ClinicalTrials.gov Posted Document Cover Page

Clinical Study Statistical Analysis Plan (SAP)

PROTOCOL TITLE: A **M**ulticenter, Randomized, Double-blind, Placeb**O**-controlled, Parallel-group Study to E**V**aluate the **E**fficacy and Safety of JTE-051 Administered for 12 Weeks to Subjects with Active **R**heumatoid **A**rthritis (MOVE-RA)

PROTOCOL NUMBER: AE051-G-13-003

SAP DATE: 10 August 2018

NCT NUMBER: NCT02919475

CONFIDENTIAL

Akros Pharma Inc.

Protocol AE051-G-13-003

A Multicenter, Randomized, Double-blind, Placebo-controlled, Parallel-group Study to Evaluate the Efficacy and Safety of JTE-051 Administered for 12 Weeks to Subjects with Active Rheumatoid **A**rthritis (MOVE-RA)

Statistical Analysis Plan

Final V1.0
10 Aug 2018

Signature Page

Protocol Number: AE051-G-13-003
Protocol Title: A Multicenter, Randomized, Double-blind, Placebo controlled, Parallel-group Study to Evaluate the Efficacy and Safety of JTE-051 Administered for 12 Weeks to Subjects with Active Rheumatoid Arthritis (MOVE-RA)

The signatures below indicate approval of this Statistical Analysis Plan.

Prepared by:

Reviewed by:

Reviewed by:

Approved by:

Approved by:

Approved by:

Approved by:

Table of Contents

1. INTRODUCTION

Final version of the protocol for this phase 2 double-blind (DB), placebo-controlled, parallel-group study evaluating efficacy, safety and pharmacokinetic (PK) of JTE-051 in subjects with active rheumatoid arthritis (RA) on stable disease-modifying anti-rheumatic drug (DMARD) therapy, including methotrexate (MTX) was completed on April 24, 2015. The protocol was amended on 24 March 2016. The protocol Amendment 1 and the electronic case report form (eCRF) contain details on the study conduct and data collection.

This document describes the statistical methods and data presentations to be used in the summary tables, listings and figures.

2. OBJECTIVES

- To evaluate the efficacy of JTE-051 in subjects with active RA receiving background non-biologic DMARD therapy
- To evaluate the safety and tolerability of JTE-051 administered for 12 weeks to subjects with active RA receiving background non-biologic DMARD therapy
- To evaluate the PK of JTE-051 in plasma of subjects with active RA

3. STUDY DESIGN

This is a multicenter, randomized, double-blind, placebo-controlled, parallel-group, 12-week study in biologic and kinase inhibitor treatment-naive RA subjects on stable background non-biologic DMARD therapy, including MTX.

Eligible subjects are randomized at Visit 2 to receive JTE-051 50 mg, 100 mg, 150 mg, 200 mg or placebo once daily for 12 weeks. Subjects continue to receive up to two non-biologic DMARDs, including MTX throughout the study (at a stable background dose and route of administration for at least 12 weeks prior to Visit 2 and throughout the study) and return for a Follow-up Visit approximately four weeks after the last dose of study drug is administered. Subject randomization are stratified by geographical region (Europe, the Middle East and Africa [EMEA], Latin America, [LA] and the United States [US]) and by the screening high-sensitivity C-reactive protein (hs-CRP) (≤2 x ULN, >2 x ULN).

The study duration will be of approximately 20 weeks per subject:

- Up to a 28-day Screening Period
  Note: Certain pre-defined exceptions that would allow for a longer duration of the screening period to ensure a 28-day pre-randomization stabilization of concomitant medications arepermitted, as described in Section 3.5.4.1 of the protocol.
- A 12-week double-blind Treatment Period
- Approximately a 4-week Follow-up Period

3.1 Study Procedures

The schedule of study procedures is described in the following Figure 1 and Table 1.

