STATISTICAL ANALYSIS PLAN

Protocol title:	**Phase 3, randomized, double-blind efficacy and safety studies comparing SAR442168 to teriflunomide (Aubagio®) in participants with relapsing forms of multiple sclerosis (GEMINI 1 and GEMINI 2)**
Protocol number:	**EFC16033 and EFC16034**
Compound number (INN/Trademark):	**SAR442168 (tolebrutinib)**
Study phase:	**Phase 3**
Short Title:	**RMS studies of BTK inhibitor tolebrutinib (SAR442168) (GEMINI 1 and GEMINI 2)**
Statistician:	[redacted]
Statistical project leader:	[redacted]
Date of issue:	**10-Jul-2024**
Regulatory agency identifier number(s):	
IND:	140884
EudraCT:	2020-000637-41
NCT:	NCT04410978
WHO:	U1111-1238-1418
EUDAMED:	Not applicable
Other:	Not applicable

Date: 10-Jul-2024 Total number of pages: 46

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

VERSION HISTORY

This statistical analysis plan (SAP) for studies EFC16033 and EFC16034 is based on the EFC16033 amended protocol 10 dated 20-Dec-2023 and EFC16034 amended protocol 11 dated 20-Dec-2023. The first participant was randomized in EFC16033 on 27-Jul-2020 and in EFC16034 on 25-Jun-2020. This SAP is approved before the first database lock of the two studies.

Major changes in statistical analysis plan

SAP Version	Approval Date	Changes	Rationale
1 for review	30-Mar-2022	Not Applicable	Original version
1	17-Aug-2023	• Add non-parametric testing for endpoints where there is a potential for deviation from the normality assumption.	Address the feedback provided by the FDA
		• Change offset variable from MRI scans to patient years.	Change offset to patient years rather than MRI scans given that scans are not equally spaced.
		• Apply multiple imputation for the primary analysis of 6-month CDW only for participants that end the study with 3-month CDW, but without 6-month confirmation.	Address the feedback provided by the FDA
2	23-Apr-2024	• Update and clarify the multiplicity testing procedure to control type-1 error in each individual study and separately for endpoints assessed in the integrated dataset.	Separate testing of individual study endpoints from those planned to be tested in the integrated dataset while maintaining a control of the type 1 error for EFC16033 (GEMINI 1), a control for EFC16034 (GEMINI 2), and a control for the integrated studies.
		• Include imputation of missing baseline data of Gd-enhancing T1-hypertintense lesion presence or absence. Impute as absent if all non-missing post baseline values for a participant are 0, otherwise assign randomly based on the baseline proportions in the overall population.	Baseline Gd-enhancing T1-hypertense lesion presence or absence is used as a covariate in the primary analysis of the primary endpoint and also for analysis of other disability related endpoints. Baseline imputation is included so that all participants are included in the analyses. The number of participants affected is minimal with no impact on results.
3	2-Jul-2024	• Revert back to the original multiplicity testing procedure where the key secondary endpoint of 6M-CDW using the combined	FDA feedback

SAP Version	Approval Date	Changes	Rationale
		data of EFC16033 and EFC16034 can only be tested if both null hypotheses for the primary endpoint (ARR) within each study have been rejected.	
		• Include SAS code to impute baseline Gd-enhancing T1 lesion status and specify random seed.	FDA feedback

Additional minor changes have been made to improve clarity and consistency in versions 1, 2 and 3.

1 INTRODUCTION

1.1 STUDY DESIGN

EFC16033 and EFC16034 are Phase 3, randomized, double-blind (DB), double-dummy, 2-arm, active-controlled, parallel-group, multicenter, event-driven trials comparing tolebrutinib (SAR442168) to teriflunomide in participants with relapsing forms of multiple sclerosis (RMS). The treatment duration is variable ranging from approximately 24 to 48 months. See Section 4 for details.

After a screening phase of up to 28 days (up to 35 days in exceptional situations), participants will be randomized within study at a 1:1 ratio to receive a daily dose of 60 mg oral tolebrutinib as well as placebo to match the teriflunomide tablet or 14 mg oral teriflunomide as well as placebo to match the tolebrutinib tablet. Randomization is stratified by the Expanded Disability Status Scale (EDSS) score at screening (<4 versus ≥4) and geographic region (US versus non-US).

A month is defined as a period of 28 days by convention. As such, 6 months is defined as 24 weeks and 24 months as 96 weeks.

