Integrated Analysis Plan

Clinical Trial Protocol Identification No.	MS700568_0022
Title	**A 2-year prospective study to evaluate the onset of action of Mavenclad® in subjects with highly active relapsing multiple sclerosis**
Trial Phase	IV
Investigational Medicinal Product(s)	Mavenclad®
Clinical Trial Protocol Version	12 February 2019 / Version 2.0
Integrated Analysis Plan Author	**Coordinating Author** Biostatistics, Merck PPD **Function** / **Author(s) / Data Analyst(s)** Biostatistics, PPD / PPD
Integrated Analysis Plan Date and Version	13 November 2020 / Version 2.0 (final)
Integrated Analysis Plan Reviewers	**Function** / **Name** Lead Biostatistician / PPD Medical Responsible / PPD Medical Responsible (back-up) / PPD PPD / PPD PPD / PPD PPD / PPD Clinical Trial Leader, PPD / PPD PPD PPD / PPD Biostatistician, PPD / PPD PPD / PPD

CCI

Signature Page

Integrated Analysis Plan: MS700568_0022

A 2-year prospective study to evaluate the onset of action of Mavenclad® in subjects with highly active relapsing multiple sclerosis

Approval of the IAP by all Merck Data Analysis Responsible is documented within ELDORADO. With the approval within ELDORADO, the Merck responsible for each of the analysis also takes responsibility that all reviewers' comments are addressed adequately.

Merck responsible	Date	Signature
PPD, Lead Biostatistician	Via ELDORADO approval process	

1 Table of Contents

2 List of Abbreviations and Definition of Terms

9HPT	9-Hole Peg Test
AE(s)	Adverse Event(s)
ARR	Annualized Relapse Rate
ARR_{qual}	Qualifying relapses in ARR
CC	Complete-Case
CDP	Confirmed Disability Progression
CI	Confidence Interval
CRF	Case Report Form
CCI	CCI
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed Tomography
CUA	Combined Unique Active
DMD	Disease Modifying Drug
dp	Decimal place
CCI	CCI
eCRF	Electronic Case Report Form
EDSS	Expanded Disability Status Scale
FAS	Full Analysis Set
CCI	CCI
FS	Functional System
GCP	Good Clinical Practice
HIV	Human Immunodeficiency Virus
HRA	High-relapse activity
IAP	Integrated Analysis Plan
ICH	International Council for Harmonisation
IEC	Independent Ethics Committee
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IMP	Investigational Medicinal Product
IRB	Institutional Review Board
iRT	Interactive Response Technology
ITT	Intention-To-Treat
KFS	Kurtzke Functional System
MAR	Missing at Random
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MTR	Magnetization Transfer Ratio
NCI-CTCAE	US National Cancer Institute Common Terminology Criteria for Adverse Events
NEDA	No Evidence of Disease Activity
NEPAD	No Evidence of Progression or Active Disease
CCI	CCI

CCI	CCI
ON	Optic Neuritis
PBVC	Percentage brain volume change
PDGMVC	Percentage deep grey matter volume change
PGMVC	Percentage grey matter volume change
PWMVC	Percentage white matter volume change
PML	Progressive Multifocal Leukoencephalopathy
PT	Preferred Term
RMS	Relapsing Multiple Sclerosis
RRMS	Relapsing Remitting Multiple Sclerosis
SAE	Serious Adverse Event
SD	Standard Deviation
SDMT	Symbol Digit Modalities Test
SOC	System Organ Class
SUSAR	Suspected Unexpected Serious Adverse Reaction
T1 Gd+	T1 gadolinium-enhancing
T25FW	Timed 25-Foot Walk
TB	Tuberculosis
TEAE	Treatment-Emergent Adverse Event
VZV	Varicella Zoster Virus
WHO-DD	World Health Organization Drug Dictionary
WOCBP	Women of Childbearing Potential

3 Modification History

Unique Identifier for Version	Date of IAP Version		Author	Changes from the Previous Version
1.0	Date of last signature		PPD	New Document
2.0	Date of last signature		PPD	• Author and reviewers list updated • Tertiary endpoint: Ratio of Post-Baseline vs Baseline T1 Gd+ lesion count added to Interim and Final Analysis (see Section 14.3.3) • Tertiary Endpoint: T1 Gd+ lesion count > 0 added to Interim and Final Analysis (see Section 14.3.3) • Covariates: Time between scans (in Months) added (see Section 9) • Tertiary endpoint: Alternative covariance structures added (see Section 14.3)

4 Purpose of the Integrated Analysis Plan

The purpose of this Integrated Analysis Plan (IAP) is to document technical and detailed specifications for the primary analysis at the Month 6 interim analysis and the final analyses of data collected for protocol MS700568_0022. 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 may 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 IAP is based upon Section 8 (Statistics) of the study protocol and protocol amendments and is prepared in compliance with ICH E9. It describes analyses planned in the protocol and protocol amendments. An additional Baseline analysis is planned once all subjects have completed baseline assessments, which is not mentioned in the protocol and will be described and reported separately.

CCI [redacted]

5 Objectives and Endpoints

	Objective	Endpoint	IAP section
Primary Objective	To determine the onset of action of Mavenclad® in subjects with highly active relapsing multiple sclerosis (RMS).	Differences in the counts of combined unique active (CUA) magnetic resonance imaging (MRI) lesions during the first 6 months (i.e., during periods months 1-6, 2-6, 3-6) compared to baseline (i.e., the period screening to baseline).	14.1
Secondary Objective	To assess the effect of Mavenclad® on different immune system composites in particular cell subtypes count and repopulation.	Characterization of immune cell subsets count at the end of 3, 6, 12, 15, 18 and 24 months compared to baseline.	14.2
Tertiary Objectives	To assess the safety and tolerability of Mavenclad®.	• Occurrences of treatment-emergent adverse events (TEAEs) and serious adverse events (SAEs) including and up to 24 months. • Lymphocyte count up to and including Month 24.	15
	To assess the effect of Mavenclad® on the progression of disability, cognition and incidence of relapse.	• Symbol Digit Modality Test (SDMT) • Expanded Disability Status Scale (EDSS) • Kurtzke Functional System (KFS) • 9-Hole Peg Test (9HPT) • Timed 25-Foot Walk (T25FW) • Annualized relapse rate (ARR) • Changes in CUA lesions • Number of CUA lesions • Changes in active T1 Gd+ (T1 gadolinium-enhancing) lesion count • Volume changes of T1 Gd+ lesions • Number of T1 hypointense lesions • Change in volume of T1 hypointense lesions • Changes in new T2 lesion count • Responder rate during the different periods as defined above with responder being defined as subjects with a CUA lesion count reduction of at least 1 • Changes in T2 lesion volume • Changes in Magnetization Transfer Ratio (MTR) • Changes in brain volume • No Evidence of Disease Activity (NEDA) • No Evidence of Progression or Active Disease (NEPAD)□ • CCI [redacted]	14.3
	CCI [redacted]		

6 Overview of Planned Analyses

The following analyses are planned for the study:

- Interim analysis after all subjects have completed the 6-month assessment,
- Final analysis at the end of the study.

In addition, a Baseline analysis has been performed. This Baseline analysis is not specified in the study protocol and is described in a separate Analysis Plan.

CCI

CCI

CCI

6.1 Interim Analysis

An interim analysis for primary endpoint that includes all planned sensitivity analyses will be undertaken after all subjects have completed the 6-month assessment.

In addition, the Interim Analysis will describe the following characteristics of the study population up to the cutoff:

- Disposition of Participants and Discontinuations (Section 10.1),
- Major Protocol Deviations (Section 10.2),
- Demographics and Other Baseline Characteristics (Section 11):
    - Baseline Characteristics (Section 11.3),
    - Medical History (Section 11.2),
    - MS Disease Characteristics (Section 11.3),
    - Vital Signs (Section 11.3).
- Previous or Concomitant Medications/Procedures (Section 12),
- Treatment Compliance and Exposure (Section 0),

and will evaluate the following tertiary and safety endpoints up to 6 months visit or the cut-off point:

- Number of CUA lesions during the Baseline and post-Baseline periods (Section 14.3.2),
- T1 Gd+ lesion count by visit and mean T1 Gd+ lesion count during the Baseline and post-Baseline periods (Section 14.3.3),
- Changes in Standardized total active T2 lesion count during the post-Baseline periods compared to the Baseline period (Section 14.3.7),

- Responder rate during the different periods (Section 14.3.8),
- Relapse reports including and up to the cut-off point (Number and Percentage of Subjects with (qualifying) relapses),
- Adverse Events (Section 15.1),
- Clinical Laboratory Evaluation (Section 15.3),
- Vital signs (Section 15.4).

The characteristics of the population as well as relapses will be summarized descriptively up to Month 6 Visit in the same way as for the final analysis. The tertiary MRI endpoints and Safety endpoints will be analyzed up to the Month 6 Visit or cut-off date in the same way as for the final analysis, as applicable.

Table 1 lists all assessments which will be included in the interim analysis.

Table 1: List of assessments included in the interim analysis

Assessments	Screening	Baseline	Month 1	Month 2	Month 3	Month 6
Informed Consent	X					
Demographics	X					
Weight		X	X			
Vital Signs	(X)	X				
Medical History	X					
Disease History	X					
Relapse History		X				
MRI	X	X	X	X	X	X
Disability (EDSS/KFS)	X	X				
Relapse count			X	X	X	X
Lymphocyte Count	X			X		X
Treatment Administration		X	X			
AEs		X	X	X	X	X
Concomitant medications and procedures	X	X	X	X	X	X
Relevant previous medications	X					
Serology (incl. Varicella and TB)	X					
Pregnancy test	X	X				

Assessments	Screening	Baseline	Month 1	Month 2	Month 3	Month 6
Study termination		X	X	X	X	X

Data cleaning will be undertaken on all data involved in interim reporting. Details and level of data cleaning will be documented in data management plans.

6.1.1 Cut-off date

Only data up to the Month 6 Visit of each subject will be analyzed in the interim analysis. Therefore, the cut-off date is defined as the maximum Month 6 Visit date (i.e., date of the last assessment for the Month 6 Visit). For subjects with a missing Month 6 Visit the cut-off date will be 180 days after start of study medication. This includes subjects who discontinued before Day 180.

- Cut-off date (Month 6 Visit done) = Maximum Month 6 Visit date,
- Cut-off date (Month 6 Visit not done) = Start of study medication + 180 days - 1.

All occurrences of death will be included in the interim analysis regardless of the timepoint (i.e., any death that is included in the database at the time of the snapshot will be reported in the interim analysis).

6.1.2 Data handling after cut-off date

Data after cut-off do not undergo the cleaning process.

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

If the stop date for an AEs is after the date of cut-off, the AEs will be considered as ongoing during the Interim Analysis.

6.2 Final Analysis

The final analysis will include all described analyses of this IAP. The final analysis will be based on all data available in the database at the time of the final database lock.

7 Changes to the Planned Analyses in the Clinical Trial Protocol

The following decision was made prior to the last enrolled subject starting study treatment, with post-treatment data unavailable at the time of the decision making:

Primary analysis will be complemented by a non-parametric analysis methods as a data review of baseline data has shown that the primary endpoint (Differences in the counts of CUA MRI lesions during the first 6 months, that is, during period months 1-6, 2-6, 3-6, compared to Baseline, from Screening to end of Baseline assessment), does not follow Normal distribution.

8 Protocol Deviations and Analysis Sets

8.1 Definition of Protocol Deviations

Major protocol deviations are protocol deviations (PD) 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.