Figure 1. Planned Study Schema

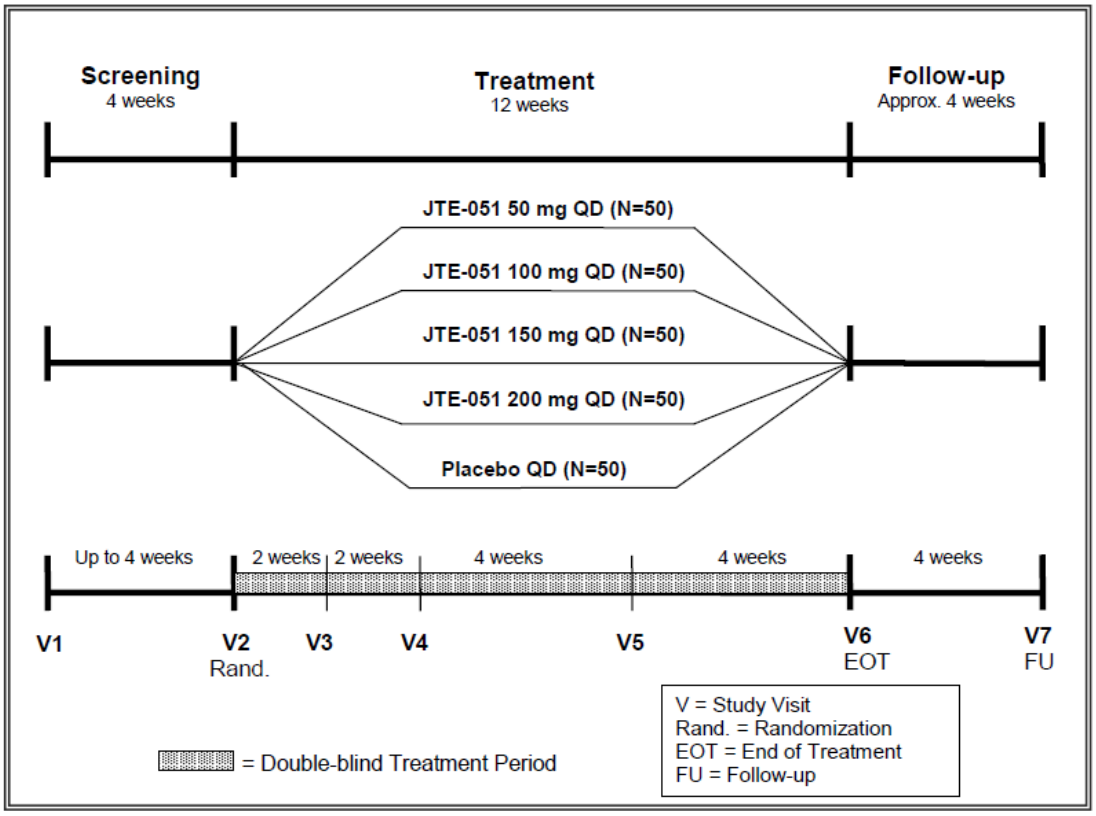


Table 1. Planned Schedule of Study Procedures

	Screening Period	Treatment Period					Follow-up Period
Duration/ Study Week (Day)[a]	Up to 4 weeks	Week 0 (Day 1)	Week 2 (Day 14±2)	Week 4 (Day 28±2)	Week 8 (Day 56±2)	Week 12 (Day 84±2)	Week 16 (Day 112±2)
Visit	1	2	3	4	5	6	7
Informed Consent	X						
Inclusion/ Exclusion Criteria	X	X					
Medical History	X						
Demographic Information	X						
Review Prior/ Concomitant Medications	X	X	X	X	X	X	X
Physical Exam	X	X	X	X	X	X	
Vital Signs, including Weight	X	X	X	X	X	X	X
Height and Calculate BMI	X						
12-Lead ECG	X	X		X		X	
Chest Radiography[c]	X						
TB test (PPD or quantiFERON gold)	X						
Drugs of Abuse Screen	X						
Viral Serology	X						
FSH (post-menopausal females only)	X						
Pregnancy Test (all females)[d]	X	X	X	X	X	X	X
Serum Biochemistry	X	X	X	X	X	X	X
Hematology	X	X	X	X	X	X	X
Urinalysis	X	X	X	X	X	X	X
Coagulation	X	X		X		X	X
Lipid Profile	X	X		X		X	X
25-hydroxyvitamin D	X						
Serum IgG, IgM and IgA		X				X	
RF[e] and Anti-CCP antibodies[f]		X					
Bone-specific ALP	X	X				X	
hs-CRP	X	X	X	X	X	X	X
CRP		X				X	
Tender Joint Count (68)	X	X	X	X	X	X	X
Swollen Joint Count (66)	X	X	X	X	X	X	X
Subject's Assessment of Arthritis Pain	X	X	X	X	X	X	X
Subject's Global Assessment of Arthritis	X	X	X	X	X	X	X
Physician's Global Assessment of Arthritis	X	X	X	X	X	X	X
HAQ-DI	X	X	X	X	X	X	X
Randomization using IWRS		X					
Access IWRS and Dispense Study Drug[g]		X	X	X	X		
Collect Study Drug and Check Compliance[h]			X	X	X	X	
Study Drug Administration[i]		X	X	X	X		
JTE-051 Trough PK Blood Samples			X	X	X	X	
Document Adverse Events[j]	X	X	X	X	X	X	X

a. The target day for each visit timepoint will be calculated relative to the date of Randomization Visit (Visit 2) and not relative to the date of the previous visit. All visits should be performed within the windows specified in the table. Every attempt should be made to have the subject attend each visit as scheduled. The investigational site is encouraged to a make a reminder phone call to the subject approximately a day or two before the scheduled visit. However, if a subject is unable to attend a

visit within the specified windows, the visit should be scheduled as closely as possible to these windows. A subject should not skip a protocol-specified visit due to scheduling difficulties.

[redacted]

c. Unless a chest radiography had been performed within 3 months (12 weeks) of the Screening Visit, and documentation is available for review by the Investigator and inclusion in the subject's file.

d. At Visit 1 serum pregnancy test will be performed; at all other visits, urine pregnancy tests will be performed. Pregnancy tests may also be repeated as needed or per request from institutional review boards (IRB)/independent ethics committees (IEC) or if required by local regulations.

e. Rheumatoid Factor

f. Anti-cyclic citrullinated peptide antibodies

g. At Visit 2, a back-up drug blister card will be dispensed to all subjects, in addition to the regular study drug blister cards. Subjects will be instructed not to use the back-up blister card unless all drug in the regular blister cards has been used and to bring all used and unused (including back-up) blister cards to each study visit for accountability purposes. At Visit 6, study drug will not be dispensed. If a study subject discontinues study drug prematurely, the appropriate systems will be accessed to record termination as soon as possible after the decision has been made.

h. Study drug compliance will be calculated by the site at each visit during the treatment period starting at Visit 3, based on the number of tablets dispensed/returned by the subject. The subject is to maintain the back-up blister card for the duration of the treatment period.

i. Oral administration for 12 weeks starting on the day of randomization at Visit 2, QD in the morning, regardless of meals. On study visit days, subjects should take their scheduled study treatment from their existing study drug supply (if available) at the site, under the supervision of the investigator or designee after all study-related procedures have been performed (except for Visit 6, when no study drug will be administered, as the last dose will be taken the day prior to the visit). If no tablets from the previously supplied non-back-up blister cards are available, then the subject should dose from the new blister card supplied at that visit. Following completion of accountability and compliance assessments, the back-up blister card will be redisposed at Visits 3 through 5 and will be collected at Visit 6 along with all non-back-up drug blister cards. The back--up blister card should only be utilized by subjects if all study drug from the regularly supplied (i.e., non-back-up) blister cards has been used.

j. Adverse event information will be collected at the specified time points as well as at any time when a clinical research unit staff member becomes aware of an AE after the subject signs the informed consent for the study. However, stable or improving pre-existing conditions that are detected through the screening procedures throughout the Screening Period (e.g., abnormalities in ECG, physical examination, [redacted] vital signs and laboratory tests) are medical history and should be documented accordingly.

4. SUBJECT POPULATIONS

4.1 Randomized Population

The randomized population consists of all subjects who are randomized at Visit 2 to one of the five treatment groups: JTE-051 50 mg QD, 100 mg QD, 150 mg QD, 200 mg QD and placebo QD.

4.2 Safety Population

Safety population consists of the randomized subjects who receive at least one dose of JTE-051 or placebo after the randomization, including those who did not complete the study.

