Integrated Analysis Plan

Clinical Trial Protocol Identification No.	MS200527-0086
Title:	A Randomized, Double-Blind, Placebo-Controlled Phase II Study of Evobrutinib with a Parallel, Open-Label, Active Control Group (Tecfidera), in Patients with Relapsing Multiple Sclerosis to Evaluate Efficacy, Safety, Tolerability, Pharmacokinetics, and Biological Activity.
Trial Phase	II
Investigational Medicinal Product(s)	Evobrutinib
Clinical Trial Protocol Version	05 July 2016 / Version 1.0

Integrated Analysis Plan Authors

Coordinating Author	
Biostatistics, Merck	
Function	**Author(s) / Data Analyst(s)**
Biostatistician, PPD	PPD
Pharmacokineticist, PPD	PPD
Clinical Pharmacokineticist, PPD	PPD

Integrated Analysis Plan Date and Version	14 February 2018 / Version 1.0

Integrated Analysis Plan Reviewers

Merck/EMD-Serono Reviewers:

Function	Name
Lead Program Statistician	PPD
Global Head Biostat Neurology/Immuno	PPD
Lead Statistical Programmer	PPD
Global Patient Safety Product Lead	PPD
Global Patient Safety	PPD
Medical Writing Lead	PPD
Medical Director	PPD
Principal Clinical Data Sciences Lead	PPD
Trial Data Manager	PPD
Clinical Biomarkers and Companion Diagnostics Lead	PPD
Clinical Pharmacology Consultant	PPD
PPD	PPD
PPD	PPD
Clinical Trial Lead	PPD

PPD Reviewers:

Function	Name
Statistical Scientist	PPD
Medical Responsible	PPD
Lead Medical Writer	PPD
Lead Statistical Programmer	PPD
Pharmacokineticist	PPD
Pharmacokineticist	PPD
Statistical Scientist	PPD

Signature Page

Statistical Analysis Plan: MS200527-0086
A Randomized, Double-Blind, Placebo-Controlled Phase II Study of Evobrutinib with a Parallel, Open-Label, Active Control Group (Tecfidera), in Patients with Relapsing Multiple Sclerosis to Evaluate Efficacy, Safety, Tolerability, Pharmacokinetics, and Biological Activity.

Approval of the IAP by all Merck Data Analysis Responsible is documented within ELDORADO.

Wet ink signature to be provided by PPD responsible for PPD use only.

Merck responsible	**Date**	**Signature**
PPD, Lead Program Statistician		
PPD **responsible**		
PPD, Biostatistician		

1 Table of Contents

Integrated Analysis Plan 1
1 Table of Contents 4
2 List of Abbreviations and Definition of Terms 7
3 Modification History 9
4 Purpose of the Integrated Analysis Plan 9
5 Objectives and Endpoints 10
6 Overview of Planned Analyses 12
6.1 Independent Monitoring Committee review 12
6.2 Interim Analysis 13
6.3 Primary Analysis 13
6.4 Blinded Extension Analysis 13
7 Changes to the Planned Analyses in the Clinical Trial Protocol 14
8 Protocol Deviations and Analysis Sets 16
8.1 Definition of Protocol Deviations and Analysis Sets 16
8.2 Definition of Analysis Sets and Subgroups 17
9 General Specifications for Statistical Analyses 19
10 Trial Subjects 34
10.1 Disposition of Subjects and Discontinuations 34
10.2 Protocol Deviations 35
10.2.1 Important Protocol Deviations 35
10.2.2 Reasons Leading to the Exclusion from an Analysis Set 36
11 Demographics and Other Baseline Characteristics 38
11.1 Demographics 38
11.2 Medical History 39
11.3 Other Baseline Characteristics 39
11.3.1 Disease history 39
11.3.2 Other 39
11.3.3 Columbia- Suicide Severity Rating Scale (C-SSRS) 40
12 Previous or Concomitant Medications/Procedures 41
13 Treatment Compliance and Exposure 42

14 Efficacy Analyses 45
14.1 Primary Endpoint Analyses 45
14.2 Secondary Endpoint Analyses 48
14.2.1 Key secondary endpoint analyses 48
14.2.2 Additional secondary endpoint analyses 52
15 Safety Evaluation 58
15.1 Adverse Events 58
15.1.1 All Adverse Events 59
15.1.2 Adverse Events Leading to Treatment Discontinuation 60
15.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events 61
15.2.1 Deaths 61
15.2.2 Serious Adverse Events 61
15.3 Clinical Laboratory Evaluation 62
15.4 Vital Signs 64
15.5 12-Lead Electrocardiogram (ECG) 65
15.6 Chest X-ray Evaluations 66
15.7 Physical Examination 66
15.8 Safety/Pharmacodynamic Endpoints: CCI and Immunoglobulin levels 66
15.9 Pregnancy test 67
16 Analyses of Other Endpoints 67
16.1 CCI
16.2 CCI
16.2.1 CCI
16.2.2 CCI
16.2.3 CCI
16.2.4 CCI
16.2.5 CCI
16.3 CCI
17 References 74
18 Appendices 75

18.1 Appendix 1: Worst on treatment value based on normal range of laboratory evaluations.....75

18.2 CCI [redacted]

18.3 CCI [redacted]

18.3.1 CCI [redacted]

18.3.2 CCI [redacted]

2 List of Abbreviations and Definition of Terms

ADaM	Analysis Data Model
AE	Adverse Event
AIC	Akaike Information Criterion
ALT	Alanine Amino-Transferase
ARR	Annualized Relapse Ratio
AST	Aspartate Amino-Transferase
ATC	Anatomical Therapeutic Class
BE	Blinded Extension
BEA	Blinded Extension Analysis
BID	Twice daily
BOA	Biostatistics Outputs Assembly
CCI [redacted]	
CFB	Change From Baseline
CI	Confidence Interval
CDF	Cumulative distribution function
CS	Clinically Significant
CSR	Clinical Study Report
C-SSRS	Columbia- Suicide Severity Rating Scale
CXR	Chest X-Ray
EAIR	Exposure Adjusted Incidence Rate
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
EEA	European Economic Area
ETA	Inter-individual random error estimate
FOCE(I)	First order conditional estimation (with interaction)
FSFD	First subject first dose
FWER	Family-wise Type I error rate
Gd+	Gladolinium-positive
GI	Gastrointestinal

GLMM	Generalized Linear Mixed Model
CCI [redacted]	[redacted]
IA	Interim Analysis
IAP	Integrated Analysis Plan
IDMC	Independent Data Monitoring Committee
Ig	Immunoglobulin
IMP	Investigational Medical Product
IOV	Interoccasion variability
IV	Intravenous
IWRS	Interactive Web Response System
LOCF	Last Observation Carried Forward
MAR	Missing At Random
MCAR	Missing Completely At Random
MCS	Mental component summary
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intention to Treat
MMRM	Mixed-effect Model for Repeated Measures
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
M&S	Modeling and simulation
NB	Negative binomial
NCI-CTCAE	National Cancer Institute – Common Terminology Criteria for Adverse Events
NONMEM	Non-linear mixed effects modeling software
(Δ)OFV	(Difference in) Objective function value
PCS	Physical component summary
CCI [redacted]	
PiC	Powder in capsule
CCI [redacted]	
PsN	Perl-speaks-NONMEM
PT	Preferred Term

QD	Once daily
Q1	25th Percentile
Q3	75th Percentile
RoW	Rest of the World
RRMS	Relapsing-Remitting Multiple Sclerosis
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SCR	Screening Analysis Set
SDTM	Study Data Tabulation Model
SD	Standard Deviation
SE	Standard Error
CCI	[redacted]
SLDR	Subject Level Data Review
SOC	System Organ Class
SPMS	Secondary progressive multiple sclerosis
TEAE	Treatment Emergent Adverse Event
TLF	Table /Listing/Figure
ULN	Upper Limit of Normal
VPC	Visual predictive check
VS	Vital Signs
WHO-DD	World Health Organization Drug Dictionary

3 Modification History

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
1.0	14 Feb 2018	PPD	NA – first version

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 MS200527-0086 dated 05 July 2016 (version 1.0). The IAP is based upon Section 8 (Statistics) of the trial protocol and is prepared in compliance with International Conference on Harmonization E9.

The first version (version 1.0) of the IAP includes details for the primary and blinded extension statistical analyses.

Results of the analyses described in this IAP will be included in the Clinical Study Report (CSR). Additionally, the planned analyses 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 SAP document (MS200527-0086_IDMC_SAP_v1.0) detailed specifications for the Independent Data Monitoring Committee (IDMC) analyses.

5 Objectives and Endpoints

	Objective	Endpoint	IAP section
Primary Objective	The primary objective is to evaluate the efficacy and dose-response of Evobrutinib on the number of gadolinium-positive (Gd+) T1 magnetic resonance imaging (MRI) lesions versus placebo after 24 weeks of treatment.	Primary Endpoint: • Total number of gadolinium-enhancing T1 lesions at Weeks 12, 16, 20, and 24	14.1
Key Secondary Objectives	To evaluate the efficacy and dose-response of Evobrutinib on clinical endpoints over 24 weeks versus placebo	Secondary Endpoints: • Annualized relapse rate (ARR), based on protocol-defined qualified relapses, at Week 24 • Qualified relapse-free status at Week 24 • Change from Baseline in Expanded Disability Status Scale (EDSS) at Week 24	14.2.1
	To evaluate the safety of Evobrutinib	Secondary Endpoint: • Safety as assessed by the nature, severity, and incidence of adverse events (AEs); vital signs; electrocardiograms (ECGs); absolute concentrations and change from Baseline in immunoglobulin (Ig) levels; absolute numbers and change from Baseline in B cells; and clinical laboratory safety parameters (duration of placebo treatment group is limited to 24 weeks).	15
Other Secondary Objectives	To evaluate the efficacy of Evobrutinib on additional MRI parameters over 24 weeks versus placebo	Secondary Endpoints: • Total number of new Gd+ T1 lesions at Weeks 12, 16, 20, and 24 • Mean per-scan number of Gd+ T1 lesions at Weeks 12, 16, 20, and 24 • Total number of new or enlarging T2 lesions at Weeks 12, 16, 20, and 24	14.2.2

		• Change from Baseline in the volume of Gd+ T1 lesions at Week 24 • Change from Baseline in the volume of T2 lesions at Week 24	
	To evaluate the efficacy of Evobrutinib on clinical and MRI endpoints from Weeks 24 to 48	Secondary Endpoints: • Number of Gd+ T1 lesions at Week 48 • Number of new Gd+ T1 lesions at Week 48 • Annualized relapse rate, based on protocol-defined qualified relapses, at Week 48 • Qualified relapse-free status at Week 48 • Change from Baseline in EDSS at Week 48 • Number of new or enlarging T2 lesions at Week 48 • Change from Baseline in the volume of Gd+ T1 lesions at Week 48 • Change from Baseline in the volume of T2 lesions at Week 48	14.2.2
	To evaluate the efficacy of Tecfidera on clinical and MRI endpoints over 24 weeks	Secondary Endpoints: • Total number of gadolinium-enhancing T1 lesions at Weeks 12, 16, 20, and 24 • Mean per-scan number of Gd+ T1 lesions at Weeks 12, 16, 20, and 24 • Annualized relapse rate, based on protocol-defined qualified relapses at Week 24 • Qualified relapse-free status at Week 24 • Change from Baseline EDSS at Week 24 • Total number of new Gd+ T1 lesions at Weeks 12, 16, 20, 24 • Total number of new or enlarging T2 lesions at Weeks 12, 16, 20, 24 • Change from Baseline in the volume of Gd+ T1 lesions at Week 24 • Change from Baseline in the volume of T2 lesions at Week 24	14.2.2
	To evaluate the efficacy of Tecfidera on clinical and MRI endpoints from Weeks 24 to 48	Secondary Endpoints: • Number of Gd+ T1 lesions at Week 48 • Number of new Gd+ T1 lesions at Week 48 • Annualized relapse rate, based on protocol-defined qualified relapses, at Week 48 • Qualified relapse-free status at Week 48 • Change from Baseline in EDSS at Week 48 • Number of new or enlarging T2 lesions at Week 48 • Change from Baseline in the volume of Gd+ T1 lesions at Week 48 • Change from Baseline in the volume of T2 lesions at Week 48	14.2.2

	To evaluate the safety of Tecfidera	Secondary Endpoint: • Safety as assessed by the nature, severity, and incidence of AEs; vital signs; ECGs; absolute concentrations and change from Baseline in Ig levels; absolute numbers and change from Baseline in B cells; and clinical laboratory safety parameters	15
Exploratory Objective	CCI		14.2.5 and 14.2.6 14.2.8 14.2.9 14.2.3

6 Overview of Planned Analyses

This IAP (Version 1.0) covers the analyses for futility, efficacy, CCI [redacted], and safety based on the data cut-offs for the primary and blinded extension analyses. Section 9 describes how data collected after the cut-off date will be handled.

6.1 Independent Monitoring Committee review

An IDMC will be set up to continually review available 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. IDMC meetings will take place every 4 months or as requested by the IDMC members. Details of the statistical analyses of the IDMC are provided in the IDMC SAP v1.0 and the IDMC charter v1.0.

6.2 Interim Analysis

If performed, the aim of the Interim Analysis (IA) was to evaluate overall futility based on the highest dose of Evobrutinib to determine whether or not to continue the study. The IA was to be triggered when the first 50% of subjects enrolled out of the planned enrollment reach Week 24 of treatment. It was planned that placebo subjects would continue at the 25 mg QD dose after Week 24, with consideration given to changing this dose based on data from the IA. CCI [redacted] analyses were not to be included in the IA.

When the last subject was enrolled in the study in July 2017, it was determined that only approximately 14 weeks would elapse between the first 50% of enrolled subjects reaching Week 24 (IA trigger point) and 100% of enrolled subjects reaching Week 24 (primary analysis trigger point). It was decided that the IA would not be conducted, as it would occur too late to support early decision-making, and IA conduct would interfere with preparations for the primary analysis. Thus this IAP does not provide specifications for the analysis of IA endpoints.

6.3 Primary Analysis

When the last subject reaches 24 weeks of treatment or discontinues from treatment prematurely, the protocol violations are determined, and the database is partially locked for the primary analysis, the drug codes will be broken and made available for the primary data analysis. All endpoints based on Baseline to Week 24 data will be evaluated (including CCI [redacted]).

6.4 Blinded Extension Analysis

The blinded extension analysis will occur only when the last subject completes the Week 52 safety follow-up or discontinues early, the protocol violations are determined, and the database is locked for the blinded extension analysis. All endpoints based on Baseline to Week 52 data will be evaluated (except for CCI [redacted] endpoints collected to Week 24 which will have been fully evaluated at the Primary Analysis).