Major protocol deviations and any action to be taken regarding the exclusion of subjects, or affected data from specific analyses, are defined in the project-specific Protocol Deviation Specification.

All protocol deviations are documented in Study Data Tabulation Model (SDTM) datasets whether identified through site monitoring, medical review or programming. Tables and listings of PDs are defined in Section 10.2.

8.2 Definition of Analysis Sets and Subgroups

Enrolled Set (ES)

All subjects enrolled in the study, i.e., all subjects with a signed informed consent.

Intent-to-treat Set (ITT)

All subjects classified as eligible (i.e., medication assigned by iRT at the Baseline Visit).

Full Analysis Set (FAS)

All subjects in the ITT treated with at least one dose of study medication**.**

Per Protocol Set (PP)

All subjects from the FAS who fulfil the highly active relapsing MS inclusion criteria (i.e.; all subjects from the FAS without a PD for the highly active relapsing MS inclusion criteria).

Treatment Completer Set – Year 1 (TCS-1)

All subjects from the FAS who completed the full Treatment Course of the first year (Treatment Course 1). Subjects will be classified as Treatment Completer for the first year if they have a compliance greater than or equal to 100% for Treatment Course 1.

Treatment Completer Set – Year 2 (TCS-2)

All subjects from the FAS who completed the full Treatment Course of the first and the second year. Subjects will be classified as Treatment Completer for the second year if they have a compliance greater than or equal to 100% for Treatment Course 1 and Treatment Course 2 and received the treatment within 90 days (end of study medication – start of study medication in Treatment Course 2 +1 ≤ 90 days).

Safety Analysis Set (SAF)

All subjects treated with at least one dose of study medication.

Unless otherwise specified, all efficacy analyses will be performed on the FAS. The primary and secondary analyses will be repeated on the Treatment Completer Sets and the primary analysis, on the ITT (if different from FAS), to evaluate the sensitivity of the primary results to deviations of the planned treatment regimen. Further details are given in the respective sections. Selected analyses planned for the end of the study will be repeated on the Per-Protocol Set. Details will be specified in an amendment to the IAP. All safety analyses will be performed on the SAF.

Additional Subgroup or Subset Analysis Sets

Subgroup analyses will be performed on subgroups as defined below.

- High-relapse activity (HRA)
    - HRA: at least 2 relapses in the previous year (i.e. the number of historical relapses within 12 months prior to Baseline ≥2) regardless of prior use of DMDs
    - Non-HRA: otherwise

  Unless otherwise indicated, all descriptive statistical analyses will be presented separately for the HRA subgroups.

- Previous treatment with Disease Modifying Drugs (DMDs):
    - DMD pre-treated: Subjects will be categorized as DMD pre-treated if they have taken DMDs any time before start of study medication. Details of the definition of DMDs are specified in Appendix 1 (Section 18.1).
    - Pre-treatment naïve: all subjects not classified as DMD pre-treated.

  The primary analysis and selected demographic and background characteristics as well as selected tertiary analyses will be repeated for this subgroup.

The following subset(s) are defined:

- Baseline Period CUA count > 0 (i.e. all subjects with a CUA lesion count > 0 during the Baseline Period)

  As subjects with a CUA count of 0 during the Baseline Period cannot show an improvement in the number of CUA lesions during the study, some of the efficacy analyses will be repeated for this subset that has at least one CUA lesion during the Baseline Period.

9 General Specifications for Data Analyses

Treatment groups

The treatment group will be labeled as Mavenclad®.

Significance level

The three hypotheses of the primary analysis (see Section 0) will be tested as one -sided, on a 2.5% significance level. The testing procedure will stop as soon as one of the hypotheses cannot be rejected following a pre-specified order. Due to this sequential order of tests an adjustment for a potential type -I -error inflation due to the multiple testing is not required.

All other statistical tests mentioned in this IAP are to be regarded as exploratory. All statistical tests will be performed two-sided. If confidence intervals are to be calculated, these will be two -sided with a confidence probability of 95%, unless otherwise specified in this IAP.

Presentation of continuous and qualitative variables

Continuous variables will be summarized using descriptive statistics, i.e., number of subjects, number of subjects with missing and non-missing values, mean with standard deviation, median, 25th Percentile - 75th Percentile (Q1-Q3), minimum, and maximum.

Qualitative variables will be summarized by counts and percentages.

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

In case where the analysis refers only to certain Visits, percentages will be based on the number of subjects still present in the study at that Visit, unless otherwise specified.

Definition of Baseline

The last non--missing measurement prior to start of study medication will serve as the Baseline measurement. Measurements performed on the day of first intake of study medication are assumed to be “prior to start of study medication”, as long as it has not been confirmed that the procedure was performed afterwards. This assumption is justified as following the sequence of procedures as specified in the study protocol. The intake of study medication should be the last procedure on the visit day.

The MRI scans will be reconciled during central MRI review process. Therefore, MRIs will be considered as Baseline MRIs if assigned to the Baseline Visit by the central imaging review.

Definition of Study Periods

- Baseline Period: The Baseline Period is defined as the period between the Screening and the Baseline MRI scan as provided by the central MRI review.
- Post-Baseline Period: The post-Baseline Periods that are used in the primary analysis are defined as follows:
  - Period 1 (p_1) is defined as the period between Month 1 and Month 6 MRI scans.
  - Period 2 (p_2) is defined as the period between Month 2 and Month 6 MRI scans
  - Period 3 (p_3) is defined as the period between Month 3 and Month 6 MRI scans

  In addition, the following periods are defined for tertiary analyses:
  - Period of Month 6 to Month 12,
  - Period of Month 1 to Month 12,
  - Period of Month 18 to Month 24,
  - Period of Month 1 to Month 24,

Additional periods might be added with an Amendment to this IAP after the respective MRI Charters are finalized.

Definition of change from Baseline

Change from Baseline = Visit value – Baseline value

Percent Change from Baseline = 100 × (Visit value – Baseline value) / Baseline value

Definition of visit dates

The visit start date is the date of the first assessment which belongs to the corresponding visit.

The visit end date is the date of the last assessment which belongs to the corresponding visit.

Definition of duration

Duration will be calculated by the difference of start and stop date + 1, in days (unless specified otherwise).

Treatment Period

The treatment period begins with the start of study medication on Study Day 1 and continues through to the completion of the treatment period at the Month 24 Visit (see CSP, Section 7.1.2).

Start of study medication is the date of first intake of study medication. Start of study medication in Year 2 is the date of first intake of study medication at or after the Month 12 Visit.

Treatment Period:

- Start = Start of study medication,
- End = Visit end date of Month 24 Visit, or Early Termination (ET) Visit, in case of early discontinuation.

Course 1 Treatment Period:

- Start = Start of study medication,
- End = Start of study medication at or after the Month 12 Visit - 1 day, or End of Treatment Period, if subject discontinued prior to receiving Treatment Course 2.

Course 2 Treatment Period:

- Start = Start of study medication at or after the Month 12 Visit,
- End = End of Treatment Period.

Definition of Delayed Treatment Course 2

The Treatment Course 2 will be defined as delayed, if the start of study medication is delayed by more than 3 months (=90 days) after the Month 12 Visit, i.e.:

- Start of Study medication of Treatment Course 2 > visit start date of Month 12 Visit Date + 90 days

Data handling in case of a Delayed Treatment Course 2

For subjects with a Delayed Treatment Course 2, only data observed during the Course 1 Treatment Period will be analyzed in all efficacy analyses. Efficacy data after the Course 1 Treatment Period will be counted as missing in the statistical analyses and marked in the listings. For subjects without a delay the whole Treatment Period will be analyzed.

For all other analyses (except efficacy) the data is not limited to the Course 1 Treatment Period for subjects with a Delayed Treatment Course 2.

Conversion factors

The following conversion factors will be used to convert days into months or years for all observations except Visit related dates:

- 1 month = 30.4375 days,
- 1 year = 365.25 days.

Time window

- Day 1 is the day of start of study medication, the day before is Day - 1 (no Day 0 is defined),

- Study day / Treatment day is defined relative to Day 1.

The visit schedule is defined as follows:

- 1 month = 30 days,
- 1 year = 360 days,

relative to the start date of the Baseline Visit for Treatment Course 1 and relative to the start date of the Month 12b Visit for Treatment Course 2.

In each analysis, should assessments be delayed, subjects' assessment data will be treated as though it occurred at the time point specified (i.e., out of visit windows will be not excluded or adjusted for any analyses) and inferences will be conducted accordingly. Thus, further time windows will not be specified.

All observations from unscheduled Visits will be included only in the analyses of Visit independent summaries (e.g. counts of events, progressions or worst value during Treatment Period) unless for the last assessment available. In this missing case the ET visit will be imputed by the unscheduled visit.

The assessments of the ET visit will be mapped to the nearest planned post-baseline visit with a missing assessment before or after the ET visit of the individual assessment for all statistical analyses and descriptive summaries. As not all assessments of the ET are performed at each visit, the different assessments of the ET can be assigned to different visits. If the duration (in days) between the ET visits and two missing planned visits are equal, then the assessment will be assigned to the later visit.

Handling of missing data

Unless otherwise specified in respective sections, missing data will not be replaced.

The handling of missing covariate and stratification factor information is described below.

If the medication start date for Treatment Course 1 Cycle 1, Treatment Course 1 Cycle 2 or Treatment Course 2 is missing in the clinical database even though it is confirmed that study medication was administered, the date from the iRT call will be used as start date of study medication.

Incomplete or missing onset dates of relapses from "MS Relapse Report" will be imputed as follows:

- In cases where the onset date is partially missing but the onset month and year, or the onset year are equal to the start of study medication or the date is missing then the onset date will be replaced by the start date of study medication.
- In all other cases the missing onset day or missing onset month will be replaced by 1 (but not before start of study medication).

In all subject data listings, imputed values will be presented and flagged.

Missing statistics, e.g. when they cannot be calculated, should be presented as "nd". For example, if n=1, the measure of variability (e.g., SD) cannot be computed and should be presented as "nd".

Where tables are presented over time, the total number of subjects with missing and non-missing observations at each time-point should reflect the population still in the study at that time. This does not apply when imputations are made beyond study withdrawal.

Data Handling for MRI data

Primary endpoint(s) and its components:

The MRI lesions will be evaluated by independent central review, which will provide the following measurements for the primary endpoints (see Independent Review Charter for Counting Brain MRI Lesions):

- Total number of T1 Gadolinium enhancing lesions (T1 Gd+) (or Not Evaluable [NE], if applicable) at the first and last MRI of the respective period,
- Total number of new T2 lesions with T1 Gd+ (or NE, if applicable) of the respective period (comparing the last with the first MRI of the respective period),
- Total number of enlarging T2 lesions with T1 Gd+ (or NE, if applicable) of the respective period (comparing the last with the first MRI of the respective period),
- Total number of new T2 lesions without T1 Gd+ (or NE, if applicable) of the respective period (comparing the last with the first MRI of the respective period),
- Total number of enlarging T2 lesions without T1 Gd+ (or NE, if applicable) of the respective period (comparing the last with the first MRI of the respective period).

The statistical analysis will use the results from the Global assessment (i.e. Readtype=GLOBAL).

The assignment of MRIs to the respective visits (e.g. assignment of EOT MRIs or delayed MRIs due to steroid treatment) will be ensured by the central reading process. MRIs taken while on Steroid Treatment as identified as a major PD will be excluded from the analysis.