4.3 Intent-To-Treat (ITT) Population

The Intent-To-Treat (ITT) population consists of the randomized subjects who receive at least one dose of the study drug and have at least one post-baseline efficacy assessment during the DB treatment period.

4.4 Per Protocol (PP) Population

The Per Protocol (PP) population is a subset of the ITT population in which subjects do not incur any of the major protocol deviations below.

Major protocol deviation criteria:

1. Dosing compliance not in 80-120% (based on actual dosing period)
2. Subject's last dose not between Days 68-100
3. Subject did not take 272-400 tablets in total
4. >4 days between last dose and last non-follow-up efficacy assessment
5. Not meeting Inclusion Criterion #3 or #4
6. Not meeting Exclusion Criterion #1, #2 or #28
7. Subject took "key" prohibited concomitant medications or had unstable background therapy that could potentially affect efficacy results/interpretation
8. Subject took incorrect study medication, e.g. due to incorrect drug dispense
9. Significant procedural deviations/non-compliance that could potentially affect efficacy results/interpretation

A pre-analysis meeting will take place after all data have been entered into the database and cleaned, but before the release of the randomization code, to identify the PP population. The decisions made to select the PP population will be documented.

4.5 Pharmacokinetic (PK) Population

The PK population consists of the randomized subjects who receive at least one dose of JTE-051 and have at least one usable JTE-051 plasma concentration measurement.

5. ANALYSIS PARAMETERS

5.1 Demographic and Baseline Characteristics

Demographic variables include age, gender, ethnic group, geographical region, race, body weight, height and body mass index (BMI).

Baseline (last non-missing value prior to first dose) subject characteristics include the following parameters: duration (years) of active RA, tender/painful joint counts (68 and 28), swollen joint counts (66 and 28), subject's assessment of arthritis pain, subject's global assessment (SGA) of disease activity, physician's global assessment of disease activity (PGA), Health Assessment Questionnaire Disability Index (HAQ-DI), hs-CRP, Disease Activity Score 28 with CRP (DAS28-CRP), Simple Disease Activity Index (SDAI), Clinical Disease Activity Index (CDAI), Rheumatoid Factor (RF), Methotrexate (MTX) dose, Concomitant Steroids, % positive anti-CCP antibody and % patients who took DMARDs (i.e., Sulfasalazine, Hydroxychloroquine or Chloroquine). In addition, hs-CRP at screening will be summarized. Where both screening and screening re-test values are available, the screening re-test value will be used in the summary.

Summary tables for demographic and baseline characteristics will be generated for randomized, safety, ITT and PP populations, respectively.

5.2 Efficacy Parameters

5.2.1 Primary Efficacy Parameter

The primary efficacy parameter is the ACR20 response rate at end of treatment (EOT), defined as percentage of subjects achieving at least 20% improvement from baseline in tender and swollen joint counts and at least 20% improvement from baseline in 3 of the 5 remaining ACR core set measures.

5.2.2 Secondary Efficacy Parameters

The secondary efficacy parameters (evaluated at Weeks 2, 4, 8, 12, 16 and EOT, respectively, unless otherwise stated):

- Percentage of subjects achieving ACR20/50/70 response rate (except ACR20 at EOT which is the primary parameter)
- Change from baseline in SDAI
- Percentage of subjects who achieved remission based on SDAI ($\leqslant$3.3)
- Percentage of subjects who achieved low disease activity based on SDAI ($\leqslant$11)
- Change from baseline in CDAI
- Percentage of subjects who achieved remission based on CDAI ($\leqslant$2.8)
- Percentage of subjects who achieved low disease activity based on CDAI ($\leqslant$10)
- Percentage of subjects achieving Boolean Remission
- Change from baseline in DAS28-CRP
- Percentage of subjects with DAS28-CRP of <2.6
- Percentage of subjects with DAS28-CRP of <3.2

- Percentage of subjects with good European League Against Rheumatism (EULAR) response at Week 12
- Percentage of subjects with moderate EULAR response at Week 12
- The lowest percentage change from baseline of 3 measures: tender joint count, swollen joint count and median change of the remaining 5 core measures (numeric ACR [ACR-N] index)
- Change from baseline in the Health Assessment Questionnaire Disability Index (HAQ-DI)
- Change from baseline in the number of tender and swollen joint counts (68/66 joints will be counted)
- Change from baseline in subject pain score by numeric rating scale (NRS)
- Change from baseline in subject global assessment (SGA) of disease activity by NRS
- Change from baseline in physician global assessment (PGA) of disease activity by NRS
- Change from baseline in hs-CRP

[redacted]

5.2.5 Efficacy Endpoints Calculation

5.2.5.1 Numeric ACR (ACR-N) Index

The numeric ACR (ACR-N) will be derived by the sponsor for Visits 3 through 7 based on the ACR core set measures collected by the Investigator. See Section 3.5.6.1 of the protocol for the description of the ACR core set measures assessments by the Investigator.

This is a continuous variable expressing the percentage improvement from baseline a subject has experienced, defined as the minimum of percent improvement in swollen joints, tender joints and the median percent change in the other five ACR core set measures:

1. Subject's Assessment of Arthritis Pain.
2. Subject's Global Assessment of Disease Activity.
3. Physician's Global Assessment of Disease Activity.
4. Health Assessment Questionnaire Disability Index
5. Acute phase reactant (hs-CRP)

Thus, an ACR-N of X means that the subject has achieved a change of at least X% in the tender and swollen joint counts and a change of at least X% in the median of the other five other parameters. Numeric ACR will take a negative value if a there is a worsening in either the swollen joint count, tender joint count and the median of the other five ACR core set measures (i.e., a subject can have a negative ACR-N despite improving in two of the three components).

5.2.5.2 The ACR20/50/70 Response Rate

The ACR20/50/70 response rate will be calculated by the sponsor for Visits 3 through 7 based on the ACR core set measures collected by the Investigator.

This composite parameter measures improvement in tender (68) and swollen (66) joint counts and improvement in at least three of the following five ACR core set measures as listed above.

The ACR20/50/70 response rate is a dichotomous variable with a positive (= responder) or negative (= non-responder) outcome.