7 Changes to the Planned Analyses in the Clinical Trial Protocol

CTCAE Grades for Relevant Laboratory Parameters

The Grade 1 definitions for the laboratory parameters AST, ALT, bilirubin, amylase, and lipase in Table 4 in the protocol Version 1.0 were cited incorrectly. All laboratory abnormalities will be graded as specified in CTCAE (2), version 4.03. This table is also now removed from protocol version 2.0.

CCI [redacted]

[redacted]

Region Randomization Stratum

Due to extreme imbalance in enrollment between Eastern Europe and Western Europe, region will not be used for adjustment or stratification in the different analyses.

Secondary Endpoints wording

In the description of lesion endpoints in this protocol, the word "total" refers to summing over assessments at weeks 12, 16, 20, 24. So "total" does not apply to the lesion count observed at a single timepoint, such as week 48. To avoid confusion, wording of MRI secondary endpoints at Week 48 was changed by removing the word "total".

Wording of "Proportion of subjects who remain qualified relapse-free at Week 24" endpoint was changed to "Qualified relapse-free status at Week 24". The same correction was applied to this endpoint at Week 48. This update is also performed in protocol version 2.0.

Per-Protocol (PP) Analysis Set definition

The protocol defined the PP analysis set as all subjects who belong to the mITT Analysis Set, complete 24 weeks of treatment, and do not have any important protocol deviations. The definition has been clarified to specify "do not have any clinically important protocol deviations. "Clinically important" protocol deviations comprise the subset of important protocol deviations that lead to exclusion of a subject from an analysis set.

CCI [redacted]

CCI [redacted]

Analysis of Secondary Endpoints

In the analysis of the key secondary endpoint, change from baseline (CFB) in EDSS at Week 24, the rank ANCOVA analysis decribed in Section 8.5.3 of the protocol will be replaced with Hodges-Lehmann estimation of the shift in distribution location, and a stratified Wilcoxon rank-sum test [Healy *et al* 2011], one test for each Evobrutinib versus placebo comparison, where the stratification will at a minimum adjust for categorical baseline EDSS. Categories such as ≤ 3.0 versus > 3.0, corresponding to a clinically meaningful threshold close to the median baseline (BL) value, will be considered. Descriptive statistics for Week 24 EDSS CFB values treated as continuous variables (i.e., mean, median, and range) will be reported. Per EMA Guideline on Clinical Investigation of Medicinal Products for the Treatment of Multiple Sclerosis (2015), mean EDSS CFB is not considered an appropriate efficacy parameter. Therefore, descriptive statistics for categorical Week 24 EDSS CFB, i.e., number and proportion improving, stable, or worsening, will also be reported. Here improvement is defined as a decrease of 1.0 point or more, stable condition as a change of no more than half a point in either direction, and worsening as an increase of 1.0 point or more.

In the analysis of the secondary endpoint CFB in T2 lesion volume at Week 24, the ANCOVA model of the appropriately transformed variable, described in Section 8.5.3 of the protocol, will be replaced with a Mixed-effect Model for Repeated Measures (MMRM) approach for the appropriately transformed variable. This approach will better exploit the data available at earlier timepoints (i.e., weeks 12, 16, 20) for subjects missing the Week 24 assessment, assuming data are missing at random (MAR). At a minimum, the model will adjust for baseline T2 lesion volume.

In the analysis of the secondary endpoint CFB in Gd+ T1 lesion volume at Week 24, the ANCOVA model of the appropriately transformed variable, described in Section 8.5.3 of the protocol, will be replaced with Hodges-Lehmann estimation of the shift in distribution location, and a stratified Wilcoxon rank-sum test, one test for each Evobrutinib versus placebo comparison, where the stratification will adjust for categorical baseline disease activity. Categories such as presence versus absence of Gd+ T1 lesion at baseline will be considered. The distribution of enhancing lesion volume has considerable probability mass at zero [van den Elskamp *et al* 2011], so Gd+ T1 lesion volume CFB is not amenable to transformation to an approximately normal random variable.

CCI **Exploratory Analysis**

CCI [redacted]

[redacted]

Protocol Version 1.0 references to the Final Analysis

An open label extension has been added to the study in version 2.0 of the protocol. Assuming that there will be an analysis at the end of the open label extension period, the analysis planned for the end of the blinded extension period, identified as "final" in version 1.0 of the protocol, will not be the final analysis of the study. Therefore, the analysis planned for the end of the blinded extension period has been renamed as the blinded extension analysis (BEA) in this IAP.

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.

The following important deviations will be identified and confirmed prior to or at the Data Review Meeting at the latest.

- Subjects who are dosed on the study despite not satisfying the inclusion criteria
- Subjects who develop withdrawal criteria whilst on the study but are not withdrawn
- Subjects who receive the wrong treatment or an incorrect dose
- Subjects who receive an excluded concomitant medication
- Deviation from Good Clinical Practice
- Inclusion and exclusion criteria violations
- Concomitant medication violations

- Other violations/events that may have a relevant influence on the CCI [redacted] (e.g., adverse events (AEs), vomiting, sample processing errors, inaccurate dosing, etc.)

Any important protocol deviations that lead to the exclusion of a subject from an analysis set will be considered clinically important (see Section 10.2).

In addition to protocol deviations that lead to exclusion of a subject from an analysis set, this study will allow for protocol deviations that lead to exclusion of a specific assessment on a subject from an analysis, but not all assessments for that subject. For example, if a subject uses high dose corticosteroids in the 3 weeks prior to a scheduled MRI scan, lesion data from that scan will be excluded from certain analyses and treated as missing.

All important protocol deviations should be documented in Clinical Data Interchange Standards Consortium Study Data Tabulation Model (SDTM) whether identified through sites monitoring, medical review and/or programming based on the inclusion/exclusion criteria presented in the protocol. Important protocol deviations are listed and described in MS200527-0086_List_of_deviations_v1.0.

8.2 Definition of Analysis Sets and Subgroups

Screening Analysis Set (SCR)
The SCR Analysis Set includes all subjects who signed the informed consent.

Safety Analysis Set (SAF)
The SAF consists of all subjects who receive at least one dose of trial treatment. Subjects will be analyzed according to the actual treatment they receive.

For safety analyses based only on the second 24-week treatment period, safety analysis set will be restricted to subjects having entered this period. This restriction will be identified as SAF-BE (Blinded Extension).

Intent-To-Treat Analysis Set (ITT)
The ITT Analysis Set consists of all subjects randomly allocated to a treatment, based on the intention to treat "as randomized" principle (ie, the planned treatment regimen rather than the actual treatment given in case of any difference).

Modified Intent-To-Treat Analysis Set (mITT)
The mITT Analysis Set consists of all subjects who belong to both the ITT and Safety Analysis Sets, and who have at least one baseline and one post-baseline MRI assessment. Subjects will be analyzed according to the treatment they were randomized to.

For efficacy analyses based only on the second 24-week treatment period, mITT will be restricted to subjects having an MRI assessment in this period. This restriction will be identified as mITT-BE.

Per-Protocol (PP) Analysis Set

The PP Analysis Set consists of all subjects who belong to the mITT Analysis Set, complete 24 weeks of treatment, and do not have any clinically important protocol deviations.

CCI [redacted]

CCI [redacted] .

CCI [redacted]

The use of the analysis sets in the different analyses is summarized in the following table:

Table 1: Analysis sets

Analyses	SCR Analysis Set	ITT Analysis Set	SAF Analysis Set	SAF-BE Analysis Set	PP Analysis Set	mITT Analysis Set	mITT-BE Analysis Set	CCI [redacted]	[redacted]	CCI [redacted]
Subject Disposition status	✓									
Analysis sets	✓									
Important Protocol Deviations		✓								
Demographic and Baseline Assessments						✓				
Prior Medications			✓							
Concomitant Medications			✓	✓						
Concurrent procedures			✓	✓						
Compliance and Exposure			✓	✓						

Analyses	SCR Analysis Set	ITT Analysis Set	SAF Analysis Set	SAF-BE Analysis Set	PP Analysis Set	mITT Analysis Set	mITT-BE Analysis Set	CCI		CCI
Safety Analysis			✓	✓						
Primary Efficacy Analysis		✓			✓	✓				
Secondary Efficacy Analysis and CCI						✓	✓			
CCI										
Randomization listing		✓								

Descriptive statistics for number of Gd+ T1 lesions over time until Week 24 will be presented for the following subgroups:

- ≤ 1 versus ≥ 2 relapses in last 2 yrs (i.e., approximately low versus high disease activity)
- EDSS ≤ 3.0 vs EDSS ≥ 3.5 (i.e., approximately lower versus higher risk of transitioning from RRMS to SPMS with relapses).

9 General Specifications for Statistical Analyses

All statistical analyses will be performed by PPD.

Treatment groups and Investigational Medical Product (IMP)

Treatment groups are defined as placebo, Evobrutinib 25 mg QD, Evobrutinib 75 mg QD, Evobrutinib 75 mg BID, and Tecfidera. Unless otherwise indicated, all analyses will be presented separately for the five treatment groups. The IMPs are placebo, Evobrutinib, and Tecfidera.

Actual Treatment Assignment

For the first 24-week treatment period, a subject who received 2 different types of kits over the course of treatment, should be tabulated according to the kit received most frequently. If there is a "tie", the highest dose will be chosen.

For the second 24-week treatment period, the same rule as for the first 24-week treatment period will be applied, with a distinction between subjects taking Placebo in the first period and Evobrutinib 25 mg in the second period.

If by mistake, a subject taking placebo in the first period was to take Evobrutinib 75 mg QD in the second period, this subject would be analyzed in "Evobrutinib 75 mg QD" treatment group.

For overall tables that do not present within-group analyses, only subjects assigned to Evobrutinib or Tecfidera in the first period will be analyzed. Each subject will be tabulated according to the type of kit he/she received most frequently overall.

Presentation of Tables/Figures/Listings

For all analyses, Tables and Figures will be produced using true treatment groups. When data from only the first 24 weeks are reported, the following labels will be used: 'Placebo', 'Evobrutinib 25 mg QD', 'Evobrutinib 75 mg QD', 'Evobrutinib 75 mg BID', 'Tecfidera'. When data from week 24-48 or week 0-48 are reported, the Placebo label will be changed to 'Placebo + Evobrutinib 25 mg QD'.

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

All data recorded during the trial will be presented in individual data listings, performed on the Safety Analysis Set (SAF), unless otherwise specified. All listings will be sorted by treatment group, subject, and scheduled time point (where applicable), if not otherwise stated. Further details are provided in the appropriate section for the analysis of the specific parameter.

Listings presented on the actual treatment, will consider the overall treatment assigned for subjects taking either Evobrutinib or Tecfidera in the first period. For subjects taking placebo in the first treatment period, treatment label will be either "Placebo" or "Placebo + Evobrutinib 25 mg QD" depending on whether they have entered the second period or not.

Presentation by Time Period

The time period corresponding to the first 24 weeks of treatment is of special interest, as the primary endpoint of the study is at Week 24, and subjects randomized to placebo will be switched to Evobrutinib at Week 24. Therefore, tabular summaries may include data from the period up through Week 24, from the period after Week 24, and from the overall study period. Listings will always present all available data in the database transfer.

For previous/concomitant medications, exposure duration, cumulative dose, compliance and safety (AEs, Labs, ECGs, vital signs, Ig levels, total B cell number), tables will be presented as follows:

For the primary analysis

When applicable, three sets of tables will be made based on data overall, data through Week 24 (first 24-week treatment period, including data from Safety Follow-up where appropriate), and data after Week 24 through Week 48 (second 24-week treatment period, including data from Week 52 Safety Follow-up visit where appropriate).

For the blinded extension analysis

When applicable, two sets of tables will be made. One set considering data overall and another one only considering data from after Week 24 through Week 48 (second 24-week treatment period, including data from Week 52 Safety Follow-up visit where appropriate).

Overall tables that do not present within-group analyses will be restricted to subjects receiving either Evobrutinib or Tecfidera in the first period. Overall tables that present within-group analyses will include all treatment groups, including subjects who receive placebo in the first period, and switch to Evobrutinib in the second.

For safety analyses not including biochemistry parameters or adverse events, any assessment performed until 30 minutes after drug intake on Week 24 will considered in the first 24-week treatment period.

Data handling after cut-off date for a planned analysis

A cut-off date will be applied at the SDTM level: for the primary analysis (resp. blinded extension analysis), all data posterior to the last dose of the last patient during the first (resp. second) 24-week treatment period will be removed from the database. Details on the implementation of this cut-off date are available in the document named *SDTM Cut-off date implementation rules_BTKi(M2951)-Compound_v2.0.docx*

Subjects discontinuing treatment early will have their data from the Safety Follow-up visit included in the primary analysis (resp. blinded extension analysis) if the Safety Follow-up visit occurs prior to the planned cut-off date. If there are subjects discontinuing treatment early whose Safety Follow-up visit occurred after the planned cut-off date, the exclusion of their Safety Follow-up data due to data cut-off will be mentioned in a footnote, where appropriate.

Data other than the date of death obtained after the cut-off will not be displayed in any listings or used for summary statistics.

Presentation of continuous and qualitative variables

Continuous variables including CCI but not CCI will be summarized using the following descriptive statistics:

- 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

The number of digits for non-derived and derived data, presented in outputs or available in ADaM (Analysis Data Model) datasets, is specified in the Biostatistics Outputs Assembly (BOA) document.

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 is 100% then display 'xx (100.0)'.

Qualitative variables will be summarized by counts and percentages. The "Missing" category should always be displayed at baseline – even when there are no missing data at baseline. At timepoints other than baseline, the "Missing" category should only be displayed when there are missing data.

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.

The total of missing and non-missing observations at each timepoint will reflect the population still in the trial at that time. For example, if a subject is still in the trial at the timepoint but with missing data, that subject should be counted in the number of missing observations.

CCI [redacted]

Definition of baseline

For the purpose of statistical analysis, baseline is defined as the last non-missing measurement (including those collected at an Unscheduled visit) prior to the first dose of study drug, as described in the following table. If baseline cannot be defined, then the baseline value will be treated as missing.

Table 2: Definition of baseline.