All derived lesion counts (except for the Mean T1 Gd+ lesion count) as specified below will be classified as NE if one of the components required for the calculation has been classified as NE by the central reading. The Mean T1 Gd+ lesion count will be set to NE only, if all T1 Gd+ lesion counts that correspond to the respective period are classified as NE. In all other cases the mean will be calculated based on the available counts during the respective period.

T1 Gd+ lesion count

As with this procedure a visit is assessed multiple times for the T1 Gd+ and thus T1 Gd+ will be reported multiple times per visit, the value for a visit of the earliest available period will be selected for the statistical analysis. As the Global Review will ensure that all duplicates have identical lesion count, this rule does not have an impact on the number of T1 Gd+ but defines further attributes (e.g. date of review).

The T1 Gd+ counts for a period will be defined as the sum of the T1 Gd+ lesions from all scheduled MRI scans reported for a Visit that is included in a Period divided by the actual number of MRI scans with a reported lesion count (i.e. not NE) during a Period.

- T1 Gd+ lesion count (by Visit) = T1 Gd+ lesions count from the MRI at the respective visit

- Mean T1 Gd+ lesion count (by Period per scan) = T1 Gd+ lesion count during the Period / # of scans belonging to the respective Period with non-missing T1 Gd+ lesion count.

MRIs are assigned to a Period as follows:

- Baseline Period:
  - Screening Visit
  - Baseline Visit

- Period 1:
  - Month 1 Visit
  - Month 2 Visit
  - Month 3 Visit
  - Month 6 Visit

- Period 2:
  - Month 2 Visit
  - Month 3 Visit
  - Month 6 Visit

- Period 3:
  - Month 3 Visit
  - Month 6 Visit

Active T2 lesion count

As the new or enlarging T2 lesion count depends on the length of the period, it will be presented as reported (non-standardized) and standardized to 1 months (30.4375 days):

- Period length (days) = MRI date at end of Period – MRI date at start of Period +1,
- Standardized total T2 lesion count (by Period per 1 months) = Total number of T2 lesions per Period * 30.4375 / Period length.

The following types of T2 lesions are defined per Period (standardized and non-standardized):

- Total new T2 lesion count =
  Total number of new T2 lesions with T1 Gd+
  + Total number of new T2 lesions without T1 Gd+
- Total enlarging T2 lesion count =
  Total number of enlarging T2 lesions with T1 Gd+
  + Total number of enlarging T2 lesions without T1 Gd+,

- Total active T2 lesion count =
  Total new T2 lesion count
  + Total enlarging T2 lesion count

The total active T2 Lesion count will also be reported for the Baseline, Month 2, Month 3, and Month 6 Visit and is defined as:

- Total active T2 Lesion count = Total number of new or enlarging T2 lesions since the scan of the previous Visit

and will be calculated as follows:

- Total active T2 Lesion count at the Baseline visit
  = Total active T2 Lesion count of the Baseline Period
- Total active T2 Lesion count at the Month 2 visit
  = Total active T2 Lesion count of Period 1 (M1-M6) – Period 2 (M2-M6)
- Total active T2 Lesion count at the Month 3 visit
  = Total active T2 Lesion count of Period 2 (M2-M6) – Period 3 (M3-M6)
- Total active T2 Lesion count at the Month 6 visit
  = Total active T2 Lesion count of Period 3

For the calculation of the CUA lesion count, only T2 lesions that did not emerge from a T1 Gd+ lesion will be taken into account.

- Total active T2 lesion count (without T1 Gd+) =
  Total new T2 lesion count without T1 Gd+
  + Total enlarging T2 lesion count without T1 Gd+

and will be standardized to the length of the Period. Total active T2 lesion count (without T1 Gd+) will only be calculated if T1 Gd+ counts for beginning and end of the corresponding period are non-missing.

CUA lesion count

The number of CUA lesions for a period will be defined as the sum of the T1 Gd+ and new or enlarging T2 counts for the respective period and will be standardized to one scan (for T1) or one month (for T2) to harmonize the units of the 2 assessments.

$\#CUA(p_i)$ = Standardized CUA lesion count (by Period i per 1 month)
= Mean T1 Gd+ lesion count (by Period i per Visit)
+ Standardized total active T2 lesion count (without T1 Gd+) (by Period i per 1 months)

The CUA lesion count will only be calculated if both components (T1 and T2) are non-missing.

Tertiary endpoints

The details of the tertiary MRI endpoints will be defined later by an Amendment to this IAP after the respective MRI charter has been finalized.

Data Handling in case of major PDs

PDs will not be considered in the planned statistical analyses (i.e. all data will be analyzed independent from any reported PD) except for PD due to MRI taken during steroid treatment (MRI will be excluded from analysis) and PD regarding the inclusion criteria of highly active relapsing MS (subjects will be excluded from PP).

Covariates and stratification factors

Age

Age (years) at time of Informed Consent (loaded from iRT). Missing age will be replaced as follows if used as a covariate:

- Age will be estimated from the year of birth (year of IC – year of birth), as recorded in the electronic Case Report Form (eCRF),
- If year of birth is missing, age will be imputed by the mean age of the FAS.

EDSS at Baseline

Expanded Disability Status Scale (EDSS) at Baseline will be categorized by:

- ≤3 (reference),
- >3.

Missing EDSS at Baseline will be imputed by the median Baseline EDSS of the FAS if used as a stratification factor.

Pooled centers

Some centers will have few eligible subjects only. Therefore, the centers will be pooled together by country except for

- sites that have more than 12 subjects in the FAS (i.e. these sites will not be pooled with any other site).
- for countries that have 6 subjects or less in the FAS that will be pooled as follows:
  - Scandinavia = Finland and Sweden
  - Mediterrannea = Italy and Spain
  - Austria-Hungary = Austria and Hungary

Time between scans

Time between scans (in Months) is calculated as follows:

- Time between scans (in Months) = (MRI date of the Visit - MRI of the previous Visit + 1) /30.4375. i.e.:
    - If (Visit=Month 1) then Time between scans (in Months) = (MRI date at Month 1 - MRI date at Baseline + 1)/ 30.4375
    - If (Visit=Month 2) then Time between scans (in Months) = (MRI date at Month 2 - MRI date at Month 1 + 1)/ 30.4375
    - If (Visit=Month 3) then Time between scans (in Months) = (MRI date at Month 3 - MRI date at Month 2 + 1)/ 30.4375
    - If (Visit=Month 6) then Time between scans (in Months) = (MRI date at Month 6 - MRI date at Month 3 + 1)/ 30.4375

If the corresponding MRI is available and the time between scans is missing (due to preceding MRI being missing) it will be replaced by expected time as follows:

- Month 1/2/3 Visit =1
- Month 6 Visit =3.

Software(s)

All analyses will be performed using SAS® Software version 9.3, or higher.

10 Trial Participants

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

10.1 Disposition of Participants and Discontinuations

Subject disposition will be presented for the ES:

- Total number of subjects enrolled (i.e., subjects who gave informed consent),
- Number of subjects who discontinued from the study prior to treatment overall and grouped by the main reason (i.e., the Reason why the subject did not continue beyond screening, as recorded in the CRF. If the Subject did not meet all Eligibility Criteria, the failed specific inclusion or exclusion criteria will be provided in addition). Subjects not classified as screened failure but without treatment assignment by the iRT will be listed as “no treatment assigned”,
- Number of eligible subjects (i.e., subjects with treatment assignment by the iRT system),
- Overall and by Treatment Course, who received at least dose one of study medication,
- Who discontinued the study medication, grouped by main reason (according to the information on the Study Termination CRF page),
- Who completed the study (according to the information on the Study Termination CRF page).

Percentages will be presented with respect to the number of eligible subjects. The number of subjects in each analysis set will be provided overall, by region, by country within region, and by site.

The following subject data listings will be provided:

- Listing of discontinued subjects,
- Listing of subjects excluded from the Analysis Sets.

The number of subjects with least one available T1 GD+ lesion count, T2 lesion count and CUA lesion count for each Study Period will be presented for the FAS together with the statistics for the length of the Study Periods, and statistics of the number of available T1 Gd+ lesion counts during the Study Periods.

10.2 Protocol Deviations

10.2.1 Important Protocol Deviations

The following summary tables and listings of major protocol deviations (see Section 8.1 for details) will be provided by Inclusion/Exclusion and other deviations for the ITT:

- Frequency table summarizing major protocol deviations,
- Listing of major protocol deviations.

11 Demographics and Other Baseline Characteristics

Unless stated otherwise, summaries will be presented on the FAS.

11.1 Demographics

Demographic characteristics will be summarized using the following information from the Screening/Baseline Visit CRF pages.

Demographic characteristics:

- Gender:
  - Male,
  - Female,
  - Missing.
- Race:
  - White,
  - Black or African American,
  - Asian,
  - American Indian or Alaska Native,
  - Native Hawaiian or Other Pacific Islander,
  - Other,
  - Not collected at this site,
  - Missing.
- Ethnicity Hispanic or Latino:
  - Yes,
  - No,
  - Missing.
- Ethnicity Japanese:
  - Yes,
  - No,
  - Missing.

- Age (years)
- Age categories:
  - ≤ 40 years,
  - 40 years, < 65 years,
  - ≥ 65 years,
  - Missing.
- Region:
  - Eastern & Central Europe: Austria, Germany, Hungary, Poland, Czech Republic
  - Scandinavia: Finland, Sweden
  - Mediterranea: Italy, Spain
  - Western Europe: France, United Kingdom
  - Australia,
  - Canada,
  - Israel.
- Country
  - Country 1
  - …
- Pooled Centers

The following subject data listing will be provided:

- Demographic Data (Country, Sex, Race, Age, EDSS at Baseline, Number of relapses within 12 months prior to Baseline).

11.2 Medical History

The medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) in the latest available version. Summary table will be presented by primary system organ class (SOC) and preferred term (PT).

11.3 Other Baseline Characteristics

MS Disease Characteristics

The duration since onset and diagnosis of MS at Baseline in months, as well as the EDSS at Baseline, will be presented. In addition, the number of subjects grouped by the major systems affected (at onset of MS) will be presented.

- Time since onset of MS (months) = (visit start date of Baseline Visit - date of first symptom + 1) / 30.4375,
- Time since diagnosis (months) = (visit start date of Baseline Visit - date of diagnosis + 1) / 30.4375,
- Time since first relapse (months) = (visit start date of Baseline Visit - date of first relapse + 1) / 30.4375,
- Number of previous DMDs (i.e. number of different PTs of DMDs as entered into the "Relevant Previous Medication Form" (see Appendix 1)),
  - 0
  - 1
  - 2
  - >2
- EDSS at Baseline,
- EDSS at Baseline,
  - ≤3
  - >3
  - Missing
- Number of historical relapses within 12 months prior to Baseline (i.e., number of relapses from the "MS Relapse History Baseline" eCRF page):
  - Baseline Visit Start Date ≥ onset date ≥ Baseline Visit Start Date - 396. Days, if the day of the onset is available,
  - Month of the Baseline Visit Start Date ≥ onset month ≥ month of Baseline Visit Start Date – 13 months, if only the month of onset is available.

- Number of historical relapses within 12 months prior to Baseline categorized to:
  - 0
  - 1
  - 2
  - >2
- High-relapse activity (HRA)
  - Yes
  - No

Missing days or months of the dates of first symptom, diagnosis, first relapse will be replaced by 1.