ACR20 is ≥20% improvement

ACR50 is ≥50% improvement

ACR70 is ≥70% improvement

5.2.5.3 Simple Disease Activity Index (SDAI)

The SDAI will be calculated by the Sponsor for all study visits. See Section 3.5.6.1.2 of the Protocol for the list of 28 joints included in the SDAI scoring.

The scoring of SDAI will be done as follows:

Variable	Range
Tender joint score	(0-28)
Swollen joint score	(0-28)
Patient global score	(0-10)
Provider global score	(0-10)
C-reactive protein (mg/dL)	(0-10)
Add the above values to calculate the SDAI score	**(0-86)**

Interpretation of scoring will be done as follows:

SDAI Score Interpretation	
0.0-3.3	Remission
3.4-11.0	Low Activity
11.1-26.0	Moderate Activity
26.1-86.0	High Activity

5.2.5.4 Clinical Disease Activity Index

The CDAI will be calculated by the Sponsor for all study visits. See Section 3.5.6.1.2 of the Protocol for the list of 28 joints included in the CDAI scoring.

The scoring of CDAI will be done as follows:

Variable	Range
Tender joint score	(0-28)
Swollen joint score	(0-28)
Patient global score	(0-10)
Provider global score	(0-10)
Add the above values to calculate the CDAI score	(0-76)

Interpretation of scoring will be done as follows:

CDAI Score Interpretation	
0.0-2.8	Remission
2.9-10.0	Low Activity
10.1-22.0	Moderate Activity
22.1-76.0	High Activity

5.2.5.5 Disease Activity Score (DAS) 28

The DAS assessment is a derived measurement with differential weighting given to each component. The DAS 28 (CRP) will be calculated by the Sponsor at all study visits. See Section 3.5.6.1.2 of the Protocol for the list of 28 joints included in the DAS calculation.

The components of the DAS 28 arthritis assessment include:

1. Tender/Painful Joint Count (28)
2. Swollen Joint Count (28)
3. CRP
4. Subject's Global Assessment of Disease Activity

The DAS 28 will be calculated using the following formula:

DAS28(CRP) = 0.56 X $\sqrt{}$(TJC28) + 0.28 X $\sqrt{}$(SJC28) + 0.36 X Ln(CRP+1) + 0.014 X SGA + 0.96

5.2.5.6 Percentage of Subjects achieving Boolean Remission

This secondary efficacy parameter will be calculated by the Sponsor, based on the following definition:

- Tender joint count (TJC) (28) ⩽1
- Swollen joint count (SJC) (28) ⩽1
- C-reactive protein ⩽1 mg/dL
- Subject's global assessment (SGA) of disease activity ≤1 (on a 0 - 10 scale)

5.2.5.7 EULAR Response Criteria

The EULAR response criteria will be assessed by the Sponsor.

This parameter considers both the change in DAS28 between two time points, as well as DAS28 value at a specific time point.

DAS28-CRP will be utilized in calculating the EULAR response in the study subjects.

DAS28 improvement → Present DAS28↓	>1.2	>0.6 and ≤1.2	≤0.6
≤3.2	good response	moderate response	no response
>3.2 and ≤5.1	moderate response	moderate response	no response
>5.1	moderate response	no response	no response

5.2.5.8 HAQ-DI Score and Calculation

The HAQ-DI's scoring conventions allow for the computation of two disability indices:

1. The Standard HAQ-DI is the preferred and traditional scoring method, which considers the use of aids/devices.
2. The Alternative Disability Index does not consider the use of aids/devices.

A subject must have a score for at least six of the eight categories. Otherwise a HAQ-DI score cannot be validly computed. For this study, the standard computing method will be applied.

Computing the Standard HAQ-DI Score (With Aids/Devices): There are three steps:

1. Sum the 8 category scores by using the highest sub-category score from each category (for example, in the category "EATING" there are three sub-category items. If a subject responds with a 1, 2 and 0, respectively; the category score is 2).
2. Adjust for use of aids/devices and/or help from another person when indicated. Table 2 identifies the Aid/Devices companion variable for each HAQ-DI category.

Table 2. Companion Aids/Device Items for HAQ-DI Categories

HAQ-DI Category	Companion Item
Dressing & Grooming	Devices used for dressing
Arising	Built up or special chair
Eating	Built up or special utensils
Walking	Cane walker, crutches, wheelchair
Hygiene	Raised toilet seat, bathtub seat, bathtub bar Long handled appliances in bathroom
Reach	Long handled appliances for reach
Grip	Jar opener (for jars previously opened)

- Adjust the score for a category by increasing a zero or a one to a two.
- If a subject's highest score for that sub-category is a two it remains a two, and if a three, it remains a three.

3. Divide the summed category scores by the number of categories answered (must be a minimum of 6) to obtain a HAQ-DI score of 0-3 (3=worst functioning)
4. If any component in HAQ-DI derivation is missing and HAQ-DI cannot be derived, no imputation will be done, and HAQ-DI will be left as missing.

[redacted]

5.3 Safety Parameters

5.3.1 Adverse Events (AEs)

The Medical Dictionary for Regulatory Activities (MedDRA) Version 21.0 will be used to map verbatim AEs to preferred terms and their respective primary system organ classes. However, a different (non-primary) system organ class may be selected where appropriate after a medical review of the mapping. Adverse events will be characterized as either pre-treatment or treatment-emergent according to the criteria described in the following sections.

5.3.2 Pre-treatment Adverse Events

A pre-treatment AE is one with an onset date prior to the date of the first dose of study drug.

For an AE occurring on the day of first dose, all AEs occurring on the day of first dose will be considered as TEAE when the time of AE occurrence relative to the dosing is unknown.

5.3.3 Treatment-Emergent Adverse Events

A treatment-emergent AE is defined as any AE with an onset date on or after the first dose of study drug, provided that this AE did not previously occur during the pre-treatment period (i.e., before the first dose of study drug taken after randomization). However, a pre-treatment AE can become treatment-emergent if it continued from the pre-treatment period but worsened after the first dose of study drug or it stopped and re-occurred (a new onset time) with a worsen severity after the first dose of study drug.

5.3.4 Medical History

A complete medical history will be obtained at the Screening Visit (Visit 1) and will include evaluations for past or present conditions. [redacted] Stable or improving pre-existing conditions that are detected as part of the screening procedures (e.g., ECG, physical examination, [redacted] vital signs and laboratory test abnormalities) are medical history.