Randomized	Time Period	Baseline
Placebo	Initial 24-week treatment period	Last non-missing measurement prior to the first dose of study drug in the initial 24-week treatment period
	Blinded Extension (second 24-week treatment period)	Last non-missing measurement prior to the first dose of study drug in the Blinded Extension.
	Overall treatment period	Baseline definition will depend on whether the assessment being compared to baseline is from the initial 24-week treatment period or the Blinded Extension.
Evobrutinib/Tecfidera	All treatment periods	Last non-missing measurement prior to the first dose of study drug in the initial 24-week treatment period

Definition of change from baseline (CFB)

CFB and percent CFB at a given post-baseline visit will be computed as follows:

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

At the baseline visit, the CFB will be equal to zero and the percent CFB will be missing. For a given post-baseline visit, when the baseline value and the absolute value are both equal to zero, then the CFB and the percent CFB will be equal to zero as well.

Definition of duration

Duration will be calculated as the difference between start and stop dates plus 1 (eg, AE duration (days) = AE end date - AE start date + 1). Durations will be calculated only when both dates are available (imputed dates cannot be used for the duration computation) unless otherwise specified.

The time since an event will be calculated as:

- reference date minus date of event +1 (eg, days in study at onset of AE = AE start date - date of randomization + 1) if date of event is equal or greater than reference date
- reference date minus date of event (eg, days in study at onset of AE = AE start date - date of randomization) otherwise.

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.

Presentation of missing data

In all subject data listings, partial dates, which are not to be imputed according to this IAP, will be presented using the format "_____YYYY".
When presented, imputed dates will be flagged (ie, D for day, M for month).

Missing statistics, eg, when they cannot be calculated, should be presented as 'nd', with 'nd' standing for 'not done'. For example, if n = 1, the measure of variability (SD) cannot be computed and should be presented as 'nd'.

In case of zero records available for presentation in a given TLF, an empty output with 0 occurrence or a sentence stating that there are no data will be provided. For tables of Adverse Events and Deaths (outputs required for EudraCT and/or clinicaltrial.gov), if there is no

observation, the output must contain the first line 'Subject with…' or 'Subject who died' displayed with 0 occurrence.

If a 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 ADaM level. When a Preferred Term (PT) is missing, it will be set to 'Uncoded PT:' TEAE verbatim text.

Handling of missing efficacy and CCI data

Total number of lesions at Weeks 12, 16, 20, 24

The imputation approach for the total number of lesions endpoint depends on whether the analysis is model-based or nonparametric.

When analyzing the total number of lesions endpoint using a NB model, no imputation is needed for subjects missing 1-3 of the 4 planned scans, as the offset parameter adjusts for the number of available scans. If a subject lacks any post-baseline evaluable MRI assessment, then scan count and lesion count are imputed using median values for scan count and lesions per-scan 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. For example, a subject's imputed lesion count would be the product of the median number of scans (i.e., 4) and median lesions per-scan (i.e., 0.2), where the median values are based on subjects in the same treatment group, with the same value for the categorical baseline covariate. The imputed count (i.e., 0.8) will be rounded to an integer value (i.e., 1) as the NB model requires that the response variable be integer-valued.

In the longitudinal Poisson analysis of number of lesions at a single scan, which makes use of assessments at Weeks 12, 16, 20, and 24, no imputation is needed for subjects having at least one post-baseline score assessment prior to Week 24. If a subject lacks any post-baseline evaluable MRI assessment, then lesion count will be imputed for only the first post-baseline timepoint used in the GLMM analysis (i.e., Week 12), using the median lesion count at that timepoint among subjects with a scan at that timepoint, who are in the same treatment group, and who have the same value for the categorical baseline covariate.

When analyzing the total number of lesions endpoint using a nonparametric method (H-L estimate, Wilcoxon rank-sum test, Jonckheere test), the method assumes that the response variable arises from the same observation effort for each subject, so missing data must be imputed. If a subject is missing 1-3 of the 4 planned scans, missing scan values will be imputed with the average of available scan values. For example, if a subject has a lesion count of 3 from the week 12 scan, 1 from the week 16 scan, 1 from the week 20 scan, and missing value for the week 24 scan, a count of 1.67 will be imputed for week 24, leading to a total count of 6.67 for weeks 12-24. 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. The imputed count will not be rounded to an integer value as the nonparametric analysis does not require that the response variable be integer-valued.

Mean lesions per-scan at Weeks 12, 16, 20, 24

No imputation is needed for subjects missing 1-3 of the 4 planned scans, as the variable amount of scans experienced by a subject is accounted for in the mean lesions per-scan response variable. If a subject lacks any post-baseline scan, then the endpoint is imputed using the median value for mean 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.

Number of lesions at Week 48

In the analysis of lesion count at Week 48, a missing assessment at Week 48 will be imputed by the median Week 48 value among subjects having a Week 48 scan, who are in the same treatment group, and who have the same value for the categorical baseline covariate.

CFB in volume of T2 lesions at Week 24

In the MMRM analysis of the Week 24 CFB in volume of T2 lesions, appropriately transformed, which makes use of assessments at Weeks 12, 16, 20, and 24, no imputation is needed for subjects having at least one post-baseline volume assessment prior to Week 24. If a subject lacks any post-baseline evaluable MRI assessment, then Week 24 CFB in transformed T2 lesion volume will be imputed for only the first post-baseline timepoint used in the MMRM analysis (i.e., Week 12), using the mean CFB in transformed T2 lesion volume at that timepoint among subjects with an assessment at that timepoint, who are in the same treatment group, and who have a baseline covariate in the same quartile.

CFB in volume of T2 lesions at Week 48

In the nonparametric analysis of CFB in T2 lesion volume at Week 48, a missing scan value at Week 48 will be imputed by the median Week 48 CFB value among subjects having a Week 48 scan, who are in the same treatment group, and who have the same value for the categorical baseline covariate.

CFB in volume of Gd+ T1 lesions at Week 24 or Week 48

In the nonparametric analysis of CFB in T1 Gd+ lesion volume at Week 24, a missing scan value at Week 24 will be imputed by the median Week 24 CFB value among subjects having a Week 24 scan, who are in the same treatment group, and who have the same value for the categorical baseline covariate. The same approach will be used for CFB in T1 Gd+ lesion volume at Week 48.

ARR at Week 24 or Week 48

Subjects discontinuing treatment before Week 24 or Week 48 will be followed for relapse through Safety Follow-up visit.

When analyzing the ARR endpoint using a NB model, no imputation is needed, as the offset parameter specifies the time under observation, which adjusts for subjects who discontinue treatment before Week 24 or Week 48.

When analyzing the ARR endpoint using the Jonckheere test applied to subject-specific ARR (i.e., number of relapses experienced by the subject divided by follow-up experienced by the subject), there is again no need for imputation, as the variable amount of follow-up experienced by a subject is accounted for in the subject-specific ARR response variable.

Qualified relapse-free status at Week 24 or Week 48

In the analysis of proportion qualified relapse-free at Week 24, subjects who discontinue study prior to Week 24, without having a qualified relapse are counted as not being qualified relapse-free at Week 24. This imputation approach will be used for both the logistic model analysis and the Cochran-Armitage test. The same approach used for qualified relapse-free status Week 0-24 will be used for qualified relapse-free status Week 25-48.

Sensitivity analyses of this endpoint will apply other methods of imputation, as described in Section 14.2.1.

CFB in EDSS at Week 24 or Week 48

In the analysis of change from baseline (CFB) in EDSS at Week 24 using nonparametric analyses (H-L estimate, Wilcoxon rank-sum test, Jonckheere test), a missing assessment at Week 24 will be imputed by the median Week 24 CFB value among subjects who have a Week 24 assessment, who are in the same treatment group, and who have the same value for the categorical baseline covariate. The same approach will be used for CFB in EDSS at Week 48. A sensitivity analysis of this endpoint will apply another method of imputation, as described in Section 14.2.1.

CCI [redacted]

[redacted]

[redacted]

CCI [redacted]

CCI [redacted]

Descriptive Statistics

Missing data will not be imputed for descriptive statistics.

Handling of missing or partial adverse events dates

For defining the TEAE flag, missing or partial adverse event dates will be imputed 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.

Imputed dates will be used for defining the TEAE flag only.

Handling of partially missing MS onset or diagnosis date

For time since MS onset or MS diagnosis, a missing onset day/month will be replaced by 1 for the duration derivation.

Trial day / Treatment day

Trial day is defined relative to the date of randomization. Treatment day is defined relative to the date of start of treatment.

In the case of subjects initially randomized to placebo, there will be a second start of treatment on the day of first administration of Evobrutinib in the second 24-week treatment period.

The day before the start date of treatment is defined as treatment day -1, i.e., there is no treatment day zero.

Repeated and unscheduled measurements

An assessment (safety or efficacy) at an unscheduled time point is linked to the previous scheduled visit in the SDTM dataset. Per schedule of assessments, EDSS and relapse data are collected at an unscheduled visit for neurological worsening and relapse assessment. Per the

eCRF package for the study, EDSS, relapse, and CCI data may be collected at any unscheduled visit.

Any relapse assessment for a subject at an unscheduled visit will contribute to the ARR estimate for that subject's treatment group, for the period including that timepoint.

EDSS and CCI assessments at unscheduled visits will contribute to by-visit summaries and inferential analyses (i.e., analysis of Week 24 CFB in EDSS, analysis of lesions summed over visits at Weeks 12-24, analysis of Week 24 CFB in CCI via MMRM with visit as a covariate) based on analysis visit windowing, as defined in Tables 3-5.

For safety shift tables, assessments associated with unscheduled time points will be used in the definition of the worst assessment during the study.

Unique to this study, additional biochemistry assessments were associated with Unscheduled visits, necessitating the definition of windows for by-visit tabulations. Apart from biochemistry parameters, for by-visit tabulations, in case of multiple assessments (including unscheduled) linked to the same visit, the first available assessment (in chronological order) will be included in the summary of a visit.

Repeated and unscheduled measurements will be reported in the listings.

Early treatment and early trial termination assessments

EDSS, relapse, MRI, and CCI data are collected at the early treatment termination visit. In addition, relapse is assessed at the early trial termination visit.

By-visit summary statistics should be programmed such that the summary for the Early Treatment Termination visit is distinct from the summary for the Week 48 visit, and the summary for the Early Trial Termination visit is distinct from the summary for the Week 52 visit. This will support convenient comparison between treatment completers (at week 48) and discontinuers, and trial completers (at week 52) and discontinuers.

Any relapse assessment at an early treatment termination or early trial termination visit will contribute to the ARR estimate for the period including that timepoint.

EDSS, MRI, and CCI assessments at early treatment termination visits will contribute to inferential analyses (i.e., analysis of Week 24 CFB in EDSS, analysis of lesions summed over visits at Weeks 12-24, analysis of Week 24 CFB in CCI via MMRM with visit as a covariate) based on analysis visit windowing, as defined in Tables 3-5.

For analyses by time period, data from safety follow-up will be included in the period in which the subject has discontinued.

Analysis visit windows

Assessments may be made at times other than the nominal times of planned visits, due to patient scheduling issues, unscheduled visits for neurological worsening or relapse assessment, or early treatment termination visits.

Explicit analysis visit windows for efficacy and CCI assessments, defined below, will be used to incorporate unscheduled assessments in by-visit summaries of EDSS, MRI endpoints, and CCI, or inferential analyses that depend on assessments at specific visits. (Relapse will be summarized by time period, not by visit, so analysis visit windows are not needed.)

In by-visit summaries, assessments collected at early treatment termination visits will be summarized separately (i.e., visit windowing will not be applied), to support convenient comparison between treatment completers and discontinuers. However, in inferential analyses, analysis visit windows will be used to incorporate assessments collected at early treatment termination visits.

Table 3: Analysis visit windows for EDSS through Week 48

Analysis visit	Nominal analysis visit day	Analysis visit window (days)
		All treatment groups
Day 1	1	[1, 1]
Week 12	84	[2, 126)
Week 24	168	[126, 189)
Week 36	252	[189, 294)
Week 48	336	[294, 339)

Table 4: Analysis visit windows for MRI through Week 48

Analysis visit	Nominal analysis visit day	Analysis visit window (days)
		All treatment groups
Week 12	84	[2, 98)
Week 16	112	[98, 126)
Week 20	140	[126, 154)
Week 24	168	[154, 189)
Week 48	336	[189, 339)

Table 5: Analysis visit windows for CCI through Week 48

Analysis visit	Nominal analysis visit day	Analysis visit window (days)
		All treatment groups
Day 1	1	[1, 1]
Week 4	28	[2, 42)
Week 8	56	[42, 70)
Week 12	84	[70, 96)
Week 16	112	[96, 126)
Week 20	140	[126, 154)
Week 24	168	[154, 189)
Week 36	252	[189, 294)

Week 48	336	[294, 339)

After the first IDMC meeting, additional biochemistry assessments were added to the Schedule of Assessments. Sites were instructed to record these additional assessments as Unscheduled visits. Analysis visit windows, defined below, will be used to incorporate unscheduled assessments in by-visit summaries of the biochemistry parameters. By-visit summaries will summarize assessments collected at early treatment termination visits separately (i.e, visit windowing will not be applied to assessments collected at the early treatment termination visit), to support convenient comparison between treatment completers (at week 48) and discontinuers.

Table 6: Analysis visit windows for biochemistry parameters through Week 48

Analysis visit	Nominal analysis visit day	Analysis visit window (days)
		All treatment groups
Day 1	1	[1, 1]
Week 4	28	[2, 42)
Week 8	56	[42, 63)
Week 10	70	[63, 77)
Week 12	84	[77, 91)
Week 14	98	[91, 105)
Week 16	112	[105, 119)
Week 18	126	[119, 133)
Week 20	140	[133, 147)
Week 22	154	[147, 161)
Week 24	168	[161, $X^{[1]}$]
Week 26	182	($X^{[1]}$, $X^{[1]}$ + 21)
Week 28	196	[$X^{[1]}$ + 21, $X^{[1]}$ + 35)
Week 30	210	[$X^{[1]}$ + 35, $X^{[1]}$ + 49)
Week 32	224	[$X^{[1]}$ + 49, $X^{[1]}$ + 63)
Week 34	238	[$X^{[1]}$ + 63, $X^{[1]}$ + 77)
Week 36	252	[$X^{[1]}$ + 77, $X^{[1]}$ + 91)
Week 38	266	[$X^{[1]}$ + 91, $X^{[1]}$ + 105)
Week 40	280	[$X^{[1]}$ + 105, $X^{[1]}$ + 119)
Week 42	294	[$X^{[1]}$ + 119, $X^{[1]}$ + 133)
Week 44	308	[$X^{[1]}$ + 133, $X^{[1]}$ + 147)
Week 46	322	[$X^{[1]}$ + 147, $X^{[1]}$ + 161)
Week 48	336	[$X^{[1]}$ + 161, $X^{[1]}$ + 171)
$X^{[1]}$ corresponds to the time + 30 minutes where subjects are exposed for the first time to Evobrutinib in the second 24-week treatment period.		

