria will be censored at the last visit with Mayo score assessment for endpoint of time to loss of remission (flare).

Descriptive summary will also be presented based on observed data for efficacy data.

For the subgroup of subjects who are initially assigned to tofacitinib 5 mg BID and later are retreated with tofacitinib 10 mg BID, missing data will not be imputed.

For fecal calprotectin and hs-CRP, the missing values will be handled in a linear mixed effects model where missing values are assumed to be missing at random.

In addition, for hs-CRP, assayed values below the limit of quantification (BLQ) will be set to 10% less than the given assay lower limit for the specific biomarker. In listings BLQ values will be reported as "<LLQ" where LLQ will be replaced with the value for the lower limits of quantification. Assay values will be set to missing if a value has been collected as ND (ie, not done) or NS (ie, no sample).

5.3.2. Safety Endpoints

Missing data for safety endpoints will not be imputed.

For dose groups, data collected after the dose escalation will not be included in the analyses. For regimen group, data from subjects with dose escalation back to tofacitinib 10 mg BID will be treated as it is.

5.3.3. Method of Handling the Data after Dose Escalation or Dose Reduction

For this study, subjects may have a dose escalation to 10 mg BID due to flare or dose reduction to 5 mg BID due to PE risk factors. The data after the dose change for the dose group and regimen group are handled as follows.

- For Tofacitinib 10 mg BID dose group: For subjects with dose reduction due to PE risk factor, data collected after the dose reduction will not be included in the safety and efficacy analyses, and last observation carried forward (LOCF) will be used for binary and continuous efficacy endpoints. The subjects will be considered as censored at the dose reduction time for endpoint of time to loss of remission (flare).

- For Tofacitinib 5 mg BID dose group: data collected after the dose escalation (with or without dose reduction due to PE risk factor after the dose escalation) from tofacitinib 5 mg BID to 10 mg BID will not be included in the safety and efficacy analyses. The subjects will be treated as non-responders for any visit after the dose escalation visit for binary efficacy endpoints. For continuous efficacy endpoints, the data collected for any visit after the dose escalation visit will be considered as missing. For time to loss of remission (flare), the subject will be considered as having an event at the visit with a dose escalation if the event criteria are met or having a censored time at the visit with a dose escalation if the event criteria are not met.

- For Tofacitinib 10 mg BID regimen group: For subjects with dose reduction due to PE risk factor, data collected after the dose reduction will be included in the safety and efficacy analyses.

- For Tofacitinib 5 mg BID regimen group: Data collected after the dose escalation will be included in the analyses. For subjects with dose reduction due to PE risk factor after the dose escalation, data collected after the dose reduction will be included in the safety and efficacy analyses.

The following table provides some examples on how to handle the data after the dose escalation for flare and dose reduction for risk factors for PE:

Table 2 Examples on How to Handle Data After Dose Change

Examples	**Dose group**		**Regimen group**	
	Safety analysis	**Efficacy analysis**	**Safety analysis**	**Efficacy analysis**
5 mg->10 mg at Month X	Data after Month X will not be included	Data after Month X will not be included. Subjects will be treated as non-responders after Month X for binary, and missing for continuous endpoints	Data after Month X will be included	Data after Month X will be included
10mg->5 mg at Month X	Data after Month X will not be included	Data after Month X will not be included. LOCF will be used	Data after Month X will be included	Data after Month X will be included.
5mg->10mg at Month X->5mg at Month Y	Data after Month X will not be included	Data after Month X will not be included. Subjects will be treated as non-responders after Month X for binary, and missing for continuous endpoints	Data after Month X(including data after Month Y) will be included	Data from Month X (including data after Month Y) will be included.

At the time of finalizing SAP V2, all subjects completed Month 6 and no subjects had dose reduction before Month 6. The data handling rules for dose reduction will not be applied for the primary analysis performed at Month 6, however they will be applied to the final analysis performed at the end of the study.