5.3.5 Study Medication Compliance

Total number of tablets taken will be recorded in the eCRF for the corresponding intervals as scheduled in Table 1. Compliance of the study medication will be calculated as follows:

$$\text{Compliance (\%)} = \frac{\text{Total tablets taken}}{\text{Treatment duration} \times 4} \text{ x } 100\%$$

Compliance will be calculated for both the treatment duration between visits and the entire treatment period.

5.3.6 Clinical Laboratory Results

Hematology, coagulation, serum biochemistry, bone-specific alkaline phosphatase, lipid profile and urinalysis are collected as scheduled in Table 1. The following information will be

derived for each of the hematology, coagulation, serum biochemistry, urinalysis and other parameters:

- Change from baseline at all post-baseline visits
- Classification relative to the normal range ('Low', 'Normal' or 'High')
- Treatment-Emergent potentially clinically significant (PCS) abnormalities, if applicable

Abnormalities in liver tests will be identified.

In addition, 25-hydroxyvitamin D, rheumatoid factor (RF), anti-cyclic citrullinated peptide (anti-CCP) antibodies and serum immunoglobulins G, M and A (IgG, IgM and IgA), viral serology, drugs of abuse and alcohol screen, pregnancy test and follicle stimulating hormone will also be obtained at the corresponding scheduled visit, respectively.

5.3.7 Vital Signs

Vital signs and body weight are collected as scheduled in Table 1. Vital signs include systolic/diastolic blood pressure, pulse rate, respiratory rate and body temperature (°C). Subjects must rest in a sitting position for at least 5 minutes in preparation for blood pressure and pulse rate assessments. The following information will be derived for all vital signs and body weight:

- Change from baseline at all post-baseline visits
- Treatment-emergent PCS abnormalities, if applicable

5.3.8 12-Lead ECG

Twelve-lead ECG recordings and conduction intervals including RR, PR, QRS, QT are obtained according to the schedules summarized in Table 1. The Fridericia-corrected QT interval (QTcF) (Fridericia's formula, $QTcF = QT/RR^{1/3}$) will be calculated and reported. The following information will be derived for each of the 12-lead ECG measurements:

- Change from baseline to all post-baseline visits
- Treatment-emergent PCS abnormalities, if applicable

The overall clinical interpretation of an ECG is assessed as follows: normal, abnormal not clinically significant or abnormal clinically significant.

5.3.9 Physical examination

Physical examination is conducted as scheduled in Table 1. The physical examinations include examination of the following body systems: general appearance, skin (including hair and nails), HEENT (head, ears, eyes, nose, throat), neck/thyroid, chest/lungs, cardiovascular, gastrointestinal (GI), [redacted] psychiatric/emotional, lymphatic and musculoskeletal. Clinically significant abnormal physical examination findings will be recorded as medical history or AE where appropriate.

5.4 Pharmacokinetic (PK) Parameters

Trough concentrations of JTE-051 are measured from visit 3 to visit 6 during the treatment period.

The relationship between the JTE-051 dose (exposure) and response will be assessed (exploratory).

5.5 Medications/Procedures

At each study visit, investigators will review concomitant medications and procedures with the patients. The World Health Organization (WHO) Drug Dictionary March 2018 will be used to classify medications by preferred term, chemical ingredient names and WHO Level 3 Anatomical Therapeutic Chemical (ATC) classification of trade name. Medication records will be characterized as prior medication and/or concomitant medication based on the subjects' start and end date of the study drug.

5.4.1 Prior Medications/Procedures

Prior medications/procedures are defined as any medications/procedures started and ended before the first dose of the study drug.

5.4.2 Concomitant Medications/Procedures

Any medications/procedures that begin before the start of the first dose of study drug and continue past that, or that begins on or after the first dose of study drug and up to 28 days post the date of the last dose of the DB medications will be considered as concomitant medications/procedures.

6. STATISTICAL METHODOLOGY

Titles and headers of all statistical analysis tables will indicate the corresponding study population; the number of subjects for the population and for each treatment will be presented in the tables. Summary statistics for continuous parameters will include the number of subjects (N), mean, standard deviation (SD), median, minimum and maximum. For plasma concentrations, summary statistics will also include the coefficient of variation (CV%). Summary statistics for discrete parameters will include frequency count and percentage of which the denominator will be clearly indicated in the analysis table.

For efficacy and safety summaries, mean and median will be presented to 1 more decimal than the source data; SD will be presented to 2 more decimals than the source data; minimum and maximum will be presented in the same precision as the source data. Percentages including response rate will be presented to 1 decimal, e.g., xx.x%. P-values will be presented to 3 decimals. P-values greater than 0.999 will be presented as >0.999; p-values less than 0.001 will be presented as <0.001.

For figures, the corresponding study population will be included in the titles. All collected observations will be included in the related listings.

Baseline value is defined as the last non-missing value prior to the first dose.

For the safety and efficacy analyses, only assessments from scheduled visits will be used for summary tables, however, data collected from both scheduled and unscheduled visits will appear in the subject listings.

When calculating percent change from baseline in efficacy analyses, measures scored as 0 at baseline will be managed as missing values when they did not change from 0, and as 0% change when they were increased (i.e., worsened).

6.1 Sample Size

Approximately 250 eligible subjects (50 in each treatment group) are randomized into the DB treatment period of this study.

The primary efficacy evaluation will be to compare the ACR20 response rate between JTE-051 with placebo at EOT. From one study comparing mavrilimumab with placebo in combination with stable MTX, the ACR20 of mavrilimumab and placebo at Week 12 were 0.69 and 0.40, respectively. Assuming the ACR20 of at least one of the JTE-051 doses to be at least 0.30 above placebo, the following sample size per group will be needed to detect a statistical significant difference at the 5% level with about 80% power (using 2-sided Fisher's exact test):

Placebo ACR20	0.25	0.35	0.45
Sample size per group	47	48	47

About 50 randomized subjects per group would therefore give at least an 80% power for a range of placebo responses. The sample size calculation was carried out using the software SiZ 6.2 (Cytel Inc., 2013).