Control of multiplicity

The truncated Hochberg test and a multistage testing algorithm (Dmitrienko *et al* 2011) will be used to control the family-wise error rate (FWER) due to multiple comparisons associated with evaluating 3 dose groups based on one primary endpoint and 3 key secondary efficacy endpoints.

The 3 hypotheses associated with the primary endpoint will be considered family F_1.

Consider the ordered p-values that arise from comparing the 3 Evobrutinib treatment groups to placebo on the basis of the primary endpoint: $p_{(1)} \leq p_{(2)} \leq p_{(3)}$. The truncated Hochberg test is a step-up test based on the following critical values:

$$a_i = \left[\frac{\gamma}{4-i} + \frac{1-\gamma}{3}\right]\alpha, i = 1, \ldots, 3$$

where $0 \leq \gamma < 1$ is the truncation fraction ($\gamma = 1$ implies the regular Hochberg test, while $\gamma = 0$ implies Bonferroni). Starting with F_1, the family of 3 hypotheses associated with the primary endpoint, the truncated Hochberg's step-up method proceeds as follows:

- Step 1. If $p_{(3)} < a_3$, reject $H_{(1)}$, $H_{(2)}$, and $H_{(3)}$ (i.e., conclude all 3 dose groups are more efficacious than placebo on the basis of the primary endpoint) and then stop; otherwise go to Step 2.
- Step 2. If $p_{(2)} < a_2$, reject $H_{(1)}$ and $H_{(2)}$, and then stop; otherwise go to Step 3.
- Step 3. If $p_{(1)} < a_1$, reject $H_{(1)}$ (i.e., conclude that only the dose group associated with the smallest p-value is more efficacious than placebo on the basis of the primary endpoint) and stop.

Given 2-sided $\alpha = 0.05$ for family F_1, the critical values (a_1, a_2, a_3) equal (0.0167, 0.0242, 0.0467) and (0.0167, 0.025, 0.05) for $\gamma = 0.9$ (truncated Hochberg) and $\gamma = 1.0$ (regular Hochberg), respectively.

The error rate function e(I) of the truncated Hochberg test, where I is the index set {1, …, |I|}, and |I| is the size of the index set, is given by:

$e(\{1\}) = 1\text{-}P(a_1) = a_1$

$e(\{1, 2\}) = 1\text{-}P(a_1, a_2) = 1 - (1-a_2)(1-2a_1+a_2)$

$e(\{1, 2, 3\}) = 1\text{-}P(a_1, a_2, a_3) = 1\text{-} (1-a_3)(1-3a_1+a_3\text{-}3(a_2)^2 + 6a_1a_2 \text{-}3\ a_1a_3 + (a_3)^2)$

Using the critical values calculated for family F_1, the error values ($e(\{1\})$, $e(\{1, 2\})$, $e(\{1, 2, 3\})$) are equal to approximately (0.0167, 0.0332, 0.0495) for both $\gamma = 0.9$ (truncated Hochberg) and $\gamma = 1.0$ (regular Hochberg), respectively.

Let F_2 be the family of hypotheses associated with the key secondary endpoint ARR at Week 24, F_3 be the family of hypotheses associated with the key secondary endpoint qualified relapse-free status at Week 24, and F_4 be the family of hypotheses associated with the key secondary endpoint EDSS CFB at Week 24. Let A_i denote the index set corresponding to the retained hypotheses in F_i and $e_i(I)$ denote the error rate used in F_i, i=1, .., 3. The multistage testing algorithm for testing primary and key secondary endpoints in this study is as follows:

- Family F_1: The 3 hypotheses (concerning the primary endpoint) are tested at an α_1 level using the truncated Hochberg test ($\gamma < 1$) described above, with $\alpha_1 = \alpha = 0.05$.
- Family F_2: The 3 hypotheses (concerning the key secondary endpoint ARR at 24 weeks) are tested at an α_2 level using the truncated Hochberg test ($\gamma < 1$) described above, with $\alpha_2 = \alpha_1 - e_1(A_1)$.

- Family F_3: The 3 hypotheses (concerning the key secondary endpoint qualified relapse-free status at Week 24) are tested at an α_3 level using the truncated Hochberg test ($\gamma < 1$) described above, with $\alpha_3 = \alpha_2 - e_2(A_2)$.
- Family F_4: The 3 hypotheses (concerning the key secondary endpoint EDSS CFB at Week 24) are tested at an α_4 level using the regular Hochberg test ($\gamma = 1$) described above, with $\alpha_4 = \alpha_3 - e_3(A_3)$.

With this algorithm, the type 1 error rate is controlled over the multiple families of hypotheses due to the primary and key secondary endpoints at the 0.05 2-sided level. Other endpoints specified in the protocol will be assessed at a nominal significance level of 0.05, 2-sided, but the results will be viewed as exploratory.

Note that the portion of α to be carried from F_1 to F_2 depends on the set of primary null hypotheses retained by the 2-stage gatekeeping procedure, and is quantified via the error rate function of the first component of the 2-stage procedure. For this portion to be positive when at least one primary null hypothesis is rejected, it is required that the procedure be separable, i.e., $e_1(I_1|\alpha) < \alpha$ for all proper subsets I_1 of N_1, where N_1 is the index set of hypotheses included in family F_1. The Bonferronni procedure is separable, but the regular Hochberg procedure ($\gamma = 1$) is not. The truncated Hochberg procedure is separable and allows the multistage testing algorithm to be able to carry positive α from F_1 to F_2, from F_2 to F_3, and from F_3 to F_4. For the F_4 family, the regular Hochberg procedure (γ=1) may be used.

It is prespecified that $\gamma = 0.9$ will be used in the multistage testing algorithm whenever the truncated Hochberg test is used, although γ could be any value ≥ 0 and < 1, and could differ from family to family.

The penalty paid for performing multiple inferences in the families corresponding to the key secondary endpoints, F_i, i=2, 3, 4, depends on the number of hypotheses rejected at earlier stages. The sequence α_1, α_2, α_3, α_4 is non-increasing, so a hypothesis tested later in the sequence faces a higher hurdle unless all hypotheses are rejected in previously examined families.

A table summarizing the results of the multi-stage testing procedure will be provided that includes the p-values for each comparison of Evobrutinib treatment group to placebo treatment group, for the hypothesis family corresponding to the primary endpoint, and for the hypothesis families corresponding to each of the 3 key secondary endpoints. The critical value used to assess each p-value for significance will be reported next to the p-value, with an annotation indicating significance, illustrating the point in the procedure at which hypothesis testing halted. This summary table will be in addition to tables that summarize the analysis of a given endpoint.

Software

All statistical analyses will be performed using SAS® (Statistical Analysis System, SAS-Institute, Cary, North Carolina Windows Version 9.4 or higher). Graphics may be prepared with SAS Version 9.4, or higher; or Sigmaplot® 12.5, or higher (Systat Software, Inc., San Jose, California).

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 of following disposition categories will be produced by treatment group:

- 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 started treatment
- Number of randomized subjects with on-going treatment prior to Week 24 (applicable for dry run of the primary analysis)
- Number of randomized subjects who completed the first 24 weeks of treatment
- Number of randomized subjects who permanently discontinued treatment prior to Week 24 and reason
- Number of randomized subjects who were treated after Week 24
- Number of randomized subjects with on-going treatment after Week 24
- Number of randomized subjects who completed 48 weeks of treatment
- Number of randomized subjects who permanently discontinued treatment after Week 24 and prior to Week 48 and reason
- Number of randomized subjects in the follow-up phase who have not completed the trial (ie, treatment completed or discontinued and lack Safety Follow-up/End of Trial Visit)
- Number of randomized subjects who completed the trial
- Number of randomized subjects who discontinued from trial after randomization and reason

A table of randomized subjects not treated as randomized in the first (resp. second) double blind treatment period, including reason for not being treated as randomized will be produced by randomized treatment group for the primary (resp. blinded extension) analysis.

A table based on screened subjects describing the number and percent of subjects in each analysis set by treatment groups, will be produced:

- Number of screened subjects
- Number of subjects included in the ITT
- Number of subjects included in the SAF
- Number of subjects included in the SAF-BE
- Number of subjects included in the mITT
- Number of subjects in the mITT-BE
- Number of subjects included in the PP
- Number of subjects included in the CCI [redacted]
- Number of subjects included in the CCI [redacted]
- Number of subjects included in the CCI [redacted]

A table based on screened subjects describing the number of subjects by region, country within region and site will be produced by analysis set.

For the primary analysis, 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.

Subjects' information on informed consent, screening and randomization will be listed.

Subjects who discontinued/completed treatment or study will be listed with their reason for discontinuation (if applicable).

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. The first identifier will not be considered in the disposition counts.

10.2 Protocol Deviations

10.2.1 Important Protocol Deviations

For the primary analysis and the blinded extension analysis, the following summary tables and listings of important protocol deviations will be provided (separately for pre-/post inclusion deviations):

- Table providing frequency for each type of important protocol deviation
- Listing of important protocol deviations

10.2.2 Reasons Leading to the Exclusion from an Analysis Set

For the primary and blinded extension analyses, for subjects excluded from the mITT or PP analysis sets, the reasons for exclusion will be summarized, and the subjects excluded will be listed:

- By-reason table of number of subjects excluded from the mITT or PP analysis set, where a subject is counted for every reason that excludes that subject from the respective analysis set.
- By-subject listing of all reasons that excluded the subject from the mITT or PP analysis set

High dose corticosteroid use

If a subject uses high dose corticosteroids in the 3 weeks prior to a scheduled MRI scan, lesion data from that scan will be excluded from the primary analysis based on the mITT analysis set and treated as missing. A by-subject listing will be provided of all subjects who have at least one scan excluded from the primary analysis on the basis of high dose corticosteroid use. The listing will indicate the date of use, identity of the corticosteroid used, and dose and frequency of use.

Corticosteroid use is considered "high dose" when the cumulative dose in prednisone equivalent over the 21 days preceding the MRI assessment is ≥ 210 mg.

The worst case scenario will be applied in case missing or partial corticosteroid use dates are not definitively outside the 3 weeks windows preceding the MRI scan.

In such a situation, missing or partial start dates will be treated as follows:

- Same month as MRI scan and day missing, then corticosteroid use day start will be imputed to 01.
- Same year as MRI scan and month missing, then corticosteroid day/month start will be imputed to 01.
- Date completely missing, then corticosteroid start date is considered 3 weeks before the MRI scan.

Missing end dates will be imputed to the day of the MRI scan.

If the dose of the corticosteroid is missing then a dose of 10 will be considered. In case unit is missing then mg will be used. When frequency is missing QD will be considered.

A concomitant medication will be defined as a corticosteroid if the ATC code begins with "H02A"

The tables below will be used to define the cumulative dose:

Table 7: Conversion factors in prednisone equivalent

WHO-drug Term	Conversion factor in prednisone equivalent
METHYLPREDNISOLONE	1,25
METHYLPREDNISOLONE SODIUM SUCCINATE	1,25
METHYLPREDNISOLONE ACETATE	1,25
HYDROCORTISONE	0,25
BETAMETHASONE	6,667
BETAMETHASONE DIPROPIONATE	6,667
BETAMETHASONE VALERATE	6,667
DEXAMETHASONE	6,667
TRIAMCINOLONE	1,25
CORTISONE	0,2
PREDNISOLONE	1
PREDNISONE	1
DEFLAZACORT	0,83333
MEPREDNISONE	1,25
PARAMETHASONE	2,5
CRONOLEVEL	8,333
BETROSPAM	8,333

Table 8: Frequency conversions for corticosteroids

Frequency	Conversion factor	Numerical Conversion factor
OAM	1/30	0,0333
QOD	1/2	0,5000
QW	1/7	0,1429
QWK	1/7	0,1429
BID	2	2,0000
TID	3	3,0000
QD	1	1,0000
ONCE	1	1,0000
QAM	1	1,0000
Q2H	12	12,0000
Q3H	8	8,0000
Q4H	6	6,0000
Q6H	4	4,0000
Q8H	3	3,0000
QHS	1	1,0000
QID	4	4,0000
QPM	1	1,0000
QH	24	24,0000
Q3W	1/21	0,0476
TIW	3/7	0,4286
Q4W	1/28	0,0357
BIW or Twice a week	2/7	0,2857
EVERY 3 HOURS	8	8,0000
X1	1	1,0000
EVERY OTHER DAY	1/2	0,5000
EVERY OTHER DAY ALTERNTATLY WITH30 MG Q0D	1/2	0,5000
EVERY OTHER DAY ALTERNTATLY	1/2	0,5000

Frequency	Conversion factor	Numerical Conversion factor
WITH 20 MG QOD		
ONCE A WEEK	1/7	0,1429
QD(EVERY MORNING)	1	1,0000
SINGLE DOSES	1	1,0000
ALTERNATE DAYS	1/2	0,5000

11 Demographics and Other Baseline Characteristics

For the primary analysis, demographics, baseline disease characteristics, and other baseline characteristics will be summarized based on the mITT analysis set, presented by treatment group and overall.

Supportive listings will be based on the ITT analysis set, with subjects flagged according to membership in the mITT and SAF analysis sets.

11.1 Demographics

- Demographic characteristics
  - Sex: male, female
  - Race: White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other
  - Ethnicity: Hispanic or Latino, Not Hispanic or Latino
  - Age (years) at informed consent: summary statistics
- Pooled Region (US or Western Europe, Eastern Europe, Rest of the World)
- Geographic/Capability Region (US, Western Europe, Eastern Europe and CCI [redacted], Eastern Europe and not CCI [redacted], Rest of the World)
- European Economic Area (EEA)

Specifications for computation:

- Age (years):
  1. (date of given informed consent - date of birth + 1) / 365.25
  2. In case of missing day for at least one date, but month and year available for both dates, the day of informed consent and the day of birth will be set to 1 and the formula above will be used.
  3. In case of missing month for at least one date, but year available for both dates, the month of informed consent and 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.