MRI Baseline Characteristics

The results of the independent MRI review for the Baseline Period will be presented:

T1 Gd+ lesion count:

- Mean T1 Gd+ lesion count for the Baseline Period
- Mean T1 Gd+ lesion count for the Baseline Period
  - 0
  - >0
  - Non-evaluable
  - Missing
- T1 Gd+ lesion count at the Screening Visit
- T1 Gd+ lesion count at the Screening Visit
  - 0
  - 1
  - >1
  - Non-evaluable
  - Missing
- T1 Gd+ lesion count at the Baseline Visit

- T1 Gd+ lesion count at the Baseline Visit
    - 0
    - 1
    - >1
    - Non-evaluable
    - Missing

Active T2 lesion count:

- Total active T2 lesion count at the Baseline Visit
- Total active T2 lesion count at the Baseline Visit
    - 0
    - 1
    - >1
    - Non-evaluable
    - Missing
- Standardized total active T2 lesion count for the Baseline Period
- Standardized total active T2 lesion count for the Baseline Period
    - 0
    - >0
    - Non-evaluable
    - Missing
- Standardized total active T2 lesion count (without T1 Gd+) for the Baseline Period
- Standardized total active T2 lesion count (without T1 Gd+) for the Baseline Period
    - 0
    - >0
    - Non-evaluable
    - Missing

CUA lesion count:

- Standardized CUA lesion count for the Baseline Period
- Standardized CUA lesion count for the Baseline Period

- 0
- >0
- Non-evaluable
- Missing

Vital Signs

The following Baseline characteristics will be provided in summary tables:

- Height, weight, body surface area, body mass index
- Body temperature, heart rate, Systolic Blood Pressure, Diastolic Blood Pressure

The body surface area (BSA) and body mass index (BMI) will be calculated with the following equations:

- BSA(m^2) = ([height(cm) * weight(kg)] / 3600) ^ 0.5
- BMI (kg/m^2) = [weight (kg) / height ($cm)^2$] * 10000

12 Previous or Concomitant Medications/Procedures

Concomitant medications are medications, other than study medications, which are taken by subjects during the study.

Concomitant medications will be summarized from the "Concomitant medication" eCRF page. ATC-1st level and preferred term will be tabulated as given from the WHO-DD dictionary current version. In case multiple ATC's are assigned to a drug, all ATC-1st level will be used for reporting. All medications entered into the "Concomitant medication" eCRF page will be assumed to be "concomitant" irrespective of the actual start and end dates. In addition, all concomitant medications as well as concomitant medications given for disease related conditions will be presented by PT and by decreasing frequency.

Relevant previous medications are medications, other than study medications, which stopped before enrolment (i.e. date of informed consent [IC]). DMDs (see Section 8.2) administered between IC and start of study medication are major PDs that will be reported as relevant previous medications as well.

Relevant previous medications will be summarized from the "Relevant previous medication details" eCRF page. ATC-1st level and preferred term will be tabulated as given from the WHO-DD dictionary current version. In case multiple ATC's are assigned to a drug, all ATC-1st level will be used for reporting. All medications entered into the "Relevant previous medication details" eCRF page will be assumed to be "previous" irrespective of the actual start and end dates. In addition, all relevant previous medications and relevant previous medications given for disease related conditions will be presented by PT and decreasing frequency.

All relevant previous medications which are classified as DMDs will be presented by PT. DMDs which are taken within 6 months prior to start of study medication (i.e. DMD end date ≥ start of study medication – 182.625) will be presented by PT. The length of the washout period (start of study medication - DMD end date + 1) of the last DMD within 6 months prior to start of study medication will be summarized. The last DMD will be defined as the DMD with the latest stop date. If the DMD end date is partially missing, the day will be imputed by the 15th of the month but the date will be truncated to IC date -1 if the imputation results in a later date. If the month is missing as well or the date is completely missing the end date will not be replaced and it assumed that the DMD stopped earlier than 6 months prior to start of study medication.

Concomitant procedures which were undertaken any time on study will be listed by subject.

13 Treatment Compliance and Exposure

Two years of treatment are planned in this study. Subjects will take tablets in week 1 (Cycle 1) and week 5 (Cycle 2) of each year. All dosing calculations and summaries will be based on "Treatment Administration" and "Treatment Termination" eCRFs pages.

The following summary tables will be provided for the SAF:

- Duration of therapy,
- Compliance and Drug Accountability

The duration will be presented for Treatment Course 1 Treatment Period, Course 2 Treatment Period and the Treatment Period in days.

- Duration (days)= (date of last tablet in period – date of first tablet in period + 1)

Counts of tablets and compliance will be calculated for each treatment week (cycle), year (course) and total for all subjects that started the respective treatment period (i.e. taken at least one tablet).

- Compliance (%) = 100 * number of tablets taken/ number of planned tablets

In addition, compliance will be categorized as follows:

- < 60%
- [60%-80%[
- [80%-90%[
- [90%-110%]
- >110%

The number of planned tablets is based on the weight at the beginning of each treatment year (Table 2).

Table 2: Dose of Mavenclad® per treatment week by subject weight in each treatment year

Weight range	Dose in mg (number tablets) per treatment week	
	Treatment week 1	Treatment week 5
40 to <50	40 mg (4 tablets)	40 mg (4 tablets)
50 to <60	50 mg (5 tablets)	50 mg (5 tablets)
60 to <70	60 mg (6 tablets)	60 mg (6 tablets)
70 to <80	70 mg (7 tablets)	70 mg (7 tablets)
80 to <90	80 mg (8 tablets)	70 mg (7 tablets)
90 to <100	90 mg (9 tablets)	80 mg (8 tablets)
100 to <110	100 mg (10 tablets)	90 mg (9 tablets)
110 and above	100 mg (10 tablets)	100 mg (10 tablets)

The number of subjects with the same dosing schema in all Treatments Courses and those with a different dosing schema in Treatment Course 1 Cycle 2, and in Treatment Course 2 compared to planned dose based on weight range calculated at start of Treatment Course 1 Cycle 1 will be presented for all subjects who started the respective treatment period.

The time difference between the start of the Course 1 Treatment Period and the start of the Course 2 Treatment Period will be calculated as follows:

- Relative Start of Treatment Course 2 (days) = Start of Treatment Course 2 - Start of Treatment Course 1 + 1.

The Relative Start of Treatment Course 2 will be presented together with the number of subjects in the following groups for the SAF:

- Treatment Course 2 not started,
    - and discontinued before or at the Month 12 Visit,
    - and discontinued after Month 12 Visit
        - Lymphocyte count < 800 cells/mm³ at Month 12 Visit,
- Treatment Course 2 started,
    - within 3 months after Month 12 Visit,
    - delayed by more than 3 months but less than 6 months
        - and any Lymphocyte count < 800 cells/mm³ at or after Month 12 Visit and before start of Treatment Course 2.
    - delayed by more than 6 months,
        - and any Lymphocyte count < 800 cells/mm³ at or after Month 12 Visit and before start of Treatment Course 2.

The delay of Treatment Course 2 will be related to the visit start date of the Month 12 Visit (see Section 9).

The cumulative dose is defined as the sum of the number of tablets taken multiplied by 10 mg divided by the weight for each visit:

Cumulative dose = number of tablets taken in Treatment Course 1 Cycle 1 * 10mg / weight in kg at Baseline + number of tablets taken in Treatment Course 1 Cycle 2 * 10mg / weight in kg at Month 1 Visit + number of tablets taken in Treatment Course 2 * 10mg / weight in kg at Month 12b Visit.

The cumulative dose will be calculated only if the weight and number of tablets are available. Otherwise it will be missing. If the Treatment Course 2 or Treatment Course 1 Cycle 2 wasn't started (i.e. no tablet taken), the cumulative dose will be calculated only for the Treatment Course 1 or Treatment Course 1 Cycle 1.

A subject data listing of the treatment compliance will be provided for the SAF.

14 Efficacy Analyses

The following sections describe the planned analyses for each of efficacy endpoints.

For the detailed definition of covariates and stratification factors that are used in the statistical models see Section 9. Details about the derived MRI endpoints and general handling of MRI data is described in Section 9 (Data Handling of MRI data).

Descriptive summary statistics will be presented for the absolute values and the changes from baseline as applicable by visit, or period, for all efficacy endpoints for the FAS and respective subgroups or subsets for which the variables are analyzed. Details about the presentation of continuous and categorical variables are given in Section 9.

Wherever possible, all estimated results from the statistical models will be calculated with the same observed margin which reflects the population of the whole corresponding analysis set, regardless of missing data.

14.1 Primary Endpoints: Differences in the counts of CUA MRI lesions during the first 6 months compared to Baseline

The primary endpoints for the trial are defined as the differences in counts of CUA MRI lesions during the first 6 months compared to Baseline, i.e., the following differences will be evaluated:

1. Period of Month 1 to Month 6 compared to the Baseline period.
2. Period of Month 2 to Month 6 compared to the Baseline period,
3. Period of Month 3 to Month 6 compared to the Baseline period,

For the detailed definition of Periods, please see Section 9 “Data Handling of MRI data”.

14.1.1 Primary Objective: Analysis of primary endpoints

The differences in counts of CUA lesions will be evaluated using a mixed-effects linear model, accounting for within-center/region correlation through a hierarchical model. The analysis will adjust for baseline factors deemed to be prognostic for the primary endpoint, as judged by clinical experts:

- CUA lesion count during the Baseline period,
- Age,
- EDSS at Baseline (≤3/>3).

The results of the mixed-effects linear model will be used to describe the effect size. A review of baseline data has shown that the distribution of the primary endpoint data does not follow Normal distribution. Consequently, the hypotheses will be tested by a Wilcoxon signed-rank test which employs ranks of absolute differences from baseline for each testing period (i.e., p_1, p_2, and p_3) using average ranks for tied values. .

The analysis will also adjust for the difference in length and number of MRIs of the periods to be compared (see Section 9 “Data Handling of MRI data” for details of lesion count standardization).

The difference in the counts of CUA lesions during the three periods compared to the baseline period will be tested in a sequential order. The three hypotheses are defined as follows:

$$H_{0i}: \Delta\,\mathrm{CUA}(\mathrm{p}_i) \geq 0 \quad \text{vs.} \quad \Delta\,\mathrm{CUA}(\mathrm{p}_i) < 0, \qquad i = 1,2,3$$

$$\Leftrightarrow H_{0i}: \#\mathrm{CUA}(\mathrm{p}_i) \geq \#\mathrm{CUA}(\mathrm{p}_0) \text{ vs. } \#\mathrm{CUA}(\mathrm{p}_i) < \#\mathrm{CUA}(\mathrm{p}_0), \qquad i = 1,2,3$$

with

$$\Delta\mathrm{CUA}(\mathrm{p}_i) = \#\mathrm{CUA}(\mathrm{p}_i) - \#\mathrm{CUA}(\mathrm{p}_0), \qquad i = 1,2,3,$$

$\#\mathrm{CUA}(\mathrm{p}_i)$: standardized CUA lesion count during period p_i, $\quad i = 0,1,2,3,$

p_0: Baseline period (period from Screening to Baseline),

p_i: period Month i to Month 6, $\quad i = 1,2,3.$

The sequential testing procedure will start with period p_3, followed by period p_2 and p_1, i.e., the hypotheses will be tested in the following order:

1. H_{03}
2. H_{02}
3. H_{01}

The three hypotheses will be tested one-sided on a 2.5% significance level. The testing procedure will stop as soon as one of the hypotheses cannot be rejected following the pre-specified order. Due to this sequential order of tests an adjustment for a potential type-I-error inflation due to the multiple testing is not required.

An estimate for the difference in CUA lesions during the specified periods compared to the baseline period will be reported, together with one-sided p-values. In addition, 95% CI and 2-sided p-values will be presented to allow for comparison with the 2-sided secondary and tertiary analyses.