6.2 Subject Disposition Summary

Patient disposition will be summarized and tabulated by treatment group and in total. The number and percentage of patients in the randomized, safety, ITT, PP and PK populations will be tabulated. Discontinued patients and the reasons for discontinuation during the screening, DB treatment, or follow-up periods will be tabulated for the randomized population by treatment group and in total.

6.3 Demographics and Baseline Characteristics

Demographics and baseline characteristics as defined in Section 5.1 will be presented by treatment group with descriptive statistics for the randomized, safety, ITT and PP populations.

6.4 Efficacy Analysis

Analyses of primary and secondary efficacy parameters will be performed for both the ITT and PP populations, respectively.

For efficacy analyses, the DB treatment period is defined as the time period from the first dose of study drug to the last dose of study drug plus 4 days. The end of treatment (EOT) value is defined as the last value taken during the DB treatment period.

6.4.1 Analyses of Primary Efficacy Endpoint

For the primary efficacy parameter (ACR20 response rate at EOT), the Fisher's exact test will be used to compare each JTE-051 dose group with placebo. Furthermore, the exact Cochran-Armitage trend test will also be performed, using the ordinal score for treatment, e.g., score 1 for placebo, score 2, 3, 4 and 5 will be used to represent JTE-051 50 mg, 100 mg, 150 mg and 200 mg, respectively.

```
Model (Fisher's exact test, will be performed by k=2, 3, 4, 5 separately)
    proc freq data=xxx
        table ARC20*trtn/chisq exact;   ***trtn is the ordinal score for treatment***;
       where trtn in (1, k);
   run;
Model (exact Cochran-Armitage trend test)
    proc freq data=xxx;
        table ARC20*trtn/trend measures ci
              plots=freqplot(twoway=stacked);
        test smdrc;
        exact trend;
     run;
```

In addition, logistic regression model-based analyses will be performed for the pairwise comparisons between JTE-051 and placebo groups (Model 1) and for the evaluation of dose response relationship (Model 2). The logistic regression models will include the baseline PGA score, hs-CRP and region as a covariate.

```
Model 1 (pairwise treatment vs placebo comparison)
    proc logistic data=xxx descending;
       class regionn hscrpn;
       model ACR20 = PGA Dose_50 Dose_100 Dose_150 Dose_200 regionn hscrpn;
    run;

Model 2 (dose-response analysis)
    proc logistic data=xxx descending;
    class regionn hscrpn;
    model ACR20 = PGA Dose regionn hscrpn;
    oddsratio Dose;
run;
```

Note:

- In Model 1, Dose_x = 1 if dose=x; otherwise. Dose_x=0; i.e., Dose_x is a dummy variable.
- In Model 2, Dose = 0, 50, 100, 150, 200 for placebo and each JTE-051 dose level, respectively.
- ACR20 is a 0, 1 variable; i.e., ACR20=1 represents a responder; ACR20=0 represents a non-responder.
- Regionn=0 represents 'EMEA'; Regionn=1 represents 'Latin America' or 'North America';
- Hscrpn=0 represents screening hs-CRP<=2 x upper limit of normal (ULN); Hscrpn=1 represents screening hs-CRP>2 x ULN.
- From the model fitting, one obtains adjusted odds ratios and corresponding p-values for the respective pairwise treatment comparison versus placebo.

6.4.2 Analyses of Secondary Efficacy Endpoints

For dichotomous secondary efficacy parameters, similar analyses as the primary efficacy parameter will be used by time point, i.e., the Fisher's exact test for pairwise comparisons with placebo and the Cochran-Armitage test for trend.

For continuous efficacy parameters, the mixed model with fixed effects for treatment, time, treatment by time interaction, the stratification factors and the baseline, and measurements within each subject as repeated measures will be used.

6.4.3 Supportive Analyses of Efficacy Parameters

For dichotomous efficacy parameter analysis with multiple time points, a supportive analysis will be performed using a generalized estimating equation (GEE) method on the logit scale. The model includes effects for treatment, time, treatment by time interaction, stratification factors and subject as the repeated factor. The odds ratio (OR) of each JTE-051 dose group relative to placebo at the same time point will be estimated.

Model: (generalized estimating equation (GEE) method on the logit scale)

```
proc genmod data=xxx;
    class usubjid trtn(ref=1) region(ref=0) hscrpn(ref=0);
    model SAEP=trtn time trtn*time regionn hscrpn;
    repeated subject=usubjid/type=un corrw;
run;
```