11.2 Medical History

For the primary analysis, 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

For the primary analysis, information on MS baseline disease characteristics based on data on Day 1 predose and during screening will be summarized in total and listed. Descriptive statistics will be presented for:

- Type of MS, either relapsing-remitting MS (RRMS) or secondary progressive MS (SPMS)
- Time (years) since MS onset (first symptom)
- Time (years) since MS diagnosis
- Major systems affected
- Number of relapses in the year prior to randomization
- Number of relapses in the last 2 years
- EDSS score
- Scores for each of the 7 Functional Systems (Visual, Brainstem, Pyramidal, Cerebellar, Sensory, Bowel/Bladder, Cerebral) and score for Ambulation used to derive EDSS score.
- Presence of at least 1 Gadolinium-positive (Gd+) T1 lesion within 6 months prior to randomization
- Number of Gd+ T1 lesions
- Volume of Gd+ T1 lesions
- Volume of T2 lesions
- CCI [redacted]
- CCI [redacted]

CCI [redacted]

11.3.2 Other

For the primary analysis, descriptive statistics and listings will be presented for:

- Physical examination as continuous variables: Height (cm), Weight (kg) and BMI (kg/m²)

- CXR interpretation category (Normal, Abnormal Not Clinically Significant, Abnormal Clinically Significant, and Abnormal Overall).

Other screening baseline characteristics like viral serology, β-human chorionic gonadotropin, follicle stimulating hormone, quantiferon TB test, hepatitis B surface antigen and prior surgeries will be listed only.

Baseline characteristics such as ECG, vital signs and laboratory values will be presented in the safety summary tables by visit.

11.3.3 Columbia- Suicide Severity Rating Scale (C-SSRS)

The Columbia-Suicide Severity Rating Scale (C-SSRS) is a numerical score derived from 10 categories. The C-SSRS assesses the suicidal behavior and suicidal ideation in subjects.

Occurrence of suicidal behavior is defined as having answered "yes" to a least 1 of the 4 suicidal behavior subcategories (actual attempt, interrupted attempt, aborted attempt, and preparatory acts or behavior).

Occurrence of suicidal ideation is defined as having answered "yes" to at least one of the suicidal ideation sub-categories (wish to be dead, non-specific active suicidal thoughts, active suicidal ideation with any methods [not plan] without intent to act, active suicidal ideation with some intent to act [without specific plan], and active suicidal ideation with specific plan and intent).

Occurrence of suicidality is defined as having at least one occurrence of suicidal ideation or at least one occurrence of suicidal behavior.

For the primary analysis, the number and percentage of subjects with occurrence of suicidal behavior, occurrence of suicidal ideation and occurrence of suicidality at screening will be summarized.

12 Previous or Concomitant Medications/Procedures

Medications/procedures will be presented according to table 9.

Table 9: Data handling for medications/procedures

Analysis	Period covered	Data summarized	Treatment groups	Analysis sets
Primary analysis	Initial 24-week treatment period + safety follow-up	Previous medications + concomitant medications + concomitant procedures	ALL	SAF
Blinded Extension analysis	Second 24-week treatment period + safety follow-up	Concomitant medications + concomitant procedures	ALL	SAF-BE
Blinded Extension analysis	Overall	Concomitant medications + concomitant procedures	Evobrutinib 25 mg QD Evobrutinib 75 mg QD Evobrutinib 75 mg BID Tecfidera	SAF

Concomitant medications are medications, other than IMPs, which are taken by subjects any time on-trial (on or after the first day of IMP treatment for each subject). Concomitant medications include those started prior to and continued during administration of IMP, as well as those that were started after first administration of IMP. If the date values do not allow a medication to be classified as a non-concomitant medication, the medication will be considered as a concomitant medication.

Previous medications are medications, other than trial medications, which started before first administration of IMP. Previous medications include those that were continued during administration of IMP, as well as those discontinued prior to first administration of IMP. If the date values do not allow a medication to be classified as a non-previous medication, the medication will be considered as a previous medication.

The ATC-second level and preferred term 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 ATC-2nd level terms will be used for reporting.

Tables 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.

Concomitant procedures are procedures which were undertaken any time on trial.

The number and proportion of subjects with previous medications or concomitant medications will be separately summarized by treatment group and listed.

Concomitant procedures will be classified by medical review. Number of subjects with concomitant procedures (prior, on or after the first day of IMP) and by type of procedure (as classified by medical review) will be summarized by treatment group and listed.

13 Treatment Compliance and Exposure

Exposure will be presented according to table 10.

Table 10: Data handling for exposure

Analysis	Period covered	Data summarized	Treatment groups	Analysis set
Primary analysis	Initial 24-week treatment period	Exposure time + Cumulative dose + Compliance	ALL	SAF
	Second 24-week treatment period	Exposure time	ALL	SAF-BE
	Overall	Exposure time	Evobrutinib 25 mg QD Evobrutinib 75 mg QD Evobrutinib 75 mg BID Tecfidera	SAF
Blinded Extension Analysis	Second 24-week treatment period	Exposure time + Cumulative dose + Compliance	ALL	SAF-BE
	Overall	Exposure time + Cumulative dose + Compliance	Evobrutinib 25 mg QD Evobrutinib 75 mg QD Evobrutinib 75 mg BID Tecfidera	SAF

Planned administration of Placebo and Evobrutinib

Each treatment kit contains 17 wallets of 12 tablets which is enough medication for the administration of 6 tablets per day for 4 weeks and 6 days (ie, 204 tablets). From Day 1 until Week 20, subjects will receive one treatment kit per planned visit. At the Week 24 and Week 36 visits, subjects will receive 3 treatment kits at each visit. At each trial visit, subjects will be given treatment kits containing the number of tablets needed up to the next planned trial visit. Subjects will self-administer IMP in a blinded manner at a set time each day (every 12 hours ± 2 hours) as three matching tablets BID for 48 weeks. Subjects must take their daily dose more than 1 hour prior to a meal or snack and more than 2 hours after a meal or snack.

- Subjects randomized to placebo arm will take 6 placebo tablets filled with mannitol for the first 24 weeks. For the 24 weeks following, it is planned that these subjects will be switched to 5 placebo tablets and one tablet filled with 25 mg of Evobrutinib.

- Subjects randomized to Evobrutinib 25 mg QD arm will take 5 tablets filled with mannitol and one tablet filled with 25 mg of Evobrutinib.

- Subjects randomized to Evobrutinib 75 mg QD arm will take 3 tablets filled with mannitol and 3 tablets filled with 25 mg of Evobrutinib.

- Subjects randomized to Evobrutinib 75 mg BID arm will take 3 tablets BID filled with 25 mg of Evobrutinib.

Planned administration of Tecfidera

At the Day 1 visit, the treatment kit contains 14 capsules of 120 mg and 112 capsules of 240 mg which is enough medication for administration for 9 weeks. For the subsequent trial visits, subjects randomized to Tecfidera arm will be given treatment kits wherein the quantity of capsules will not be fixed due to the possibility of down titration in case of adverse event. Subjects randomized to Tecfidera arm will be given 120 mg of Tecfidera BID orally for the first 7 days. Following this, and for the duration of treatment, Tecfidera is given 240 mg BID orally.

On trial visit days, IMP will be administered during the trial visit after the scheduled trial visit procedures (other than post treatment CCI sampling) are completed.

Exposure

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

$$\text{exposure (weeks)} = (\text{date of last dose} - \text{date of first dose of the corresponding period} + 1)/7$$

If the end date of the last dose is missing or after the cut-off date then the date of last dose will be replaced by the cut-off date.
If a subject is lost to follow up and hasn't performed end of treatment (EOT) visit, then end of study date will be considered as the end of treatment date.

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

- ≤ 1 week
- > 1 to 8 weeks
- > 8 to 16 weeks
- > 16 to 24 weeks
- > 24 to 36 weeks
- > 36 to 48 weeks

For subjects randomized to placebo, and switched to Evobrutinib after Week 24, any exposure time after Week 24 represents exposure to Evobrutinib.

Cumulative actual dose (mg) per subject will be summarized by treatment groups.

- For Evobrutinib treatment groups, the cumulative actual dose is the number of tablets taken during that period, multiplied by the fraction of tablets that contain Evobrutinib 25 mg according to the actual treatment group, multiplied by 25 mg (dosage of one tablet). The number of tablets ingested will be deduced from the BATCH1 electronic case report form (eCRF).
- For the Tecfidera treatment group, the cumulative actual dose is the number of capsules ingested during that period multiplied by the appropriate dosage of one capsule (either 120 mg or 240 mg). The number of capsules ingested will be deduced from start date, end date, and dose, as recorded on the EXPOSUREDT2 eCRF.
- For subjects randomized to placebo, and switched to Evobrutinib after Week 24, the Evobrutinib cumulative actual dose will be calculated in the same manner as for the Evobrutinib treatment groups, using data from the BATCH1 eCRF, starting with Week 24.
- 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).

Compliance

For Evobrutinib or Placebo groups

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 $= 100*(\frac{(N_1)}{6 * N_2})$

where N_1 = number of tablets given minus number of tablets returned over N_2 days, deduced from the BATCH1 eCRF,

and N_2 = number of days between treatment start and treatment termination visit (or last visit before cut-off if treatment is ongoing).

For Tecfidera dose group

Compliance with treatment is defined as the actual cumulative dose divided by the planned cumulative dose, multiplied by 100 to yield a percentage, ie:

Compliance with treatment $= 100*(\frac{N_3}{N_4})$

where N_3 = actual cumulative dose for the given time period,

and N_4 = planned cumulative dose for the given time period.

Compliance with treatment and number of ingested tablets/capsules will be tabulated.

The following listings will be provided:

- listing of Evobrutinib/placebo kit numbers with date of dispense, and number of tablets returned (from BATCH1), and Tecfidera kit numbers with date of dispense, and number of capsules returned (from BATCH2) limited to Week 24 for the primary analysis and Week 48 for the blinded extension analysis.
- listing of Evobrutinib/placebo start/end dates with number ingested tablets (from EXPOSUREDT1), and Tecfidera start/end dates with dose (120 or 240 mg) (from EXPOSUREDT2) limited to Week 24 for the primary analysis and Week 48 for the blinded extension analysis.
- listing with exposure time, cumulative dose, and compliance by time period limited to Week 24 for the primary analysis and Week 48 for the blinded extension analysis.

14 Efficacy Analyses

Efficacy endpoints will be included in listings for the primary analysis and the blinded extension analysis.

14.1 Primary Endpoint Analyses

The primary endpoint, total number of Gd+ T1 lesions at Weeks 12, 16, 20, and 24, will be analyzed at the time of the primary analysis. The primary analysis will be based on the mITT analysis set. This analysis will be repeated for the ITT and PP analysis sets only if the respective analysis set differs from the mITT analysis set by > 4 subjects.

The primary analysis of total number of Gd+ T1 lesions, at Weeks 12, 16, 20, and 24, will be an estimate of lesion rate ratio, comparing each Evobrutinib dose group to placebo, together with associated 95% confidence interval (CI) and p-value, based on a negative binomial (NB) model, with offset equal to the log of number of scans. In general, missing scan assessments will be handled via the offset in the NB model, as the offset reflects the number of scans available for a

given subject. Details on missing data handling for subjects lacking any post-baseline scan assessments are provided in Section 9.

In this model, Evobrutinib dose group or placebo group will be a factor and there will be adjustment for baseline disease activity (i.e., presence/absence of Gd+ T1 lesions at baseline). The placebo treatment group and the lower category of baseline disease activity, will be used as references in the model. The p-value reported for each Evobrutinib treatment group tests the null hypothesis that adjusted lesion rate ratio, comparing Evobrutinib treatment to placebo, is equal to 1.

In the primary analysis, if a subject uses high dose corticosteroids in the 3 weeks prior to a scheduled MRI scan, lesion data from that scan will be excluded from the analysis and treated as missing.

The NB regression will be computed with the SAS® GENMOD procedure, using the DIST=NEGBIN option. In addition to the estimates of lesion rate ratio due to treatment and baseline disease activity, the estimates of the intercept and dispersion parameters will be reported.

Sensitivity Analyses

The first sensitivity analysis will analyse the Gd+ T1 lesion count at a single scan using a longitudinal Poisson approach (generalized linear mixed model) via PROC GLIMMIX (i.e., MODEL statement with DIST=POISSON) , based on the mITT analysis set. The model for the lesion rate corresponding to a single scan will include treatment (Evobrutinib dose group or placebo group) and visit (weeks 12, 16, 20, 24) as fixed effects, and a covariate for presence/absence of Gd+ T1 lesions at baseline. No offset parameter is needed as the lesion count data used to fit the model correspond to a single scan. A treatment-by-visit interaction will not be included (i.e., treatment effect assumed constant over time). As in the primary analysis, if a subject uses high dose corticosteroids in the 3 weeks prior to a scheduled MRI scan, lesion data from that scan will be excluded from the analysis and treated as missing. In general, missing scan assessments will be handled via the MAR assumption of the GLMM model. Details on missing data handling for subjects lacking any post-baseline scan assessments are provided in Section 9.

It is planned that both AR(1) correlation between repeated measures (i.e., RANDOM visit / SUBJECT=subjid TYPE=AR(1) RESIDUAL) and a random effect for subject (i.e., RANDOM subjid) will be modeled, yielding estimated lesion rate ratios that have a subject-specific, not population-average, interpretation. However, if there are convergence issues with this approach, the random effect for subject will be omitted, and the correlation between repeated measures will be modeled using a compound symmetric (i.e., RANDOM visit / SUBJECT=subjid TYPE=CS RESIDUAL) or variance components structure.

For each Evobrutinib dose group, the estimated lesion rate ratio relative to placebo will be reported, with 95% CI, and p-value. Assuming a random effect for subject is retained in the model, the p-value reported for each Evobrutinib treatment group tests the null hypothesis that adjusted lesion rate ratio, comparing Evobrutinib treatment to placebo, conditional on random effects, is equal to 1. Other statistics to be reported include the estimated lesion rate (adjusted for

visit and baseline), with 95% CI, for each dose group (Evobrutinib or placebo), estimated lesion rate ratio due to baseline disease activity, intercept, and covariance structure parameters.

The second sensitivity analysis will be based on the same NB model as in the primary analysis, applied to the mITT analysis set, but where all scans will be included in this analysis, regardless of subject use of high dose corticosteroids.

Supplemental Analyses

As a supplemental nonparametric analysis, the stratified Hodges-Lehman estimate of the shift in location of the lesion count distribution, comparing each Evobrutinib dose group to placebo, will be reported, together with associated 95% CI and p-value, based on a stratified Wilcoxon rank-sum test. Strata will be defined according to categorical baseline disease activity, following the approach to covariate adjustment in the NB model. In general, missing scan values for a subject will be imputed with the average of available scan values for that subject. Details on missing data handling for subjects lacking any post-baseline scan assessments are provided in Section 9.

