Integrated Analysis Plan

Clinical Trial Protocol Identification No.	MS200527-0060
Title:	A Phase IIb, Randomized, Double-blind Study in Subjects with Rheumatoid Arthritis Evaluating the Safety and Efficacy of Evobrutinib Compared with Placebo in Subjects with an Inadequate Response to Methotrexate
Trial Phase	IIb
Investigational Medicinal Product(s)	Evobrutinib
Clinical Trial Protocol Version	12 July 2018/Version 3.0 16 November 2017/Version 2.1 (for US only)

Integrated Analysis Plan Authors

Coordinating Author

PPD, Merck KGaA / EMD Serono Research & Development Institute, Inc.

Function	Author(s) / Data Analyst(s)
PPD	PPD
PPD	PPD
PPD, EMD Serono Research & Development Institute, Inc.	PPD

Integrated Analysis Plan Date and Version	14 October 2019 / Version 2.0

Integrated Analysis Plan Reviewers

Merck KGaA / EMD Serono Research & Development Institute, Inc. Reviewers:

Function	Name
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD

PPD Reviewers:

Function	Name
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD
PPD	PPD

Signature Page

Integrated Analysis Plan: MS200527-0060

A Phase IIb, Randomized, Double-blind Study in Subjects with Rheumatoid Arthritis Evaluating the Safety and Efficacy of Evobrutinib Compared with Placebo in Subjects with an Inadequate Response to Methotrexate

Approval of the IAP by Merck KGaA / EMD Serono Research & Development Institute, Inc. Responsible is documented within CARA. Wet ink signature outside CARA – for PPD use only.

Merck KGaA / EMD Serono Research & Development Institute, Inc. responsible	**Date**	**Signature**
PPD		

PPD **responsible**	**Date**	**Signature**
PPD		

1 Table of Contents

Table of In-Text Tables

2 List of Abbreviations and Definition of Terms

Δ	Difference
ACR	American College of Rheumatology
ADaM	Analysis Data Model
AE	adverse event
AESI	adverse event of special interest
ANCOVA	analysis of covariance
ANOVA	analysis of variance
anti-CCP	anti-cyclic citrullinated peptide
ATC	anatomical therapeutic class
BID	twice daily
BMI	body mass index
BOA	biostatistics outputs assembly
BTK	Bruton's tyrosine kinase
CARLOS	cartilage loss scale
CDAI	clinical disease activity index
CFB	change from baseline
CI	confidence interval
CTP	Clinical Trial Protocol
CV	coefficient of variation
DAS28-hsCRP	disease activity score based on 28 joints and high-sensitivity C-reactive protein
CCI [redacted]	
DMARD	disease-modifying anti-rheumatic drug
DNA	deoxyribonucleic acid
DR	distal radius
DRM	data review meeting
DTS	data transfer specifications
EAIR	exposure adjusted incidence rates
eCRF	Case Report Form, electronic
ECG	electrocardiogram

ESR	erythrocyte sedimentation rate
EULAR	The European League Against Rheumatism
FACIT	functional assessment of chronic illness therapy
HAQ-DI	Health Assessment Questionnaire - Disability Index
HBV	Hepatitis B Virus
HL	Hodges-Lehmann
hrQoL	health-related quality of life
CCI	[redacted]
hsCRP	high-sensitivity C-reactive protein
IAP	Integrated Analysis Plan
IDMC	Independent Data Monitoring Committee
Ig	immunoglobulin
IMP	investigational medicinal product
ITT	intent-to-treat
IWRS	Interactive Web Response System
JSW	Joint Space Width
LFT	Liver Function Test
LOCF	last observation carried forward
MAR	missing at random
MCID	minimal clinically important difference
MCMC	Markov chain Monte Carlo
MCP	metacarpophalangeal
MCP-MOD	multiple comparison procedures - modeling
MCS	mental component summary
MedDRA	medical dictionary for regulatory activities
mITT	modified intent-to-treat
MMRM	Mixed model with repeated measures
MN	Miettinen-Nurminen
MNAR	missing not at random
MRI	magnetic resonance imaging
MTX	methotrexate
N	number of subjects

NCI-CTCAE	National Cancer Institute - Common Terminology Criteria for Adverse Events
NK	natural killer
NSAID	nonsteroidal anti-inflammatory drug
OLE	open-label extension
OMERACT	outcome measures in rheumatology clinical trials
OR	odds ratio
PCR	polymerase chain reaction
PCS	physical component summary
CCI	[redacted]
CCI	[redacted]
PP	per protocol
[redacted]	CCI [redacted]
PT	preferred term
Q1	25th percentile
Q3	75th percentile
QD	once daily
QoL	quality of life
RA	rheumatoid arthritis
RAMRIQ	rheumatoid arthritis MRI quantitative
RAMRIS	rheumatoid arthritis magnetic resonance imaging scoring system
RF	rheumatoid factor
ROW	rest of the world
SAE	serious adverse event
SAF	safety analysis set
SCR	screening analysis set
SD	standard deviation
SDAI	simplified disease activity index
SDTM	Study Data Tabulation Model
SF-36	36-item short form health survey
SJC	swollen joint count

SOC	system organ class
TEAE	treatment-emergent adverse event
TJC	tender joint count
TLF	table/listing/figure
ULN	upper limit of normal
VAS	visual analog scale
WHO-DD	World Health Organization drug dictionary

3 Modification History

Unique Identifier for IAP Version	Date of IAP Version	Authors	Changes from the Previous Version
Version 1.0	18 April 2019	PPD	NA – first version

Version 2.0	14 October 2019	PPD	• Magnetic Resonance Imaging (MRI) substudy participation removed from the list of covariates in statistical models
			• Wording for subgroup analyses clarified (See Section 14.3)
			• Correction of Hochberg procedure terminology in Appendix 18.5
			• Correction of SAS Sample code for:
			  • Stratified Miettinen-Nurminen (MN) 95% Confidence Interval (CI),
			  • Mixed model with repeated measures (MMRM),
			  • Logistic regression with repeated measures.
			• Correction of Table 43
			• Correction of Table 53
			• Addition of MRI Per-Protocol (PP) Analysis Set
			• Update of PP Analysis Set
			• Clarification of the wording in Section 15.11
			• The cumulative proportion of subjects achieving the primary endpoint will be presented as a bar chart instead of a curve (See Section 14.1)
			• The term “point estimate” has been removed throughout the document and replaced by difference of response rate compared to placebo
			• The term “Hyperuricemia” has been removed from Table 53, since the parameter will not be analyzed based on Grade values, only normal ranges.
			• Addition of hsCRP at baseline as a covariate in multiple imputations models as well as joint counts
			• Addition of a paragraph for actual treatment assignment (See Section 9.1)
			• Update time window tables for MRI endpoints and ePRO questionnaires
			• Remove adjusted p-values for supportive analyses (except tipping point analysis)
			• Add a specific imputation rule for number of missing returned tablets in Section 13
			• Clarification of the multiple imputation process
			• Update of total inflammation score for RAMRIQ (See Section 14.2.3).
			• Addition of imputation rule in case of discrepancies between the two independent readers of MRI scan with regards to RAMRIS data (See Section 14.2.3).
			• Removal of compliance during the OLE (See Section 13.2)

4 Purpose of the Integrated Analysis Plan

The purpose of this Integrated Analysis Plan (IAP) is to document technical and detailed specifications for analyses of data collected for protocol MS200527-0060 version 3.0 dated 12 July 2018 and version 2.1 dated 16 November 2017 (US only). The IAP is based upon Section 8 (Statistics) of the trial protocols and is prepared in compliance with International Council for Harmonisation E9.

This version (version 2.0) of the IAP includes details for the final analysis planned at the end of study.

Results of the analyses described in this IAP will be included in the Clinical Study Report (CSR). Additionally, the planned analysis identified in this IAP will be included in regulatory submissions or future manuscripts. Any post-hoc, or unplanned analyses performed to provide results for inclusion in the CSR but not identified in this prospective IAP will be clearly identified in the CSR.

The final clinical database cannot be locked until the final IAP has been approved and signed.

Another document (ms200527-0060-idmc-sap-version1-0-06aug2018) detailed specifications for the Independent Data Monitoring Committee (IDMC) analysis.

5 Objectives and Endpoints

Objectives and Endpoints are summarized in Table 1:

Table 1: Objectives and Endpoints

	Objective	Endpoint	IAP section
Primary Objective	To evaluate the efficacy and dose response of 12 weeks of treatment with evobrutinib compared with placebo in subjects with rheumatoid arthritis (RA) with inadequate response to methotrexate (MTX-IR) on stable MTX therapy by assessment of the signs and symptoms of RA, as measured by American College of Rheumatology (ACR) 20% (ACR20) response assessed using high-sensitivity C-reactive protein (hsCRP) at Week 12.	Primary Endpoint: • ACR20 at Week 12.	Section 14.1
Key Secondary Objectives	To further evaluate the efficacy and dose response of 12 weeks of treatment with evobrutinib in MTX-IR subjects with RA on stable MTX therapy by assessment of the signs and symptoms of RA, as measured by Disease activity score based on 28 joints and high-sensitivity C-reactive protein (DAS28-hsCRP) low disease activity (DAS28 < 3.2) rate at Week 12	Secondary Endpoint: • DAS28-hsCRP < 3.2 at Week 12.	Section 14.2
	To further evaluate the efficacy of 12 weeks of treatment with evobrutinib compared to placebo in MTX-IR subjects with RA on stable MTX therapy by assessment of the signs and symptoms of RA, as measured by DAS28-hsCRP remission (DAS28 < 2.6) rate at Week 12	Secondary Endpoint: • DAS28-hsCRP < 2.6 at Week 12.	Section 14.2
	To further evaluate the efficacy and dose response of 12 weeks of treatment with evobrutinib in MTX-IR subjects with RA on stable MTX therapy by assessment of the signs and symptoms of RA, as measured by ACR50 and ACR70 at Week 12	Secondary Endpoints: • ACR50 at Week 12 • ACR70 at Week 12.	Section 14.2

	To evaluate the safety of evobrutinib in MTX-IR subjects with RA on stable MTX therapy	Safety Endpoints:
		• The nature, severity, and occurrence of Adverse Events (AEs) and serious AEs (SAEs)
		• Absolute values and change from Baseline (CFB) in:
		  • Vital signs
		  • Electrocardiogram (ECG) parameters including RR interval,
			Section 15

		PR interval, QRS duration, QT interval, and QTcF interval • Serum Immunoglobulin (Ig) levels (IgG, IgA, IgM) • Total B cell counts • Clinical laboratory parameters.	
Other secondary Objectives	To further evaluate the efficacy of evobrutinib on the signs and symptoms of RA with inadequate response to MTX	Secondary Endpoints: • Clinical Disease Activity Index (CDAI) • Simplified Disease Activity Index (SDAI) • European League Against Rheumatism (EULAR) Responder Index • EULAR Boolean Remission • ACR hybrid score • Changes and percentage changes from Baseline of individual components of the ACR Core Set Exploratory Endpoints: CCI [redacted]	Section 14.2
	To evaluate the effect of evobrutinib on joint structures and inflammation, at Week 4 and Week 12 in the MRI substudy, as assessed by MRI	Secondary Endpoints: • Change from Baseline in Outcome Measures in Rheumatology Clinical Trials (OMERACT) Rheumatoid Arthritis Magnetic Resonance Imaging Scoring System (RAMRIS) scores at Week 12 for subjects in the MRI substudy • Synovitis score • Bone marrow edema (osteitis) score Exploratory Endpoints: CCI [redacted]	Section 14.2

		CCI [redacted]	
	To evaluate the effect of evobrutinib on physical function in RA subjects	Secondary Endpoint: • Health Assessment Questionnaire – Disability Index (HAQ-DI).	Section 14.2
	To evaluate the effect of evobrutinib on subject-reported health-related quality of life (HRQoL)	Secondary Endpoint: • Medical Outcomes Study 36-Item Short Form Health Survey (SF-36).	Section 14.2
	To evaluate the effect of evobrutinib on subject-reported fatigue (using Functional Assessment of Chronic Illness Therapy [FACIT]-Fatigue)	Secondary Endpoint: • FACIT-Fatigue.	Section 14.2
Exploratory Objectives	CCI [redacted]		

	CCI [redacted]

6 Overview of Planned Analyses

6.1 Independent Data Monitoring Committee review

An IDMC will be set up to continually review available efficacy, safety and tolerability data and will be mandated to make immediate decisions regarding the conduct of the trial; no additional risk minimization measures are proposed. Further details are described in the IDMC charter. Details of the statistical analyses and data presentations for the IDMC meetings are provided in the IDMC Statistical Analysis Plan.

6.2 Final Analysis

The final analysis will be triggered when 100% of subjects randomized either complete the Week 12 visit or discontinue prematurely prior to Week 12, the protocol violations are determined, and the database is locked. All data collected from the Safety follow-up visit during the double-blind 12-week treatment period will be included.

Data collected during the OLE period from earlier versions of the protocol will not be part of the Final analysis for the 12 Weeks treatment period but be summarized separately (See Section 7).

7 Changes to the Planned Analyses in the Clinical Trial Protocol

- The CTP states a consistency analysis is also planned to be performed as part of the final analysis to evaluate the consistency of efficacy at Week 12 between the Japan and non-Japan subjects.

  However, the decision has been taken by the Sponsor not to recruit in Japan. As a result, this consistency analysis will not be performed. Subgroup analysis on Ethnicity Japanese/Non-Japanese will not be performed either.

- In CTP version 2.0, subjects completing the Treatment Period could have entered an optional Open-Label Extension (OLE) period, but the Sponsor made the decision to not initiate the OLE (CTP version 3.0). As a result, some data have been recorded during the OLE for a subset of subjects. Safety and efficacy data from the OLE will then be summarized at the time of final analysis, as detailed in Section 12, 13, 14 and 15.

- The MRI Analysis Set has been redefined to include a pre-dose assessment.

CCI

- The CTP states that the primary endpoint will be evaluated based on a logistic regression model for the odds of ACR20 response, with evobrutinib dose group or placebo as a factor and adjustment for covariates based on randomization strata (region, MRI-substudy participation). This assumed a study design in which subjects could choose whether or not they would participate in the MRI substudy on an individual basis. However, as participation in the MRI substudy was shifted from a per-subject decision to a per-site decision (sites are either MRI or non-MRI), the selection bias is reduced and no clinically significant differences between MRI and non-MRI subjects is expected. Therefore, the adjustment for the covariate of participation in the MRI substudy has been removed. The same change in planned analysis applies for all efficacy endpoints.

8 Protocol Deviations and Analysis Sets

8.1 Definition of Protocol Deviations and Analysis Sets

Important protocol deviations are protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being. Important protocol deviations are defined in the latest version of document named MS200527-0060_List_of_IPDs_Version 2.0_24JUL2019.docx.

All deviations will be identified and confirmed prior to or at the Data Review Meeting (DRM), which will occur before the database lock, including clinically important deviations if leading to the exclusion of a subject from PP analysis set (see Section 8.2).

The outcome of the DRM will document the important protocol deviations as well as the finalization of the analysis populations in a memo. Important protocol deviations will be documented in Clinical Data Interchange Standards Consortium Study Data Tabulation Model (SDTM) whether identified through sites monitoring, medical review and/or programming.

8.2 Definition of Analysis Sets and Subgroups

8.2.1 Sample Size

The primary endpoint of the trial is the ACR20 response at Week 12. A sample size of 80 evaluable subjects per group will provide approximately 90% power at an α of 0.05, 2-sided significance level to detect a difference in Week 12 ACR20 response proportion between evobrutinib- and placebo-treated subjects, assuming a placebo response proportion of 30% to 45% and an expected treatment benefit of 30%, after adjusting for multiple comparisons of 3 dose comparisons with placebo. For the same range of placebo response rate, the sample size of 80 evaluable subjects per group will also provide adequate power (approximately 80%) for the expected difference of 25%. To account for reduced information provided by the Modified Intent-to-Treat (mITT) analysis set, due to subject dropout and protocol noncompliance, occurring at a rate of approximately 11% over a 12-week period, approximately 90 subjects per group will be enrolled.

8.2.2 Analysis Sets

Table 2 provides the definitions of the Analysis Sets and their purposes.

Table 2: Analysis Sets

Analysis Set	Definition	Purpose
Screening Analysis Set (SCR)	The SCR includes all subjects who signed the informed consent.	The purpose of this analysis set is to count how many subjects signed the informed consent
Intent-To-Treat (ITT) Analysis Set	The ITT analysis set will include all subjects randomly allocated to a treatment, based on the intention to treat "as randomized" principle (i.e., the planned treatment regimen rather than the actual treatment given in case of any difference).	The ITT Analysis Set will be used for TLFs based on randomized subjects, such as randomization listing and TLFs on protocol deviations.
mITT Analysis Set	The mITT analysis set will include all randomized subjects who received at least one dose of IMP (evobrutinib or placebo). Subjects in the mITT analysis set will be included in the treatment group as randomized.	The mITT Analysis Set will be used for demographic/baseline characteristics and efficacy endpoints.
Quality of Life (QoL) Analysis Set	The QoL Analysis Set will include all randomized subjects who have received at least one dose of IMP (evobrutinib or placebo) and have at least one Baseline and one post-Baseline QoL assessment (among the following: HAQ-DI, SF-36, FACIT, CCI [redacted]). Subjects will be analyzed according to randomized treatment.	The QoL Analysis Set will be used for QoL endpoints.
Safety Analysis Set (SAF)	The SAF analysis set will include all subjects who receive at least 1 dose of IMP (evobrutinib or placebo). Subjects will be analyzed according to the actual treatment they receive. The SAF will be used for safety analyses.	The SAF will be used for medications, exposure/compliance and safety endpoints.
PP Analysis Set	The PP analysis set will include all mITT subjects who do not have any clinically important protocol deviations before or at the time of Week 12 visit. Subjects in the PP analysis set will be included in the treatment group as randomized. Clinically important protocol deviations related to MRI assessments (protocol deviation PDEV63 from the List of Important Protocol Deviations) will not be considered.	The PP analysis set will be used for the primary and key secondary efficacy endpoints.
MRI Analysis Set	The MRI analysis set will include the mITT subjects who have at least 1 pre-dose and 1 post-dose MRI assessment. Subjects in the MRI analysis set will be included in the treatment group as randomized.	The MRI Analysis Set will be used for all the analyses of the MRI endpoints.

MRI PP Analysis Set	The MRI PP analysis set will include all MRI subjects who do not have any clinically important protocol deviations before or at the time of Week 12 visit. Subjects in the MRI PP analysis set will be included in the treatment group as randomized. All clinically important protocol deviations (including protocol deviation PDEV63 from the List of Important Protocol Deviations) will be considered.	The MRI PP analysis set will be used for the MRI endpoints.

CCI [redacted]

8.2.3 Subgroup Analysis

Table 3 below defines the subgroups to be used for subgroup analysis as described in Section 14 and Section 15.