The LS Means will be calculated with observed margin which reflects the distribution of age, EDSS categories at Baseline and respective Baseline Period CUA Lesion of the FAS. An unstructured covariance matrix will be assumed for the model. If the model doesn't converge a variance components covariance structure will be used instead or the results of sensitivity analysis with center effect as a fixed effect will be used instead.

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Primary endpoints: Differences in the counts of CUA MRI lesions during the first 6 months (i.e., during periods months 1-6, 2-6, 3-6) compared to baseline (i.e., the period screening to baseline)			
Primary (FAS)	1. Period 3 (Month 3-6): $\Delta CUA(p_3)$ (Hypothesis 3)	P-values for the three Hypotheses from the Wilcoxon signed-rank test will be presented. The testing procedure will stop as soon as one of the hypotheses cannot be rejected following the pre-specified order (H3, H2, H1). Due to this sequential order of tests an adjustment for a potential type-I-error inflation due to the multiple testing is not required. The Wilcoxon signed-rank test will be complemented by the estimated Least Square Means (LS Means) and corresponding Standard Errors (SE) including one-sided p-values as well as two-sided p-value and 95% Cis from the mixed-effects linear model with the following adjusting factors: • Baseline score, • Age (in years), • EDSS at Baseline (≤3, >3), • Within-pooled center correlation. In addition, the distribution of $CUA(p_i)$ and $\Delta CUA(p_i)$ will be displayed graphically using box plots.	Missing observations: Subjects with a missing CUA lesion count at either the Baseline period or the respective post-baseline period will not be included in the analysis.
	2. Period 2 (Month 2-6): $\Delta CUA(p_2)$ (Hypothesis 2)		
	3. Period 1 (Month 1-6): $\Delta CUA(p_1)$ (Hypothesis 1)		
$\Delta CUA(p_i)$ = Difference in the standardized CUA lesion count between Period i and the Baseline Period			

14.1.2 Primary Objective: Sensitivity Analyses of the primary endpoints

The robustness of the results of the primary analysis will be evaluated by the following

- Multiple Imputation to handle missing data using different assumptions,
- Non-parametric factorial model for longitudinal data,
- Generalized model assuming negative binomial distribution,
- Mixed-effects linear model with pooled-centers as fixed effects.

To evaluate further the impact of the potential zero inflation, the primary endpoint will be classified in different response categories for each Period as follows:

- Presence of MRI activity
    - Yes (CUA lesion count > 0)
    - No (CUA lesion count = 0)
- Standardized CUA lesion count classified to:
    - 0 (standardized CUA lesion count = 0)
    - >0 – 1 (standardized CUA lesion count > 0 and less or equal to 1)
    - >1 (standardized CUA lesion count > 1)

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Primary endpoints: Differences in the counts of CUA MRI lesions during the first 6 months (i.e., during periods months 1-6, 2-6, 3-6) compared to Baseline Period (i.e., the period from Screening to Baseline).			
Sensitivity (FAS by subgroup "High-relapse activity")	Period 3 (Month 3-6): $\Delta CUA(p_3)$	Same as in the Primary Analysis (see Section 14.1.1)	Same as in the Primary Analysis (see Section 14.1.1)
	Period 2 (Month 2-6): $\Delta CUA(p_2)$		
	Period 1 (Month 1-6): $\Delta CUA(p_1)$		
Sensitivity (FAS by subgroup "Previous treatment with DMDs")	Period 3 (Month 3-6): $\Delta CUA(p_3)$		
	Period 2 (Month 2-6): $\Delta CUA(p_2)$		
	Period 1 (Month 1-6): $\Delta CUA(p_1)$		
Sensitivity (FAS for subset "Baseline Period CUA count > 0")	Period 3 (Month 3-6): $\Delta CUA(p_3)$		
	Period 2 (Month 2-6): $\Delta CUA(p_2)$		
	Period 1 (Month 1-6): $\Delta CUA(p_1)$		
Sensitivity (FAS)	Period 3 (Month 3-6): $\Delta CUA(p_3)$	Same as in the Primary Analysis (see Section 14.1.1)	Multiple Imputation (see below for details)
	Period 2 (Month 2-6): $\Delta CUA(p_2)$		
	Period 1 (Month 1-6): $\Delta CUA(p_1)$		
Sensitivity (FAS)	Period 3 (Month 3-6): $\Delta CUA(p_3)$	Non-parametric factorial model for longitudinal data. Fixed factors: • Age in years (≤ 40,40 - < 65), • EDSS at Baseline (≤3, >3), (see below for details)	Same as in the Primary Analysis (see Section 14.1.1).
	Period 2 (Month 2-6): $\Delta CUA(p_2)$		
	Period 1 (Month 1-6): $\Delta CUA(p_1)$		
Sensitivity (FAS)	Period 3 (Month 3-6): $\#CUA(p_3)/\#CUA(p_0)$	Generalized model assuming a negative binomial distribution with the following adjusting factors: • Period • Age (in years), • EDSS at Baseline (≤3, >3), • Within-pooled center correlations. Estimated standardized CUA for the periods as well as the ratios of each post-baseline Period versus the Baseline Period, and corresponding Standard Errors (SE) will be presented including estimated fixed effects including the SE, 2-sided p-values and 95% CI	Same as in the Primary Analysis (see Section 14.1.1).
	Period 2 (Month 2-6): $\#CUA(p_2)/\#CUA(p_0)$		
	Period 1 (Month 1-6): $\#CUA(p_2)/\#CUA(p_0)$		

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Sensitivity (FAS)	Period 3 (Month 3-6): $\Delta CUA(p_3)$ Period 2 (Month 2-6): $\Delta CUA(p_2)$ Period 1 (Month 1-6): $\Delta CUA(p_1)$	Mixed-effects linear model with the following adjusting factors: • Baseline score, • Age (years), • EDSS at Baseline (≤3, > 3), • Period, • pooled center. Estimated Least Square Means (LS Means) and corresponding Standard Errors (SE) will be presented by Visit including 2-sided p-values and 95% CI	Same as in the Primary Analysis (see Section 14.1.1).
Categorization of the Primary Endpoints: Presence of MRI activity (at least one CUA lesion) during the first 6 months (i.e., during Periods 1-3 compared to Baseline Period (i.e., the period from Screening to Baseline).			
Sensitivity (FAS) Sensitivity (FAS by subgroup High-relapse activity) Sensitivity (FAS by subgroup "Previous treatment with DMDs") 	Cross-tabulation for each Period i, i=1,2,3 • No: #CUA(p_i) = 0 • Yes: #CUA(p_i) > 0 vs Baseline Period: • No: #CUA(p_0) = 0 • Yes: #CUA(p_0) >0	McNemar Test.	Same as in the Primary Analysis (see Section 14.1.1).

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Categorization of the Primary Endpoints: Classified CUA lesion count during the first 6 months (i.e., during Periods 1-3) compared to Baseline Period (i.e., the period from Screening to Baseline).			
Sensitivity (FAS)	Cross-tabulation for each Period i, i=1,2,3 • 0: #CUA(p_i) = 0 • >0-1: 0 < #CUA(p_i) ≤ 1 • >1: #CUA(p_i) > 1 vs Baseline Period: • 0: #CUA(p_0) = 0 • >0-1: 0 < #CUA(p_0) ≤ 1 • >1: #CUA(p_0) >1	Bowker's test of symmetry.	Same as in the Primary Analysis (see Section 14.1.1).
Sensitivity (FAS by subgroup "High-relapse activity")			
Sensitivity (FAS by subgroup "Previous treatment with DMDs")			
ΔCUA(p_i) = Difference in the standardized CUA lesion count between Period i and the Baseline Period, i=1 (Month1-6), 2 (Month 2-6), 3 (Month 3-6) #CUA(p_i) = Standardized CUA lesion count during Period I, i=0 (Baseline Period), 1 (Month1-6), 2 (Month 2-6), 3 (Month 3-6)			

Multiple imputation:

In order to evaluate the robustness of the assumption applied in the primary analysis that missing data are observed as missing at random (MAR), the primary analysis will be repeated applying a pattern mixture model. If missingness can be considered MAR (e.g., MRI missing for technical reasons), data will be imputed by multiple imputation from post-baseline data while if it is considered as missing not at random (MNAR) (e.g., drop-out due to lack of efficacy), data will be imputed from baseline data.

The multiple imputation procedure will use a pre-specified seed number repeating the imputation procedure 100 times.

The following reasons for missing MRI results will be considered as MAR:

- MRI classified as non-evaluable by the central MRI reading
- MRI missing at Baseline
- Subject discontinued the study before the Month 6 Visit with primary reason
    - Lost to follow-up
    - Any Protocol non-compliance reported as "Other" and classified as non-related to study treatment
    - Any withdrawal of consent classified as non-related to study treatment
    - Any pregnancy (Discontinuation Reason "Adverse Event" and "Pregnancy" reported as AE leading to discontinuation.
    - Any "Other" reason classified as non-related to study treatment

All other missing MRI results will be handled as MNAR.

The following procedure for the different types of missing data will be performed:

- CUA lesion counts missing at the Baseline Period will be imputed from the distribution of available CUA lesion counts at the Baseline Period.
- CUA lesion counts missing at any post-baseline Period that are considered MAR will be imputed from the distribution of available CUA lesion counts of the same post-baseline Period.
- CUA lesion counts missing at any post-baseline Period that are considered MNAR will be imputed from the distribution of CUA lesion counts of the Baseline Period

The changes from the Baseline Period will be calculated for all periods and subjects, and the primary analysis (i.e. the Wilcoxon test as well as the mixed-effects linear model) will be repeated for each of the imputation repetition (i.e. for each of the imputed datasets).

Results combined using Rubin's rule will be presented.

Non-parametric factorial model:

The primary analysis will be repeated using a factorial model for longitudinal data according to Brunner, Domhoff and Langer (2002). Following the authors' terminology a Fx-LD-F1 model will be assumed with x=2 fixed factors (age, EDSS at Baseline) and one longitudinal factor ('LD') which is reflecting Baseline and one of the primary post-Baseline Periods.

The underlying nonparametric hypothesis is defined in terms of the distribution functions. In a simple model with two time points and no fixed factors one would assume that the pairs of lesion count data (Baseline, post Baseline) would have marginal distributions F_s, s=1, 2, and the null hypothesis would be H_0: $F_1 = F_2$, the alternative accordingly H_0: $F_1 \neq F_2$. In a simple shift model one would assume that these marginal distribution functions only differ through a 'location parameter' μ, namely one would assume that $F_2(x) = F_1(x-\mu)$ and the null hypothesis would be $\mu=0$. For multiple fixed factors and one 'LD' factor the notation of hypotheses is more complicated and can be found in Brunner, Domhoff and Langer (2002) in section 8.3 and especially in 8.3.1 for the null hypothesis of 'no time effect'.

Midranks over all CUA lesions from the Baseline Period and respective post-Baseline Period will be calculated. The ranks will be analyzed with a heteroscedastic factorial designs model with repeated measures (Baseline and one post-Baseline Period) and unspecified covariance matrices. The following factors will be used:

- Age groups (≤ 40,>40),
- EDSS at Baseline (≤3/>3).

The p-value for the time effect will be calculated as described by Brunner, Domhoff and Langer (2002).

Generalized linear model assuming negative binomial distribution:

The standardized CUA lesion count will be evaluated using a generalized mixed-effects linear model with log link function. Period (i.e Baseline Period, each post-Baseline Period i, i=1,2,3) will be analyzed as a repeated factor accounting for within subject and within-center/region correlation using an unstructured covariance matrix. Age and EDSS at Baseline (≤3, >3) will be used as fixed effect covariates.