An unstratified version of this nonparametric analysis will be reported as a sensitivity analysis. The second sensitivity analysis will be a stratified analysis that will include all scans, regardless of subject use of high dose corticosteroids.

The p-value reported for each Evobrutinib treatment group tests the null hypothesis that $P(X < Y) + 0.5*P(X = Y) = 0.5$, where X denotes the Gd+ T1 lesion count for a subject in a given Evobrutinib treatment group, Y denotes the Gd+ T1 lesion count for subjects in the placebo group, and P denotes probability.

If the NB model fails to converge, the results of the supplemental nonparametric analysis will be considered as the primary analysis. The supplemental analysis will be repeated for the ITT and PP analysis sets only if the respective analysis set differs from the mITT analysis set by > 4 subjects.

Multiplicity

The truncated Hochberg procedure will be used to adjust for multiplicity in testing the primary endpoint for the 3 Evobrutinib dose groups, to preserve alpha for testing key secondary endpoints (see Section 9). Both raw and multiplicity-adjusted p-values will be reported.

Dose-Response

A monotonic dose-response relationship, between ordered dose groups (placebo, Evobrutinib 25mg QD, Evobrutinib 75mg QD, and Evobrutinib 75mg BID) and ordered categories of Gd+ T1 lesion count for wks 12, 16, 20, 24, will be assessed via the Jonckheere-Terpstra trend test, both without and with stratification by BL factors (i.e., presence/absence of Gd+ T1 lesions at baseline). Prior to defining suitable categories for Gd+ T1 lesion count for wks 12, 16, 20, 14, missing scan data for a subject will be imputed using the same approach as used in the supplemental nonparametric analysis of this endpoint.

After imputation, the categories of Gd+ T1 lesion count (summed over wks 12, 16, 20, 24) will be defined by the quartiles of the distribution as observed when all subjects from the 4 treatment groups are combined. Thus, the unstratified table will be 4x4, and the stratified table will be 4x4x2, assuming a binary stratum covariate. The p-value reported tests the null hypothesis that the lesion count distribution is the same across dose groups, against the alternative that the location of the lesion count distribution is ordered (from largest value to smallest) according to increasing dose group.

In addition to the nonparametric test of monotonic dose-response, a test of linear trend in log lesion rate with increasing dose, adjusted for baseline, will be based on an appropriate contrast from the NB model (i.e., (3, 1, -1, -3)). The p-value reported tests the null hypothesis that the log lesion rate is the same across dose groups, against the alternative that the log lesion rate decreases linearly with increasing dose group order.

Descriptive Analyses

Descriptive statistics for the total number of Gd+ T1 lesions by timepoint through Week 24 will be provided by treatment group (including Tecfidera), based on the mITT analysis set. This analysis will be repeated for the ITT and PP analysis sets only if the respective analysis set differs from the mITT analysis set by > 4 subjects.

Mean number of Gd+ T1 lesions will be presented as a by-visit line plot for each treatment group, with a vertical line segment at each visit representing ± SE (jittered if needed for legibility), and with all treatment groups included in a single figure. Proportion of subjects who are Gd+ T1 lesion-free will be presented as a by-visit line plot for each treatment group (vertical line segment at each visit representing 95% CI may be omitted for legibility), and with all treatment groups included in a single figure. Mean total number of Gd+ T1 lesions from Weeks 12, 16, 20, and 24, will be presented as a by-treatment group bar chart, with a vertical line segment for each bar representing 95% CI, and with the p-value for the comparison with placebo (based on NB model) displayed above each evobrutinib dose group bar.

14.2 Secondary Endpoint Analyses

The analysis of secondary endpoints will be based on the mITT analysis set.

14.2.1 Key secondary endpoint analyses

The key secondary endpoints will be analyzed at the time of the primary analysis. The key secondary efficacy endpoints are:

1. Annualized relapse rate (ARR), based on protocol-defined qualified relapses, at Week 24. The unadjusted ARR is defined as the total number of qualified relapses divided by the number of subject-years of observation on treatment.

2. Qualified relapse-free status at Week 24

3. Change from baseline in EDSS score at Week 24

Collectively, safety endpoints are also considered to be a key secondary endpoint, but are not part of multiplicity adjustment required to control overall type I error rate at the two-sided α=0.05 level. Analyses of safety endpoints are described in Section 15. The multi-stage testing procedure for controlling family-wise error rate (FWER) across the 4 hypothesis families defined by the primary endpoint and the 3 key secondary efficacy endpoints, is described in Section 9.

Inferential analyses of key secondary endpoints will not include the Tecfidera arm. In modeling, the placebo group will be the reference treatment group. If an appropriate categorical BL MRI/relapse/EDSS assessment is adjusted for in the analysis of a key secondary endpoint, the lower disease-activity category will be the reference. For each of the key secondary endpoints, a test for a monotonic dose-response relationship and descriptive statistics will be provided. If the key secondary endpoint is analyzed via a parametric model, then a test of linear trend will also be reported.

ARR at Week 24

The comparison of a Evobrutinib treatment group to the placebo group using ARR at Week 24 will be based on a NB model for qualified relapse count, with offset equal to the log of years on study, with Evobrutinib dose group or placebo group as a factor and adjustment for categorical number of relapses in the last 2 years before randomization (≤ 1 relapse, > 1 relapse). Subjects discontinuing early are analyzed according to number of years of follow-up on treatment and number of qualified relapses observed at the time of discontinuation, including data (both follow-up time and relapse events) from the 4-week Safety Follow-up period. The NB regression will be computed with the SAS® GENMOD procedure, using the dist=NB option in the MODEL statement.

Two sensitivity analyses will be reported, based on early discontinuers, as described below:

- In the first sensitivity analysis, a subject discontinuing early without any qualified relapse in the 30 days prior to discontinuation will be counted as having a relapse on day of discontinuation.
- In the second sensitivity analysis, a subject discontinuing early without any qualified relapse in the 30 days prior to discontinuation will be counted as having a relapse on day of discontinuation if and only if the subject belongs to an Evobrutinib treatment group.

In the table summarizing inferential results (Evobrutinib dose group versus placebo group), the following will be reported for each group: number of qualified relapses, number of total relapses, follow-up on treatment (in subject-years), unadjusted ARR (number of qualified relapses divided by follow-up on treatment), and point estimate of adjusted ARR with 95% CI. For each treatment comparison (Evobrutinib group versus placebo group), the following will be reported: adjusted qualified relapse rate ratio with 95% CI, and p-value for the test that the qualified relapse rate ratio, adjusted for categorical BL relapse activity, is equal to 1.

A monotonic dose-response relationship, between ordered dose groups and ordered categories of ARR estimated on a subject level (i.e., number of qualified relapses experienced by a subject divided by the subject's follow-up), will be assessed via the

Jonckheere-Terpstra trend test, both without and with stratification by categorical BL relapse count (i.e., ≤ 1 relapse, > 1 relapse in the last 2 years before randomization). The categories of subject-level ARR will be defined by the quartiles of the distribution as observed when all subjects from the 4 treatment groups are combined. Thus, the unstratified table will be 4x4, and the stratified table will be 4x4x2, assuming a binary stratum covariate. The p-value reported tests the null hypothesis that the relapse count distribution is the same across dose groups, against the alternative that the location of the relapse count distribution is ordered (from largest value to smallest) according to increasing dose group.

In addition to the nonparametric test of monotonic dose-response, a test of linear trend in log relapse rate with increasing dose, adjusted for baseline, will be based on an appropriate contrast from the NB model (i.e., (3, 1, -1, -3)). The p-value reported tests the null hypothesis that the log relapse rate is the same across dose groups, against the alternative that the log relapse rate decreases linearly with increasing dose group order.

Descriptive statistics for the ARR at Week 24 will be provided by treatment group (including Tecfidera). In particular, at the treatment-group-level, number of qualified relapses, number of total relapses, follow-up on treatment (in subject-years), and unadjusted ARR (number of qualified relapses divided by follow-up on treatment) will be reported. In addition, descriptive statistics of subject-level ARR (number of qualified relapses experienced by a subject by week 24, divided by the follow-up experienced by a subject by week 24) will be summarized for each treatment group, in terms of mean, SD, median, Q1, Q3, min, and max.

ARR at Week 24 will be presented as a by-treatment group bar chart, with a vertical line segment for each bar representing 95% CI, and with the p-value for the comparison between Evobrutinib dose group and placebo (based on NB model) displayed above each Evobrutinib dose group bar.

Qualified relapse-free status at Week 24

The comparison of a Evobrutinib treatment group to the placebo group using proportion qualified relapse-free at Week 24, will be based on a logistic model for the odds of qualified relapse-free status at Week 24, with Evobrutinib dose group or placebo group as a factor and adjustment for categorical number of relapses in the last 2 years before randomization (≤ 1 relapse, > 1 relapse). Subjects who discontinue treatment prior to Week 24 without having a qualified relapse will be counted as not being qualified relapse-free at Week 24.

Two sensitivity analyses will be reported , based on early discontinuers, as described below:

- In the first sensitivity analysis, a subject who discontinues treatment prior to Week 24 without having a qualified relapse is considered as qualified relapse-free at Week 24.
- In the second sensitivity analysis, a subject who discontinues treatment prior to Week 24 without having a qualified relapse is considered as not qualified relapse-free if member of Evobrutinib group, and qualified relapse-free if member of placebo group.

In the table summarizing inferential results (Evobrutinib dose group versus placebo group), the following will be reported for each group: number (proportion) of subjects with ≥ 1

relapse through Week 24 (either treatment discontinuer or completer of 24 weeks of treatment), number (proportion) of subjects with 0 relapses through Week 24, number (proportion) of subjects discontinuing treatment prior to Week 24 among subjects with 0 relapses in that period., and number (proportion) of subjects who are relapse-free at Week 24, using imputation for treatment discontinuers, with 95% CI. For each treatment comparison (Evobrutinib group versus placebo group), the following will be reported: adjusted odds ratio (OR) with 95% CI, and p-value for the test that the OR, adjusted for categorical BL relapse count, is equal to 1.

A monotonic dose-response relationship, between ordered dose groups and proportion qualified relapse-free at Week 24, will be assessed via the Cochran-Armitage trend test, both without and with stratification by BL factor (i.e., ≤ 1 relapse, > 1 relapse in the last 2 years before randomization). The p-value reported tests the null hypothesis that the proportion qualified relapse-free is the same across dose groups, against the alternative that the proportion is ordered (from smallest value to largest) according to increasing dose group. In the Cochran-Armitage test, missing data are handled the same as in the main logistic modeling analysis.

In addition to the nonparametric test of monotonic dose-response, a test of linear trend in log odds of qualified relapse-free status with increasing dose, adjusted for baseline, will be based on an appropriate contrast from the logistic model (i.e., (3, 1, -1, -3)). The p-value reported tests the null hypothesis that the log odds of being qualified relapse-free is the same across dose groups, against the alternative that the log odds increases linearly with increasing dose group order.

Descriptive statistics for qualified relapse-free status up to Week 24 will be provided by treatment group (including Tecfidera), with and without stratification by BL factor (i.e., ≤ 1 relapse, > 1 relapse in the last 2 years before randomization). In particular, number (proportion) of subjects with ≥ 1 relapse, number (proportion) of subjects with 0 relapses, number (proportion) of subjects discontinuing treatment prior to Week 24 among subjects with 0 relapses will be reported.

For each treatment group, the Kaplan-Meier estimate of proportion surviving qualified relapse-free as a function of time (i.e., proportion qualified relapse-free between Week 0 and Week 24) will be presented, with all KM curves in a single figure. A subject discontinuing study prior to Week 24 without relapse will have his/her time to first relapse right-censored at the time of study discontinuation (i.e., the last time at which he/she is at risk of relapse on study). A subject completing 24 weeks of treatment without relapse will have his/her time to first relapse right-censored at 24 weeks. Below the horizontal axis, number of events and number of subjects at risk for the event will be depicted at each event time. The vertical axis may be restricted to 0.50 – 1.0, if none of the curves reach 50%. This KM estimate may be biased if the reason for study discontinuation is informative for relapse.

Change from baseline in EDSS score at Week 24

The comparison of a Evobrutinib treatment group to placebo group using CFB in EDSS at Week 24 will be based on a stratified Wilcoxon rank-sum test, with stratum defined by

categorical BL EDSS score (≤ 3.0 and > 3.0). A subject's missing value for EDSS CFB at Week 24 will be imputed by the median value among subjects having a Week 24 assessment, who are in the same treatment group, and who have the same value for the categorical baseline covariate.

Two sensitivity analyses will be reported as described below:

- Same analysis without any adjustment for strata.
- Same analysis using last observation carried forward (LOCF) post baseline to impute a missing value at Week 24. (Note that the only scheduled assessment of EDSS between baseline and Week 24 is at Week 12.)

In the table summarizing inferential results (Evobrutinib dose group versus placebo group), the following will be reported for each group: mean, median, and range of raw EDSS and EDSS CFB at Week 24. For each treatment comparison (Evobrutinib group versus placebo group), the following will be reported: p-value for the test that the shift in location of the distribution of the Week 24 EDSS CFB value, adjusted for categorical BL EDSS, is zero.

A monotonic dose-response relationship, between ordered dose groups (placebo, Evobrutinib 25mg QD, Evobrutinib 75mg QD, and Evobrutinib 75mg BID) and ordered categories of Week 24 EDSS CFB, will be assessed via the Jonckheere-Terpstra trend test, both without and with stratification by categorical BL EDSS. Prior to defining suitable categories for Week 24 EDSS CFB, missing values will be imputed as described in the main analysis for this endpoint. The categories of EDSS CFB will be defined by the quartiles of the distribution as observed when all subjects from the 4 treatment groups are combined. Thus, the unstratified table will be 4x4, and the stratified table will be 4x4x2, assuming a binary stratum covariate. The p-value reported tests the null hypothesis that the Week 24 EDSS CFB distribution is the same across dose groups, against the alternative that the location of the Week 24 EDSS CFB distribution is ordered (from largest signed CFB value to smallest signed CFB value) according to increasing dose group.

Descriptive statistics for the EDSS score at Week 24 will be provided by treatment group (including Tecfidera). In particular, mean, median, and range of raw EDSS and EDSS CFB at Week 24 will be reported. Mean CFB in EDSS will be presented as a by-visit line plot (BL, Week 12, Week 24) for each treatment group, with a vertical line segment at each visit representing ± SE (jittered if needed for legibility), and with all treatment groups included in a single figure.