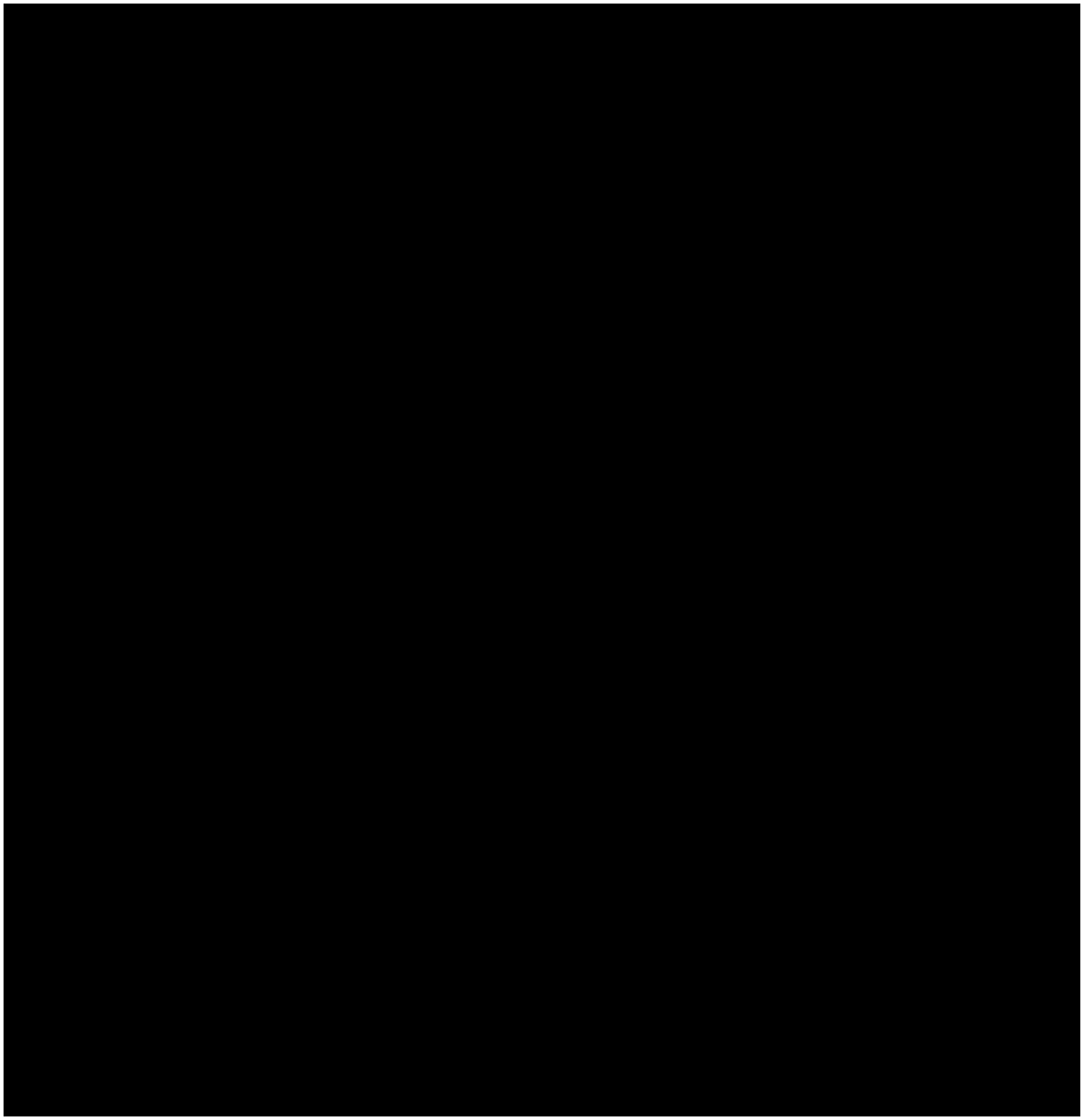
Note: SAPE denotes any dichotomous efficacy parameter, regionn and hscrpn are stratification factors, we can add other stratification when needed.

The supportive efficacy parameters utilizing the non-hs-CRP will be analyzed in a similar manner as their secondary efficacy counterparts using hs-CRP.

All analyses will be performed two-sided at the 5% significance level. No formal multiple comparison adjustment will be made; therefore, results should be interpreted with the multiplicity in mind. The estimated treatment effect (relative to placebo), e.g., odds ratio, from the model, along with a two-sided 95% confidence interval (CI) and p-value, will be tabulated where appropriate. The Newcombe's CI of the rate difference between JTE-051 dose and placebo will be computed when performing the Fisher's exact test. Graphical presentations of the treatment profile will be depicted for the primary and key secondary analyses. Sensitivity analysis and model fit assessment may be conducted, and data transformation may be employed if appropriate.

6.4.4 Descriptive Summary Statistics for Efficacy Parameters

Descriptive statistics of efficacy parameters over time will be presented by treatment. They will include N, arithmetic mean, SD, median, minimum and maximum for continuous parameters and in frequency tabulation form for dichotomous parameters.



6.4.7 Treatment by Center Effect

The randomization in this study is stratified by region (US, LA and EMEA) and by the screening hs-CRP level. Center effect is not considered due to the expectation that many centers will be used and only a small number of subjects will be enrolled in most centers. The region effect will be assessed with respect to the primary efficacy parameter and may be included in the analysis model where appropriate.

6.4.8 Treatment by Baseline Covariate Effect

The treatment-by-baseline effect is not considered; however the baseline will be included in the respective parametric analysis model of the efficacy parameters where appropriate.

6.4.9 Subgroup Analysis

Subgroup analyses may be performed as appropriate upon reviewing the results of the primary analyses.

6.4.10 Handling of Dropouts and Missing Data

For the analyses of dichotomous efficacy parameters by time point, subjects with missing data will not be included in the analysis.

For the analyses of dichotomous efficacy parameters with multiple time points, any missing data will be imputed using the last observation carried forward (LOCF) method except for subjects who withdraw for treatment-related AEs, in which case, any value after the treatment-related withdrawal will be imputed as non-responsive.

For missing data handling in the calculation of ACR20/50/70 responses, because the ACR20 is a composite value, it is sometimes possible to determine it even if there are any missing components. For example, if there is non-responsiveness in at least one of the tender/painful joint count or the swollen joint count, then the ACR20/50/70 is non-responsive. However, if there are missing values in 1 or 2 of the 5 remaining ACR core set components, the method of last observation carried forward (LOCF) will be used to carry forward any of the missing components, and from that mix of actual and carried-forward values, the ACR20/50/70 will be determined. If 3 or more of 5 components are missing, then the corresponding ACR response will not be calculated, and the value of ACR response is considered missing. Missing component at the follow-up visit (Week 16) will not be imputed.

For continuous efficacy parameters, no imputation will be employed.

For the calculation of duration (years) of active RA at Screening, January will be imputed if month of diagnosis is missing.

Other missing data imputation methods may be employed for additional sensitivity analysis if deemed necessary.

6.5 Safety Analysis

All safety analysis will be performed on the Safety population.

Potentially clinically significant (PCS) abnormalities will be defined according to the Akros Safety Reporting Standards (Standard72.doc, 10/18/2004). The PCS values for vital signs, ECG and laboratory data will be flagged in data listings.

Treatment-emergent PCS values will be summarized, presenting the number and percentage of subjects who had at least one PCS value on or after the first dose of study drug (JTE-051 or

placebo). The denominator is always the number of subjects in the safety population, regardless of whether the subject had any post-baseline measurements. The numerator is any PCS value in the treatment-emergent period if the baseline is normal or missing; and is any worsening PCS value in the treatment-emergent period if the baseline is also a PCS value. A worsening PCS is defined as a > (<) 0 change from baseline if a high (low) value indicating undesirable outcome; or the PCS value is in the opposite direction of the baseline PCS value. Note that for the purpose of determining PCS abnormalities, baseline is always the last non-missing value before the first dose of study drug.