Table 3: Subgroups

Subgroup	Content	Definition
A	MRI sub-study participation	Yes; No (from IWRS – Is the subject participating in the MRI substudy?)
B	Gender	Male; Female (from DEMO form of electronic case report form [eCRF])
C	Region	US; Western Europe; ROW (from IWRS)

9 General Specifications for Data Analyses

All statistical analyses will be performed by PPD, except the modeling and simulation analysis, which will be performed by Merck KGaA / EMD Serono Research & Development Institute, Inc.

9.1 Actual Treatment Assignment

A subject who received 2 different types of treatment regimen over the course of treatment should be tabulated according to the treatment regimen received most frequently. If there is a "tie", the highest dose will be chosen.

9.2 Presentation of Tables/Listings/Figures (TLFs)

Tables and figures will be sorted by treatment group (in the order stated in Table 4) and chronological time point (where applicable).

Table 4: Treatment Groups and Regimens

Treatment Group	Regimen	Presentation for the 12-week treatment period	Presentation for the OLE (*)
1	Placebo	Placebo	Placebo/Evobrutinib 50 mg BID
2	Evobrutinib 25 mg QD (once daily)	Evobrutinib 25 mg QD	Evobrutinib 25 mg QD/Evobrutinib 50 mg BID
3	Evobrutinib 75 mg QD	Evobrutinib 75 mg QD	Evobrutinib 75 mg QD/Evobrutinib 50 mg BID
4	Evobrutinib 50 mg BID (twice daily)	Evobrutinib 50 mg BID	Evobrutinib 50 mg BID/Evobrutinib 50 mg BID

(*): As detailed in Section 7, some subjects have entered in the OLE, even if the Sponsor decided not to initiate the OLE in CTP V3.0. Subjects who agreed to participate in the OLE were administered evobrutinib 50 mg BID.

All data recorded during the trial will be presented in individual data listings (in the order stated in Table 4). All listings are sorted by treatment group, subject, and time point (where applicable), if not otherwise stated.

9.3 Data handling for the planned analyses

As the database lock will be based on data up to and including the last patient's last visit, all data from screening up to and including the Safety Follow-up visit for all subjects will be locked and available in the database for the final analysis.

9.4 Presentation of continuous and qualitative variables

Continuous variables will be summarized using descriptive statistics, ie,

- number of subjects (N)
- number and percentage of non-missing values
- number and percentage of missing values
- mean, standard deviation (SD)
- median, 25th Percentile - 75th Percentile (Q1-Q3)
- minimum and maximum

Unless otherwise specified, the number of digits for non-derived and derived data, presented in outputs or available in ADaM datasets, is specified in the Biostatistics Outputs Assembly (BOA) document. For efficacy endpoints from Section 14, median and SD will be presented with 1 more digit compared to the original data, whereas mean, min, max, Q1-Q3 will be presented with the same number of digits as the original data.

For both continuous and qualitative variables, percentages such as 0% or 100% should be reported with the same format used for the column, together with the count of observations. For example, if the count of observations is zero, then display '0 (0.0)'; if the count of observations represents 100%, then display 'xx (100.0)'.

Qualitative variables will be summarized by counts and percentages. Unless otherwise stated, the calculation of proportions will be based on the number of subjects in the analysis set of interest. Therefore, counts of missing observations will be included in the denominator and presented as a separate category.

In case the analysis refers only to certain visits, percentages will be based on the number of subjects in the analysis set of interest, unless otherwise specified. Total of missing and non-missing observations at each time-point will reflect the population still in the trial at that time. For example, if a subject is still in the trial at the time-point but with missing data, it should be counted in the number of missing observations.

CCI [redacted]

CCI [redacted]

[redacted]

[redacted]

Mean, Min, Median, Max:	3 significant digits
SD:	4 significant digits
CV%:	1 decimal place

9.5 Definition of baseline

For the purpose of statistical analysis, baseline is defined as the last non-missing measurement prior to the first dose of study drug as described in Table 5:

Table 5: Definition of Baseline

Randomized	Time Period	Baseline
• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	12-week treatment period	Last non-missing measurement prior to first dose of study drug of the 12-week treatment period
	OLE Period	Last non-missing measurement prior to the first dose of evobrutinib of the OLE period

9.6 Other Specifications

Difference between the randomization strata (MRI site) and MRI substudy participation
In case of difference between the two variables "Was the subject enrolled in an MRI site?" and "Did the subject participate in the MRI substudy?", the MRI substudy participation will be used for the statistical analysis.

Definition of Change from Baseline (CFB):
CFB and percent CFB will be computed as follows:

- CFB = visit value – baseline value
- Percent CFB = 100 * (visit value – baseline value) / baseline value

Definition of duration:
Unless otherwise specified, duration will be calculated as the difference between start and stop date + 1 (e.g. duration of AE (days) = AE stop date – AE start date + 1). Unless otherwise specified, missing dates will not be imputed.

Conversion factors:
The following conversion factors will be used to convert days into months or years: 1 month = 30.4375 days, 1 year = 365.25 days. For time windows calculation, 1 month is expressed as 30 days.

Definition of Body Mass Index (BMI) (kg/m²):
BMI will be computed as weight / (height²), where weight is expressed in kg and height in m.

Handling of missing data:
For efficacy analysis, methods of missing data handling are detailed in Section 14.

Unless otherwise specified, missing data will not be replaced.

In all subject data listings, imputed values will be presented. In all listings imputed information will be flagged. Non-imputed partial dates will be presented in a format such as "_____YYYY". Where imputation is defined, an imputed date with flag (ie, D for day, M for month) will be reported.

Missing statistics, eg when they cannot be calculated, should be presented as 'NE' (denoting "Not Evaluable"). For example, if n=1, the measure of variability (SD) cannot be computed and should be presented as 'NE'.

In case of zero records, an empty output with 0 occurrences, or a sentence mentioning that there are no data, will be presented. For tables of AEs and Deaths (outputs required for EudraCT and/or clinicaltrial.gov), if there are no observations, the output must contain the first line 'Subjects with…' or 'Subjects who died' displayed with 0 occurrence.

If the System Organ Class (SOC) or Anatomical Therapeutic Class (ATC) term is missing/not coded yet, then 'Uncoded SOC' (or 'Uncoded ATC') will be indicated at the TLF level. When a Preferred Term (PT) is missing, it will be set to 'Uncoded PT: verbatim text'.

Treatment day:
Treatment day is defined as relative day to the start date of treatment. Treatment day will be calculated accordingly, ie:

- For the 12 weeks treatment period, treatment day 1 is defined as the date of first administration of any IMP.
- For the OLE period, treatment day 1 is defined as the date of first administration of evobrutinib 50 mg BID during OLE.

The day before is defined as Treatment day –1 (no Treatment day 0 is defined).

Treatment day will only be presented in subject safety data listings. In these listings, the difference (in days) between an event (eg, AE or laboratory assessment) and the first administration will be calculated as follows:

- If the event is posterior to the first IMP administration, the difference is equal to Date of event – Date of first IMP administration + 1
- Otherwise, the difference is equal to Date of first IMP administration – Date of event

Time windows
As a measurement may have been performed out of window, data will be re-allocated according to time windows, so that each measurement is allocated to the closest time point.

Appendix 18.6 provide further details of time window allocations for safety and efficacy endpoints.

For efficacy and safety analyses, each measurement will be assigned an analysis visit number first. The analysis visit will then be used for all missing data imputations, analysis variable derivations, statistical calculations and presentations.

For patient data listings by time point, the nominal visit (as collected in the database) as well as the analysis visit will be displayed.

For time windows calculation, 1 month is expressed as 30 days.

Unblinding:
Details regarding the unblinding process are available in the unblinding plan (MS200527-0060_UnblindingPlan_v2.0_21Feb2019.pdf)

Software:

All statistical analyses will be performed using SAS® (Statistical Analysis System, SAS-Institute, Cary North Carolina, USA, Windows Version 9.4 or higher), or R software (R Foundation for Statistical Computing, Vienna, Austria, Version 3.4.1 or higher) if certain statistical procedures are not available in SAS.

10 Trial Subjects

The subsections in this section include specifications for reporting subject disposition and treatment/trial discontinuations. Additionally, procedures for reporting protocol deviations are provided.

10.1 Disposition of Subjects and Discontinuations

A table on screened subjects describing the number and percent of subjects in each disposition category will be produced by treatment group. These categories are summarized below:

- Total number of screened subjects, ie subjects that gave informed consent (overall summary only)
- Number of subjects who discontinued prior to randomization and reason (overall summary only)
- Number of randomized subjects
- Number of randomized subjects who did not start treatment
- Number of randomized subjects who permanently discontinued treatment during 12-week treatment period and reason (as reported in the eCRF)
  - Number of randomized subjects who completed safety follow-up period
  - Number of randomized subjects who discontinued from trial during safety follow-up and reason
- Number of randomized subjects who completed 12-week treatment period
  - Number of randomized subjects who completed safety follow-up period
  - Number of randomized subjects who discontinued from trial during safety follow-up and reason
- Number of randomized subjects who signed OLE informed consent under previous versions of the protocol

A table based on screened subjects, describing the number and percent of subjects set by treatment group, will be produced, with the following categories:

- Number of screened subjects
- Number of subjects included in the ITT Analysis Set
- Number of subjects included in the SAF
- Number of subjects included in the mITT Analysis Set
- Number of subjects in MRI Analysis Set
- Number of subjects included in the PP Analysis Set
- Number of subjects included in the MRI PP Analysis Set

- Number of subjects included in the QoL Analysis Set

CCI [redacted]

- Number of subjects who received at least 1 dose in the OLE under previous versions of the protocol

A table based on screened subjects, describing the number and percent of subjects in each region (Western Europe, US, ROW), country within region, site within country, by analysis set, will be produced.

Another table based on randomized subjects will provide the number and percent of subjects by randomization strata and country. The different randomization strata are listed below:

- MRI substudy participation
  - Yes
  - No
- Region
  - US
  - Western Europe
  - Rest of the world

Corresponding individual listings will be prepared:

- Discontinued subjects (from treatment or study) will be listed with their reason for withdrawal (from treatment or study).
- A listing of the subjects screened but not randomized will be produced with the reason for not being randomized.
- A listing of randomized subjects with subject number, randomization date, and randomized treatment group, ordered by randomization number within randomization stratum, will be produced for the purpose of assessing whether randomization was conducted as planned.
- The list of re-screened subjects and corresponding subject identifiers will be provided. Only the second subject identifier will be used in statistical descriptions and analyses, while the first identifier will not be considered from the disposition counts.

10.2 Protocol Deviations

The Table 6 summarizes how protocol deviations will be handled for 12-week treatment period. Protocol deviations in the OLE period will be listed.

Table 6: Protocol Deviations

Period covered for protocol deviations reporting	Treatment Group	Observation Period	Analysis Set
12-week treatment period	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	First dose of 12-week treatment period up to last observed dose during the 12-week treatment, (safety follow-up posterior to the 12-week treatment period included)	ITT

10.2.1 Important Protocol Deviations

The following summary tables and listings of important protocol deviations will be provided:

- Tables providing frequency for each type of important protocol deviation/clinically important protocol deviation
- Listing of important protocol deviations

10.2.2 Clinically Important Protocol Deviations

Clinically important protocol deviations, as defined in Section 8.1, will be summarized and listed.

For subjects excluded from the PP, the reasons for exclusion will be summarized and listed:

- Frequency table per reason of exclusion from the PP population
- Listing of reasons of exclusion from the PP population

For subjects excluded from the MRI PP, the reasons for exclusion will be summarized and listed:

- Frequency table per reason of exclusion from the MRI PP population
- Listing of reasons of exclusion from the MRI PP population

11 Demographics and Other Baseline Characteristics

Demographic and other baseline characteristics will be summarized on mITT and presented by treatment group and overall. All data will also be listed on mITT.

11.1 Demographics

The demographic characteristics to be summarized are presented in Table 7.

Table 7: Demographic Characteristics

Demographic Characteristics	Modalities
Gender (from DEMO form of eCRF)	• Male, • Female.
Race (from DEMO form of eCRF)	• White, • Black or African American, • Asian, • American Indian or Alaska Native, • Native Hawaiian or Other Pacific Islander, • Not collected at this site, • Other.
Ethnicity (from DEMO form of eCRF)	• Hispanic or Latino, • Not Hispanic or Latino.
Age (defined below)	• < 65 years, • ≥ 65 years.
Geographic Region (from IWRS)	• US, • Western Europe, • ROW.
European Economic Area (EEA) (from DEMO form of eCRF)	• Yes, • No.
Has the subject been enrolled into an MRI site?	• Yes, • No.
Is the subject participating in the MRI substudy?	• Yes, • No.

Specifications for computation:

- Age (years):
  - (date of given informed consent - date of birth + 1) / 365.25
  - In case of missing day for the date of birth, but month and year available for both dates:
    - For the derivation of age, the day of birth will be set to 1 and the formula above will be used
  - In case of missing month for the date of birth, but year available for both dates:
    - For the derivation of age, the day and month of birth will be set to 1 and the formula above will be used
- Site codes will be used for the determination of the subject's geographic region and EEA membership.

11.2 Medical History

The medical history will be summarized using Medical Dictionary for Regulatory Activities (MedDRA), current version, PT as event category and MedDRA SOC body term as Body System category. The MedDRA version used will be indicated in footnote. Medical history will be tabulated by SOC and PT. SOC and PT will be alphabetically sorted. Medical history will be also listed.

11.3 Other Baseline Characteristics

11.3.1 Disease History

Descriptive statistics will be presented for disease history endpoints as indicated in Table 8. The disease history will be repeated for the randomization stratification with marginal total for each category of MRI substudy participation and Region by treatment group.

Table 8: Disease History

Disease History Characteristics	Modalities
Time (months) since confirmed diagnosis of RA according to 2010 ACR/The EULAR RA classification criteria.	• Summary statistics will be presented. • Time will be computed as (Date of Informed Consent Signature – Date of confirmed diagnosis of RA) / 30.4375. If the date of confirmed diagnosis of RA is entirely missing, then time since confirmed diagnosis of RA will not be computed. In case of partial missing date of confirmed diagnosis of RA, this date will be imputed as follows: o If only the day is missing, then it will be replaced by the first day of the month o If both day and month are missing, then it will be replaced by the first of January

Components from the ACR Core Set: • Tender Joint Count ([TJC] among the 68-joint count), • Swollen Joint Count ([SJC] among the 66-joint count), • Subject's Global Assessment of Disease Activity (100-mm VAS), • Subject's Assessment of Pain (100-mm VAS), • HAQ-DI total score, • Physician's Global Assessment of Disease Activity (100-mm VAS), • hsCRP (mg/L)	• Summary statistics
Erythrocyte sedimentation rate (ESR) (mm/Hr)	By category (based on local normal ranges collected in electronic case report form [eCRF]): • Low • Normal • High
hsCRP (mg/L)	By category: • Normal • High
Acute-phase reactant	At least one test result is needed for the classification: • Normal hsCRP and normal ESR, • Abnormal hsCRP or abnormal ESR Abnormal refers to either low or high normal range classification.
Rheumatoid factor (RF) (expressed in IU/mL)	By category: • Negative (≤ Upper Limit of Normal [ULN]) • Positive (> ULN)
Serum Ig levels (IgG, IgA, IgM) (mg/dL)	By category: • Negative (≤ ULN) • Positive (> ULN)
Anti-cyclic citrullinated peptide (Anti-CCP) (Units)	By category: • Negative (≤ ULN) • Positive (> ULN)
Serology	By category (At least one test result is needed for the classification): • Negative RF and negative anti-CCP (≤ ULN), • Low-positive RF or low-positive anti-CCP (> ULN, ≤ 3*ULN), • High-positive RF or high positive anti-CCP (> 3*ULN) • Missing RF and missing anti-CCP

14-3-3η (ng/mL)	By category: • Negative (≤ ULN) • Positive (> ULN)
DAS28-hsCRP (See Appendix 18.1 for definition)	By category: • < 2.6 [remission], • ≥ 2.6 to < 3.2 [low], • ≥ 3.2 to ≤ 5.1 [moderate], • > 5.1 [high].
CDAI (See Appendix 18.1 for definition)	By category: • ≤ 2.8 [remission], • > 2.8 to ≤ 10.0 [low], • > 10.0 to ≤ 22.0 [moderate], • > 22.0 [high].
SDAI (See Appendix 18.1 for definition)	By category: • ≤ 3.3 [remission], • > 3.3 to ≤ 11.0 [low], • > 11.0 to ≤ 26.0 [moderate], • > 26.0 [high].
EULAR/ACR Boolean Remission (See Appendix 18.1 for definition)	• Yes, • No.

11.3.2 Other

Chest X-ray evaluations will be listed and tabulated by treatment group using the number and percentage of subjects for each interpretation category (Normal, Abnormal Not Clinically Significant, Clinically Significant, and Abnormal Overall).

Other characteristics like viral serology, Quantiferon TB test, thyroid-stimulating hormone and prior surgeries will be listed only.

12 Previous or Concomitant Medications/Procedures

Previous or concomitant medications during the 12 weeks treatment period will be summarized using the SAF.

As detailed in Section 7, some subjects have entered in the OLE, even if the Sponsor decided not to initiate the OLE in protocol V3.0. Therefore, concomitant medications during OLE will also be summarized for subjects who received at least 1 dose in the OLE under previous versions of the protocol.

The definition of previous or concomitant medications is presented in Table 9:

Table 9: Definition of Previous/Concomitant Medications

Analysis Set	Period covered	Treatment groups	Definition
SAF	12-week treatment period	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	**Previous medications** are medications, other than trial medications, which either: 1. started and stopped before first administration of any IMP (placebo or evobrutinib) or 2. started prior to the first administration of IMP (placebo or evobrutinib) and are taken by subjects on or after the first administration of IMP.
			Concomitant medications are medications, other than IMPs, which either: 1. started on or after the first administration of any IMP (placebo or evobrutinib) or 2. started prior to the first administration of IMP (placebo or evobrutinib) and are taken by subjects on or after the first administration of IMP.
Subjects with at least 1 dose during OLE	OLE	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	**Concomitant medications** are medications, other than IMPs, which either: 1. started on or after the first administration of evobrutinib during OLE or 2. started prior the first administration of evobrutinib during OLE and are taken by subjects on or after the first administration of evobrutinib during OLE.

Partial dates will be handled as follows:

- For previous medications, in case the date values will not allow a medication to be unequivocally allocated to previous medication, the medication will be considered as previous medication.
- For concomitant medications, in case the date values will not allow a medication to be unequivocally allocated to concomitant medication, the medication will be considered as concomitant medication.

The ATC-2nd level and PT will be tabulated as given from the World Health Organization Drug Dictionary (WHO-DD) current version. In case multiple ATCs are assigned to a drug, all ATC2nd level will be used for reporting.

The number and proportion of subjects with previous or concomitant medications (previous medications will not be summarized for final analysis) will be separately summarized by treatment group and will be presented by descending frequency of ATC 2nd level term and then by descending frequency of PT in total column. If multiple ATCs/PTs have the same frequency, they will be sorted alphabetically. The WHO-DD version used will be indicated in footnote.