The ratio of the standardized CUA lesion count of each post-Baseline Period versus the Baseline Period will be calculated as the inverse log of the difference of the estimated contrast between each post-baseline Period and the Baseline Period.

The CUA lesions are already standardized for the observation time and therefore no offset will be used. The estimated standardized CUA lesion count will be calculated based on the observed margin of the FAS at Baseline.

The estimates, standard deviation and 95% confidence intervals for the standardized CUA lesion count at each Period and the ratio between Periods will be present together with the estimated fixed effects.

If the model doesn't converge a variance components covariance structure will be used instead of the unstructured covariance matrix.

14.2 Secondary Endpoints: Characterization of immune cell subsets count at the end of 3, 6, 12, 15, 18 and 24 months compared to Baseline

The statistical analysis of the immune cell subsets counts will be specified in a separate Biomarker SAP.

14.3 Tertiary Endpoints

14.3.1 Changes in CUA lesions during pre-defined post-Baseline periods compared to the Baseline period (i.e., the period from Screening to Baseline)

The changes in CUA lesions counts from Baseline Period for the following post -Baseline periods will be analyzed as tertiary endpoints:

- Period of month 6 to month 12
- Period of month 1 to month 12
- Period of month 18 to month 24
- Period of month 1 to month 24

The statistical analysis of these periods will be defined in an amendment to the IAP, as it is not known yet how the data will be assessed by the Central MRI review.

Further periods and comparisons may be defined in the pending MRI Charter and will then be amended to the IAP after finalization of the corresponding MRI Charter.

14.3.2 Number of CUA lesions during the Baseline and post-Baseline periods

Please see section 14.3.1 for Periods 1-3. Further periods or comparisons may be added by an amendment to the IAP.

14.3.3 Changes in T1 Gd+ lesion count compared to Baseline

The absolute count and change from Baseline in the T1 Gd+ lesion count will be presented by visit.

Repeated mixed-effects linear model:

The change from Baseline will be also analyzed by a repeated mixed-effects linear model, which adjusts for the Baseline count, Age (in years) and EDSS at Baseline (≤3, >3), visit, time between scans (in months) and within subject correlation. An unstructured covariance matrix will be assumed for the model. If the model doesn't converge an alternative covariance structure will be used instead.

Generalized linear model assuming negative binomial distribution:

The T1 Gd+ lesion count will be evaluated using a generalized mixed-effects linear model with log link function assuming a negative binomial distribution. Visit (i.e. Baseline Visit (excluding Screening Visit), each post-Baseline Visit i, i=1,2,3,4) will be analyzed as a repeated factor accounting for time between scans (in years) and within subject correlation using an unstructured covariance matrix. Age and EDSS at Baseline (≤3 [as reference], >3) will be used as fixed effect covariates.

The ratio of the T1 Gd+ lesion count of each post-Baseline Visit versus the Baseline Visit will be calculated as the inverse log of the difference of the estimated contrast between each post-baseline Visit and the Baseline Visit.

No offset will be used. The estimated T1 Gd+ lesions count will be calculated based on the observed margin of the FAS at Baseline.

The estimates, standard deviation and 95% confidence intervals for the T1 Gd+ lesion count at each Visit and the ratio between Visits will be present together with the estimated fixed effects.

If the model doesn't converge an alternative covariance structure will be used instead of the unstructured covariance matrix.

To evaluate further the course of T1 GD+ lesion appearance over time, T1 Gd+ lesions will be further classified as follows for each Visit:

- T1 Gd+ lesion count > 0

- Yes (T1 Gd+ lesion count > 0)
- No (T1 Gd+ lesion count = 0)

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: Change from Baseline in T1 Gd+ lesion count.			
Tertiary (FAS) Tertiary (FAS by subgroup HRA subgroup) Tertiary (FAS by subgroup "Previous treatment with DMDs") Tertiary (FAS for subset "Baseline Period CUA count > 0")	Change from Baseline in T1 Gd+ lesion count at the end of Month 1, 2, 3, 6, 12, 15, 18 and 24 visits.	Repeated mixed-effects linear model for the change from Baseline with the following adjusting factors: • Baseline score, • Age (in years), • EDSS at Baseline (≤3, > 3), • Visit, • Time between scans (in months), • Within-subject correlation. Estimated Least Square Means (LS Means) and corresponding Standard Errors (SE) will be presented by Visit including 2-sided p-values and 95% CI. In addition, the profile over time of the absolute values (including screening visit) as well as changes from Baseline will be displayed graphically.	Missing observations at Baseline: Subjects with a missing Baseline T1 Gd+ lesion count will not be included in the linear model. Missing post-Baseline observations: Missing post-Baseline assessments will not be imputed. The repeated measurement model handles missing post-Baseline assessments under the assumption of MAR.
Tertiary endpoint: Ratio of Post-Baseline vs Baseline T1 Gd+ lesion count			
Tertiary (FAS) Tertiary (FAS by subgroup HRA subgroup) Tertiary (FAS by subgroup "Previous treatment with DMDs") Tertiary (FAS for subset "Baseline Period CUA count > 0")	T1 Gd+ lesion count at Baseline and at the end of Month 1, 2, 3, 6, 12, 15, 18 and 24 visits.	Generalized model assuming a negative binomial distribution with the following adjusting factors: • Age (in years), • EDSS at Baseline (≤3, >3), • Visit, • Time between scans (in months), • Within-subject correlation. Estimated T1 Gd+ lesion count for each Visit as well as the ratios of each post-baseline Visit versus the Baseline Visit, and corresponding Standard Errors (SE) will be presented including estimated fixed effects including the SE, 2-sided p-values and 95% CI.	Missing observations at Baseline: Subjects with a missing Baseline T1 Gd+ lesion count will not be included in the linear model. Missing post-Baseline observations: Missing post-Baseline assessments will not be imputed. The repeated measurement model handles missing post-Baseline assessments under the assumption of MAR.

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary Endpoint: T1 Gd+ lesion count > 0			
Tertiary (FAS) Tertiary (FAS by subgroup HRA subgroup) Tertiary (FAS by subgroup "Previous treatment with DMDs")	Cross-tabulation for each Visit Month 1, 2, 3, 6, 12, 15, 18 and 24 visits vs the Baseline Visit.	McNemar Test.	Missing observations at Baseline: Subjects with a missing Baseline T1 Gd+ lesion count will not be included in McNemar Test, however separate cross-tabulation table including missing/not-evaluable lesions counts will be presented and thus providing percentage related to all patients. Missing post-Baseline observations: Same as above but adapted for missing post-Baseline observations (see "Missing observations at Baseline").
Tertiary endpoint: Change from Baseline in Mean T1 Gd+ lesion count.			
Tertiary (FAS)	Period 1 (Month 1-6): $\Delta T1(p_1)$ Period 2 (Month 2-6): $\Delta T1(p_2)$ Period 3 (Month 3-6): $\Delta T1(p_3)$	Same as in the Primary Analysis (see Section 14.1.1)	Missing observations: Subjects with a missing Mean T1 Gd+ lesion count at the Baseline period or the respective post-baseline period will not be included in the analysis.
$\Delta T1(p_i)$ = Difference in Mean T1 Gd+ lesion count between Period i and the Baseline Period Further periods or comparisons may be added by an amendment to the IAP			

14.3.4 Volume changes of T1 Gd+ lesions compared to Baseline

The statistical reporting of the changes in T1 Gd+ lesions volume compared to Baseline evaluated by the central reading will be defined later by an Amendment to this IAP after the respective MRI Charter has been finalized.

14.3.5 Number of T1 hypointense lesions compared to Baseline

The statistical reporting of the changes in T1 hypointense lesions compared to Baseline evaluated by the central reading will be defined later by an Amendment to this IAP after the respective MRI Charter has been finalized.

14.3.6 Change in volume of T1 hypointense lesions compared to Baseline

The statistical reporting of the changes in T1 hypointense lesions volume compared to Baseline evaluated by the central reading will be defined later by an Amendment to this IAP after the respective MRI Charter has been finalized.

14.3.7 Changes in active T2 lesion count

The number and changes from Baseline of the different types of active T2 lesions will be presented by Period or visit.

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: Changes in the standardized total active T2 lesion count (without T1 Gd+) during pre-defined post-Baseline periods compared to the Baseline Period (i.e., the period from Screening to Baseline).			
Tertiary (FAS)	Change from Baseline of the standardized total active T2 lesion count without T1 Gd+ to Periods 1-3.	Descriptive statistics	
Tertiary endpoint: Changes in standardized total active T2 lesion count during pre-defined post-Baseline periods compared to the Baseline Period (i.e., the period from Screening to Baseline).			
Tertiary (FAS)	Period 3 (Month 3-6): $\Delta T2(p_3)$	Same as in the Primary Analysis (see Section 14.1.1)	Same as in the Primary Analysis (see Section 14.1.1)
	Period 2 (Month 2-6): $\Delta T2(p_2)$		
	Period 1 (Month 1-6): $\Delta T2(p_1)$		
Tertiary (FAS for subgroup HRA subgroup	Period 3 (Month 3-6): $\Delta T2(p_3)$		
	Period 2 (Month 2-6): $\Delta T2(p_2)$		
	Period 1 (Month 1-6): $\Delta T2(p_1)$		

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary (FAS by subgroup “Previous treatment with DMDs	Period 3 (Month 3-6): $\Delta T2(p_3)$		
	Period 2 (Month 2-6): $\Delta T2(p_2)$		
	Period 1 (Month 1-6): $\Delta T2(p_1)$		
Tertiary (FAS for subset “Baseline Period CUA count > 0”)	Period 3 (Month 3-6): $\Delta T2(p_3)$		
	Period 2 (Month 2-6): $\Delta T2(p_2)$		
	Period 1 (Month 1-6): $\Delta T2(p_1)$		

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: Total active T2 Lesion count by Visit			
Tertiary (FAS)	Change from Baseline in total active T2 Lesion count at the end of Month 2, 3, 6 visits.	Same as in the Tertiary Analysis of Change from Baseline in T1 Gd+ lesion count (see Section 14.3.3)	Same as in the Tertiary Analysis of Change from Baseline in T1 Gd+ lesion count (see Section 14.3.3)
Tertiary (FAS for subgroup HRA subgroup			
Tertiary (FAS by subgroup "Previous treatment with DMDs			
Tertiary (FAS for subset "Baseline Period CUA count > 0")			
$\Delta T2(p_i)$ = Difference in the standardized total active T2 lesion count between Period i and the Baseline Period Further periods or comparisons may be added by an Amendment to the IAP			

14.3.8 Responder rate during the different periods

Subjects will be classified as Responder for a post-Baseline period if the standardized CUA lesion count decreased by at least 1 in 6 months in comparison to the Baseline period:

- Responder: $\Delta\mathrm{CUA}(p_i) = \left(\#\mathrm{CUA}(\mathrm{p}_i) - \#\mathrm{CUA}(\mathrm{p}_0)\right) * 6 \geq -1,\ i = \text{Period } 1, 2, 3,$
- Non-Responder: otherwise.

Subjects with missing assessments will be classified as "Non-Responder".