6. ANALYSES AND SUMMARIES

Data will be summarized by dose group, regimen group and the subgroup of subjects who are initially assigned to tofacitinib 5 mg BID and later are retreated with tofacitinib 10 mg BID.

Efficacy analyses will be based on FAS and safety analyses will be based on the safety analysis set. Demographics and baseline characteristics will be summarized for the FAS.

6.1. Primary Efficacy Endpoint

The primary efficacy endpoint is remission based on modified Mayo score at Month 6.

As detailed in Section 5.2.1, the number and proportion of responders will be presented by dose group. The stratified estimation of the treatment difference between dose groups will be presented along with its 95% CI.

6.2. Secondary Efficacy Endpoints

Time to loss of remission (flare) based on modified Mayo score will be analyzed as described in Section 5.2.4.

Binary secondary efficacy endpoints will be analyzed using the same approach as described for the primary efficacy endpoint.

Change from baseline (of Study A3921288) in modified Mayo score, modified partial Mayo score, total Mayo score, and partial Mayo score will be analyzed as described in Section 5.2.2.

Fecal calprotectin and hs-CRP, and their change from baseline will be summarized descriptively for each visit by dose group and by regimen group.

The fecal calprotectin and hs-CRP data will also be log-transformed (natural logarithm) for the analyses. The change from baseline will be analyzed using the same approach as the continuous efficacy.

CCI [redacted]

6.4. Subset Analyses

Subgroup analyses will be conducted for remission based on modified Mayo score, remission based on total Mayo score, remission based on partial mayo score and MH for Endoscopic subscore at baseline (0, 1), prior TNF failure (Y/N), prior TNF exposure (Y/N). The number and proportion of responders will be presented by dose group, regimen group in each

subgroup. The difference between groups will be presented along with its 95% confidence interval using the normal approximation for the difference in binomial proportions. P-value based on the Chi-square test will be used to test for the group difference within a subgroup unless the cell frequencies are too small. In this case, the Fisher's exact test will be used.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

Demographics and baseline characteristics will be summarized by dose group based on FAS.

6.5.2. Study Treatment Exposure

Subjects were expected to take 4 tablets per day (2 tablets on a BID schedule) during the double-blind treatment period. However, since the number of tablets taken could be for a full day or only for a half day on the first treatment date or last treatment date, for simplicity, we assume that the first treatment date and last treatment date are days that the subjects took the expected number of tablets. The study drug compliance is defined as:

Drug compliance = (Actual Number of Tablets taken / Expected number of tablets)*100%.

Where expected number of tablets is defined as (Treatment Duration – 2)*expected number of tablets per day+ the actual number of tablets taken on the first and the last treatment dates if treatment duration>=2 days. If a subject only took drug on the first day, we assumed he/she took the expected number of tablets on that day, and define the compliance=100%.

Treatment duration in days is calculated as last treatment date – first treatment date +1.

6.6. Safety Summaries and Analyses

The safety data will be summarized in accordance with Pfizer Data Standards.

All safety data will be summarized descriptively by dose group, regimen group and the subgroup of subjects who are initially assigned to tofacitinib 5 mg BID and later are retreated with tofacitinib 10 mg BID in the safety analysis set.

6.6.1. Adverse Events

The analyses of adverse events under the 3-tier approach described in Section 3.5 is considered exploratory. There will be no adjustment for multiple comparisons or stratification factors in the analyses. For tier-1 events, the proportion of AEs observed in each group will be presented along with the point estimates and associated 95% confidence intervals of the risk difference, and p-value for dose comparison or regimen comparison using the exact methods proposed by Chan and Zhang (1999)[2] described in Section 5.2.5. For tier-2 events, the proportion of AEs observed in each group will be presented along with the point estimates and associated 95% confidence intervals of the risk difference using the normal approximation for the difference in binomial proportions. In case of zero event rate for tier-2 events in any of group, the 95% confidence intervals will not be computed. AEs will be arranged in the output sorted in descending point estimate of the risk difference within system organ class. Tier 3 AEs will not be summarized separately, they are included

in the overall standard AE summary. Footnotes in the outputs will include the methods used to derive any p-values and confidence intervals as per Pfizer standards.