6.5.1 Adverse Events (AEs)

The number and percentage of subjects experiencing TEAEs will be tabulated by treatment according to the following categories:

- Overall summary of TEAEs.
- TEAEs by system organ class and preferred term.
- TEAEs leading to study discontinuation by system organ class and preferred term.
- TEAEs by system organ class, preferred term and AE severity.
- TEAEs by system organ class, preferred term and AE causality.
- TEAEs by preferred term (in order of descending frequency in the total JTE-051 group).
- Serious AEs (SAEs) by system organ class and preferred term.
- Serious AEs leading to study discontinuation by system organ class and preferred term.

Listings will be provided for SAEs and AEs leading to study discontinuation.

6.5.2 Medical History

Subjects with significant medical history conditions and/or surgeries, the corresponding system organ class and preferred term coded from the MedDRA (version 21.0) will be presented in the data listing.

6.5.3 Clinical Laboratory Tests

Descriptive statistics of observed results and change from baseline for hematology, coagulation, serum biochemistry, bone-specific alkaline phosphatase, lipid panel and continuous urinalysis will be presented by time point and treatment group. Frequency and percentage of the distribution of discrete urinalysis tests will be presented by time point. Number and percentage of subjects with treatment-emergent PCS abnormal values will be summarized by treatment group.

Number and percentage of subjects with post-baseline abnormalities in liver tests will be summarized by treatment group according to Table 3.

Table 3. Criteria for Categorical Analysis in Liver Tests

Parameter	Specified Value			
ALT	ALT ≥ULN×3	ALT ≥ULN×5	ALT ≥ULN×10	ALT ≥ULN×20
AST	AST ≥ULN×3	AST ≥ULN×5	AST ≥ULN×10	AST ≥ULN×20
ALT and AST	ALT and AST ≥ULN×3	ALT and AST ≥ULN×5	ALT and AST ≥ULN×10	ALT and AST ≥ULN×20
ALT or AST	ALT or AST ≥ULN×3	ALT or AST ≥ULN×5	ALT or AST ≥ULN×10	ALT or AST ≥ULN×20
T-Bil	T-Bil >ULNx1.5		T-Bil >ULN×2	
ALP	ALP > ULN×1.5			
ALT and T-Bil	ALT ≥ULN×3 and T-Bil ≥ULN×1.5		ALT ≥ULN×3 and T-Bil ≥ ULN×2	
AST and T-Bil	AST ≥ULN×3 and T-Bil ≥ULN×1.5		AST ≥ULN×3 and T-Bil ≥ ULN×2	
AT, ALP and T-Bil	ALT ≥ULNx3, AST ≥ULNx3, ALP ≥ULN×2 and T-Bil ≥ULN×2			
ALT and AST	ALT or AST > ULN×3 at any treatment visit and any of the following AEs: Nausea, Vomiting, Anorexia, Abdominal pain or Fatigue, reported within +/- 14 days of the abnormal AT values.			

Other clinical laboratory tests (e.g., drug abuse, alcohol screen, viral serology) will be listed in the data listings.

Qualitative urine parameters are generally reported by a descriptive score, which may differ between laboratories. For pooling the data, a four-point scale is used. The five-point scale will be converted into a four-point scale by combining the lowest two positive results as shown in the table below. Furthermore, a clinical review will be performed on the mapping of the descriptive scores to the four-point scale; and the descriptive scores will be remapped where appropriate. The four-point scale will be used in the summary of the PCS abnormality analysis.

Five-point scale	Four-point scale
Negative	Negative
Trace	Rare/Mild
1+	Rare/Mild
2+	Moderate
3+	Severe

The mean profile may be presented for selected laboratory parameters.

6.5.4 Vital Signs

Descriptive statistics of observed results and change from baseline for vital signs and body weight will be presented by time point and treatment group. Number and percentage of subjects with treatment-emergent PCS abnormal values will be presented by treatment group.

6.5.5 12-Lead ECG

Descriptive statistics of observed results and change from baseline will be presented by time point and treatment group for RR, PR, QRS, QT and QTcF intervals. The number and percentage of subjects with treatment-emergent PCS abnormal values will be presented by treatment group. Clinical interpretation of overall ECG findings will be summarized by time point and treatment group.

6.5.6 Concomitant Medications Analysis

Prior medications and concomitant medications will be listed in the data listings. Concomitant medications will be summarized by ATC Level 3 term, drug preferred term and treatment group.

6.5.7 Study Medication Exposure and Compliance

Summary statistics will be presented for % compliance by visit and by treatment group. Summary statistics for total duration of study medication exposure and overall % compliance will be presented by treatment group.

6.6 Pharmacokinetic Analysis

6.6.1 Method of PK Data Representation

The PK data including individual data and descriptive statistics for plasma drug concentrations will be represented with 3 significant values except for CV% which will be represented to 1 decimal place. The ratio values (see Section 6.6.2) will be displayed to 2 decimal places.

6.6.2 Descriptive Statistics of Trough Plasma Concentrations

Descriptive statistics of trough plasma concentration data for JTE-051 will be presented by treatment and visit (nominal time point; week) with N, arithmetic mean, geometric mean, SD, CV%, median, minimum and maximum. The mean ratio of trough concentration relative to the lowest dose will be calculated. Mean plasma concentration-time profiles will be plotted by treatment on linear scales. Below the limit of quantification (BLQ) concentrations will be treated as zero for descriptive statistics.

6.6.3 Exploratory Exposure-Response Analysis

The relationship between JTE-051 trough concentration and ACR20 response rate will be explored based on the superimposed time course by treatment.

7. SUBJECT DATA LISTING

Subject listings of all the data collected in the study database, including eCRF data and any external data transfers (such as clinical laboratory results), will be presented by treatment group and subject number.

If a subject prematurely discontinues the study, additional subjects may be enrolled at the discretion of the Sponsor, as appropriate.

8. SOFTWARE AND QUALITY CONTROL

All analysis will be performed using SAS® Version 9.1.3 or higher. The SOPs of [redacted] will be followed in the creation and quality control of all tables, listings and analysis.

9. REFERENCES

Please see the references in protocol.

10. APPENDICES

10.1 Mock Summary Tables

The mock summary tables will be created in a separate document.

10.2 Mock Data Listings

The mock summary listings will be created in a separate document.

10.3 Mock Figures

The mock summary figures will be created in a separate document.