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: Responder rate during the different post-Baseline periods			
Tertiary (FAS)	Responder rates for Periods 1-3.	Descriptive summary statistics and 95% confidence intervals	Missing Baseline Period: Subjects with a missing CUA lesion count for the Baseline Period will not be included in analysis. Subjects with a missing post-Baseline Period will be classified as Non-Responder for the respective Period.
	Further periods or comparisons may be added by an Amendment to the IAP		

14.3.9 Changes in T2 lesion volume compared to Baseline

The statistical reporting of the changes in T2 lesion volume compared to Baseline evaluated by the central reading will be defined later by an Amendment to this IAP after the respective MRI Charter has been finalized.

14.3.10 Changes in Magnetization Transfer Ratio (MTR) compared to Baseline

The statistical reporting of other Magnetization Transfer Ratio (MTR) evaluated by the central reading will be defined later by an Amendment to this IAP after the respective MRI Charter has been finalized.

14.3.11 Changes in brain volume

The statistical reporting of other Magnetization Transfer Ratio (MTR) evaluated by the central reading will be defined later by an Amendment to this IAP after the respective MRI Charter has been finalized.

14.3.12 No Evidence of Disease activity

No Evidence of Disease activity (NEDA), also referred to as freedom from disease activity, is a new goal that is emerging in MS treatment. NEDA is a composite measure of disease activity, including relapses, cognition and disability progression, and MRI activity.

NEDA

NEDA at Month 24 Visit is defined by the absence of:

- Qualifying Relapses,
- Confirmed Disability Progression (CDP),
- MRI activity,

during the whole Treatment Period. If the information for one of the components is unknown, and none of the other components are observed, NEDA will be classified as "unknown". Thus, NEDA can be derived from its components as follows:

- Yes: all components = "No",
- No: at least one of the components =" Yes",
- Unknown: Otherwise.

Qualifying Relapse

- Yes: qualifying relapses reported during the Treatment Period,
- No: study completed, and no qualifying relapse reported during Treatment Period,
- Unknown: otherwise.

Confirmed Disability Progression (CDP)

A Sustained Increase in EDSS score, is defined by an increase of:

- at least 1.5 points if the Baseline EDSS score was 0, or
- at least 1 point if Baseline EDSS score was between 0.5 and 4.5 inclusively, or
- at least 0.5 point if the Baseline EDSS score was at least 5,

that occurs over a 6-month time period.

The increase will be defined as sustained when it occurs on two post-Baseline Visits, which are at least 166 days (6 months × 30 - 14 days Visit Window) apart and no observations at any other Visit (including unscheduled Visits) in between are less than the defined increase.

The occurrence of a CDP is defined as:

- Yes: sustained increase in EDSS score that started during the Treatment Period,
- No: at least 2 post-Baseline EDSS assessments that are 166 days apart are available and no sustained EDSS progression that started during the Treatment Period,
- Unknown: Otherwise.

Start date of the CDP is the start date of the first sustained increase in EDSS.

MRI activity

- Yes: At least one CUA lesion count greater 0 in any post-baseline period,
- No: all post-baseline periods have no CUA lesion,
- Unknown: Otherwise.

Analysis of NEDA

NEDA and its components will be analyzed with a logistic regression with adjusting factors age (in years) and EDSS at Baseline (≤3, >3).

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: NEDA at 24 months			
Tertiary (FAS)	NEDA during Treatment Period	Logistic regression: Adjusting factors: • Age (in years), • EDSS at Baseline (≤3, >3). Estimated probability and corresponding Standard Errors (SE) will be presented including 2-sided p-values and 95% CIs.	Unknown status: Subjects with an unknown status will be excluded from the logistic regression. .
Tertiary (FAS)	CDP during Treatment Period		
Tertiary (FAS)	MRI activity during the Treatment Period		
Tertiary (FAS)	Qualifying Relapse during the Treatment Period		

All components of NEDA will be listed by subject.

14.3.13 No Evidence of Progression or Active Disease

No Evidence of Progression or Active Disease (NEPAD) is a composite measure, evolution of NEDA, including relapses and different measurements of disability progression and MRI activity.

NEPAD at Month 24 Visit is defined as the composite endpoint of NEDA and the absence of:

- 20% progression on T25-FW,
- 20% progression on 9HPT,

during the Treatment Period.

20% confirmed T25FW progression

20% confirmed progression on T25FW is defined by an increase of at least 20% from T25FW Baseline that is sustained over a 6-month time period.

The increase will be defined as sustained when it occurs on two post-Baseline Visits, which are at least 166 days (6 months *30 -14 days Visit Window) apart and no score at any other Visit including unscheduled Visits in between is less than the defined increase.

The occurrence of a 20% confirmed T25FW progression is defined as:

- Yes: sustained T25FW progression that started during the Treatment Period.
- No: at least 2 post-Baseline T25FW assessments that are 166 days apart are available and no sustained T25FW progression that started during the Treatment Period.
- Unknown: Otherwise.

Only T25FW assessments with the same assistance devices as at Baseline will be analyzed. T25FW trials which couldn't be finished by the subjects will be imputed with 3 minutes.

20% confirmed 9HPT progression

20% confirmed progression on 9HPT is defined by an increase of at least 20% from 9HPT Baseline that is sustained over a 6-month time period.

The increase will be defined as sustained when it occurs on two post-Baseline Visits, which are at least 166 days (6 months *30 - 14 days Visit Window) apart and no score at any other Visit including unscheduled Visits in between is less than the defined increase.

The occurrence of a 20% confirmed 9HPT progression is defined as:

- Yes: sustained 9HPT progression that started during the Treatment Period.
- No: at least 2 post-Baseline 9HPT assessments that are 166 days apart and no sustained 9HPT progression that started during the Treatment Period.
- Unknown: Otherwise.

Analysis of NEPAD

NEPAD will be analyzed with a logistic regression with adjusting factors age (in years) and EDSS at Baseline (≤3, >3).

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: NEPAD at Month 24			
Tertiary (FAS)	NEPAD during the Treatment Period	Same as in the Tertiary Analysis of NEDA (Section 14.3.12)	Same as in the Tertiary Analysis of NEDA (Section 14.3.12)
Tertiary (FAS)	Confirmed T25FW progression during the Treatment Period		
Tertiary (FAS)	Confirmed 9HPT progression during the Treatment Period		

14.3.14 Changes in SDMT outcome at the end of 6, 12, 18 and 24 months compared to baseline

The SDMT (Symbol Digit Modalities Test) presents a series of nine symbols, each paired with a single digit in a key at the top of a standard sheet of paper. An adapted version of the test is presented in the protocol. Subjects are asked to voice the digit associated with each symbol as rapidly as possible for 90 sec. The SDMT score is the number of correct digits over the 90 sec time span.

The observed results and changes from Baseline in the SDMT and scaled scores of the SDMT will be tabulated by Visit for the FAS, see Section 18.2.

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: Changes in SDMT outcome at the end of 6, 12, 18 and 24 months compared to Baseline			
Tertiary (FAS)	Change from Baseline in SDMT (raw and scaled scores) at 6, 12, 18 and 24 months.	Descriptive summary statistics.	

14.3.15 Changes in level of disability as measured by EDSS

The Expanded Disability Status Scale (EDSS) ranges from 0 to 10 in increments of 0.5 on an ordinal scale. Higher scores represent a higher level of disability. Details of the EDSS scoring are included in the appendix of the study protocol.

The observed EDSS and the change from Baseline will be presented by Visits using summary statistics (median, quartiles, min, max) and by the following categorization for each Visit:

- ≤3,
- >3-<6,
- ≥6.

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Secondary endpoint: Changes in level of disability as measured by EDSS at Month 6, 12, 18, and 24 compared to Baseline			
Tertiary (FAS)	Changes from Baseline of EDSS at 6,12, 18 and 24 months.	Descriptive summary statistics for the median and quartiles.	Same as in the Primary Analysis (see Section 14.1.1).
Tertiary (FAS)	EDSS categories (<3, 3-<6, ≥6) at Baseline, Month 6, 12, 18 and 24 Visits.	Descriptive summary statistics.	Missing values will be included in the denominator and presented as a separate category.

14.3.16 Changes in level of disability as measured by 9HPT

The 9HPT should be completed at Baseline and then at Month 6, 12 18, and 24 Visit to measure the level of disability over time. The subjects have to stick 9 pegs in 9 holes and place them back. That will be done twice for each hand. The mean of the 4 stop times will be analyzed. At least one stop time must be recorded, otherwise, the Visit will be considered as missing. Details of the 9HPT are included in the appendix of the study protocol.

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: Changes in level of disability as measured by 9HPT at Month 6, 12, 18, and 24 Visit compared to Baseline			
Tertiary (FAS)	Change from Baseline in mean time of the 9HPT to the Month 6, 12, 18, and 24 Visit.	Same as in the Tertiary Analysis of Change from Baseline in T1 Gd+ lesion count (see Section 14.3.3) except that no graphical displays will be generated.	Same as in the Tertiary Analysis of Change from Baseline in T1 Gd+ lesion count (see Section 14.3.3)

14.3.17 Changes in level of disability as measured by T25FW

T25FW is a quantitative mobility and leg function performance test based on a timed walk over 25 feet (approximately 7.6 meters). Administration time will vary depending upon the ability of the subject. Subjects have to twice walk a trial of 25 feet. The time limit for each trial is 3 minutes. At least one trial has to be completed to calculate the mean time for a Visit. If the subjects couldn't finish a trial, the trial will be scored with 3 minutes. T25FW performed with a different assistance device as compared to what was used at Baseline will be considered as "changed device" and will not be included in the statistical analysis. Details of the T25FW are included in the appendix of the study protocol.

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: Changes in level of disability as measured by T25FW at Month 6, 12, 18, and 24 Visit compared to Baseline			
Tertiary (FAS)	Change from Baseline in mean time of the T25FW to the Month 6, 12, 18, and 24 Visit.	Same as in the Tertiary Analysis of Change from Baseline in T1 Gd+ lesion count (see Section 14.3.3) except that no graphical displays will be generated.	Same as in the Tertiary Analysis of Change from Baseline in T1 Gd+ lesion count (see Section 14.3.3)

14.3.18 Annualized Relapse Rate (ARR) between Baseline and 12/24 months

The unadjusted ARR will be calculated for the Treatment Period, Course 1 Treatment Period and Course 2 Treatment Period.

The unadjusted ARR for a subject *i* is defined as the number of relapses per year:

- ARR for subject *i* = number of relapses during the period / person-years of subject *i*,
- Person-years for subject *i* = period (in days) / 365.25.

The unadjusted population ARR is the mean of all subject's ARRs. The unadjusted ARR will be calculated for all relapses, as well as the qualifying relapses (ARR_{qual}), as reported in the eCRF.

Relapses will be counted for the defined time period, if the date of onset of the relapse is within the defined time period. Incomplete or missing onset dates will be imputed as specified in Section 9.

The number of subjects with (qualifying) relapses, the number of (qualifying) relapses, person-years and the unadjusted population ARR will be presented for the Treatment Period, Course 1 Treatment Period and Course 2 Treatment Period. Also, the number of subjects with relapses and number of relapses which were treated with steroids or leading to hospitalization will be presented separately. For the Interim Analysis the different relapse counts up to the month 6 visit will be presented without the ARR.

In addition, the relapse rate will be analyzed using a Poisson regression model with count of relapses as dependent variable and with fixed effects for age (in years) and EDSS at Baseline (≤3, >3). The log of the duration of the treatment period (precision in days) will be the offset variable in the model.