Previous or concomitant medications will be also listed.

All concurrent procedures, which were undertaken any time on trial, will be summarized. Number of subjects with concurrent procedures (prior, on or after the first day of IMP) overall and by SOC and PT will be summarized by treatment group, using current version of MedDRA dictionary.

Concurrent procedures will be also listed.

13 Treatment Compliance and Exposure

13.1 Exposure

As detailed in Section 7, some subjects have entered in the OLE, even if the Sponsor decided not to initiate the OLE in CTP V3.0. As a result, exposure during the OLE period will be separately summarized for subjects who received at least 1 dose in the OLE period under previous versions of the protocol.

The Table 10 summarizes how exposure time will be computed:

Table 10: Exposure Time

Period covered	Treatment groups	Exposure time	Analysis sets
12-week treatment period	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	From first dose of any IMP during 12-week treatment period to last observed dose of 12-week treatment period	SAF
OLE	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	From first dose of evobrutinib during OLE to last dose of evobrutinib during OLE	Subjects who receive at least 1 dose in the OLE period

12-week treatment period

Exposure for the 12-week treatment period will be presented on SAF. For the 12-week treatment period, subjects will receive one of the following IMPs:

- Placebo
- Evobrutinib 25 mg QD
- Evobrutinib 75 mg QD
- Evobrutinib 50 mg BID

Dosage and administration are summarized in Table 11:

Table 11 : Dosage and Administration

Product Description	Dosage Form	Evobrutinib Dose	Number of Tablets (Morning Dose)		Number of Tablets (Evening Dose)	
			25 mg		25 mg	
			Evobrutinib	Placebo	Evobrutinib	Placebo
Evobrutinib	Oral	25 mg QD	1	2	0	2
Evobrutinib	Oral	75 mg QD	3	0	0	2
Evobrutinib	Oral	50 mg BID	2	1	2	0
Placebo	Oral	0 mg	0	3	0	2

Note: To maintain the blind, all subjects will receive BID treatment. Subjects will self-administer 3 tablets in the morning and 2 tablets in the evening (every 12 hours ± 2 hours). Subjects who participate in the optional OLE Period will take 50 mg of evobrutinib twice daily, administered as two 25 mg tablets twice daily.

On trial visit days, IMP will be administered during the trial visit after the scheduled trial visit procedures are completed.

Exposure time in weeks will be calculated according to the following formula:

$$exposure\ (weeks) = \frac{(date\ of\ last\ dose\ -\ date\ of\ first\ dose\ +\ 1)}{7}$$

First dose refers to the first administration of any IMP in 12-week treatment period. Last dose refers to the last administration of any IMP in 12-week treatment period. Both dates of first and last dose will be retrieved from SDTM EC domain.

Exposure time will be presented by summary statistics and according to the following categories:

- ≤ 1 week
- > 1 to 2 weeks
- > 2 to 4 weeks
- > 4 to 8 weeks
- > 8 weeks

The calculated total dose (mg) per subject for the 12-week treatment period will also be summarized for the active treatment groups, based on the actual treatment the subject receives. The calculated total dose is defined according to the following formula:

The number of ingested tablets will be retrieved from the BATCH (Study Treatment Box Number) panel of the CRF, the number of tablets dispensed being equal to 40 for a given kit during the treatment period and 32 during the OLE period:

$$Number\ of\ tablets\ ingested = Number\ of\ tablets\ dispensed - number\ of\ tablets\ returned$$

Cumulative actual dose cannot be calculated for placebo treatment group.

For subjects with missing number of returned tablets (e.g. lost to follow-up), the cumulative actual dose will be calculated by imputing the missing returned amount with the number of tablets the subjects should have returned if 100% compliance was observed while on treatment for the associated dispensed kit.

Subject data listings:

Study drug administrations will also be listed by treatment group, and subject, with start/end dates of administration and reason for dose change (if applicable).

13.2 Compliance

Compliance will be analyzed for the 12-week treatment period on the SAF Analysis set. As the Sponsor made the decision to not initiate the OLE, the compliance will not be analyzed during the OLE.

12-week treatment period

For the 12-week treatment period, compliance with treatment is defined as the number of tablets taken during a period divided by the number of tablets that should have been taken during that period, multiplied by 100 to yield a percentage, ie:

$$Compliance\ with\ \ treatment\ = 100 \times (\frac{N_1}{5 \times N_2})$$

where

- N_1 = number of tablets given minus number of tablets returned over N_2 days,
- N_2 = number of days between treatment start and treatment termination visit.

, where N_1 will be computed using the BATCH form of the eCRF and N_2 corresponds to the exposure time expressed in days (see Section 13.1 for exposure time and number of tablets ingested formulas).

CCI [redacted]

For subjects with missing number of returned tablets (e.g. lost to follow-up), the compliance will be calculated by imputing the missing returned amount with the number of tablets the subjects should have returned if 100% compliance was observed while on treatment for the associated dispensed kit.

Compliance with treatment will be tabulated by treatment group from first intake to last intake of 12-week treatment period.

Compliance with treatment will also be presented into categories as follows:

- < 60%
- ≥ 60% to < 80%
- ≥ 80% to ≤ 100%
- > 100% to ≤ 110%
- > 110%

Subject data listings:

The following listings will be provided:

- listing of kit numbers with date of dispense, and number of tablets returned (from BATCH).
- listing of start/end dates with number of ingested tablets (from EXPOSUREDT).
- listing with exposure time, cumulative dose, and compliance.

14 Endpoint Evaluation

Table 12 details data handling for efficacy analysis.

Table 12 : Data Handling for Efficacy Analysis

Period covered	Treatment groups	Data to be analyzed
12-week treatment period	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	All data collected up to Week 12 visit
OLE period	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/ Evobrutinib 50 mg BID	Data from OLE treatment period

14.1 Primary Endpoint: ACR20 Response Rate at Week 12

The primary endpoint, ACR20 response rate at Week 12, is defined in Appendix 18.1.

The analysis of the primary endpoint and missing data handling are summarized in Table 13.

Table 13 : Analysis of the Primary Endpoint – ACR20 Response Rate at Week 12

Analysis Set	Analysis method	Description	Missing data handling
• mITT set • PP set	Primary analysis: logistic regression	Presentation of response rate for each treatment group, difference [Δ] in response rate compared to placebo, 2-sided 95% CI of Δ (non-stratified MN), OR, 2-sided 95% CI of OR, 2-sided nominal p-value, and adjusted p-value (using Hochberg procedure). Adjusted p-values reported only for primary analysis based on mITT set.	LOCF-NR*
• mITT set • PP set	Supportive analysis no. 1: Stratified difference of response rate	Presentation of response rate for each treatment group, Δ in response rate compared to placebo, 2-sided 95% CI of Δ and 2-sided nominal p-value.	LOCF-NR*
• mITT set	Supportive analysis no. 2: Tipping Point	Presentation of p-values and response rate for each treatment group at the tipping point.	Tipping point analysis

• mITT set	Supportive analysis no. 3: Stratified difference of response rate using multiple imputation with Missing At Random (MAR) pattern	Presentation of response rate for each treatment group, Δ in response rate compared to placebo, 2-sided 95% CI of Δ and 2-sided nominal p-value.	The ACR20 will be derived using imputed ACR Components. These components will be imputed using Multiple Imputation Procedure with MAR pattern for discontinuations due to reasons other than safety or efficacy
• mITT set	Supportive analysis no. 4: Stratified difference of response rate using multiple imputation with Missing Not At Random (MNAR) pattern	Presentation of response rate for each treatment group, Δ in response rate compared to placebo, 2-sided 95% CI of Δ and 2-sided nominal p-value.	The ACR20 will be derived using imputed ACR Components. These components will be imputed using Multiple Imputation Procedure with MNAR pattern, using control-based approach, regardless of reasons for treatment discontinuation.
• mITT set	Supportive analysis no. 5: Non-parametric method: Cochran-Armitage test	Presentation of response rate for each treatment group, Δ in response rate compared to placebo, 2-sided 95% CI of Δ (stratified MN), and 2-sided p-value from Cochran-Armitage test	LOCF-NR*
• mITT set	Supportive analysis no. 6: Multiple Comparison Procedures - Modelling (MCP-MOD) approach to dose response	Presentation of T-statistic and p-value from multiple contrast test for each model tested, the optimal model selected, and the estimated dose for the target effect using the selected model.	LOCF-NR*
• mITT set	Descriptive statistics	Number and percentage of subjects with response status (Yes/No/Missing) over time	As observed

* LOCF-NR: post-baseline last observation carried forward. Patients who discontinued from treatment are considered as non-responders.

Primary analysis: logistic regression

For the primary analysis, the estimate of the odds ratio (OR), together with the associated 2-sided 95% CI, nominal p-value and adjusted p-value, comparing each evobrutinib dose group to placebo, will be provided based on a logistic regression model for the odds of ACR20 response, with treatment group as a factor and adjustment based on region. An example of SAS code is available in Appendix 18.5.

Missing data will be handled using LOCF-NR as follows: each component will be imputed using post-baseline LOCF (ie, baseline value cannot be used for missing data imputation). Then the ACR20 will be re-calculated using imputed components. Subjects who prematurely discontinued from treatment will be considered as non-responders.

A family-wise Type I error rate (FWER) at the primary analysis will be controlled at the 2-sided 0.05 significance level for the 3 pairwise comparisons of evobrutinib dose versus placebo, using the Hochberg procedure (see Hochberg 1990).

The statistical test comparing each evobrutinib group to placebo can be written as follows:

$$\begin{cases} H0: OR = 1.0 \\ H1: OR \neq 1.0 \end{cases}$$

Consider the ordered two-sided p-values (from lowest to highest) that arise from comparing the 3 comparisons on the basis of the primary endpoint: $p_{(1)} \leq p_{(2)} \leq p_{(3)}$. The Hochberg's step-up method proceeds as follows:

1. Step 1: If $p_{(3)} < 0.05$ reject $H_{(1)}$, $H_{(2)}$ and $H_{(3)}$ (i.e., conclude all evobrutinib dose groups are more effective than placebo under the condition the OR are > 1) and then stop; otherwise go to Step 2.
2. Step 2: If $p_{(2)} < 0.025$ reject $H_{(1)}$ and $H_{(2)}$ (i.e., conclude doses associated with $H_{(1)}$ and $H_{(2)}$ are more effective than placebo under the condition the OR are > 1) and then stop; otherwise go to Step 3.
3. Step 3: If $p_{(1)} < 0.0167$ reject $H_{(1)}$ (i.e., conclude that only one of the dose groups associated with $H_{(1)}$ is more effective than placebo under the condition the OR is > 1) and stop

Hochberg procedure will be computed with SAS® (see Appendix 18.5). Adjusted p-values will not be presented for any analyses based on PP analysis set.

In addition, the number and percentage of ACR20 responders (Yes/No) at Week 12 will be summarized by treatment group.

- For the summary of ACR20 responders (Yes) at Week 12, the number and percentage of subjects from each treatment group who have met the responder criterion (at least 20% improvement from baseline) for each ACR component will be provided.
- For the summary of ACR20 non-responders (No) at Week 12, the number and percentage of subjects from each treatment group who are imputed as non-responders due to different reasons (i.e., discontinued due to adverse event, lack of efficacy, etc) will be provided.

A bar chart will be generated by treatment group to present the cumulative proportion of subjects achieving ACRx where x is the cut-off of percentage reduction from baseline for ACR as a composite endpoint. The bar chart will start with x=20%, incremented by 10%, and up to the maximum observed percentage of improvement. Vertical lines will be drawn at 20%, 50% and 70%.

The number of responders/non-responders (Yes, No) at Week 2, Week 4 and Week 8 will also be presented in the logistic regression table using the LOCF-NR method.

Supportive analysis no. 1: Stratified difference of response rates 95% CI

In order to further characterize the dose-response, the proportion of ACR20 response at Week 12 will also be presented for each treatment group, along with the stratified difference in ACR20 response proportion between each evobrutinib dose and placebo, and the 2-sided 95% CI for the difference [Evobrutinib – Placebo].

The following will be presented: stratified Δ between response rates for pairwise comparisons versus placebo, the 2-sided 95% CI, and nominal p-value. As there are 3 regions (Western Europe, US, ROW), the number of strata is equal to K=3.

For each pairwise comparison, the difference of response rate will be computed using the formula no. (7), available in Lu 2008. By denoting A the 1st treatment group (evobrutinib 25 mg QD, evobrutinib 75 mg QD or evobrutinib 75 mg BID) and B the treatment group of comparison (placebo), the difference in response rate is the weighted average of differences within strata. It can be written as

$$\hat{\delta} = \frac{\sum_{j=1}^{K} W_j (r_{Aj} - r_{Bj})}{\sum_{j=1}^{K} W_j}$$

, where r_{Aj} represents the response rate of the treatment group A in the strata j, and W_j represent the CMH weight of the stratum j. This CMH weight must be calculated using the formula no. (9), available in Lu 2008, and then standardized to have a sum equal to 1:

$$W_j^{CMH} = \left[\frac{1}{n_{Aj}} + \frac{1}{n_{Bj}}\right]^{-1} \Bigg/ \sum_{j=1}^{K} \left[\frac{1}{n_{Aj}} + \frac{1}{n_{Bj}}\right]^{-1}$$

, where n_{Aj} represents the sample size of the strata j for the treatment group A (n_{Bj} for treatment B respectively).

Once $\hat{\delta}$ is calculated, the chi-square statistic is given by the formula no. (1), available in Lu 2008. Consider the following hypotheses $H(\delta)$: $\pi_{Aj} - \pi_{Bj} = \delta$ for j = 1, …, K, we obtain

$$X^2(\delta) = \frac{\left[\sum_{j=1}^{K} W_j(r_{Aj} - r_{Bj} - \delta)\right]^2}{\sum_{j=1}^{K} W_j^2 \tilde{V}_{r_{Aj} - r_{Bj}}}$$

where

$$\tilde{V}_{r_{Aj} - r_{Bj}} = \left\{\frac{\tilde{R}_{Aj}(1 - \tilde{R}_{Aj})}{n_{Aj}} + \frac{\tilde{R}_{Bj}(1 - \tilde{R}_{Bj})}{n_{Bj}}\right\}\frac{N_j}{(N_j - 1)}$$

denotes the variance of $r_{Aj} - r_{Bj}$ under the hypothese $H(\delta)$, $\tilde{R}_{Aj}$ and $\tilde{R}_{Bj}$ are the constrained maximum likelihood estimates of π_{Aj} and π_{Bj} so that $\tilde{R}_{Aj} - \tilde{R}_{Bj} = \delta$, W_j is the CMH weight for stratum j, and N_j is the total number of subjects in stratum j. The MN weights will not be used to compute the chi-square statistic, in order to avoid an additional iterative process.

For each stratum j, the constrained maximum likelihood estimates are computed as follows, as described in Farrington 1990:

$$\tilde{R}_{Aj} = 2u\cos(w) - {}^{b}/_{3a}$$

$$\tilde{R}_{Bj} = \tilde{R}_{Aj} - \delta$$

, where

$$w = (\pi + cos^{-1}\left({}^{\vartheta}/_{u^3}\right))/3$$

$$\vartheta = ({}^{b^3}/_{3a})^3 - {}^{bc}/_{6a^2} + {}^{d}/_{2a}$$

$$u = sign(\vartheta)\sqrt{{}^{b^2}/_{(3a)^2} - {}^{c}/_{3a}}$$

$$a = 1 + \theta$$

$$b = -(1 + \theta + r_{Aj} + \theta r_{Bj} + \delta(\theta + 2))$$

$$c = \delta^2 + \delta(2r_{Aj} + \theta + 1) + r_{Aj} + \theta r_{Bj}$$

$$d = -r_{Aj}\,\delta(1 + \delta)$$

$$\theta = {}^{n_{Bj}}/_{n_{Aj}}$$

Finally, the confidence limits based on MN method can be obtained by solving the equation (3), available in Lu 2008. Indeed, considering $\alpha = 0.05$, if we denote $\chi^2_{1,0.95}$ the 95th percentile of the

chi-square distribution with one degree of freedom, then we reject $H(\delta)$ if $X^2(\delta) > \chi^2_{1,0.95}$. The 2-sided 95% confidence limits of δ can be obtained by solving

$$X^2(\hat{\delta}_L) = X^2(\hat{\delta}_U) = \chi^2_{1,1-\alpha}$$

To solve this equation, an iterative process must be implemented to reach the lower and upper bounds of the 95% CI, ie the δ must vary between $(-1\ ;\ 1)$, bounds excluded, until the equation is verified. This will provide the values of the lower bound $\hat{\delta}_L$ and the upper bound $\hat{\delta}_U$.

A sample code to calculate this 2-sided 95% CI is available in Appendix 18.5. It is based on the example provided in Garner 2016.

For each pair-wise comparison, the statistical hypothesis test can be written as follows:

$$\begin{cases} \text{H0: there is no significant difference of response rate between treatment A and B, ie, } \delta = 0 \\ \text{H1: there is significant difference of response rate between treatment A and B, ie, } \delta \neq 0 \end{cases}$$

If we denote x the result the equation (1) in Lu 2008, then the p-value is given by $P(X > x)$ where X follows a chi-square distribution with one degree of freedom.

Supportive analysis no. 2: Tipping Point

Subjects with reason of treatment discontinuation related to safety or efficacy will be considered as non-responders.

Tipping point analysis will be performed as in Yan 2009, Section 2.3.2.

Let

$$n_1 = \textit{number of responders out of missing data due to various reasons for placebo}$$

$$n_2 = \textit{number of responders out of missing data due to various reasons for evobrutinib } 25\ \textit{mg QD}$$

$$n_3 = \textit{number of responders out of missing data due to various reasons for evobrutinib } 75\ \textit{mg QD}$$

$$n_4 = \textit{number of responders out of missing data due to various reasons for evobrutinib } 50\ \textit{mg BID}$$

The same analysis method as the primary analysis of the ACR20 at Week 12 will be performed using the mITT Analysis Set for each possible combination of n_1, n_2 , n_3 and n_4.

Using the Hochberg procedure, the values of n_1, n_2 , n_3 and n_4 where the study result will be tipped from statistically significant (not statistically significant) to not statistically significant (statistically significant) will be assessed for plausibility based on the corresponding response rate from each treatment.

Supportive analysis no. 3: Stratified difference of response rates with imputed ACR components using multiple imputations (MAR pattern)

The same method for estimating stratified difference of response rates will be applied, but the ACR components will be imputed using multiple imputations process (MI SAS procedure) with MAR pattern, as if they follow their original assignment of treatment group.

Data will be imputed for subjects with reason of discontinuation unrelated to safety or efficacy. Subjects with reason of treatment discontinuation related to safety or efficacy will be considered as non-responders.

The SAS code is available in Appendix 18.5.