Participants completing the planned DB treatment period can enroll in a separate long-term safety (LTS) study with 60 mg tolebrutinib oral daily dosing. For participants that complete the DB treatment and do not enter the LTS study, a final follow-up visit to collect safety data is performed 4 weeks after the last dose of investigational medicinal product (IMP). A participant is considered to have completed the study if he/she has completed all periods of the study including the end of study (EOS) visit, whether remaining on IMP or not.

1.2 OBJECTIVES AND ENDPOINTS

Table 1 - Objectives and endpoints

Objectives	Endpoints
Primary	
• To assess efficacy of daily SAR442168 compared to a daily dose of 14 mg teriflunomide (Aubagio) measured by annualized adjudicated relapse rate in participants with relapsing forms of MS	• Annualized relapse rate during the study period assessed by confirmed protocol-defined adjudicated relapses
Secondary	
• To assess efficacy of SAR442168 compared to teriflunomide (Aubagio) on disability progression, magnetic resonance imaging (MRI) lesions, cognitive performance, and quality of life	• Time to onset of confirmed disability worsening (CDW), confirmed over at least 6 months, defined as follows: - increase of ≥1.5 points from the baseline Expanded Disability Status Scale (EDSS) score when the baseline score is 0, OR - increase of ≥1.0 point from the baseline EDSS score when the baseline score is 0.5 to ≤5.5, OR

Objectives	Endpoints
	- increase of ≥0.5 point from the baseline EDSS score when the baseline score is >5.5 • Time to onset of CDW, assessed by the EDSS score and confirmed over at least 3 months • Total number of new and/or enlarging T2-hyperintense lesions as detected by MRI, defined as the sum of the individual number of new and/or enlarging T2 lesions at all scheduled visits starting after baseline up to and including the End-of-Study (EOS) visit and number of new and/or enlarging T2-hyperintense lesions by visit over time • Total number of new gadolinium- (Gd-) enhancing T1-hyperintense lesions as detected by MRI, defined as the sum of the individual number of new Gd-enhancing T1-hyperintense lesions at all scheduled visits starting after baseline up to and including the EOS visit • Time to confirmed disability improvement (CDI), defined as a ≥1.0 point decrease on the EDSS from the baseline EDSS score confirmed over at least 6 months • Percent change in brain volume loss as detected by brain MRI scans at the EOS compared to Month 6 • Change in cognitive function at the EOS compared to baseline as assessed by the Symbol Digit Modalities Test (SDMT) • Change in cognitive function at the EOS compared to baseline as assessed by the CVLT-II, where available • Change in Multiple Sclerosis Quality of Life 54 (MSQoL-54) questionnaire score at the EOS compared to baseline
• To evaluate the safety and tolerability of daily SAR442168	• Adverse events (AEs), serious AEs, AEs leading to permanent study intervention discontinuation, AEs of special interest, and potentially clinically significant abnormalities in laboratory tests, safety scales, ECG, and vital signs during the study period
• To evaluate population pharmacokinetics (PK) of SAR442168 and relevant metabolite(s) and its relationship to efficacy and safety (EFC16033 only)	• Plasma concentration of SAR442168 and relevant metabolite(s) (population PK assessment) at Months 6, 9, and 12
• To evaluate pharmacodynamics (PD) of SAR442168	• Change in plasma neurofilament light chain (NfL) levels at the EOS compared to baseline, where available • Change in lymphocyte phenotype subsets in whole blood at the EOS compared to baseline in a subset of participants (EFC16033 only) • Changes in serum immunoglobulin level at the EOS compared to baseline • Change in serum Chi3L1 levels at the EOS compared to baseline

Objectives	Endpoints
Tertiary/exploratory	
• To evaluate the efficacy of SAR442168 on disease activity as measured by additional clinical, brain MRI, and composite measurements	• EDSS score change from baseline at scheduled visits starting after baseline and including the EOS visit • Proportion of adjudicated relapse-free participants from randomization until the EOS visit • Time to onset of 20% worsening in the 9-hole peg test (9-HPT) confirmed over at least 3 and 6 months • Time to onset of 20% worsening in the timed 25-foot walk (T25-FW) test confirmed over at least 3 and 6 months • Time to onset of 4-point decrease in Symbol Digit Modalities Test (SDMT) confirmed over at least 3 and 6 months • Change from baseline of total volume of T2-hyperintense lesions as detected by brain MRI at Months 18, 24, and the EOS • Magnetization transfer ratio recovery at the EOS in new magnetization transfer ratio lesions detected at Months 6 and 12 • Change in number of phase rim lesions in susceptibility-weighted imaging (SWI) MRI from baseline by visit over time (subset of centers with capacity of 3T MRI) • Proportion of participants with no evidence of disease activity (NEDA-3) at Months 18, 24, and the EOS • Change from baseline to Months 12, 18, and 24 and to the EOS in modified Multiple Sclerosis Functional Composite 3 (MSFC-3), assessed as the composite of the T25-FW test, 9-HPT, and SDMT • Change from baseline by visit over time in volume of T1-hypointense lesions, and cumulative number of new T1-hypointense lesions • Number and volume of slowly evolving lesions (SELs) • Normalized T1 intensity evolution in SELs
• To evaluate the treatment effect of SAR442168 via changes in participants' health-related quality of life (HRQoL), and working capacity	• Change in EuroQol 5-dimension 5-level questionnaire (EQ-5D-5L) from baseline by visit over time