A listing of subjects with dose reduction due to PE risk factors will be provided.

Incidence Rates (IR) of Adverse Events of Special Interest will be provided. Confidence intervals (95%) will be computed using the exact Poisson method.

For regimen groups, IR will be calculated only for adverse event data for events of special interest only. The numerator will be the number of unique subjects with a treatment emergent event as defined above during the interval from the first dose in A3921288 study to last dose of tofacitinib + 28 days at risk period. The denominator of the incidence rate is person time accruing from the subject's first dose to last dose + 28 days, or to the date of the first event, whichever occurs earlier. For a subject who died, the time will be the minimum of (the date of the last dose + 28 days, or death date).

For dose groups, IR will be calculated only for adverse events of special interest only. The numerator will be the number of unique subjects with a treatment emergent event as defined above during the interval from the first dose in A3921288 study to dose adjustment for dose escalated or dose reduced subjects otherwise to last dose + 28 days at risk period. The denominator of the incidence rate is person time accruing from the subject's first dose to day of dose escalation, or the first dose to day of dose reduction, or last dose + 28 days, or to the date of the first event, whichever occurs earlier. For a subject who died, the time will be the minimum of dose escalation day, dose reduction day, the day of the last dose + 28 days, or death day.

However, for mortality (death), malignancies (including NMSC), malignancies (excluding NMSC) and MACE related endpoints, all events regardless of their onset within 28 days or beyond after the last dose of tofacitinib will be included. The concept of 28 days is applied to subjects without events that discontinued or completed the study.

7. INTERIM ANALYSES

7.1. Introduction

The data collected from this study will be analyzed at two time points.

The primary analysis will be performed to assess efficacy and safety when all subjects enrolled complete their Month 6 study visit (or drop out before Month 6 visit) and data are cleaned and locked. The primary analysis will include efficacy data up to Month 6 and safety data up to the data cutoff date, and will be performed only for dose group comparison. This will be considered as the primary completion date. The initial treatment assignments at baseline will remain double-blinded to the site and subject (i.e., Pfizer will not provide the individual subject treatment/randomization codes to the sites until the study is completed). The Pfizer study team will become unblinded at the time of the Month 6 primary analysis and will remain unblinded for the duration of the study. The Argus Safety Database will not be unblinded for the primary analysis. Results from the primary analysis may be shared with Regulatory Authorities and may be published for scientific presentation.

The final analysis will be performed when all subjects complete the study and the database is cleaned and locked. This secondary analysis will be performed for all efficacy endpoints and safety endpoints and for dose group, regimen group and the subgroup of subjects who escalate their dose from tofacitinib 5 mg BID back to tofacitinib 10 mg BID.

An external data monitoring committee (E-DMC) will be responsible for ongoing monitoring of the safety of subjects in the study according to the DMC charter. 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.

7.2. Interim Analyses and Summaries

No interim analysis is planned for this study.

8. REFERENCES

1. Yan X., Su X.G. (2010). "Stratified Wilson and Newcombe confidence intervals for multiple binomial proportions." *Statistics in Biopharmaceutical Research* 2(3), 329-335.

2. Chan ISF, Zhang Z. (1999). Test-based exact confidence intervals for the difference of two binomial proportions. Biometrics, 55:1201–1209.

9. APPENDICES

Appendix 1. Definition of Analysis Visit Windows

Analysis visit windows will be used for efficacy and safety data that are summarized by visit.