The multiple imputations will be performed as follows:

1. Monotone missing data structure will be created as follows: intermediate (non-monotone) missing data will be multiply imputed using the Markov chain Monte Carlo (MCMC) method and assuming MAR and multivariate normality. Transformation of data will only be used if there is a clear deviation from normality. The SAS procedure PROC MI with the MCMC option will be used with seed number=PPD. The number of burn-in iterations will be set to 200, which is the default value. Nevertheless, if diagnosis plots show that the convergence has not yet occurred, this will be adjusted. The ACR components must be pre-specified in the following order, the last one being the most significant component:
    i. Physical function (as assessed using the HAQ-DI) from Week 2 to Week 12
    ii. Patient's Global Assessment of RA Disease Activity from Week 2 to Week 12
    iii. Physician's Global Assessment of RA Disease Activity from Week 2 to Week 12
    iv. Pain (as assessed using the Patient's Assessment of Arthritis Pain VAS) from Week 2 to Week 12
    v. hsCRP from Baseline to Week 12
    vi. Tender Joint Count (of 68 joints assessed) from Baseline to Week 12
    vii. Swollen Joint Count (of 66 joints assessed) from Baseline to Week 12
2. Then, each component will be imputed with treatment group and data at previous visit from all ACR20 components as covariates**.**
3. Imputation will be repeated N=1000 times. The ACR20 will then be calculated for each of the data sets from the imputed components.
4. Single proportion and associated 95% CI of each treatment arm will be computed using EXACT BINOMIAL from FREQ procedure with SAS. The 95% CI will be used to compute the standard error of each proportion for the N=1000 imputations.
5. The difference of response rate compared to placebo and associated 95% CI based on stratified MN method will be computed for each of the N=1000 imputations. The 95% CI

will be used to compute the standard error of the difference of response rate compared to placebo.

6. Results will be combined using MIANALYZE SAS procedure with the Rubin's rules.

Supportive analysis no. 4: Stratified difference of response rates with imputed ACR components using multiple imputations (MNAR pattern)

The same method as in Supportive analysis no. 3 will be considered, but this time with MNAR pattern. Subjects' missing data due to early discontinuation will be imputed, using data available from the placebo arm, regardless of reason for study discontinuation.

Supportive Analysis no. 5: Non-parametric method – Cochran-Armitage test

A monotonic dose-response relationship, between ordered treatment groups and proportion of responders will be assessed via the Cochran-Armitage trend test (Dörner 2017). The p-value reported tests the null hypothesis that the proportion of ACR20 responders is the same across treatment groups, against the alternative that the proportion is ordered (from smallest value to largest) according to increasing treatment group. In the Cochran-Armitage test, missing data are handled the same as in the main logistic modeling analysis.

For this Cochran-Armitage trend test, the proportion of ACR20 response rate, the p-value (asymptotic) will be presented.

An example of SAS code is available in Appendix 18.5.

Supportive analysis no. 6: MCP-MOD approach to dose response

The MCP-MOD approach to dose response (Bretz 2005) will be implemented using R software, in particular the DoseFinding package (see Bornkamp 2007), based on the mITT Analysis Set.

First of all, the following shapes of models will be considered:

- Linear model
- Quadratic model
- E_{max} model (with several parameters)
- Logistic model

For the primary MCP-MOD approach, the highest dose (50 mg BID) is assumed to be equivalent to 100 mg QD. Further assumptions of the highest dose in QD regimen may be explored as post hoc analyses to assess the robustness of the primary result.

The expected minimal response rate that can be obtained (with the placebo group, given a dose of evobrutinib equal to zero) has been estimated at 30%. If we denote $f^0(d, \theta^*)$ the standardized dose-response model function (based on notations from Section 2.1 of Bornkamp 2009), the parameters to estimate are as follows:

- For the linear model, there are no parameters to estimate.

- For the E_{max} model, the equation is $f^0(d,\theta^*) = {d}/{(ED_{50} + d)}$
- For the quadratic model, the equation is $f^0(d,\theta^*) = d + \delta d^2$
- For the logistic model, the equation is $f^0(d,\theta^*) = {1}/{\{1 + \exp[(ED_{50} - d)/\delta]\}}$

Table 14 describes the final models to be used for the MCP-MOD Analysis. The estimated parameter value(s) for each model are calculated based on efficacy results of the primary endpoint from study MS200527-0086 (See appendix 18.9 for rationale and calculations of the estimated parameters in Table 14).

Table 14 : MCP-MOD Approach: Selected Candidates Models

Model	$f^0(d,\theta^*)$	Parameters estimated
Linear model	d	None
E_{max} model	${d}/{(ED_{50} + d)}$	ED_{50} = 26.35
Quadratic model	$d + \delta d^2$	δ = -0.0067
Logistic model	${1}/{\{1 + \exp[(ED_{50} - d)/\delta]\}}$	$ED_{50} = 73.27$; $\delta = 1.66$

Note: the function $f^0(d,\theta^*)$, which is the standardized model function driven by the equation $f^0(d,\ \theta^*) = \theta_0 + \theta_1 f^0(d,\theta)$, denotes the standardized model parameterized by the vector θ^*.
In this parameterization, θ_0 is a location and θ_1 a scale parameter such that only the parameter-vector θ^* determines the shape of the model function.

For each of the dose-response models in the candidate set, the null hypothesis that the sum of the components of the optimal contrast vector is equal to zero will be tested. To do so, the contrasts coefficients of each model are chosen so they maximize the power to detect the absence of dose response. The methodology of the MCP-MOD is detailed in Section 2.2 of Bornkamp 2009.

Under the null hypothesis, if the maximum statistic of the optimal contrasts for the candidate models exceeds the critical value (corresponding to a unilateral test with alpha = 0.05), then the null hypothesis will be rejected and the proof of concept of a non-null dose-response will be demonstrated.

Once the proof of concept has been established, the next step is to estimate the dose-response curve and the target doses of interest:

- The minimum effective dose (MED), defined as the smallest dose ensuring a clinically relevant and statistically significant improvement over placebo, estimated at 20%.
- The target dose will be estimated for the difference in ACR20 response over placebo of 25%, 30% and 35% using the optimal model selected above.

An example of R code is available in Appendix 18.5.

Descriptive statistics:

Number and percentage of subjects with ACR20 response status (Yes/No/Missing) will be presented over time for each treatment group, including the delta from placebo group. Data will be analyzed as observed, ie, without any imputation.

14.2 Secondary Endpoints

Statistical analyses of the secondary endpoints are supportive in nature. All p-values and 95% CIs from comparisons with placebo are at nominal levels.

14.2.1 Key Secondary Endpoints at Week 12

The following key secondary endpoints will be presented on both mITT and PP analysis sets for the primary analysis and mITT analysis set only for supportive analysis:

- DAS28-hsCRP low disease activity (DAS28 < 3.2) rate
- DAS28-hsCRP remission (DAS28 < 2.6) rate
- ACR50 response assessed using hsCRP
- ACR70 response assessed using hsCRP.

These endpoints are defined in Appendix 18.1 and will be analyzed according to Table 15.

Table 15 : Analysis of the Key Secondary Endpoints at Week 12

Endpoints	Analysis Set	Analysis method	Description	Missing data handling
• DAS28-hsCRP < 3.2 • DAS28-hsCRP < 2.6 • ACR50 • ACR70	• mITT set	Primary analysis: Stratified MN for Δ	• Presentation of response rate for each treatment group, Δ in response rate compared to placebo, 2-sided 95% CI of Δ, nominal p-value • Summary Stats over time	LOCF-NR*
	• mITT set	Descriptive statistics	Number and percentage of subjects with response status (Yes/No/Missing) over time	As observed

* LOCF-NR: post-baseline last observation carried forward. Patients who discontinued from treatment are considered as non-responders

14.2.2 Other Secondary Endpoints at Week 12

The following secondary endpoints, defined in Appendix 18.1, will be presented on mITT analysis set:

- Achieving ACR EULAR Boolean remission
- Achieving CDAI score ≤ 2.8 (CDAI remission)
- Achieving SDAI score ≤ 3.3 (SDAI remission)
- EULAR Responder Index (based on DAS-hsCRP)
- ACR hybrid scores computed using hsCRP
- Change from Baseline in DAS28-hsCRP
- Change from Baseline in CDAI and SDAI
- Changes and percentage changes from Baseline of individual components of the ACR Core Set
- ESR (mm/Hr)

These endpoints will be analyzed according to Table 16.

Table 16 : Analysis of the Other Secondary Endpoints at Week 12

Endpoints	Analysis Set	Analysis method	Description	Missing data handling
• EULAR Boolean remission • CDAI ≤ 2.8 • SDAI ≤ 3.3	• mITT set	Primary analysis: Stratified Δ	• Presentation of response rate for each treatment group, Δ in response rate compared to placebo, 2-sided stratified MN 95% CI of Δ • Summary statistics over time	LOCF-NR*
• EULAR Responder Index	• mITT set	Primary analysis for binary data: Stratified for Δ	• Presentation of response rate for each treatment group, Δ in response rate compared to placebo, 2-sided stratified MN 95% CI of Δ • Summary statistics over time	LOCF-NR*
• EULAR Responder Index	• mITT set	Primary analysis for all categories: descriptive statistics	Summary statistics over time	LOCF-NR*
• ACR hybrid score • DAS28-hsCRP • CDAI • SDAI	• mITT set	Primary analysis: Analysis of Covariance (ANCOVA) model	Presentation of adjusted means, adjusted Δ of means, associated 2-sided 95% CI of Δ.	LOCF (components)
		Supportive analysis: Hodges-Lehmann (HL) estimator	Presentation of difference median estimate and 95% CI (Asymptotic and exact)	
		Summary statistics	Summary statistics over time	
• ACR components	• mITT set	Summary statistics	Summary statistics over time	LOCF (components)
• ESR	• mITT set	Summary statistics	Summary statistics over time	As observed

* LOCF-NR: post-baseline last observation carried forward. Patients who discontinued from treatment are considered as non-responders.

ANCOVA model:

Δ in response rate compared to placebo and corresponding 95% CI will be estimated with SAS® GLM procedure using LSMEANS statement (see SAS code in Appendix 18.5). The ANCOVA will include baseline, region as covariates and treatment group as an independent variable.

Validity of the model will be verified as follows:

- The homogeneity of variances between treatment groups will be checked using the Levene's test. A scatter plot of residuals versus fitted values will be also performed.

- In order to assess the independence of error terms, a residual lag plot will be performed, ie constructed by plotting residual (i) against residual (i-1).
- A scatter plot of the response variable against the covariates, using separate symbols for each treatment group will be performed to verify the presence of linear relationship between the covariates and the response variable, and that all treatment regression lines have the same slope.
- Homogeneity of regression slopes will be verified by first running the model with the interaction between the covariates and the treatment group. This assumption will be validated if this interaction is not significant. If this assumption is verified, then the interaction will be removed from the model.
- Normality of residuals will be checked by computing Skewness and Kurtosis, which should fall within the interval [-2;2]. Normal QQ-plot of residuals will be performed as well.

HL estimator:

A sensitivity analysis will be performed using a non-parametric method. The Hodges-Lehmann difference median estimate and 2-sided 95% CI will be calculated with SAS® NPAR1WAY procedure using HL statement (see SAS code in Appendix 18.5). Asymptotic and exact CIs will be computed.

Summary statistics:

For each continuous endpoint, baseline value will be presented first. Then, for each time point, the following will be presented on subjects who have reached this time point: absolute value, CFB and percent CFB.

14.2.3 Imaging Endpoints at Week 4 and Week 12

The following imaging endpoints will be presented on MRI and MRI PP analysis sets:

- CFB in OMERACT RAMRIS scores (designated hand, minimal assessed joints: wrist, MCP joints #1 to #5):
  - Synovitis score,
  - Bone marrow edema (osteitis) score,

CCI

OMERACT RAMRIS scores for synovitis, tenosynovitis, bone marrow edema (osteitis), and erosions (Ostergaard 2004, Ostergaard 2017) will be read centrally by 2 independent readers, who are blinded to treatment and dose assignment and to chronology of each time point. Discrepancies in RAMRIS scores between the 2 readers will be reviewed at a consensus adjudication session with both readers to reach a consensus score for the designated cases. At a minimum, the images with the top 10% discrepancies in scores between the 2 readers will be reviewed (refer to Imaging Guidelines for details of Adjudication and Selection of Cases for Adjudication). In case of further discrepancies after the adjudication process, the average between the results from the 2 readers will be considered for statistical analysis. In subject data listings, both results from the 2 readers will be presented as well as the average result.

In addition, the Total Damage Score and Total Inflammation Scores for both RAMRIS and RAMRIQ will be analyzed as well (Total Damage Score will not be computed for RAMRIQ). They are defined according to the following formulas (Peterfy 2016, Brett 2018):

CCI

Table 17 : Analysis of Imaging Endpoints

Endpoints	Analysis Set	Analysis method	Description	Missing data handling
• RAMRIS scores [redacted]	• MRI Analysis set • MRI PP Analysis Set	Primary analysis: MMRM	• Presentation of adjusted means, adjusted Δ (Evobrutinib – Placebo) of means, associated 2-sided 95% CI of Δ, nominal p-value. • Summary statistics over time	All available data as observed (MAR)
		Supportive analysis #1: ANOVA model with repeated measures	• Presentation of adjusted means, adjusted Δ (Evobrutinib – Placebo) of means, associated 2-sided 95% CI of Δ, nominal p-value. • Summary statistics over time	Missing data will be imputed using linear interpolation or linear extrapolation
		Supportive analysis #2: ANOVA model with repeated measures	• Presentation of adjusted means, adjusted Δ (Evobrutinib – Placebo) of means, associated 2-sided 95% CI of Δ, nominal p-value. • Summary statistics over time	As observed
		• Lineplot	• Lineplots will be presented for RAMRIS Synovitis and Osteitis only.	As observed

MMRM:

The MAR are ignorable missingness. When missing data due to dropout are ignorable, the parameters of dropout and outcome processes are assumed to be distinct. Hence, likelihood-based methods can be used on the marginal distribution of the observed data for statistical inference. MMRM is a particular form of a mixed model analysis and is fitted within the ignorable likelihood paradigm.

The MRI endpoints (change from baseline) will be analyzed using the MMRM model, with treatment group, visit, and treatment by visit interaction as the main effects, baseline MRI assessment and region as covariates. The unstructured covariance matrix will be considered. Denominator degree of freedom will be computed using Kenward and Roger's method (Kenward 1997).

The LSMEANS estimate of MRI mean change from baseline score, difference of the LS Mean from placebo, 2-sided 95% CI of the difference and nominal 2-sided p-value of the difference from the model contrast will be presented at each visit.

An example of SAS code is available in Appendix 18.5.

In addition, summary statistics of observed values, change and percent change from baseline will be presented at each visit for each treatment group.

Supportive Analysis no. 1:

An ANOVA model for repeated measures will be performed with treatment group and region as covariates. Missing data will be imputed using linear interpolation for assessments at Week 4 and linear extrapolation for assessments at Week 12.

The LSMEANS estimate of MRI score, difference of the LS Mean from placebo, 2-sided 95% CI of the difference and nominal 2-sided p-value of the difference from the model contrast will be presented at each visit.

Validity of the model will be verified as follows:

- Repeated-measures designs have the assumption of Sphericity which means that the variance of the population difference scores for any two conditions should be the same as the variance of the population difference scores for any other two conditions. This assumption can be tested using the Mauchly's sphericity test (available with GLM SAS procedure).
- Normality of residuals by timepoint will be checked by computing Skewness and Kurtosis, which should fall within the interval [-2;2]. Normal QQ-plot of residuals will be performed as well.

An example of SAS code is available in Appendix 18.5.

In addition, summary statistics of observed values, change and percent change from baseline will be presented at each visit for each treatment group.

Supportive Analysis no. 2:

A similar analysis as the supportive analysis no. 1 will be performed, but this time data will not be imputed.

Lineplot:

Mean score CFB will be presented as a by-visit line plot for each treatment group, with all treatment groups included in a single figure, and horizontal axis extending to Week 12 (RAMRIS Synovitis and Osteitis only).

14.2.4 HRQoL Endpoints at Week 12

The following HRQoL endpoints will be presented on QoL analysis set:

- HAQ-DI
- SF-36
- FACIT-Fatigue

CCI [redacted]

HAQ-DI scoring is detailed in Appendix 18.2.

A full description of the scoring rules of each health concept of SF-36 as well as Physical Component Summary (PCS) and Mental Component Summary (MCS) is available in Appendix 18.3.

A full description of the scoring instructions of FACIT-Fatigue scale is available in Appendix 18.4.

CCI [redacted]

Table 18 : Analysis of HRQoL Endpoints at Week 12

Endpoints	Analysis Set	Analysis method	Missing data handling
• HAQ-DI Total Score	• QoL set	• MMRM	As observed
• SF-36 (Change in PCS, MCS scores, and their components) • FACIT-Fatigue (Change in Total score over time) CCI [redacted]	• QoL set	• MMRM model • Descriptive statistics • Probability of achieving %CFB value or higher • Logistic regression	As observed

MMRM model: similar as in Section 14.2.3.

Descriptive statistics:

For each treatment group, descriptive statistics will be presented for absolute value, CFB and percent CFB value for each visit, from baseline to Week 12.

In addition, for each SF-36 endpoint (domain scores, MCS, PCS), FACIT-Fatigue Total Score and CCI , the following will be provided:

- the number (proportion) of subjects not worsening, and number (proportion) of subjects worsening, between baseline and Week 12 by treatment group. Worsening is defined as a decrease of 5 for SF-36 endpoints and FACIT-Fatigue Total Score, and as an increase of 5 for CCI .
- Mean score over time will be presented as a by-visit bar plot for each treatment group, with all treatment groups included in a single figure, and horizontal axis extending to Week 12.

In addition, for the CCI , the frequency of item administration by visit will be reported (ie, the items that were selected and administered by the CAT algorithm).

Probability of achieving %CFB value or higher:

For each SF-36 endpoint (domain scores, MCS, PCS), a figure will be provided describing the distribution of % CFB at Week 12, one curve per treatment group. The curve will display proportion of subjects having a value for % CFB > x at Week 12, where the range of x depends on the data.

The same will be applied for FACIT-Fatigue Total Score and CCI .

Logistic regression:

Response on the SF-36 (PCS, PF, VT), FACIT-Fatigue total score and CCI will be further analyzed based on differences in proportion of responders, comparing each dose of evobrutinib versus placebo.

Responder definitions will be based on minimal clinically important difference (MCID) criteria for improvement or deterioration as follows:

- SF-36 - PCS:
  - Improvement is defined as an increase greater or equal to 2.5 compared to baseline
  - Deterioration is defined as a decrease greater or equal to 0.8 compared to baseline
- SF-36 – domain scores:
  - Improvement is defined as an increase greater or equal to 5.0 compared to baseline
  - Deterioration is defined as a decrease greater or equal to 2.5 compared to baseline
- FACIT-Fatigue Total Score:
  - Improvement is defined as an increase greater or equal to 4.0 compared to baseline

- Deterioration is defined as a decrease greater or equal to 4.0 compared to baseline

CCI [redacted]

Logistic regression model for repeated measures will be applied, with adjustments for the region, baseline value and with treatment as a factor. An example of SAS code is available in Appendix 18.5.