14.2.2 Additional secondary endpoint analyses

Secondary efficacy endpoints : baseline to 24 weeks

The following additional secondary endpoints will be evaluated at the time of the primary analysis. Reported p-values will be considered nominal, as these endpoints are not included in the multiple testing strategy. There will be no sensitivity analyses, nor test for trend, as was the case for the primary and key secondary analyses.

Inferential analyses will not include Tecfidera arm, and placebo will be used as a reference for parameter estimates.

Descriptive statistics will be provided for these endpoints along with inferential results, as described below. Another table will describe these endpoints by treatment group (including Tecfidera), without imputed data.

Total number of new Gd+ T1 lesions at Weeks 12, 16, 20, 24

This endpoint will be analyzed similarly to the primary endpoint, using an NB model adjusted for baseline MRI value, with the offset based on the number of scans. Here the baseline MRI value will be presence/absence of Gd+ T1 lesions at baseline, as it is not possible to define "new" Gd+ T1 lesions at baseline. The same strategy as that used for the primary endpoint will be used.

Descriptive statistics for new Gd+ T1 lesions at Weeks 12, 16, 20, 24, similar to those for the primary endpoint, will be provided by treatment group (including Tecfidera). Mean number of new Gd+ T1 lesions will be presented as a by-visit line plot for each treatment group, with a vertical line segment at each visit representing the ±SE, and with all treatment groups included in a single figure.

Total number of new or enlarging T2 lesions at Weeks 12, 16, 20, 24

This endpoint will be analyzed similarly to the primary endpoint, using an NB model adjusted for baseline MRI value, with the offset based on the number of scans. Here the baseline MRI value will be categorical BL T2 lesion volume (≤ 13 cc, > 13cc), as it is not possible to define "new or enlarging" T2 lesions at baseline. In the present study, median T2 volume at baseline was 12.6 cc among the combined Evobrutinib/placebo group.

Descriptive statistics for new and newly enlarging T2 lesions at Weeks 12, 16, 20, 24, similar to those for the primary endpoint, will be provided by treatment group (including Tecfidera). Mean number of new or newly enlarging T2 lesions will be presented as a by-visit line plot for each treatment group, with a vertical line segment at each visit representing ±SE, and with all treatment groups included in a single figure.

Mean per-scan number of Gd+ T1 lesions at Weeks 12, 16, 20, 24

This endpoint will be analyzed similarly to the key secondary endpoint EDSS score CFB at Week 24, using a nonparametric approach. With this endpoint, no imputation is required if a subject is missing 1-3 post-baseline scans, as the endpoint is based on available scans. Details on missing data handling for subjects lacking any post-baseline scan assessments are provided in Section 9. The mean per-scan value (based on scans available from Weeks 12, 16, 20, 24) will be analyzed via Hodges-Lehmann estimation of the shift in distribution location due to Evobrutinib treatment group versus placebo. Each Evobrutinib dose group will be compared to placebo via a stratified Wilcoxon rank-sum test, where the stratification will adjust for categorical baseline disease activity, defined as presence/absence of Gd+ T1 lesions at baseline.

Descriptive statistics for the mean per-scan number of Gd+ T1 lesions, with mean value for each subject based on available scans for that subject at Weeks 12, 16, 20, 24, will be provided by treatment group (including Tecfidera), overall and by baseline disease activity subgroup.

Change from baseline in volume of Gd+ T1 lesions at Week 24

This endpoint will be analyzed similarly to the key secondary endpoint EDSS score CFB at Week 24, using a nonparametric approach, as neither endpoint is expected to be able to be transformed to follow a standard parametric distribution. The distribution of enhanced lesion volume is expected to have substantial probability mass at zero (van den Elskamp *et al* 2011). In the present study, median Gd+ T1 lesion volume at baseline was zero among the combined Evobrutinib/placebo group. The CFB in Gd+ T1 lesion volume at Week 24 will be analyzed via Hodges-Lehmann estimation of the shift in distribution location due to Evobrutinib treatment group versus placebo. Each Evobrutinib dose group will be compared to placebo via a stratified Wilcoxon rank-sum test, where the stratification will adjust for categorical baseline disease activity. A categorical covariate such as zero versus nonzero Gd+ T1 lesion volume at baseline will be considered. A missing value for CFB in Gd+ T1 lesion volume at Week 24 will be imputed using the median value among subjects having an assessment at Week 24, who are in the same treatment group and stratum.

Descriptive statistics for Gd+ T1 lesion volume CFB at Week 24 treated as a continuous variable (i.e., mean, median, and range), and as a categorical variable (i.e., number and proportion with decreasing, constant, or increasing volume), will be provided by treatment group (including Tecfidera). Mean CFB in volume of Gd+ T1 lesions will be presented as a by-visit line plot for each treatment group, with a vertical line segment at each visit representing ±SE, and with all treatment groups included in a single figure.

Change from baseline in volume of T2 lesions at Week 24

The Change from Baseline (CFB) in cube root transformed T2 lesion volume (van den Elskamp et al 2011), will be analyzed using MMRM via PROC MIXED. The model for mean CFB in the cube root of T2 lesion volume will include treatment (Evobrutinib dose group or placebo group), visit (weeks 12, 16, 20, 24), and treatment-by-visit interaction as fixed effects, a covariate for the baseline value of the cube root of T2 lesion volume, and account for correlation between repeated measures (i.e., residual random effects). The AR(1) covariance structure will be considered initially for the equally spaced repeated measures, but other structures (i.e., CS, VC) will be considered if there are convergence issues. For each Evobrutinib dose group, the LSMeans estimate of an appropriate contrast expressing the effect of Evobrutinib on mean cube root of T2 lesion volume CFB at week 24, relative to placebo, will be reported, with 95% CI, and p-value. In addition, the LSMeans estimate of the cube root of T2 lesion volume CFB at week 24 (adjusted for baseline), with 95% CI, will be reported for each dose group (Evobrutinib or placebo). In general, missing scan

assessments will be handled via the MAR assumption of the MMRM model. Details on missing data handling for subjects lacking any post-baseline scan assessments are provided in Section 9.

Descriptive statistics for T2 lesion volume at Week 24, similar to those for Gd+ T1 lesion volume at Week 24, will be provided by treatment group (including Tecfidera). Mean CFB in volume of T2 lesions will be presented as a by-visit line plot for each treatment group, with a vertical line segment at each visit representing ±SE, and with all treatment groups included in a single figure.

Secondary efficacy endpoints: baseline to 48 weeks

The following additional secondary endpoints will be evaluated at the time of the blinded extension analysis. When all subjects will have reached Week 48 or discontinued treatment prematurely, inferential analyses involving Week 48 endpoints will be performed. Reported p-values will be considered nominal, as type-1 error was not controlled for comparisons involving these endpoints.

For each of these endpoints, descriptive statistics at Week 48, similar to those provided at Week 24, will be provided for each treatment group, including Tecfidera, based on the mITT and the mITT-BE analysis sets.

Number of Gd+ T1 lesions at Week 48

The within-group comparison of number of Gd+ T1 lesions at Week 24 and Week 48 will be based on a stratified Wilcoxon signed-rank test, with stratum defined by presence/absence of Gd+ T1 lesions at baseline. A missing value for Gd+ T1 lesion count at Week 24 or Week 48 will be imputed using the median value among subjects having an assessment at that timepoint, who are in the same treatment group and stratum.

For each treatment group, the following will be reported: mean, median, and range of the Gd+ T1 lesion count at weeks 24 and 48, and a p-value for the test that the shift in location of the distribution of Gd+ T1 lesion count between Weeks 24 and 48, adjusted for baseline, is zero. For each dose group, the number of subjects with imputed Week 24 value, and number of subjects with imputed Week 48 value, will be reported.

The by-visit line plots for mean number of Gd+ T1 lesions, and proportion of subjects who are Gd+ T1 lesion-free, for weeks 0-24, provided at the time of the primary analysis, will be extended to include Week 48 at the time of the blinded extension analysis.

Mean total number of Gd+ T1 lesions at Week 24 and Mean total number of Gd+ T1 lesions at Week 48 will be presented as a by-treatment group bar chart, two bars for each treatment group displayed adjacently in a “cluster” (i.e., Week 24, Week 48), with a vertical line segment representing 95% CI extending from each bar, and with the p-value for the Wilcoxon signed-rank test displayed above each dose group “cluster” of 2 bars.

Number of new Gd+ T1 lesions at Week 48, Number of new and enlarging T2 lesions at Week 48, Change from baseline in Gd+ T1 lesion volume at Week 48, Change from baseline in T2 lesion volume at Week 48, Change from baseline in EDSS score at Week 48

These endpoints will be used to compare Week 48 with Week 24 using the same nonparametric approach as with "Number of Gd+ T1 lesions at Week 48." For "Number of new Gd+ T1 lesions at Week 48", the strata will be defined by presence/absence of Gd+ T1 lesions at baseline. For "Number of new and enlarging T2 lesions at Week 48", the strata will be defined by categorical BL T2 lesion volume (≤ 13 cc, > 13cc). For "Change from baseline in Gd+ T1 lesion volume at Week 48", the covariate zero/nonzero Gd+ T1 lesion volume at baseline will be considered for strata definition. For "Change from baseline in T2 lesion volume at Week 48", the strata will be defined by categorical BL T2 lesion volume (≤ 13 cc, > 13cc). For "Change from baseline in EDSS score at Week 48", the strata will be defined by categorical BL EDSS score (≤ 3.0, > 3.0).

A missing value at Week 24 or Week 48 will be imputed using the median value among subjects having an assessment at that timepoint, who are in the same treatment group and stratum.

The descriptive statistics, relevant by-visit lineplots, and barchart figure provided for the "Number of Gd+ T1 lesions at Week 48" endpoint will also be provided for these endpoints.

Qualified relapse-free analysis status at Week 48

The within-group comparison of qualified relapse-free status for Week 0-24 and Week 25-48 will be based on a stratified McNemar (SMN) test, with strata defined by categorical number of relapses in the last 2 years before randomization (≤ 1 relapse, > 1 relapse). Subjects who discontinue treatment between Week 24 and Week 48 without having a qualified relapse will be counted as not being qualified relapse-free for the period Week 25-48.

For each treatment group, the following statistics will be reported describing the period Week 25-48: number (proportion) of subjects with ≥ 1 relapse between Week 24 and Week 48 (either discontinuer or completer of 48 weeks of treatment), number (proportion) of subjects with 0 relapses between Week 24 and Week 48, number (proportion) of subjects discontinuing treatment between Week 24 and Week 48, among subjects with 0 relapses in that period, and number (proportion) of subjects who are relapse-free during Week 25 - 48, using imputation for treatment discontinuers, with 95% CI.

For each treatment group, the following statistics will be reported comparing the period Week 0-24 with the period Week 25-48:

- number (proportion) of subjects whose relapse-free status improves from the period Week 0-24 to the period Week 25-48 (using imputation for treatment discontinuers during Week 25-48), by stratum

- number (proportion) of subjects whose relapse-free status worsens from the period Week 0-24 to the period Week 25-48 (using imputation for treatment discontinuers during Week 25-48), by stratum
- p-value for the test that the proportion of subjects whose relapse-free status improves equals the proportion of subjects whose relapse-free status worsens (using imputation for treatment discontinuers during Week 25-48) adjusted for baseline.

For treatment groups that received a consistent regimen for the entire 48 weeks, the following will be reported:

- proportion relapse-free during Week 0-48, with 95% CI, without and with imputation for treatment discontinuers.
- Kaplan-Meier estimate of proportion surviving qualified relapse-free as a function of time (i.e., proportion qualified relapse-free between Week 0 and Week 48) for each treatment group, with all KM curves in a single figure. A subject discontinuing study prior to Week 48 without relapse will have his/her time to first relapse right-censored at the time of study discontinuation (i.e., the last time at which he/she is at risk of relapse on study). A subject completing 48 weeks of treatment without relapse will have his/her time to first relapse right-censored at 48 weeks. Below the horizontal axis, number of events and number of subjects at risk for the event will be depicted at each event time. The vertical axis may be restricted to 0.50 – 1.0, if none of the curves reach 50%. This KM estimate may be biased if the reason for study discontinuation is informative for relapse.

Annualized relapse rate at Week 48

For each treatment group, the following will be reported: number of qualified relapses during Week 25-48, number of total relapses during Week 25-48, follow-up (in subject-years) during Week 25-48, and unadjusted ARR during Week 25-48 (number of qualified relapses divided by follow-up), with 95% CI.

For treatment groups that receive a consistent regimen for the entire 48 weeks, the following will be reported: number of qualified relapses during Wk 0-48, number of total relapses during Week 0-48, follow-up (in subject-years) during Wk 0-48, and unadjusted ARR at Week 48 (number of qualified relapses divided by follow-up), with 95% CI.

The estimates of ARR at Week 24 and ARR during Week 25-48 will be presented as a by-treatment group bar chart, two bars for each treatment group displayed adjacently in a "cluster" (i.e., 0-24 week, 25-48 week), with a vertical line segment representing 95% CI extending from each bar. The number of subjects contributing to the ARR estimate will be displayed below each bar.

No within-group analysis related to the endpoint ARR at Week 48 is planned.

15 Safety Evaluation

The subsections in this section include specifications for summarizing safety endpoints that are common across clinical trials such as adverse events, laboratory tests and vital signs.

15.1 Adverse Events

Pre-treatment adverse events are defined as those AEs with an onset date on or after the date of informed consent and prior to the date of first dose of IMP.

Treatment-emergent adverse events (TEAEs) are defined as those AEs with an onset date on or after the date of first IMP administration.
The initial 24-week treatment period will include the following TEAEs:

- TEAEs with an onset date on or after first IMP administration and strictly before the first administration in the second 24-week period.
- All TEAEs for subjects discontinuing during the initial 24-week treatment period.

The second 24-week treatment period will include all TEAEs with an onset date on or after the first administration in the second period.

IMP related Adverse Events are those AEs with relationship to study treatment reported by the investigator as related or those of unknown relationship.

Serious Adverse Events are those events reported on the AE eCRF form with the serious field ticked "Yes" or with unknown seriousness.

Adverse Events leading to withdrawal of IMP are those AEs with action taken regarding study treatment as "Drug withdrawn" (as recorded on the AEs eCRF page).

Adverse Events leading to study termination are those AEs with other action taken as "Led to study termination" (as recorded on the AEs eCRF page).

Adverse events 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 AEs (NCI-CTCAE version 4.03) toxicity grades.

All Adverse Events recorded during the course of 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.

Group/SOC terms will be sorted alphabetically. Preferred terms within each group/SOC will be sorted by the highest active dose of Evobrutinib descending frequency, and alphabetically if multiple preferred terms have the same frequency.