Analysis (Analysis Set)	Derivation	Statistical Analysis Methods	Missing data handling
Tertiary endpoint: Annualized relapsed rate (ARR) between Baseline and Month 12 and Month 24			
Tertiary (FAS)	ARR during the Treatment Period	Poisson regression for number of relapses. Adjusting factors: • Age (in years), • EDSS at Baseline (≤3/>3). • Log of the duration of the Treatment Period (in days) as offset. Estimated mean ARR and corresponding Standard Errors (SE) will be presented, including 2-sided p-values and 95% CIs. In addition, the unadjusted ARR will be presented descriptively.	The analysis will consider the different observation periods of early withdrawals by including the treatment period as offset in the analysis. This strategy is assuming that the ARR after drop-out is independent.
Tertiary (FAS)	ARR during the Course 1 Treatment Period		
Tertiary (FAS)	ARR during the Course 2 Treatment Period		
Tertiary (FAS)	ARR_{qual} during the Treatment Period		
Tertiary (FAS)	ARR_{qual} during the Course 1 Treatment Period		
Tertiary (FAS)	ARR_{qual} during the Course 2 Treatment Period		
Tertiary (FAS)	ARR leading to hospitalization during the Treatment Period		
Tertiary (FAS)	ARR requiring steroid treatment during the Treatment Period		

All relapses during the treatment period will be listing together with relevant MRI results (T1, T2 and CUA Lesion counts) by subject.

CCI [redacted]

15 Safety Analyses

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

Safety analyses will be done on the SAF. Unless otherwise specified, the safety analysis will be done for interim and final analysis.

15.1 Adverse Events

Treatment emergent adverse events (TEAEs) are those events with onset dates occurring within the Treatment Period (see Section 9 for definition).

All analyses described in Section 15.1 will be based on TEAEs if not otherwise specified.

Incomplete AE-related dates will be handled as follows:

- In case the onset date is missing completely or missing partially and the onset month and year or the onset year are equal to the start of study medication, then the onset date will be replaced by the start date of study medication or AE resolution date whichever is earlier.
- 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.
- In all other cases the incomplete stop date will not be imputed.

15.1.1 All Adverse Events

TEAEs will be summarized using MedDRA (latest version available) PT as event category and MedDRA (latest available version) primary SOC body term as Body System category overall and by HRA subgroups (HRA and non-HRA).

Unless otherwise stated, adverse events will be displayed in terms of frequency tables: PT and primary SOC in alphabetical order.

Adverse events related to study treatment are those events with relationship missing, unknown or related.

The following overall frequencies of subjects with the corresponding TEAEs will be prepared. In addition, the tables will be provided by PT and primary SOC in alphabetical order:

- Any AE,
- Any study treatment related AEs,
- Any serious AEs,
- Any non-serious AEs,

- Any study treatment related serious AEs,
- Any AE by severity (mild, moderate, severe),
- Any study treatment related AE by severity (mild, moderate, severe),
- Any AEs leading to death (AEs with outcome "fatal"),
- Any study treatment related AEs leading to death (AEs with outcome "fatal").

Summary table for non-serious adverse events applying frequency threshold of 5% will be provided by SOC and PT.

Summary tables for serious and non-serious adverse events applying frequency threshold of 5% sorted by decreasing frequency will be provided by PT.

All AE by worst severity (Mild or worse, moderate or worse, severe) will be presented by SOC and PT. Missing severity will be presented in all severity groups.

All AEs including non-treatment emergent AEs will be listed by subject for the SAF.

Exposure Adjusted Incidence Rate

Exposure adjusted incidence rates (EAIR) are calculated as number of subjects with AE divided by the sum of the individual times in years of all subjects in the safety population from start of study medication to first onset of AE or end of treatment period, whichever occurs first.

The exact Poisson 95% confidence intervals for the EAIR are calculated using the relationship between the Poisson and the Chi-square distribution (Ulm, 1990):

$$LCI = \frac{\chi^2_{2n,\frac{a}{2}}}{2 \times t},$$

$$UCI = \frac{\chi^2_{2(n+1),1-\frac{a}{2}}}{2 \times t},$$

where t is the sum of the individual times in years of all subjects and n is the number of subjects with a specific AE for the EAIR, which will be the basis for the number of the degrees of freedom for the chi-square quantile for the upper tail probability χ^2.

The incidence rate multiplied with 1000 would give the number of AEs expected in 1000 subjects within 1 year.

Exposure adjusted incidence rates of TEAEs will be presented overall and by SOC and PT.

The following table will be provided:

- Exposure adjusted incidence rates of AEs by SOC and PT.

15.1.2 Adverse Events Leading to Treatment Discontinuation

The following overall frequency tables will be prepared for the adverse event actions. Tables will be provided by PT and primary SOC in alphabetical order:

- Any AE leading to temporary discontinuation of study drug,.
- Any AE leading to permanent discontinuation of study drug,
- Any AE leading to dose reduction of study treatment.

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

15.2.1 Deaths

Any discontinuations due to death (see Section 10.1) and any AE leading to death will be tabulated (see Section 15.1.1).

A summary table of observed Deaths (from "Study Termination" eCRF page) will be provided together with TEAEs with fatal outcome by PT.

15.2.2 Serious Adverse Events

The statistical analysis and reporting of serious adverse events are defined in Section 15.1.1. In addition, all serious adverse events will be listed.

15.2.3 Other Significant Adverse Event

Not applicable.

15.3 Clinical Laboratory Evaluation

Lymphocyte counts from local laboratory will be used for summary statistics and shift tables. They will be classified according to the NCI-CTCAE (latest version available). The classification will be derived only from the laboratory results at a given assessment, thus ignoring the underlying syndrome.

The Treatment Course 2 can be delayed because of Lymphocyte count below 800 cells/mm³ at the Month 12 Visit. In this case, the Lymphocyte count can be measured several times between the Month 12 Visit and the start of Treatment Course 2. Lymphocyte counts at or after the visit start date of the Month 12 Visit and before or at the start date of the Treatment Course 2 will be displayed as followed:

- The earliest Lymphocyte count will be assigned to the Month 12 Visit,
- The latest Lymphocyte will be assigned to the Month 12b Visit,
- All other Lymphocyte count will be assigned to unscheduled visits.

A Lymphocyte count can be assigned to the Month 12 Visit and Month 12b Visit if just one count is available. If the Treatment Course 2 wasn't started, only the earliest Lymphocyte count will be assigned to the Month 12 Visit.

The worst on-study grade (i.e., on, or after, first study treatment administration) will be summarized by subject considering only subjects with post-Baseline laboratory samples: Laboratory tests by NCI-CTC grade (0, 1, 2, 3, 4, any).

Lymphocyte counts will be examined for trends using descriptive statistics (mean, SD, median, Q1, Q3, minimum, and maximum) of actual values and changes from Baseline to each scheduled Visit including Month 12b Visit over time. Shift tables of Baseline versus endpoint (as well as the worst value at any post-Baseline Visit) will be presented. Abnormalities classified according to NCI-CTCAE toxicity grading (latest version available) will be described using the worst grade.

All Lymphocyte counts will be listed by subject for the SAF.

Results of the tuberculosis and serology tests that do not comply with the protocol will be listed as PD.

15.4 Vital Signs

The maximum changes of vital sign measurements from Baseline to maximum change after start of 1st treatment will be grouped as follows:

Heart rate increase from Baseline <100 bpm; ≥ 100 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
Heart rate decrease from Baseline <100 bpm; ≥ 100 bpm	≤20 bpm, >20 – 40 bpm, >40 bpm
SBP increase from Baseline <140 mmHg; ≥ 140 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
SBP decrease from Baseline <140 mmHg; ≥ 140 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP increase from Baseline <90 mmHg; ≥ 90 mmHg	≤20 mmHg, >20 – 40 mmHg, >40 mmHg
DBP decrease from Baseline <90 mmHg; ≥ 90 mmHg,	≤20 mmHg, >20 – 40 mmHg, >40 mmHg

For each subject, the worst change during the treatment period will be considered. Missing values will be presented as a separate category.

The following summaries will be prepared for vital sign parameters as grouped above considering only subjects with post-Baseline values:

- Maximal Shifts (changes in categories),
- Listing of highest change per subject.

15.5 Other Safety or Tolerability Evaluations

The results of the physical examination will not be presented in any statistical analysis or listing. Abnormalities occurring, or worsening, during the study will be reported as AE.

Safety Findings during Central Review are reported back to the sites for further clinical assessment and reporting as AE, as outlined in the Review Charter. Thus, no separate statistical reporting of the Safety MRI central review data will be done.

Positive pregnancy tests or pregnancy tests not done without a reason, will be listed for the SAF.

16 Analyses of Other Endpoints

Not applicable.

17 References

Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio. American Journal of Epidemiology 1990;131(2):373-375.

Brunner, E., Domhof, S. und Langer, F. (2002) Nonparametric Analysis of Longitudinal Data in Factorial Experiments. Wiley, New York.

Independent Review Charter for Counting Brain MRI Lesions, Version 1.0 (29 Apr 2019), Magnify MS700568_0022.

Parmenter, B. A., et al. The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). J. Int. Neuropsychol. Soc. 16, 6–16 (2010).

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

18 Appendices

18.1 Appendix 1: List of DMDs

DMDs	Preferred Term
Alemtuzumab* (Campath, MabCampath, Lemtrada)	ALEMTUZUMAB
Daclizumab* (Zinbryta)	DACLIZUMAB
Dimethyl fumarate (Tecfidera)	DIMETHYL FUMARATE
Fingolimod* (Gilenya)	FINGOLIMOD FINGOLIMOD HYDROCHLORIDE
Glatiramer Acetate (Copaxone)	GLATIRAMER GLATIRAMER ACETATE
Immunoglobulins	IMMUNOGLOBULIN HUMAN NORMAL IMMUNOGLOBULINS IMMUNOGLOBULINS NOS
Interferon beta (Avonex, Rebif, Betaferon, Extavia, Plegridy)	INTERFERON INTERFERON BETA INTERFERON BETA-1A INTERFERON BETA-1B PEGINTERFERON PEGINTERFERON BETA-1A
Mitoxantrone (Novantrone)	MITOXANTRONE MITOXANTRONE HYDROCHLORIDE
Natalizumab* (Tysabri)	NATALIZUMAB
Non-approved investigational DMDs* (monoclonal antibodies, antiS1PR, laquinimod etc)	DIROXIMEL FUMARATE INVESTIGATIONAL DRUG LAQUINIMOD OFATUMUMAB OPICINUMAB OTHER ANTINEOPLASTIC AGENTS OZANIMOD
Ocrelizumab* (Ocrevus)	OCRELIZUMAB
Off-label immunosuppressants (azathioprine, mycophenolate, cyclophosphamide)	AZATHIOPRINE CYCLOPHOSPHAMIDE METHOTREXATE
Siponimod*	SIPONIMOD
Teriflunomide (Aubagio)	TERIFLUNOMIDE
DMDs marked with a * are second line DMDs, which are not allowed as a previous medication according to the study protocol.	

Subjects who have taking any of the listed DMDs according to the "Relevant Previous Medication" eCRF page will be categorized as DMD pre-treated. All other subjects will be considered as pre-treatment naïve. The list may be extended during medical review in case that new or other DMDs will be applied during the study.

18.2 Appendix 2: SDMT Scaled Score

Scaled score	SDMT
2	
3	<43
4	43-35
5	46
6	47-50
7	51-54
8	55-56
9	57-59
10	60-63
11	64-66
12	67-69
13	70-72
14	73-74
15	75-78
16	79
17	80-87
18	>87

Parmenter et al., JINS 2010;16:6-16

ELECTRONIC SIGNATURES

Document: study 700568 0022 iap v2

Signed By	Event Name	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'GMT'Z)
PPD	Task Completed (Approval eSign): Approved	Business Approval	16-Nov-2020 13:33