14.2.5 Exploratory Efficacy Endpoints at Week 12

CCI [redacted]

14.3 Subgroup Analysis

The following subgroup analyses will be considered:

- Region (Western Europe, US, ROW)
- MRI substudy participation (Yes, No)
- Gender (Male, Female)

The endpoints at Week 12 and the statistical analysis method are described in Table 21.

To assess numerically the consistency of each endpoint among subgroups of subjects, the analysis will be performed on the mITT Analysis Set for each modality using a logistic regression model. The interaction between treatment group and subgroup of interest will be calculated as well. Logistic models to be performed are summarized in Table 20.

Table 20 : Logistic regression models for subgroup analysis

Subgroup	Subjects to be included in the model	Main effects	Result to be presented in TLFs
• Region	• All subjects from mITT Set	• Treatment, region and treatment by region interaction	P-value from the interaction between treatment and region
	• Subjects from mITT Set – Region=Western Europe	• Treatment	From comparison with placebo: • Estimated OR, • Two-sided 95% CI of the estimated OR
	• Subjects from mITT Set – Region=US	• Treatment	
	• Subjects from mITT Set – Region=ROW	• Treatment	
• MRI substudy participation	• All subjects from mITT Set	• Treatment, MRI substudy participation and treatment by MRI substudy participation interaction	P-value from the interaction between treatment and MRI substudy participation
	• Subjects from mITT Set – MRI substudy participants only	• Treatment, region	From comparison with placebo: • Estimated OR, • Two-sided 95% CI of the estimated OR
	• Subjects from mITT Set – MRI substudy non-participants only	• Treatment, region	
• Gender	• All subjects from mITT Set	• Treatment, gender and treatment by gender interaction	P-value from the interaction between treatment and gender
	• Subjects from mITT Set – females	• Treatment, region	From comparison with placebo: • Estimated OR, • Two-sided 95% CI of the estimated OR
	• Subjects from mITT Set – males	• Treatment, region	

The same missing data imputation approach will be used as the primary analysis method. An example of SAS code is available in Appendix 18.5.

In addition, a forest plot will be presented for each endpoint for the overall treatment effect and within each subgroup of interest including OR from logistic model and associated 2-sided 95% CI of OR.

Summary tables will be presented for each endpoint by subgroup, comparing each evobrutinib dose group vs placebo.

Table 21 : Subgroup Analysis – Endpoints

Endpoints	Analysis	Analysis method	Description	Missing data handling
• ACR 20/50/70 • DAS28-hsCRP < 2.6 • DAS28-hsCRP < 3.2 • EULAR Boolean Remission • CDAI ≤ 2.8 • SDAI ≤ 3.3 • EULAR Responder Index [good/moderate, vs no response]	• Week 12	Logistic regression	Presentation of response rate for each treatment group, Δ in response rate compared to placebo, 2-sided 95% CI of Δ (non-stratified MN), OR, 2-sided 95% CI of OR, p-value of interaction between treatment and subgroup of interest	LOCF-NR*

* LOCF-NR: post-baseline last observation carried forward. At Week 12, patients who discontinued from treatment are considered as non-responders.

14.4 Open-label Extension

Efficacy Data from OLE will be listed on subjects from mITT set who entered in the OLE.

15 Safety Evaluation

The subsections in this section include specifications for summarizing safety endpoints that are common across clinical trials such as AEs, laboratory tests and vital signs. As detailed in Section 7, some subjects have entered in the OLE, even if the Sponsor decided not to initiate the OLE in protocol V3.0. As a result, for all safety endpoints, data selection will be handled as indicated in Table 22.

Table 22 : Data Handling for Safety Analysis

Population	Period covered	Treatment groups	Data to be analyzed
SAF	12-week treatment period	• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	All Data collected during the 12-week treatment period, safety follow-up included
Subjects with at least 1 dose during OLE	OLE period only	• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	Data from OLE period only including all Safety follow-up data posterior to OLE (Safety follow-up / End of Study)

15.1 Adverse Events

All analyses described in this section will be based on treatment-emergent adverse events (TEAEs) if not otherwise specified. TEAEs will be defined according to Table 23. AEs from the OLE period will be listed only (those with a start date after the first intake during the OLE).

Table 23: Definition of TEAE

Treatment group	Period covered	Definition
• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	12-week treatment period	TEAEs will be defined as: • AEs starting on or after first treatment administration of any IMP (placebo or evobrutinib) during the 12-week treatment period (safety follow-up included) • or if it was present prior to any IMP administration but exacerbated after. Any AE which started before study first treatment administration of any IMP (placebo or evobrutinib), but improved during treatment period, will not be counted as TEAE.

For the purpose of defining the TEAE flag, incomplete AE-related dates will be handled as follows:

- In case the onset date is missing completely or missing partially but the onset month and year, or the onset year are equal to the start of IMP then the onset date will be replaced by the minimum of start of IMP and AE resolution date.
- In all other cases the missing onset day or missing onset month will be replaced by 1.
- Incomplete stop dates will be replaced by the last day of the month (if day is missing only), if not resulting in a date later than the date of subject's death. In the latter case, the date of death will be used to impute the incomplete stop date.
- Further information collected after the cut-off for an analysis (such as a fatal outcome) may be extracted from the Safety data base and presented separately in the CSR.

15.1.1 All Adverse Events

AEs will be coded according to the latest MedDRA version available at the time of analysis. The severity of AEs will be graded using National Cancer Institute - Common Terminology Criteria for Adverse Events (NCI-CTCAE version 4.03) toxicity grades. AEs with missing classification concerning IMP relationship will be considered related to the IMP.

15.1.1.1 3-Tier Approach

The 3-tier approach is a systematic way to summarize and analyze adverse events (AEs) in clinical trials (Crowe 2009). AEs in different tiers are analyzed using different levels of statistical analyses.

Tier 1 AEs are provided in Appendix 18.10. Other AEs will be classified into Tier 2 or Tier 3 based on the Rule-of-3. If there are 3 or more subjects with the reported term in any treatment group, that term will be included in Tier 2. Otherwise, it will be included in Tier 3.

The Tier 1 and Tier 2 AEs will be assessed with the number and percentage of subjects having at least one occurrence, as well as the difference with placebo and corresponding 95% CI for between-group comparison. For the difference in crude rates, the CIs will be based on MN method (Miettinen 1985). No multiplicity adjustment will be applied.

The tier 3 AEs will be assessed via summary statistics and risk differences.

For each comparison of evobrutinib group with placebo, forest trees for Tier 1 and Tier 2 AEs will be provided as well, displaying the incidence rate and associated 2-sided 95% CI of difference.

15.1.1.2 Overview of TEAEs

Two summary tables of TEAEs will be provided as described in Table 24.

Table 24: Summary Tables of TEAEs

Summary	Modalities
Overview of TEAEs	• Any TEAE • IMP related TEAE • Serious TEAE • IMP related serious TEAE • TEAE with NCI-CTCAE grade 1 • IMP related TEAE with NCI-CTCAE grade 1 • TEAE with NCI-CTCAE grade 2 • IMP related TEAE with NCI-CTCAE grade 2 • TEAE with NCI-CTCAE grade 3 • IMP related TEAE with NCI-CTCAE grade 3 • TEAE with NCI-CTCAE grade 4 • IMP related TEAE with NCI-CTCAE grade 4 • TEAE leading to death • IMP related TEAE leading to death
Overview of TEAEs Actions: • Change in dose, • Administration of medication, • Procedure, • Study termination	• TEAE with no change of dose • IMP related TEAE with no change of dose • TEAE leading to dose reduction • IMP related TEAE leading to dose reduction • TEAE leading to dose increase • IMP related TEAE leading to dose increase • TEAE leading to interruption of IMP • IMP related TEAE leading to interruption of IMP • TEAE leading to withdrawal of IMP • IMP related TEAE leading to withdrawal of IMP • TEAE leading to administration of concomitant medication • IMP related TEAE leading to administration of concomitant medication • TEAE leading to concomitant procedure • IMP related TEAE leading to concomitant procedure • TEAE leading to study termination • IMP related TEAE leading to study termination

15.1.1.3 Tabulation of Adverse Events by SOC and PT

The TEAE tables to be prepared are listed in Table 25.

Table 25: TEAE Tables to be produced

	Overall frequency	By primary SOC and PT	By PT only	By primary SOC, PT and worst grade
Overview of TEAEs	✓	*NA*	*NA*	*NA*
Overview of TEAEs actions	✓	*NA*	*NA*	*NA*
All TEAEs	✓	✓	✓	✓
IMP-related TEAEs	✓	✓	✓	✓
AEs of Special Interest (AESI)	✓	✓		
IMP-related AESIs	✓	✓		
Serious TEAEs	✓	✓		
IMP-related serious TEAEs	✓	✓		
Non-serious TEAEs*	✓	✓		
TEAEs leading to death	✓	✓		
IMP-related TEAEs leading to death	✓	✓		
TEAEs leading to withdrawal of IMP	✓	✓		
TEAEs leading to study termination	✓	✓		
EAIR	✓	✓		

(*): A table with all non-serious TEAEs will be first provided and then only TEAEs exceeding a frequency of 5% in at least one of the treatment groups (> 5%), by SOC and PT will be provided.

Specific rules for SOC/PT tabulation

All AEs recorded during the trial (ie assessed from signature of informed consent until the end of the Follow-up/End of Trial visit) will be coded according to the MedDRA and assigned to a SOC and PT.

SOC terms will be sorted alphabetically. The combined dose of evobrutinib (25 mg QD, 75 mg QD and 50 mg BID) will be computed and displayed in the table. PTs within each SOC will be sorted by descending frequency of this combined dose of evobrutinib, and then alphabetically if multiple PTs have the same frequency.

If a subject experiences more than one occurrence of the same TEAE during the trial, the subject will be counted only once for that treatment (the worst severity and the worst relationship to trial treatment will be tabulated).

In case a subject had events with missing and non-missing grades, the maximum of the non-missing grades will be displayed.

Exposure adjusted incidence rates (EAIR) are calculated as number of subjects with TEAE divided by the sum of the individual times of all subjects in the safety population from start of treatment to first onset of TEAE or end of treatment period, whichever occurs first. The incidence rate multiplied with 1000 would give the number of TEAEs expected in 1000 subjects within 1 year. EAIR of TEAEs will be presented by SOC and PT.

Subject data listings

TEAEs will be also listed by treatment group and subject. A listing of TEAEs leading to withdrawal of IMP, a listing of TEAEs leading to study termination, if any, will be provided as well.

15.1.1.4 Adverse Events of Special Interest

AESIs will be summarized by SOC and PT as stated in Table 25. CCI [redacted]

15.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

15.2.1 Deaths

A summary of deaths will be provided including (clinicaltrials.gov requirement):

- Number and percentage of (all) deaths
- Number and percentage of the primary cause of death (categories: disease progression, adverse event, unknown, other)

The tabulation of TEAEs leading to death is described in Section 15.1.1.2. A listing of deaths, if any, will be provided.

In case there is no death in the trial, only the summary of death required by clinicaltrials.gov will be performed, neither tabulation of TEAE leading to death will be edited, nor the listing of death.

15.2.2 Serious Adverse Events

The tabulation of serious TEAEs is described in Section 15.1.1.2. A subject listing of serious TEAEs will be provided.

15.3 Clinical Laboratory Evaluation

The following laboratory parameters will be measured during the trial as part of the safety evaluation:

- Hematology
- Biochemistry, including the following supplementary liver function tests (LFT):
    - Aspartate aminotransferase,
    - Alanine aminotransferase,
    - Alkaline phosphatase,
    - γ-Glutamyl-transferase,
    - Bilirubin (total),
- Urinalysis
- Coagulation
- Lipid panel
- Hepatic panel (restricted to subjects for whom above AST or ALT withdrawal criteria are met and who permanently discontinue dosing due to elevated LFTs)

The clinical laboratory safety tests to be measured in this trial are provided in the protocol (refer to Section 7.4.3 Table 10 of the CTP). Parameters from the Section 7.4.3 of the CTP to be summarized and listed in the TLFs are provided in Appendix 18.8.

All laboratory data results will be presented using international system of units (SI).

Continuous protocol-specified clinical laboratory findings (hematology, biochemistry, urinalysis, coagulation, lipid panel, hepatic panel) will be summarized by treatment using descriptive statistics over time (see Section 9) separately for 12-week treatment period and OLE. The descriptive statistics will be presented as follows:

- The baseline will be presented first
- Then, when applicable, each scheduled time point will be presented on subjects who have reached this time point: absolute values, CFB and percent CFB will be displayed.

Laboratory results will be classified according to NCI-CTCAE Version 4.03 as provided by the central laboratory. In case a laboratory parameter has bi-directional toxicities (eg Potassium) both directions will be presented for the given parameter (ie Potassium Low and Potassium High).

Laboratory results containing a modifier such as "<" or ">" will be handled case by case for summary statistics and will be reported both as collected in the database and imputed in subject

data listings. Decision for each laboratory parameter will be documented in a Note-to-File prior to the database lock.

A shift table of baseline versus post-baseline based on the worst NCI-CTCAE grade will be presented by treatment group for hematology and biochemistry.

Subject data listings will be provided, with a flag for abnormal values, along with corresponding normal ranges:

- Laboratory gradable parameters part of NCI-CTCAE will be presented according to the categories based on normal ranges along with the grade. Abnormal values will be flagged according to the direction of toxicity as detailed in Appendix 18.8 (eg, for a parameter such as Potassium Low, only values below the LLN will be flagged).

- Laboratory parameters that are not part of NCI-CTCAE will be presented according to the categories based on normal ranges: below normal limits (Low), within normal limits (Normal), and above normal limits (High). Values that are either above ULN or below LLN will be flagged.

Boxplots of the laboratory values by treatment group and time point will be provided for the following parameters of Table 26:

Table 26: Laboratory Parameters

Category	Laboratory parameter	Conventional Unit	SI Units
Hematology	Hemoglobin	g/dL	g/L
	Reticulocyte count	$10^3/\mu L$	$10^9/L$
	White blood cell count	$10^3/\mu L$	$10^9/L$
	Neutrophil count	$10^3/\mu L$	$10^9/L$
	Lymphocyte count	$10^3/\mu L$	$10^9/L$
	Platelet count	$10^3/\mu L$	$10^9/L$
Biochemistry	Albumin	g/dL	g/L
	Alanine aminotransferase (ALT)	U/L	U/L
	Aspartate aminotransferase (AST)	U/L	U/L
	Gamma-glutamyl transferase (GGT)	U/L	U/L
	Alkaline phosphatase	U/L	U/L
	Total bilirubin	mg/dL	µmol/L
	Amylase	U/L	U/L
	Lipase	U/L	U/L
	Creatinine	mg/dL	µmol/L
	Blood urea nitrogen (BUN)	mg/dL	mmol/L

Boxplots for all laboratory parameters will be displayed using the unit of measurement. If consistent with BOA standards, the ULN and LLN will be added to the lab parameter boxplot, for any lab parameter where the normal range is the same for all subjects in the analysis set.

Listings of individual data with a flag for abnormal values will be provided, along with corresponding normal ranges.

In this study, clinically significant lab abnormalities were recorded as adverse events. In lieu of a listing of clinically significant lab abnormalities for each domain, the following by-subject lab value listings will be provided:

- Listing of Grade ≥ 3 hematology values
- Listing of Grade ≥ 3 biochemistry values
- Listing of urinalysis values with Grade ≥ 3, value ≥ 2 times ULN (excluding values for Specific Gravity or pH parameters), or an increase of "++" for non-gradable parameters when applicable.

CCI

15.4 Vital Signs

Vital signs (height (m), weight (kg), BMI (kg/m²), body temperature (°C), SBP (mmHg), DBP (mmHg), respiratory rate (breaths/min) and pulse rate (beats/min)) will be summarized by treatment group using descriptive statistics (see Section 9).

The descriptive statistics will be presented as follows:

- The baseline will be presented first
- Then each scheduled time point will be presented on subjects who have reached this time point: absolute values, CFB and percent CFB will be displayed (except Height and BMI which are only collected at baseline).

Body temperature, SBP, DBP, respiratory rate and pulse rate will be analyzed with shift tables of maximum CFB using the categories defined in Table 27:

Table 27: Vital Signs Categories

Parameter	Unit	Shift	Baseline categories	Post-baseline categories (absolute change)
Temperature	°C	Increase	<37 / ≥37 - <38 / ≥38 - <39 /≥39 - <40 / ≥ 40	≤0* / >0 - <1 / ≥1 - <2 /≥2 - <3 / ≥3
Pulse rate	bpm	Increase and decrease	<100 / ≥100	≤0* / >0 - ≤20 / >20 - ≤40 / >40
SBP	mmHg	Increase and decrease	<140 / ≥140	≤0* / >0 - ≤20 / >20 - ≤40 / >40
DBP	mmHg	Increase and decrease	<90 / ≥90	≤0* / >0 - ≤ 20 / >20 - ≤40 / >40
Respiratory rate	breaths/min	Increase and decrease	<20 / ≥20	≤0* / >0 - ≤5 / >5 - ≤10 / >10

* This category will include the subjects with no changes or decrease/increase in the increase/decrease part of the table respectively.

A listing of maximum CFB and a listing of all vital signs data will be provided.

15.5 12-Lead Electrocardiogram (ECG)

The 12-lead ECG data will be listed and summarized for observed values, CFB and percent CFB values by treatment group using descriptive statistics:

- RR interval (ms)
- PR interval (ms)
- QRS (ms)
- QT (ms)
- Fridericia corrected QT (QTcF) (ms).

QTcF values will be categorized according to their calculated values into the categories

- ≤ 430 ms,
- > 430 – 450 ms,
- > 450 – 480 ms,
- > 480 – 500 ms,
- > 500 ms

and categorized according to their CFB into the categories

- ≤ 30 ms,
- > 30 – 60 ms,
- > 60 msec.

A listing of ECG quantitative values, morphological and rhythm results will be produced.

A shift table of rhythm results, from baseline to end of treatment, of the number and percentage of subjects for each category (Sinus rhythm, Atrial fibrillation, Other, Missing and Total) will be provided.

A shift table of morphological assessments, from baseline to worst on treatment observation of the 12-week treatment period, of the number and percentage of subjects for each interpretation category (Normal, Abnormal NCS, Abnormal CS, Missing, and Total) will also be provided.

15.6 Physical Examination

No summary table will be provided since physical examination findings during screening will be recorded as medical history events and findings during the trial as AEs.

15.7 Pregnancy Test

Results of pregnancy test (serum and urine beta human chorionic gonadotropin for women only) will be listed.