Visit Label	Target Day	Analysis Visit window
Baseline	Day 1 (First dose date)	Up to Day 1
Month 1	Day 30	Days 2-60
Month 3	Day 90	Days 61-135
Month 6	Day 180	Days 136-225
Month 9	Day 270	Days 226-315
Month 12	Day 360	Days 316-405
Month 15	Day 450	Days 406-495
Month 18	Day 540	Days 496—585
Month 21	Day 630	Days 586-675
Month 24	Day 720	Days 676-765
Month 27	Day 810	Days 766-855
Month 30	Day 900	Days 856-945
Month 33	Day 990	Days 946-1035
Month 36	Day 1080	Days 1036-1125
Month 39	Day 1170	Days 1126-1215
Month 42	Day 1260	After Day 1216
Follow-up		Follow-up visit and >last dose day+5

For time to event variables, there is no need to have the windowing. The actual days will be used as the event time or censored time. But the event rate at each visit will be estimated at the corresponding protocol schedule day +10 days using the Kaplan-Meiermethod. If more than one observation from the same subject falls into the same window, the value closest to the target day will be used as the observation for that analysis visit. All observations will, however, be included in the listings.

Analysis visit window Follow-up is defined by the visit name=FOLLOW_UP and >last dose day +5. For example, if a subject had the last dose on Day 1000 and the follow-up visit at Day 1006, then the data from Day 1006 will be treated as follow-up data instead of Month 33 data. However, if a subject had the last dose on Day 1000 and the follow-up visit at Day 1003, then the data from Day 1003 will be treated as Month 33 data.

New windowing rule for retreatment group:

Visit Label	Target Day	Analysis visit window
Baseline	Day 1 (dose escalation day)	Last observation up to Day 1 (dose escalation day)
Month 3	Day 90	Days 2-135
Month 6	Day 180	Days 136-225
Month 9	Day 270	Days 226-315
Month 12	Day 360	Days 316-405
Month 15	Day 450	Days 406-495
Month 18	Day 540	Days 496-585
Month 21	Day 630	Days 586-675
Month 24	Day 720	Days 676-765

Appendix 2. Further Definition of Endpoints

Appendix 2.1. Mayo Scoring System for Assessment of Ulcerative Colitis Activity

The Mayo score ranges from 0 to 12, with higher scores indicating more severe disease. It consists of 4 subscores. Each sub-score ranges from 0 to 3. The 4 components of the Mayo score are:

Stool Frequency:
0 = Normal no. of stools for this patient.
1 = 1 to 2 stools more than normal.
2 = 3 to 4 stools more than normal.
3 = 5 or more stools more than normal.

Rectal Bleeding:
0 = No blood seen.
1 = Streaks of blood with stool less than half the time.
2 = Obvious blood with stool most of the time.
3 = Blood alone passes.

Findings on Endoscopy:
0 = Normal or inactive disease.
1 = Mild disease (erythema, decreased vascular pattern).
2 = Moderate disease (marked erythema, lack of vascular pattern, any friability, erosions).
3 = Severe disease (spontaneous bleeding, ulceration).

Physician's Global Assessment:
0 = Normal.
1 = Mild disease.
2 = Moderate disease.
3 = Severe disease.

The subject's stool diary information will be captured by an interactive voice response system (IVRS) connected through the phone system. Subjects will report stool frequency and the most severe bleeding of the day for rectal bleeding subscore daily at least 5 consecutive days prior to each study visit during the study. Stool frequency subscore will be calculated based on the average number (rounded to an integer) of stools over the prior 3 days minus the normal number of stools per day reported by the subject. The rectal bleeding subscore is the average number (rounded to an integer) of the 3 days prior to the study visit.

Diary data of stool frequency and rectal bleeding on bowel preparation date(s) will be excluded while calculating the subscores for stool frequency and rectal bleeding. If there are missing diary data, the average will be taken from the 3 most recently available days reported within 5 days prior to the study visit. If there are less than 3 available days reported within 5 days prior to the study visit, the average will be taken from the limited available data unless there is no diary data reported within 5 days. In this case, stool frequency and rectal bleeding subscores will be considered as missing.

If at least one of the 4 Mayo subscores is missing, the total Mayo score will be considered as missing.