If a subject experiences more than one occurrence of the same TEAE during the trial, the subject will only be counted 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.

All analyses described in Section 15.1 will be based on Treatment Emergent Adverse Events (TEAEs) if not otherwise specified.

15.1.1 All Adverse Events

A TEAE summary table will include a row for the overall frequency of TEAEs of the following types:

- any TEAE
- IMP-related TEAE
- serious TEAE
- IMP-related serious TEAE
- TEAE by intensity (NCI-CTCAE grade 1 to 4)
- IMP-related TEAE by intensity (NCI-CTCAE grade 1 to 4)
- TEAE leading to death
- IMP-related TEAE leading to death

Exposure adjusted incidence rates (EAIR) are calculated as the 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 in the corresponding time period. The incidence rate multiplied with 1000 would give the number of AEs expected in 1000 subjects within 1 time unit (year). EAIR of TEAEs will be presented by SOC and PT. If a subject has multiple events, the exposure period of the first event will be used. For a subject with no event, the exposure period will be censored at the last follow-up time within the AE summarization period.

The TEAE tables to be prepared are listed below:

	Table design					
	Overall frequency	By primary SOC and PT	By SOC only	By PT only	By HLGT only	By HLT only
TEAE Overview Summary	X	*NA*	*NA*			
TEAE by SOC and PT	X	X				
TEAE by SOC	X		X			
TEAE by PT	X			X		
TEAE by HLGT	X				X	
TEAE by HLT	X					X
IMP-related TEAE by SOC and PT	X	X				

Non-Serious TEAE by SOC and PT*	X	X				
TEAE by worst grade, SOC and PT		X				
IMP-related TEAE by worst grade, SOC and PT		X				
EAIR of TEAEs	X	X				

*A table with all 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.

Pretreatment and TEAEs will be listed separately by treatment group and subject.

Three-tier Approach to Summarizing and Analysing AEs

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.

This study will not include Tier 1 analyses, as no Tier 1 AEs were identified prior to the bulk of the data being collected. Tier 3 analyses will not be provided, due to the relatively small sample size per treatment group. Only Tier 2 AEs will be defined and the respective analyses provided.

AEs will be classified into Tier 2 based on the Rule-of-3. If there are 3 or more patients with the reported term in any treatment group, that term will be included in Tier 2.

For Tier 2 AEs, the difference in crude rates (between Evobrutinib dose group and placebo group), and the CI for the difference, will be reported. The CI will be based on the unconditional MN method (Miettinen 1985, G.F. Liu 2006). No multiplicity adjustment will be applied for Tier 2 AEs.

EAIR of TEAEs will be presented for Tier 2 AEs, by SOC and PT. The difference in EAIR between Evobrutinib dose group and placebo group will be summarized.

15.1.2 Adverse Events Leading to Treatment Discontinuation

A TEAE summary table will include a row for the overall frequency of TEAEs of the following types:

- 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 dose reduction of Tecfidera
- IMP-related TEAE leading to dose reduction of Tecfidera
- TEAE leading to concomitant medication
- IMP-related TEAE leading to 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

The TEAE tables to be prepared are listed below:

	Table design		
	Overall frequency	By primary SOC and PT	By PT only
TEAE leading to discontinuation of IMP/study Overview Summary	X	*NA*	*NA*
TEAE leading to IMP withdrawal by SOC and PT	X	X	
IMP-related TEAE leading to IMP withdrawal by SOC and PT	X	X	
TEAE leading to study termination	X	X	
IMP-related TEAE leading to study termination	X	X	

A listing of TEAEs leading to withdrawal of IMP, and a listing of TEAEs leading to study termination, will be provided.

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)

TEAEs leading to death and IMP-related TEAEs leading to death will be tabulated by SOC and PT along with overall frequency. 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

Serious TEAEs and IMP-related serious TEAEs will be tabulated by SOC and PT along with overall frequency.

A subject listing of serious TEAEs will be provided.

15.3 Clinical Laboratory Evaluation

The clinical laboratory safety parameters measured in this trial , and reported as part of the safety evaluation of hematology, biochemistry, and urinalysis, are specified in the protocol (Table 5 of the Clinical Trial Protocol, Section 7.4.3). Ig levels, B, CCI cell count are not considered in this section.

Summaries of Clinical Laboratory Data

Laboratory results will be classified according to the grades defined in NCI-CTCAE Version 4.03 as provided by the central laboratory. If a laboratory parameter has bi-directional toxicities (eg, Potassium), both directions will be presented for the given parameter (ie, Potassium Low and Potassium High). On-treatment values are defined as results of assessments made after the first IMP administration on Day 1. Laboratory results containing a modifier such as "<" or ">=" will be handled case by case for summary statistics and will be reported as collected in the database in subject data listings.

All parameters will also be presented according to the categories based on normal ranges: below normal limits (Low), within normal limits (Normal) and above normal limits (High).

Protocol-specified clinical laboratory parameters (hematology, biochemistry, urinalysis) will be summarized using descriptive statistics appropriate to continuous-valued random variables (see Section 9) by time point and by treatment group. At each time point, both observed value and CFB value will be summarized.

Shift tables of baseline value (low, normal, high) versus worst on-treatment value, presented by treatment group, will be provided for hematology, biochemistry, and urinalysis parameters.

Shift tables of baseline grade versus worst post-baseline grade, presented by treatment group, will be provided for hematology and biochemistry parameters.

Directional laboratory tests (hematology, biochemistry) (Appendix 1) will be summarized by worst on-treatment value and by treatment group.

Graphical Display of Clinical Laboratory Data

Boxplots of laboratory values by treatment arm and timepoint will be provided for the following hematology and biochemistry parameters:

- hemoglobin
- white blood cell count
- absolute neutrophil count
- absolute lymphocyte count
- platelet count
- Alanine Amino-Transferase (ALT)

- Aspartate Amino-Transferase (AST)
- Alkaline Phosphatase (ALP)
- total bilirubin
- Gamma-Glutamyl Transferase (GGT)
- glucose
- creatinine
- sodium
- potassium
- calcium
- lipase
- amylase

Boxplots for the laboratory parameters listed above 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.

For the primary analysis, a by-subject line plot of log ALT by time point (through Week 24), one curve per subject, (i.e., "spaghetti" plot) will be provided. One panel will include subjects from the Evobrutinib treatment groups, displayed so that curves from subjects in the same group have the same color or line type. A second panel will include subjects from the Tecfidera and placebo groups, displayed in a similar manner.

Kaplan-Meier estimates of proportion surviving event-free as a function of time will be presented for the following 3 types of events based on the ALT parameter:

- Time from first dose to first assessment of ≥ Grade 2 ALT (days)
- Time from first dose to first increase in ALT ≥ 1 grade above BL grade (days)
- Time from first dose to first increase in ALT ≥ 2 grade above BL grade (days)

For the primary analysis, the 3 time-to-event figures will be based on data from weeks 0-24; a subject reaching week 24 without experiencing the event will have his/her event time right-censored at the time of the last ALT assessment. A separate KM curve will be estimated for each group: Evobrutinib (any dose), placebo, tecfidera.

For the blinded extension analysis, the 3 time-to-event figures will be based on data from weeks 0-48; a subject reaching week 48 without experiencing the event will have his/her event time right-censored at the time of the last ALT assessment. A separate KM curve will be estimated for each group: Evobrutinib (any dose), tecfidera (i.e., no curve for the group of subjects who switch from placebo to Evobrutinib at week 24).

To be included in a time-to-event figure, a group must have at least one event. Below the horizontal axis of the time-to-event figure, # of events and # of subjects at risk will be displayed at each event time. The vertical axis may be restricted to 0.50 – 1.0, if none of the curves reach 50%.

eDISH (Evaluation of Drug-Induced Serious Hepatotoxicity) figures, as described in Merz *et al* (2014), will be presented for both the primary analysis (limited to data from weeks 0-24) and the blinded extension analysis (all data from weeks 0-48). The figure produced for the primary analysis will include 6 panels: one for each Evobrutinib dose group, one for all dose groups combined, one for placebo, and one for tecfidera. The figure produced for the blinded extension analysis will include only 5 panels, as there will not be a panel for the subjects who switch from placebo to Evobrutinib at week 24.

Listings of Clinical Laboratory Data

By-subject listings of all individual hematology, biochemistry, urinalysis and coagulation values present in the database will be provided.

Laboratory values that are outside the normal range will be flagged in the data listings, 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 (not to include values for Specific Gravity or pH parameters), or an increase of "++" for non gradable parameters when applicable.

15.4 Vital Signs

Vital signs (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) for baseline, each applicable time point and CFB to each time point.

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

Table 21: Vital sign 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)

For the 12-lead ECG parameters listed below, observed values and CFB values will be summarized by treatment group using descriptive statistics for continuous random variables:

- Ventricular rate (beats/min)
- Pulse rate interval (msec)
- QRS (msec)
- QT (msec)
- Fridericia corrected QT (QTcF) (msec).

The QTcF parameter will be categorized by observed value into the categories

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

and by CFB value into the categories

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

Number and percentage of subjects within each category listed above, based on their post-baseline QTcF data, will be presented by treatment group.

A shift table of rhythm results (Sinus rhythm, Atrial fibrillation, Other, Missing) from baseline to last on treatment category, presented by treatment group, will be provided.

A shift table of morphological assessments (Normal, Abnormal, Missing) from baseline to maximum on treatment category, presented by treatment group, will be provided.

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

15.6 Chest X-ray Evaluations

Investigator reported abnormalities in chest x-ray will be listed, refer to Section 11.3.2.

15.7 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.

15.8 Safety/Pharmacodynamic Endpoints: CCI and Immunoglobulin levels

Safety/pharmacodynamic endpoints include CCI and immunoglobulin (Ig) levels. CCI Immunoglobulin (IgM, IgA, and IgG and subclass) levels will be assessed prior to the morning dose on Day 1, Day 28, Day 112, Day 168, and Day 336 (End of Treatment).

For each SafetyCCI endpoint, observed value, change from baseline, will be summarized. Descriptive statistics for these variables will be presented as described in Section 9.

15.9 Pregnancy test

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

16 Analyses of Other Endpoints

16.1 CCI

[redacted]

16.2 CCI

16.2.1 CCI

[redacted]

CCI

16.2.3 CCI

[redacted]

16.2.3.1 [redacted]

[redacted]

16.2.3.2 [redacted]

[redacted]

CCI

[redacted]

[redacted]

16.2.3.3 CCI [redacted]

[redacted]

16.2.3.4 CCI [redacted]

[redacted]

16.2.3.5 CCI [redacted]

[redacted]

16.2.3.6 CCI [redacted]

[redacted]

16.2.4 CCI [redacted]

[redacted]

16.2.5 CCI [redacted]

[redacted]

16.3 CCI [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

[redacted]

CCI [redacted]

17 References

1. Hauser SL, Waubant E, Arnold DL, et al. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. N Engl J Med. 2008; 358(7):676-88.

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

3. 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.

4. Beal SL, Sheiner LB. NONMEM Users Guides. NONMEM project group, University of California, San Francisco, CA, 1992.

5. Dmitrienko A, Kordzakhia G, Tamhane AC. Multistage and mixture parallel gatekeeping procedures in clinical trials. Journal of Biopharmaceutical Statistics 2011; 21: 726-747.

6. Dmitrienko A, Tamhane AJ. Chapter 5: Gatekeeping Procedures in Clinical Trials. https://web.njit.edu/~wguo/Math654_2012/DTB_Chapter5.pdf

7. Healy B, Chitnis T, Engler D. Improving power to detect disease progression in multiple sclerosis through alternative analysis strategies. J Neurol 2011; 258: 1812-1819.

8. Van den Elskamp IJ, Knol DL, Uitdehaag BMJ, Barkhof F. Modeling MR imaging enhancing-lesion volumes in multiple sclerosis: application in clinical trials. Am J Neuroradiol. 2011; 32: 2093-97

9. 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.

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

11. 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.

12. Wahlby U, Bouw MR, Jonsson EN, Karlsson MO. Assessment of type I error rates for the statistical sub-model in NONMEM. J Pharmacokinet Pharmacodyn 2002;29:251-69

13. Wahlby U, Jonsson EN, Karlsson MO. Assessment of actual significance levels for covariate effects in NONMEM. J Pharmacokinet Pharmacodyn 2001;28:231-52

14. Karlsson MO, Sheiner LB. The importance of modeling inter-occasion variability in population pharmacokinetic analyses. J Pharmacokinet Biopharm 1993;21(6):735-50.

15. Merz M, Lee KR, Kullak-Ublick GA, Brueckner A, Watkins PB. Methodology to assess clinical liver safety data. Drug Safety 2014; 37 (Suppl 1):S33-S45.

16. Jonsson EN, Karlsson MO. Automated covariate model building within NONMEM. Pharm Res 1998;15(9):1463-8.

18 Appendices

18.1 Appendix 1: Worst on treatment value based on normal range of laboratory evaluations

Names of Clinical Safety Laboratory Evaluations in Protocol version 1.0, 05 July 2016	If gradable parameter, corresponding evaluation names in NCI-CTCAE 4.03	Worst on treatment value based on normal range
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
Estimated Glomerular Filtration Rate		LOW
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

Names of Clinical Safety Laboratory Evaluations in Protocol version 1.0, 05 July 2016	**If gradable parameter, corresponding evaluation names in NCI-CTCAE 4.03**	**Worst on treatment value based on normal range**
Hematocrit		LOW/HIGH
Hemoglobin	Hemoglobin increased	HIGH
Hemoglobin	Anemia	LOW
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
B, CCI cell count		LOW
Immunoglobulin and subclass concentrations		LOW
Total IgG		LOW
Total IgA		LOW
Total IgM		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
pH		NA
Nitrite		NA
Urobilinogen		NA
Bilirubin (urinalysis)		NA
Glucose (urinalysis)		NA
Ketones bodies		NA
Protein (urinalysis)		NA
Microscopy: white blood cells		HIGH
Microscopy: red blood cells		HIGH
Microscopy: casts		NA

Names of Clinical Safety Laboratory Evaluations in Protocol version 1.0, 05 July 2016	If gradable parameter, corresponding evaluation names in NCI-CTCAE 4.03	Worst on treatment value based on normal range
Protein/creatinine ratio		HIGH

18.2 CCI

[redacted]


CCI


CCI

CCI

CCI

18.3 Appendix 3: Details on Pharmacometry Evaluation

CCI

CCI


CCI

CCI


CCI

MS200527-0086 Integrated Analysis Plan v1

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'GMT'Z)
PPD	Technical Approval	14-Feb-2018 21:16 GMT+01