15.8 Serum IgG, IgA, IgM Levels

Boxplot of Serum IgG, IgA and IgM levels (mg/dL) by treatment group and time point will be provided. Descriptive statistics by treatment group and time point will be performed as well. Serum IgG, IgA and IgM levels data will be listed by treatment group, subject and time point (where applicable).

15.9 B (CD19+) Cell Count

A boxplot of B (CD19+) cell count (cells/µL) by treatment group and time point will be provided. Descriptive statistics by treatment group and time point will be performed as well. B (CD19+) cell count data will be listed by treatment group, subject and time point (where applicable).

15.10 Urinalysis Microscopic Evaluation

Urinalysis Microscopic Evaluation data will be listed by treatment group and time point (where applicable). Parameters are listed in Appendix 18.8.

15.11 HBV DNA

Hepatitis B Virus (HBV) deoxyribonucleic acid (DNA) polymerase chain reaction (PCR) testing results will be listed by treatment group and time point (where applicable). A shift table from baseline to worst post-baseline values including the categories < 20 IU/mL versus ≥ 20 IU/mL will be provided.

CCI

16 Analyses of Other Endpoints

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]


CCI

17 References

Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol. 2005;23(Suppl. 39):S100-8.

American College of Rheumatology Clinical Trial Priorities and Design Conference, July 22-23, 2010. Arthritis Rheum. 2011;63(8):2151-6.

Anderson J, Caplan L, Yazdany J, et al. Rheumatoid Arthritis Disease Activity Measures: American College of Rheumatology Recommendations for Use in Clinical Practice. Arthritis Care & Research. 2012; 64(5):640-47.

Bornkamp, B. et al (2007) Innovative Approaches for Designing and Analyzing Adaptive Dose-Ranging Trials, Journal of Biopharmaceutical Statistics, 17, 965-995

Bornkamp, B. et al (2009) MCPMod: An R Package for the Design and Analysis of Dose-Finding Studies, Journal of Statistical Software.

Brett, A., Guillard, G., and Bowes, M. (2018) Reporting and normalization of RAMRIQ data, and comparison with RAMRIS.

Bretz, F., Pinheiro, J.C., and Branson, M. (2005) Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics, 61, 738–748

Bykerk V, Massarotti E. The new ACR/EULAR remission criteria: rationale for developing new criteria for remission. Rheumatology. 2012;51(suppl 6):vi16-20.

Conaghan P, Emery P, Ostergaard M, et al. Assessment by MRI of inflammation and damage in rheumatoid arthritis patients with methotrexate inadequate response receiving golimumab: results of the GO-FORWARD trial. Ann Rheum Dis. 2011;70:1968-74.

Crowe B, Xia A, Berlin J, et al. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. Clin Trials. 2009;6:430-40.

DAS28 Website [internet]. Yousofa; 2016. [cited 2016 January 12]. Available from: http://www.das-score.nl/das28/en/difference-between-the-das-and-das28/how-to-measure-the-das28/how-to-calculate-the-das28/alternative-validated-formulae.html.

Dörner T, Weinblatt M, Van Beneden K, Dombrecht E, De Beuf K, Schoen P, et al. Poster FRI0239: results of a Phase 2B study of vobarilizumab, an anti-Interleukin-6 receptor nanobody, as monotherapy in patients with moderate to severe rheumatoid arthritis. Ann Rheum Dis (2017) 76(2):575

FACIT-Fatigue Scale: http://www.ser.es/wp-content/uploads/2015/03/FACIT-F_INDICE.pdf.

Farrington C.P, Manning G, Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. Statistics in Medicine 1990; 9: 1447-1454.

Felson D, Smolen J, Wells G, et al. American College of Rheumatology/European League Against Rheumatism Provisional Definition of Remission in Rheumatoid Arthritis for Clinical Trials. Ann Rheum Dis. 2011;70(3):404-13.

Felson D. A proposed revision to the ACR20: The hybrid measure of American College of Rheumatology response. Arthritis Rheum. 2007;57(2):193-202.

Fransen J, van Riel P. The Disease Activity Score and the EULAR response criteria. Clin Exp Rheumatol. 2005;23(Suppl. 39):S93-9.

Fries J, Spitz P, Kraines R, et al. Measurement of patient outcome in arthritis. Arthritis Rheum. 1980;23(2):137-45.

Garner W. Constructing Confidence Intervals for the Differences of Binomial Proportions in SAS. ®. Gilead Sciences, Inc., Foster City, CA. ABSTRACT. 2016.

Genovese M, Kremer J, Zamani O, et al. Baricitinib in patients with refractory rheumatoid arthritis. New England Journal of Medicine. 2016a Mar 31;374(13):1243–52.

Genovese M, Yang F, Ostergaard M, et al. Efficacy of VX-509 (decernotinib) in combination with a disease-modifying antirheumatic drug in patients with rheumatoid arthritis: clinical and MRI findings. Ann Rheum Dis. 2016b;75(11):1979-83.

Liu G.F., Wang J., Liu K. and Snavely D. B.; Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials; Stat Med. 2006 Apr 30; 25(8): 1275–1286.

Hochberg Y., Benjamini Y., (1990), "More Powerful Procedures for Multiple Significance Testing," *Statistics in Medicine*, 9, 811–818

Kenward, M. and Roger, J. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics 53, 983-997.

Lu K., Cochran-Mantel-Haenszel weighted Miettinen & Nurminen method for confidence intervals of the difference in binomial proportions from stratified 2x2 samples. JSM Proceedings 2008; Denver, CO: American Statistical Association.

Miettinen O., Nurminen M.; Comparative analysis of two rates; Stat Med. 1985 Apr-Jun; 4(2): 213–226.

Ostergaard M, Peterfy C, Conaghan P, et al. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. J Rheumatol. 2003;30(6):1385-6. Erratum in: J Rheumatol. 2004;31(1):198.

Ostergaard M, Peterfy CG, Bird P, et al. The OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging (MRI) Scoring System: Updated Recommendations by the OMERACT MRI in Arthritis Working Group. J Rheumatol. 2017;44;1706-12.

Peterfy C, et al. Ann Rheum Dis 2016;75:170–177. doi:10.1136/annrheumdis-2014-206015

CCI [redacted]

Smolen J, Aletaha D. Scores for all seasons: SDAI and CDAI. Clin Exp Rheumatol 2014;32(Suppl. 85): S75-9.

Smolen J, Breedveld F, Eberl G, et al. Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity. Arthritis Rheum. 1995;38(1):38-43.

US Department of Health and Human Services. Common Terminology Criteria for Adverse Events Version 4.0. 2009 (v4.03: June 14, 2010).

van Gestel A, Prevoo M, van 't Hof M, et al. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum. January 1996;39(1):34-40.

van Gestel A, Haagsma C, van Riel P. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. Arthritis Rheum. 1998;41(10):1845-50.

van Vollenhoven R, Felson D, Strand V, et al. American College of Rheumatology hybrid analysis of certolizumab pegol plus methotrexate in patients with active rheumatoid arthritis: Data from a 52-week phase III trial. Arthritis Care Res. 2011;63(1):128-34.

Vander Cruyssen B, Van Looy S, Wyns B, et al. DAS28 best reflects the physician's clinical judgment of response to infliximab therapy in rheumatoid arthritis patients: validation of the DAS28 score in patients under infliximab treatment. Arthritis Res Ther. 2005;7(5):R1063-71.

Ware J, Sherbourne C. The MOS 36-ltem Short-Form Health Survey (SF-36). Med Care. 1992;30(6):473-83.

Ware JE Jr, Kosinski M, Bjorner JB, Turner-Bowker DM, Gandek B, Maruish ME. User's manual for the SF-36 v2 Health Survey (2nd edition). Lincoln, RI: Quality Metric Incorporated, 2007.

Yan X, Shiowjen L, Li N. Missing data handling methods in medical device clinical trials. Journal of Biopharmaceutical Statistics, 19: 1085–1098, 2009.

18 Appendices

18.1 Definitions of Efficacy Endpoints

ACR20 - The ACR20 is a primary efficacy measure for which a subject must have at least 20% improvement in the following ACR Core Set values:

- TJC (being the number of joints assessed as "Pain/Tender only" or "Pain/Tender and Swollen" among the 68-joint count) and
- SJC (being the sum of joints assessed as "Swollen only" or "Pain/Tender and Swollen" among the 66-joint count) and
- An improvement of at least 20% in at least 3 of the following 5 assessments:
  - Subject's Global Assessment of Disease Activity
  - Subject's Assessment of Pain
  - Subject's Assessment of Physical Function as measured by the HAQ-DI
  - Physician's Global Assessment of Disease Activity
  - Acute phase reactant as measured by hsCRP.

In this trial, ACR20 response calculations will use the HAQ-DI for the subject's assessment of physical function and hsCRP as the measure of acute phase reactant.

Improvement of at least 20% of these assessments is defined as follows:

- TJC (68 joint count): a decrease of at least 20% of the number of tender joints compared to baseline
- SJC (66 joint count): a decrease of at least 20% of the number of swollen joints compared to baseline
- Subject's Global Assessment of Disease Activity: a decrease of at least 20% on the VAS compared to baseline. *Note: The scale has been scored from 100 [very well] to 0 [very poor] during the data collection. As a result, for analyses, the parameter is derived as 100 – the original value to correctly assess the subject's global assessment of disease activity and calculate the improvement from baseline.*
- Subject's Assessment of Pain: a decrease of at least 20% on the VAS compared to baseline.
- Subject's Assessment of Physical Function: a decrease of at least 20% of the HAQ-DI score compared to baseline.
- Physician's Global Assessment of Disease Activity: a decrease of at least 20% on the VAS compared to baseline
- Acute phase reactant: a decrease of at least 20% of hsCRP measurement compared to baseline

ACR50, ACR70 - ACR50 and ACR70 are defined in the same way as the ACR20 using at least 50% and 70% improvement, respectively.

DAS28-hsCRP - The DAS28-hsCRP is a measure of disease activity in 28 joints that consists of a composite numerical score of the following variables: TJC, SJC, hsCRP, and Patient's Global Assessment of Disease Activity (Vander Cruyssen 2005).

For DAS28-hsCRP, the 28 joints to be examined and assessed as tender or not tender for TJC and to be examined and assessed as swollen or not swollen for SJC include 14 joints on each side of the subject's body: 2 shoulders, 2 elbows, 2 wrists, 10 MCP joints, 2 interphalangeal joints of the thumb, 8 proximal interphalangeal joints, and 2 knees (Smolen 1995).

In this trial, the acute phase reactant for computing DAS28 is the hsCRP. DAS28-hsCRP will be derived using the following formula from the DAS28 website (DAS28 2016):

$$DAS28\text{-}hsCRP = 0.56 \times \sqrt{(TJC28)} + 0.28 \times \sqrt{(SJC28)} + 0.014 \times GH + 0.36 \times ln(hsCRP + 1) + 0.96$$

Where:

- TJC28 = 28 joint count for tenderness
- SJC28 = 28 joint count for swelling
- ln(hsCRP) = natural logarithm of hsCRP

GH = the general health component of the Disease Activity Score (DAS) (i.e., Patient's Global Assessment of Disease Activity).

ACR Hybrid - The ACR hybrid score was developed to provide a quasi-continuous score that can detect smaller changes between treatments than the categorical ACR20/50/70 scores by combining the continuous score of the mean improvement in the core set measures combining the ACR20, ACR50, and ACR70 scores with the mean percent change from Baseline in all 7 ACR response core set components (Felson 2007, American College of Rheumatology Clinical Trial Priorities and Design Conference 2010, van Vollenhoven 2011). ACR hybrid, along with the DAS28, is recommended by the FDA regulatory guidance as a way to support dose-response assessment. It is calculated according to Table 28:

Table 28: Scoring Method for the Hybrid American College of Rheumatology Response Measure

ACR Status	Mean % Change in Core Set Measures			
	< 20	≥ 20, < 50	≥ 50, < 70	≥ 70
Not ACR20	Mean % change	19.99	19.99	19.99
ACR20 but not ACR50	20	Mean % change	49.99	49.99
ACR50 but not ACR70	50	50	Mean % change	69.99
ACR70	70	70	70	Mean % change

Source: Felson 2007.

ACR = American College of Rheumatology, ACR20 = American College of Rheumatology 20% Response Criteria, ACR50 = American College of Rheumatology 50% Response Criteria, ACR70 = American College of Rheumatology 70% Response Criteria.

Note: 1) Calculate the average percentage change in core set measures. For each core set measure, subtract score after treatment from Baseline score and determine percentage improvement in each measure. Next, if a core set measure worsened by > 100%, limit that percentage change to 100% (a - 100% bound). Then average the percentage changes for all core set measures. 2) Determine whether the subject has achieved ACR20, ACR50, or ACR70. 3) Using the table above, obtain the Hybrid ACR response measure. To use the table, take the ACR20, ACR50, or ACR70 status of the subject (left column) and the mean percentage improvement in core set items; the Hybrid ACR score is where they intersect in the table.

EULAR Responder index – Assessments of subjects with RA by EULAR response criteria will be used to categorize subjects as having had no response, moderate response, good response, or any response (moderate + good responders) according to Table 29.

Table 29: Categorization of Subject Responses as No Response, Moderate Response, or Good Response Based on EULAR Response Criteria

Post-baseline Level of DAS28	Improvement Since Baseline in DAS28		
	> 1.2	≤ 1.2 and > 0.6	≤ 0.6
DAS28 ≤ 3.2	Good response		
3.2 < DAS28 ≤ 5.1		Moderate response	
DAS28 > 5.1			No response

Source: van Gestel 1996 and van Gestel 1998.
DAS28 = Disease Activity Score modified to include the 28 diarthrodial joint count.

SDAI – The SDAI is the numerical sum of 5 outcome parameters: TJC and SJC (based on a 28-joint assessment), the Patient's and the Physician's Global Assessments of disease activity (0–10 cm VAS), and level of CRP (expressed in units of mg/dL). The SDAI is a valid and sensitive assessment of disease activity (Aletaha 2005, Smolen 2014) and treatment response that is comparable with the DAS28 and ACR response criteria; it is easy to calculate and therefore a viable tool for day-to-day clinical assessment of RA treatment (Fransen 2005, Smolen 2014). The ACR/EULAR 2011 Index-based definition of remission is based on SDAI (SDAI ≤ 3.3), and is being assessed as a secondary endpoint (Felson 2011, Smolen 2014).

CDAI – The CDAI is a composite index (without acute-phase reactant) for assessing disease activity. CDAI is based on the simple summation of the count of SJC and TJC of 28 joints along with the Patient's and Physician's Global Assessments (0–10 cm VAS) for estimating disease activity. The greater advantage associated with CDAI is its potential to be employed in evaluation of subjects with RA consistently with close frequency and independently of any calculating device; therefore, it can essentially be used everywhere and anytime for disease activity assessment in subjects with RA (Aletaha 2005, Smolen 2014). Thresholds for disease severity grading are included in Anderson (Anderson 2012). Remission based on CDAI is also being assessed as a secondary endpoint.

EULAR/ACR Boolean Remission – Following the Boolean-based definition of remission of ACR/EULAR, at any time point, a subject must satisfy all of the following: tender joint count TJC ≤ 1, SJC ≤ 1, CRP ≤ 1 mg/dL and Patient's Global Assessment ≤ 1 (on a 0–10 scale). The Patient's Global Assessment must be elicited as follows: "Considering all of the ways your arthritis has affected you, how do you feel your arthritis is today? (0–10, where 0 = very well and 10 = very poorly)". The Boolean criteria appear more stringent than the DAS28 remission and have been specifically created for use in clinical trials (Bykerk 2012, Felson 2011).

18.2 HAQ-DI Scoring

The HAQ-DI is a subject-reported questionnaire that consists of 24 questions referring to 8 domains: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities.

The questionnaire scores the subject's self-perception (20 questions): 0=without any difficulty, 1=with some difficulty, 2=with much difficulty, 3=unable to do, and reports use of special aids or devices and/or the need for assistance from another person as indicated in Table 30.

Table 30: List of Questions from HAQ-DI Questionnaire

Domains	Questions
Dressing and grooming	1. Dress yourself, including tying shoelaces and doing buttons? 2. Shampoo your hair?
Arising	3. Stand up from a straight chair? 4. Get in and out of bed?
Eating	5. Cut your meat? 6. Lift a full cup or glass to your mouth? 7. Open a new milk carton?
Walking	8. Walk outdoors on flat ground? 9. Climb up five steps?
	Check any aids or devices that you usually use for any of the above activities (questions 1 to 9): Devices used for dressing (button hook, zipper pull, etc.) / Special or built up chair / Built up or special utensils / Cane / Walker / Crutches / Wheelchair / Other
	Check any categories for which you usually need help from another person (questions 1 to 9): Dressing and grooming / Arising / Eating / Walking
Hygiene	10. Wash and dry your body? 11. Take a tub bath? 12. Get on and off the toilet?
Reach	13. Reach and get down a 5 pound object (such as a bag of sugar) from just above your head? 14. Bend down to pick up clothing from the floor?
Grip	15. Open car doors? 16. Open jars which have been previously opened? 17. Turn faucets on and off?
Activities	18. Run errands and shop? 19. Get in and out of a car? 20. Do chores such as vacuuming or yard work?
	Check any aids or devices that you usually use for any of the above activities(questions 10 to 20): Raised toilet seat / bathtub seat / bathtub bar / long-handled appliances in bathroom / long-handled appliances for reach / jar opener (for jar previously opened) / Other
	Check any categories for which you usually need help from another person (questions 10 to 20): Hygiene / Reach / Gripping an opening things / Errands and Chores

The highest score for any component question of each of the eight domains determines the score for that domain. Each domain must have at least one question answered. Otherwise, the domain score is set to missing.

The non-missing domain scores are adjusted, based upon the patient's use of any aid, device or assistance (multiple aids, devices, or assistance could be checked). The relationship between aids, devices help from another person and the disability domain is shown in Table 31.

Table 31: HAQ-DI: Relationship between Aids, Devices Help from Another Person and the Disability Domain

Domains	Aids or devices	Help from another person
Dressing and grooming	Devices used for dressing / Other	Dressing and grooming
Arising	Special or built up chair / Other	Arising
Eating	Built up or special utensils / Other	Eating
Walking	Cane / Walker / Crutches / Wheelchair / Other	Walking
Hygiene	Raised toilet seat / bathtub seat / bathtub bar / long-handled appliances in bathroom / Other	Hygiene
Reach	long-handled appliances for reach / Other	Reach
Grip	jar opener / Other	Gripping and opening things
Activities	Other	Errands and Chores

If either aid or device and/or help from another person are checked for a domain or 'other' aid/device/assistance was needed, then the score for that domain is raised from 0 or 1 to 2 and unchanged if already scored at 2 or 3. If no aid, device and assistance were needed, the score for that domain will remain as raw score.

The disability index is the mean of the eight domain scores. If at least two of the domains are missing, the disability index cannot be obtained. If fewer than 2 of the domain scores are missing, divide the sum of the domains by the number of answered domains. The disability scores between 0 and 3, with higher score indicating greater disability.

18.3 36-Item Short Form Survey (SF-36) Scoring Instructions

The SF-36 is a validated 36-item, subject-reported indication of overall health status not specific to any age, disease, or treatment group (Ware 1992). The SF-36 has been extensively studied in RA, and a significant association has been shown between the physical functioning score of the SF-36 and the HAQ-DI score, as well as with other measures of disease activity and severity, and co-morbidities (Fries 1980).

The SF-36 includes multi-item scales measuring each of the following 8 health concepts: (1) physical functioning; (2) role limitations because of physical health problems; (3) bodily pain; (4) social functioning; (5) general mental health (psychological distress and psychological wellbeing); (6) role limitations because of emotional problems; (7) vitality (energy/fatigue); and (8) general health perceptions (see Table 32). These are summarized in two summary measures of physical and mental health: the PCS and MCS.

Questions in the standard version of the SF-36 refer to a 4-week time period. Scales are scored according to the Likert method. Lower scores equate to higher disability and higher scores equate to lower disability.

The SF-36v2 multi-item scales yield a health profile (8 scores) or can be aggregated into two summary scores, the PCS score and MCS score obtained through a linear combination of weighted transformed scores from the 8 subscales. PCS and MCS are standardised, with an average of 50 and a standard deviation of 10 in the general American population. PCS and MCS are computed only if all of the 8 scale scores are available. This Appendix details how these two scores should be calculated, as described in Ware 2007.

Table 32: SF-36 – Abbreviated Item Content for the SF-36v2 Health Domain Scales

Scale	Original item#	Item# in the MS200527-0060 study	Abbreviated Item Content
Physical Functioning (PF)	3a	3	Vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports
	3b	4	Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf
	3c	5	Lifting or carrying groceries
	3d	6	Climbing several flights of stairs
	3e	7	Climbing one flight of stairs
	3f	8	Bending, kneeling, or stooping
	3g	9	Walking more than a mile
	3h	10	Walking several hundred yards
	3i	11	Walking one hundred yards
	3j	12	Bathing or dressing oneself
Role-Physical (RP)	4a	13	Cut down the amount of time spent on work or other activities
	4b	14	Accomplished less than you would like
	4c	15	Limited in kind of work or other activities
	4d	16	Had difficulty performing work or other activities (eg, it took extra effort)
Bodily Pain (BP)	7	21	Intensity of bodily pain
	8	22	Extent pain interfered with normal work
General Health (GH)	1	1	Is your health: excellent, very good, good, fair, poor
	11a	33	Seem to get sick a little easier than other people
	11b	34	As healthy as anybody I know
	11c	35	Expect my health to get worse
	11d	36	Health is excellent
Vitality (VT)	9a	23	Feel full of life
	9e	27	Have a lot of energy
	9g	29	Feel worn out
	9i	31	Feel tired
Social Functioning (SF)	6	20	Extent health problems interfered with normal social activities
	10	32	Frequency health problems interfered with social activities
Role-Emotional (RE)	5a	17	Cut down the amount of time spent on work or other activities
	5b	18	Accomplished less than you would like
	5c	19	Did work or other activities less carefully than usual
Mental Health (MH)	9b	24	Been very nervous
	9c	25	Felt so down in the dumps that nothing could cheer you up
	9d	26	Felt calm and peaceful
	9f	28	Felt downhearted and depressed
	9h	30	Been happy
Self-Evaluated Transition (SET)	2	2	How health is now compared to 1 year ago

Step 1: Recoding Item Response Values

Some of the SF-36v2 items will be re-coded so that across all questions, a higher score will indicate a better health state. Questions 2, 3a-3j, 4a-4d, 5a-5c, 9b, 9c, 9f, 9g, 9i, 10, 11a, 11c will be scored as recorded; the other questions will have the scores transformed as shown in Table 33. If multiple answers are given to the same item, then the item score will be left as missing.

Table 33: SF-36 – Recoding

Question	Original code and re-code response					
Question number: 1						
Original response	1	2	3	4	5	
Re-coded response	5	4.4	3.4	2	1	
Questions numbers: 6, 11b, 11d						
Original response	1	2	3	4	5	
Re-coded response	5	4	3	2	1	
Question number: 7						
Original response	1	2	3	4	5	6
Re-coded response	6	5.4	4.2	3.1	2.2	1
Question number: 8 (if question number 7 is answered)						
Original response to #8	1	1	2	3	4	5
Original response to #7	1	2-6	1-6	1-6	1-6	1-6
Re-coded response	6	5	4	3	2	1
Question number: 8 (if question number 7 is NOT answered)						
Original response	1	2	3	4	5	
Re-coded response	6	4.75	3.5	2.25	1	
Questions number: 9a, 9d, 9e, 9h						
Original response	1	2	3	4	5	
Re-coded response	5	4	3	2	1	

Step 2: Determining Health Domain Scale Scores (0-100 Scores)

After item recoding, a total raw score is computed for each health domain scale. The total raw score is the simple algebraic sum of the final response values for all items in a given scale, as shown in Table 34. The total raw score for each scale is transformed to a 0-100 scale score using the following formula:

$$Total\ raw\ score = 100 \times \frac{(\text{Raw score} - \text{Lowest possible raw score})}{\text{Possible raw score range}}$$

Table 34: SF-36 – Values used in transforming SF-36v2 Health Survey Health Domain Scale Total Raw Scores on the 0-100 Scale

Scale	Sum of Final Response Values	Lowest and highest possible total raw scores	Possible total raw score range
PF	3a+3b+3c+3d+3e+3f+3g+3h+3i+3j	10, 30	20
RP	4a+4b+4c+4d	4, 20	16
BP	7+8	2, 12	10
GH	1+11a+11b+11c+11d	5, 25	20
VT	9a+9e+9g+9i	4, 20	16
SF	6+10	2, 10	8
RE	5a+5b+5c	3, 15	12
MH	9b+9c+9d+9f+9h	5, 25	20

Raw and transformed scale scores are not calculated for the Reported Health Transition (HT) item.

As recommended by the developers of the questionnaire, missing item responses will be treated using the "Half-scale rule", which states that a score can be calculated if the respondent answers at least 50% of the items in a multi-item scale. In such cases, the missing item data will be replaced by the mean of the answered items of its scale. If more than 50% of the items are missing within a scale, the scale score will be missing.

Step 3: Calculating Normalized Health Domain Scores
The normalized scale scores will then be calculated using the following formulas:

$$Health\ Domain\ Z_{\text{score}} = {(Health\ Domain_{0-100\ score} - a)}/{b}$$

$$Normalized\ Health\ Domain\ Score = 50 + (Health\ Domain\ Z_{\text{score}} \times 10)$$

where a and b are the Mean and Standard Deviation of the Health Domain scale in the 1998 U.S. general population as shown in Table 35.

Table 35: SF-36 – 1998 General US Population Means and Standard Deviations used to Calculate Normalized Health Domain Scores

Health Domain Scales	Mean	Standard Deviation
PF	83.29094	23.75883
RP	82.50964	25.52028
BP	71.32527	23.66224
GH	70.84570	20.97821
VT	58.31411	20.01923
SF	84.30250	22.91921
RE	87.39733	21.43778
MH	74.98685	17.75604

The advantages of the normalization of the eight health domain scales are that results for one health domain scale can be meaningfully compared with those from the other scales and that domain scores have a direct interpretation in relation to the distribution of scores in the 1998 U.S. general population.

Step 4: Scoring the Physical and Mental Component Summary Measures
The Physical Component Summary (PCS) and Mental Component Summary (MCS) measures are scored using a three-step procedure:

1. First, the 8 health domain scale scores are standardized using means and standards deviations from the 1998 U.S. general population (see Table 35).

2. Second, these Z-scores are aggregated using weights (factor score coefficient) from the 1990 U.S general population.

3. Third, aggregate PCS and MCS scores are standardized by multiplying the standardized scale by 10 and adding 50.

U.S. general population statistics used in the standardization and in the aggregation of SF-36v2 Health Survey health domain scale scores are presented in Table 36.

Table 36: SF-36 – Factor Score Coefficients used to Calculate PCS and MCS Scores for the SF-36v2

	Summary component measure factor score coefficients	
Scales	**PCS**	**MCS**
PF	0.42402	-0.22999
RP	0.35119	-0.12329
BP	0.31754	-0.09731
GH	0.24954	-0.01571
VT	0.028877	0.23534
SF	-0.00753	0.26876
RE	-0.19206	0.43407
MH	-0.22069	0.48581

Example:

Let's consider the following answers from a subject:

Table 37: SF-36 – Numerical Example – Raw Data

#	Item description	Answer
1	In general, would you say your health is:	1 – Excellent
2	Compared to one year ago, how would you rate your health in general now?	2 – Somewhat better now than one year ago
3	Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	2 – Yes, limited a little
4	Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	2 – Yes, limited a little
5	Lifting or carrying groceries	2 – Yes, limited a little
6	Climbing several flights of stairs	2 – Yes, limited a little
7	Climbing one flight of stairs	2 – Yes, limited a little
8	Bending, kneeling or stooping	1 – Yes, limited a lot
9	Walking more than a mile	2 – Yes, limited a little
10	Walking several hundred yards	2 – Yes, limited a little
11	Walking one hundred yards	2 – Yes, limited a little
12	Bathing or dressing yourself	2 – Yes, limited a little
13	Cut down on the amount of time you spent on work or other activities	2 – Most of the time
14	Accomplished less than you would like	3 – Some of the time
15	Were limited in the kind of work or other activities	2 – Most of the time
16	Had difficulty performing the work or other activities (for example, it took extra effort)	3 – Some of the time
17	Cut down on the amount of time you spent on work or other activities	2 – Most of the time
18	Accomplished less than you would like	3 – Some of the time

19	Don't do work or other activities as carefully as usual	2 – Most of the time
20	During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?	2 – Slightly
21	How much bodily pain have you had during the past 4 weeks?	3 – Mild
22	During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?	2 – A little bit
23	Did you feel full of life?	4 – A little of the time
24	Have you been a very nervous person?	3 – Some of the time
25	Have you felt so down in the dumps that nothing could cheer you up?	2 – Most of the time
26	Have you felt calm and peaceful?	4 – A little of the time
27	Did you have a lot of energy?	3 – Some of the time
28	Have you felt downhearted and low?	2 – Most of the time
29	Did you feel worn out?	4 – A little of the time
30	Have you been a happy person?	3 – Some of the time
31	Did you feel tired?	2 – Most of the time
32	During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?	4 – A little of the time
33	I seem to get ill more easily than other people	3 – Don't know
34	I am as healthy as anybody I know	2 – Mostly true
35	I expect my health to get worse	3 – Don't know
36	My health is excellent	2 – Mostly true

After having recoded the answers, the 8 domain scores are equal to:

Table 38: SF-36 – Numerical Example – Domain Scores

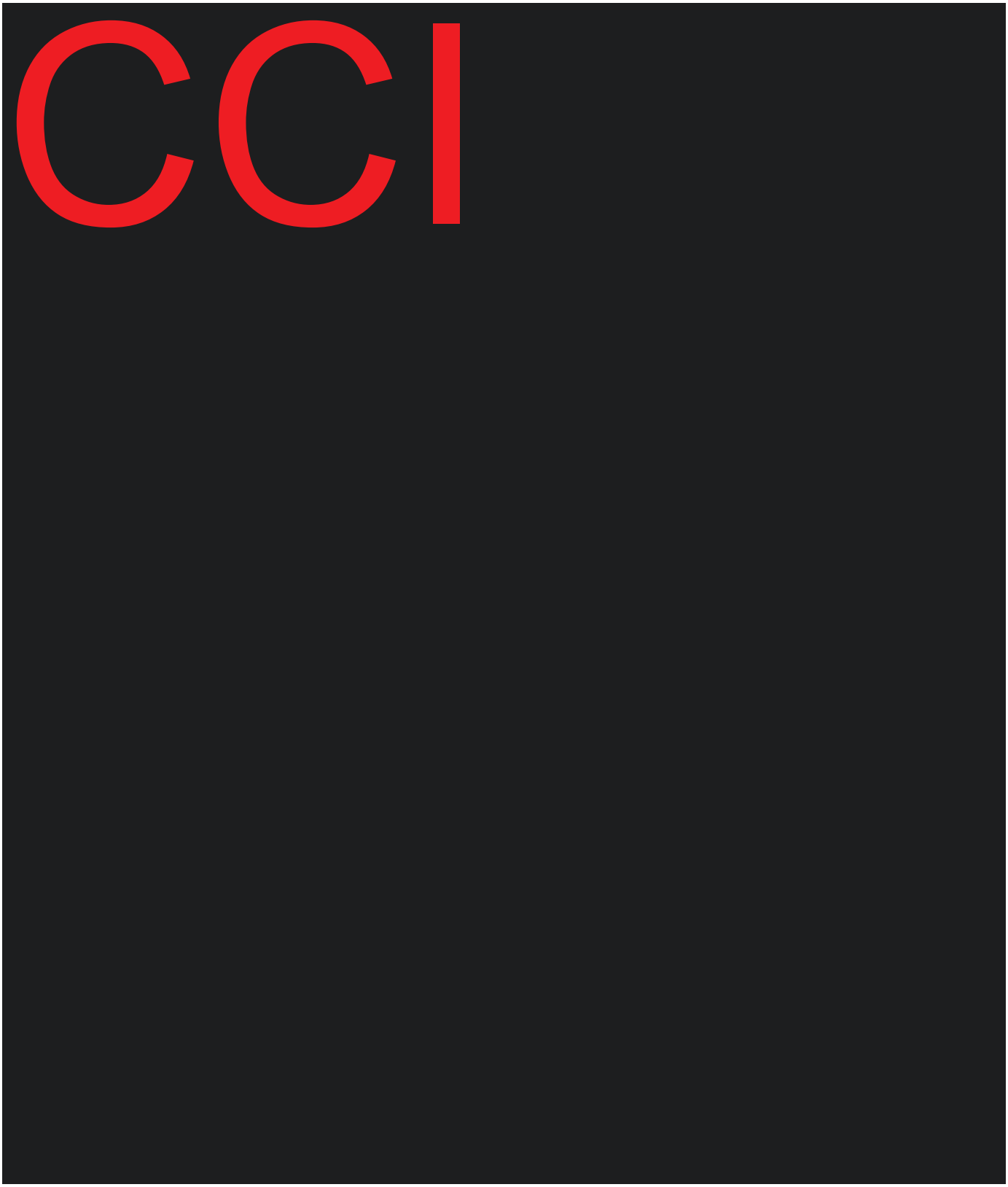
	Raw Score	Raw Score$_{0\text{-}100}$	Health Domain Z-Score	Normalized Health Domain Z-Score
PF	19	100 x (19-10) /20 = 45	45 - 83.29094 / 23.75883 = -1.61	-1.61 x 10 + 50 = 33.9
RP	10	100 x (10-4) /16 = 37.5	37.5 - 82.50964 / 25.52028 = -1.76	-1.76 x 10 + 50 = 32.4
BP	8.2	100 x (8.2-2) /10 = 62	62 - 71.32527 / 23.66224 = -0.39	-0.39 x 10 + 50 = 46.1
GH	19	100 x (19-5) /20 = 70	70 - 70.84570 / 20.97821 = -0.04	-0.04 x 10 + 50 = 49.6
VT	11	100 x (11-4) /16 = 43.75	43.75 - 58.31411 / 20.01923 = -0.73	-0.73 x 10 + 50 = 42.7
SF	8	100 x (8-2) / 8 = 75	75 - 84.30250 / 22.91921 = -0.41	-0.41 x 10 + 50 = 45.9
RE	7	100 x (7-3) / 15 = 27	27 - 87.39733 / 21.43778 = -2.82	-2.82 x 10 + 50 = 21.8
MH	12	100 x (12-5) / 20 = 35	35 - 74.98685 / 17.75604 = -2.25	-2.25 x 10 + 50 = 27.5

Finally, PCS and MCS are provided below using Table 36:

- PCS = -1.61 x 0.42402 + … + -2.25 x -0.22069 = -0.41
- Normalized PCS = 10 x PCS + 50 = 45.9
- MCS = -1.61 x -0.22999 + … + -2.25 x 0.48581 = -2.62
- Normalized MCS = 10 x MCS + 50 = 23.8

18.4 FACIT-Fatigue Scoring Instructions

FACIT-Fatigue is a 13-item questionnaire that assesses self-reported fatigue and its impact upon daily activities and function (Wolfe 1996). It uses a 5-point Likert-type scale (0 = not at all; 1 = a little bit; 2 = somewhat; 3 = quite a bit; 4 = very much). As each of the 13 items of the FACIT-Fatigue scale ranges from 0–4, the range of possible scores is 0–52, with 0 being the worst possible score and 52 the best. To obtain the 0–52 score, each negatively worded item response is recoded so that 0 is a bad response and 4 is good response. All responses are added with equal weight to obtain the total score.


CCI

18.5 SAS/R Sample Code

This Appendix details how the statistical methods will be implemented with either SAS or R software applications.

Logistic regression: PROC LOGISTIC with SAS

The logistic regression can be implemented using the following SAS code:

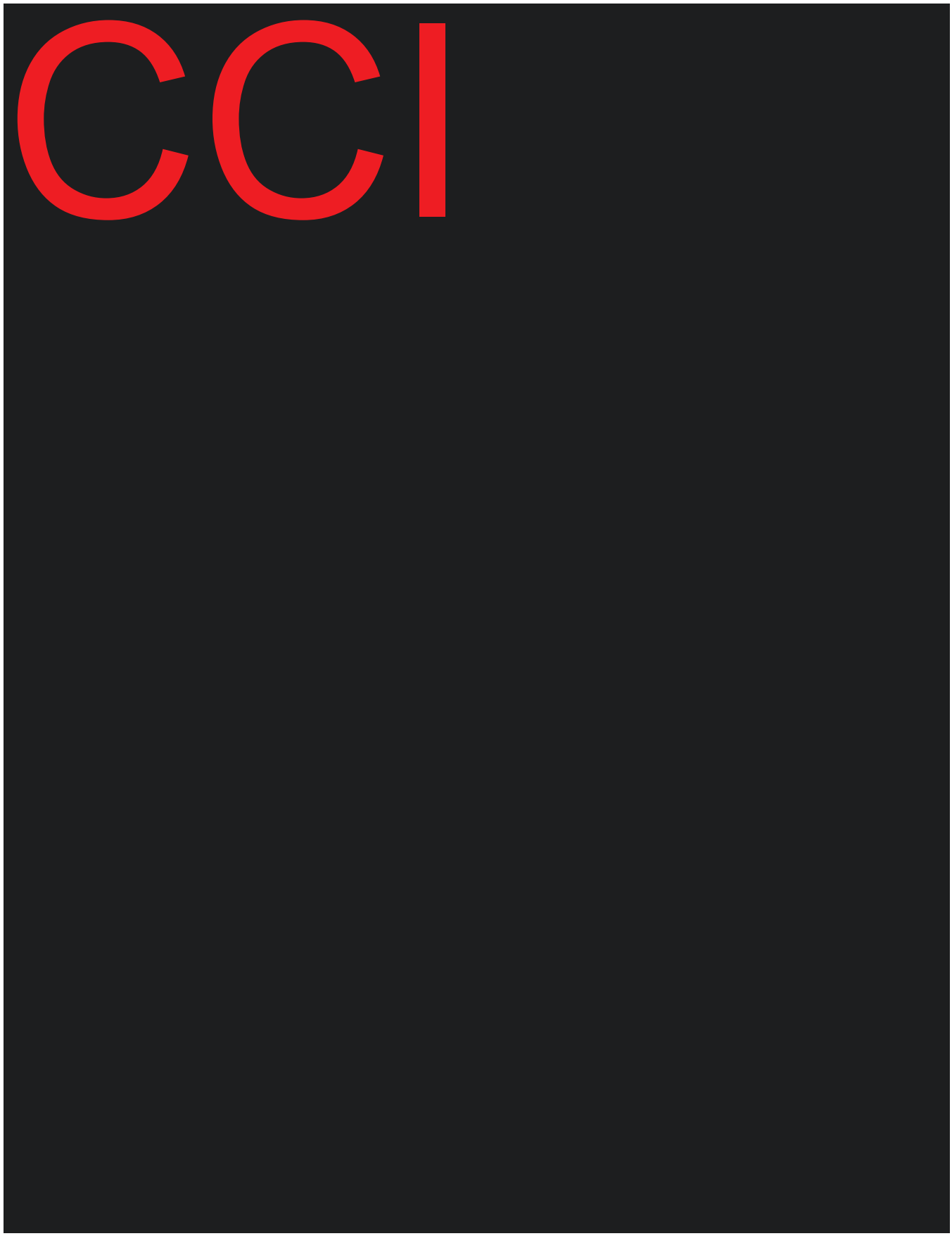
```
proc logistic data=<dataset>;
       class TreatmentGroup(ref='Placebo') Region;
       model AVALC(event="Y") = TreatmentGroup Region / alpha=0.05;
       ods output ParameterEstimates=ParameterEstimates OddsRatios=OddsRatios ;
run ;
```

CCI

CCI

CCI

CCI


CCI

CCI

MMRM: PROC MIXED with SAS

A mixed model can be performed with SAS using the following code below. This sample code assumes the treatment group is assigned in the following order:

- 1 = Placebo
- 2 = Evobrutinib 25 mg QD
- 3 = Evobrutinib 75 mg QD

- 4 = Evobrutinib 50 mg BID


CCI


CCI

Table 43: Time Windows for Urinalysis

Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	12-week treatment period	1	Day 1
		[2 ; 70)	Week 8 – Day 56
		[70 ; 91)	Week 12 – Day 84
• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	OLE period	1	Day 1 – OLE
		[2 ; 45)	Month 1 – OLE
		[45 ; 75)	Month 2 – OLE
		[75 ; 105)	Month 3 – OLE
		[105 ; 150)	Month 4 – OLE
		[150 ; 225)	Month 6 – OLE
		[225 ; 315)	Month 9 – OLE
		[315 ; 367)	Month 12 – OLE

Table 44: Time Windows for Vital Signs, Serum IgA, IgG, IgM, Total B cell counts, Plasma/serum collection for CCI , Hematology and Biochemistry

Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	12-week treatment period	1	Day 1
		[2 ; 21)	Week 2 – Day 14
		[21 ; 42)	Week 4 – Day 28
		[42 ; 70)	Week 8 – Day 56
		[70 ; 91)	Week 12 – Day 84
• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	OLE period	1	Day 1 – OLE
		[2 ; 45)	Month 1 – OLE
		[45 ; 75)	Month 2 – OLE
		[75 ; 135)	Month 3 – OLE
		[135 ; 225)	Month 6 – OLE
		[225 ; 315)	Month 9 – OLE
		[315 ; 367)	Month 12 – OLE

Note: Plasma/serum collection for CCI are not collected during OLE

Table 45: Time Windows for B cell subsets

Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• Placebo	12-week treatment period	1	Day 1
• Evobrutinib 25 mg QD		[2 ; 56)	Week 4 – Day 28
• Evobrutinib 75 mg QD • Evobrutinib 50 mg BID		[56 ; 91)	Week 12 – Day 84

Table 46: Time Windows for ECG

Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	12-week treatment period	1	Day 1
		[2 ; 70)	Week 8 – Day 56
		[70 ; 91)	Week 12 – Day 84
• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	OLE period	1	Day 1 – OLE
		[2 ; 270)	Month 6 – OLE
		[270 ; 367)	Month 12 – OLE

Table 47: Time Windows for ACR components

Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	12-week treatment period	1	Day 1
		[2 ; 21)	Week 2 – Day 14
		[21 ; 42)	Week 4 – Day 28
		[42 ; 70)	Week 8 – Day 56
		[70 ; 98]	Week 12 – Day 84
• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	OLE period	1	Day 1 – OLE
		[2 ; 270)	Month 6 – OLE
		[270 ; 367)	Month 12 – OLE

Table 48: Time Windows for SF-36, FACIT-Fatigue and CCI

Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	12-week treatment period	1	Day 1
		[2 ; 42)	Week 4 – Day 28
		[42 ; 70)	Week 8 – Day 56
		[70 ; 98]	Week 12 – Day 84
• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	OLE period	1	Day 1 – OLE
		[2 ; 270)	Month 6 – OLE
		[270 ; 367)	Month 12 – OLE


CCI


CCI

Table 51: Time Windows for MRI assessments

Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	12-week treatment period	[-35 ; 1]	Day 1
		[2 ; 56)	Week 4 – Day 28
		[56 ; 98]	Week 12 – Day 84

Table 52: Time Windows for HBV DNA assessments

Treatment groups	Period covered	Interval (days) [Date of assessment – Date of first dose + 1]	Analysis Visit
• Placebo • Evobrutinib 25 mg QD • Evobrutinib 75 mg QD • Evobrutinib 50 mg BID	12-week treatment period	[1 ; 42)	Week 4 – Day 28
		[42 ; 70)	Week 8 – Day 56
		[70 ; 91)	Week 12 – Day 84
• Placebo/Evobrutinib 50 mg BID • Evobrutinib 25 mg QD/Evobrutinib 50 mg BID • Evobrutinib 75 mg QD/Evobrutinib 50 mg BID • Evobrutinib 50 mg BID/Evobrutinib 50 mg BID	OLE period	1	Day 1 – OLE
		[2 ; 45)	Month 1 – OLE
		[45 ; 75)	Month 2 – OLE
		[75 ; 135)	Month 3 – OLE
		[135 ; 225)	Month 6 – OLE
		[225 ; 315)	Month 9 – OLE
		[315 ; 367)	Month 12 – OLE

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

CCI

CCI

[redacted]

18.8 Laboratory Parameters to be summarized in the TLFs

Table 53: Laboratory Parameters to be Summarized in the TLFs

	Names of Clinical Safety Laboratory Evaluations in Protocol version 3.0	If gradable parameter, corresponding evaluation names in NCI-CTCAE 4.03	Worst on treatment value based on normal range
Biochemistry	Albumin	Hypoalbuminemia	LOW
	Aspartate aminotransferase	Aspartate aminotransferase increased	HIGH
	Alanine aminotransferase	Alanine aminotransferase increased	HIGH
	Alkaline phosphatase	Alkaline phosphatase increased	HIGH
	γ-Glutamyl-transferase	GGT increased	HIGH
	Lactate dehydrogenase		HIGH
	Bilirubin (total)	Blood bilirubin increased	HIGH
	Protein (total)		LOW
	Creatinine	Creatinine increased	HIGH
	Amylase	Serum amylase increased	HIGH
	Lipase	Lipase increased	HIGH
	Total carbon dioxide		LOW
	Blood urea nitrogen		HIGH
	Glucose	Hyperglycemia	HIGH
	Glucose	Hypoglycemia	LOW
	Sodium	Hypernatremia	HIGH
	Sodium	Hyponatremia	LOW
	Potassium	Hyperkalemia	HIGH
	Potassium	Hypokalemia	LOW
	Chloride		NA
	Calcium	Hypercalcemia	HIGH
	Calcium	Hypocalcemia	LOW
	Magnesium	Hypermagnesemia	HIGH
	Magnesium	Hypomagnesemia	LOW
	Phosphate	Hypophosphatemia	LOW
	Uric Acid		HIGH
Hematology	Hematocrit		LOW/HIGH
	Hemoglobin	Hemoglobin increased	HIGH
	Hemoglobin	Anemia	LOW
	Hemoglobin A1C		HIGH
	Red blood cell count		NA
	Mean corpuscular volume		NA
	Mean corpuscular hemoglobin		NA
	Mean corpuscular hemoglobin concentration		NA
	Reticulocyte count		NA

	Platelet count	Platelet count decreased	LOW
	White blood cell count	Leukocytosis	HIGH
	White blood cell count	White blood cell decreased	LOW
	White blood cell differentials and absolute counts: Basophils		NA
	White blood cell differentials and absolute counts: Eosinophils		NA
	White blood cell differentials and absolute counts: Lymphocytes	Lymphocyte count increased	HIGH
	White blood cell differentials and absolute counts: Lymphocytes	Lymphocyte count decreased	LOW
	White blood cell differentials and absolute counts: Monocytes		NA
	White blood cell differentials and absolute counts: Neutrophils	Neutrophil count decreased	LOW
Urinalysis	pH		NA
	Nitrite		NA
	Protein		NA
	Blood		NA
	Glucose		NA
	Ketones bodies		NA
	Urobilinogen		NA
	Bilirubin		NA
	Specific Gravity		NA
Coagulation	International normalized ratio		NA
	Partial thromboplastin time		NA
Urine Microscopy	White blood cells		HIGH
	Red blood cells		HIGH
	Casts (Granular and Hyaline)		NA
Lipid Panel	HDL-C		
	LDL-C		
	Total Cholesterol	Total Cholesterol increased	HIGH
	Triglycerides	Hypertriglyceridemia	HIGH
Reflex Testing for HBV DNA	HBV DNA		HIGH
Hepatic Panel	International normalized ratio		
	Partial thromboplastin time	Partial thromboplastin time prolonged	HIGH
	Fibrinogen	Fibrinogen decreased	LOW
	hsCRP		HIGH
	Hepatitis serology: anti hepatitis A Virus IgG, anti-hepatitis A Virus IgM, HBsAg, anti hepatitis B core antigen, anti HBsAg, anti HCV, anti hepatitis E Virus IgG and IgM, anti viral capsid antigen IgG and IgM, anti early antigen IgG, anti Epstein		NA

	Barr nuclear antigen IgG, anti cytomegalovirus IgG and IgM		
	Antinuclear antibody, anti smooth muscle antibody, antibody to liver kidney microsomal antibody		NA
	Albumin	Hypoalbuminemia	LOW

18.9 Justification for the MCPMOD parameters

The primary endpoint of the MS200527-0086 study, total number of Gd+ T1 lesions at Weeks 12, 16, 20, and 24, was significantly lower in the evobrutinib 75 mg once daily and evobrutinib 75 mg twice daily groups when compared to placebo, and the primary objective of the study was met for these 2 treatment groups.

A monotonic dose-response relationship between ordered dose groups (placebo, evobrutinib 25 mg once daily, evobrutinib 75 mg once daily, and evobrutinib 75 mg twice daily) and ordered categories of Gd+ T1 lesion count for Weeks 12, 16, 20, and 24, was assessed via the Jonckheere-Terpstra trend test. The results demonstrated that the location of the lesion count distribution is ordered (from largest value to smallest) according to increasing dose group ($p<0.0001$). Results are provided in Table .

Table 54: Results from Jonckheere-Terpstra trend test

Response	Category	Placebo N=53 (100%)	Evobrutinib 25 mg QD N=50 (100%)	Evobrutinib 75 mg QD N=51 (100%)	Evobrutinib 75 mg BID N=53 (100%)	Tecfidera N=54 (100%)
Number of subjects with total number Gd+ T1 lesions, n (%)	0	20 (37.7)	24 (48.0)	37 (72.5)	36 (67.9)	31 (57.4)
	[1, 2)	6 (11.3)	8 (16.0)	4 (7.8)	8 (15.1)	11 (20.4)
	[2, 3)	6 (11.3)	6 (12.0)	2 (3.9)	2 (3.8)	3 (5.6)
	≥ 3	21 (39.6)	12 (24.0)	8 (15.7)	7 (13.2)	9 (16.7)
Jonckheere-Terpstra trend test*	Statistic (Z)	-4.08				
	P-Value (asymptotic)	<.0001				

Tecfidera treatment group is not included in inferential analyses.

Prior to defining suitable categories for Gd+ T1 lesion count for wks 12, 16, 20, 24, missing scan data for a subject will be imputed using average of available scan data for that subject. Scans collected within 3 weeks of high dose corticosteroid use are considered missing. If a subject lacks any post-baseline evaluable MRI assessment, lesion count is imputed as the product of "4" and the median lesions per scan, where the median is among subjects with available post-baseline scans, who are in the same treatment group, and who have the same value for the categorical baseline covariate.

*Jonckheere-Terpstra monotonic trend test: unstratified test for decreasing ordinal total number Gd+ T1 lesions with increasing ordinal dose.

In addition to the nonparametric test of monotonic dose-response, a model-based test of linear trend in log lesion rate with increasing dose suggests that the log lesion rate decreased linearly with increasing dose group order. Results from this model-based test are provided in Table .

Table 55: Results from negative binomial and Poisson models

	Placebo N=53 (100%)	Evobrutinib 25 mg QD N=50 (100%)	Evobrutinib 75 mg QD N=51 (100%)	Evobrutinib 75 mg BID N=53 (100%)	Tecfidera N=54 (100%)
Total number of Gd+ T1 lesions at Weeks 12, 16, 20, 24					
Subjects with 0 lesions, n (%)	20 (37.7)	24 (48.0)	37 (72.5)	36 (67.9)	31 (57.4)
1 lesion, n (%)	8 (15.1)	8 (16.0)	4 (7.8)	8 (15.1)	11 (20.4)
2 lesions, n (%)	5 (9.4)	6 (12.0)	2 (3.9)	4 (7.5)	3 (5.6)
3 lesions, n (%)	3 (5.7)	1 (2.0)	1 (2.0)	2 (3.8)	1 (1.9)
Total number of scans					
Subjects with 1 scan, n (%)	2 (3.8)	2 (4.0)	2 (3.9)	4 (7.5)	1 (1.9)
2 scans, n (%)	1 (1.9)	1 (2.0)	0 (0.0)	1 (1.9)	2 (3.7)
3 scans, n (%)	8 (15.1)	12 (24.0)	2 (3.9)	7 (13.2)	6 (11.1)
4 scans, n (%)	42 (79.2)	35 (70.0)	47 (92.2)	41 (77.4)	45 (83.3)
Gd+ T1 lesion rate (subject level)					
n (%)	53 (100.0)	50 (100.0)	51 (100.0)	53 (100.0)	54 (100.0)
Mean ±SD	1.02 ±1.439	1.31 ±3.130	0.42 ±1.173	0.33 ±0.961	1.45 ±7.293
Median	0.50	0.25	0.00	0.00	0.00
Q1; Q3	0.00; 1.50	0.00; 0.67	0.00; 0.25	0.00; 0.25	0.00; 0.25
Min; Max	0.00; 6.00	0.00; 19.00	0.00; 6.75	0.00; 6.25	0.00; 53.33
Lesion RR based on NB model (*) [95% CI]		1.45 [0.72, 2.91]	0.30 [0.14, 0.63]	0.44 [0.21, 0.93]	
P-value (Evobrutinib vs placebo)		0.2947	0.0015	0.0313	
Hochberg adjusted P-value		0.3157	0.0046	0.0648	
Linear trend test P-value					
Based on Likelihood Ratio Test	0.0011				
Based on Wald Test	0.0007				
Conditional Lesion RR based on longitudinal Poisson model (**) [95% CI]		1.44 [0.81, 2.58]	0.49 [0.22, 1.09]	0.31 [0.13, 0.77]	
P-value (Evobrutinib vs placebo)		0.2173	0.0793	0.0112	

Tecfidera treatment group is not included in inferential analyses.

(*) Negative binomial model for lesion count (summed over scans) includes treatment and covariate presence/absence of Gd+ T1 lesions at baseline, with offset equal to the log of number of scans performed. A lesion rate ratio (RR) < 1 favors the active treatment. Subjects with missing scans or discontinuing early are analyzed according to the available number of scans and lesion counts. Scans collected within 3 weeks of high dose corticosteroid use are considered missing. Linear trend test assesses linearly decreasing trend of log lesion rate with increasing dose.

(**) Poisson Generalized Linear Mixed Model (GLMM) for lesion count at a single scan includes fixed effects for treatment and scan time point, a covariate for presence/absence of Gd+ T1 lesions at baseline, zero offset, and CS covariance structure for repeated measures. Same handling of early discontinuers, and consideration of scans collected within 3 weeks of high dose corticosteroid use as in first model. Lesion RR is conditional on random effects.

- Lesion rate on the subject level is the Total # of Gd+ T1 lesions for the subject (summed over all scans), divided by the Total # of scans for the subject.

- Summary and inferential statistics are based on data where imputation is performed only for those subjects in the mITT analysis set who are missing all post-baseline scans.

Therefore, a dose-response relationship was observed. However, the negative binomial model showed a lower lesion rate ratio in the evobrutinib 75 mg once daily group when adjusted on the number of scans, but the conditional lesion rate ratio based on longitudinal Poisson model showed an ordered dose-response relationship, as indicated in Table .

As a result, since the evobrutinib 75 mg once daily dose is common from both studies MS200527-0060 and MS200527-0086, it has been decided to use the conditional lesion rate ratio based on longitudinal Poisson model to assess the percentage of maximum effect observed. As the maximum effect was obtained in the evobrutinib 75 mg twice daily group with a lesion rate ratio of 0.31, the percentage of maximum effect observed for the evobrutinib 75 mg once daily group was estimated as follows:

$$\%\ of\ maximum\ effect = \frac{0.51}{0.69} \times 100 = 74\%$$

Based on this percentage of maximum effect in the evobrutinib 75 mg once daily group, the parameters of the models for the MCPMOD were assessed using the following R statements:

```
R> ## load package
R> library(DoseFinding)

R> # assign doses in mg
R> doses <- c(0, 25, 75, 100)

R> emax<-guesst(d=75, p=0.74, model="emax")
R> emax[1]

R> quad<-guesst(d=75, p=1, "quadratic")
R> quad[1]

R> logi<-guesst(d=c(75,100), p=c(0.74,0.9999999), "logistic")
R> logi[1]
R> logi[2]
```

It provides the parameters described in Table 14.

CCI

ELECTRONIC SIGNATURES

Document: ctp-ms200527-0060-iap-v-2-0

Signed By	Event Name	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'GMT'Z)
PPD	Task Completed (Approval eSign): Approved	Technical Approval	15-Oct-2019 08:31