1.2.1 Estimands

Primary estimands for the primary and key secondary endpoints are defined in Table 2. Complete details are provided in Section 3. For all estimands, the comparison of interest is tolebrutinib versus teriflunomide.

Table 2 - Summary of primary estimand for main endpoints

Endpoint Category (estimand)	Estimands			
	Endpoint	**Population**	**Intercurrent event(s) handling strategy**	**Population-level summary (Analysis and missing data handling)**
Primary objective: To assess efficacy of daily tolebrutinib compared to a daily dose of 14 mg teriflunomide (Aubagio) measured by annualized adjudicated relapse rate (ARR) in participants with relapsing forms of MS				
Primary endpoint (treatment policy estimand)	Annualized adjudicated relapse rate during study period.	ITT	Regardless of study intervention discontinuation (treatment policy)	Relative reduction in the annualized adjudicated relapse rate as derived from the negative binomial model with treatment group, Gd-enhancing T1 lesions at baseline (presence, absence), geographical region (US, non-US) and EDSS strata (<4, ≥4) as covariates; log transformed observation duration as the offset variable. Missing data will not be imputed.
Secondary objective: To assess efficacy of tolebrutinib compared to teriflunomide (Aubagio) on disability progression				
Key Secondary endpoint (treatment policy estimand)	Time from randomization to onset of 6-month confirmed disability worsening	ITT (EFC16033 and EFC16034 pooled data)	Regardless of study intervention discontinuation (treatment policy)	Hazard ratio from Cox proportional hazards model with terms for treatment group, Gd-enhancing T1 lesions at baseline (presence, absence), geographical region (US, non-US), EDSS strata (<4, ≥4) and study (EFC16033, EFC16034). The statistical comparison will be performed using the log-rank test stratified by the geographical region (US, non-US), EDSS strata (<4, ≥4) and study. Incomplete data will be imputed only for participants who complete the study, meet 3-month CDW criteria, and continue to meet criteria for EDSS disability worsening through the final study assessment, but did not reach 6-month confirmation. For other participants without a confirmed 6-month CDW or with premature study discontinuation prior to 6-month confirmation after onset of an event, the participants will be censored at the date of last EDSS measurement.

ITT: intention to treat

2 ANALYSIS POPULATIONS

Table 3 - Populations for analyses

Population	Description
Enrolled	All participants who sign the ICF.
Randomly assigned to study intervention / randomized	All participants with a treatment kit number allocated and recorded in the IRT database, regardless of whether the treatment kit was used or not. Participants treated without being randomized will not be considered randomized and will not be included in any efficacy population.
ITT	The primary efficacy population will be the ITT population, defined as all randomly assigned participants. All efficacy analyses will be conducted according to the treatment group allocated by the randomization schedule, irrespective of the treatment received.
Safety	All participants randomly assigned and exposed to study treatment, regardless of the amount of exposure, analyzed according to the treatment actually received. Randomized participants for whom it is unclear whether they took the study medication will be included in the safety population as randomized. The pharmacodynamic (PD) analyses will be performed on the safety population.
PK	All participants in the safety population with at least one non-missing PK sample after first dose of the study treatment. Participants will be analyzed according to the treatment actually received.

ICF: informed consent form; IRT: interactive response technology; ITT: intention to treat; PD: pharmacodynamlic(s); PK: pharmacokinetic(s)

The PK population applies only to study EFC16033.

Participants exposed to study IMP before or without being randomized will not be considered randomized and will not be included in any analysis population. The safety experience of these participants will be reported separately.

For any participant randomized more than once, only the data associated with the first randomization will be used in any analysis population. The safety experience associated with any later randomization will be reported separately.

For participants receiving more than one study intervention during the study, the treatment group for as-treated analyses will be the treatment the participant received for the longest duration.

3 STATISTICAL ANALYSES

3.1 GENERAL CONSIDERATIONS

The baseline values of efficacy parameters are generally defined as the last available value prior to the first dose of study medication unless otherwise specified. For the EDSS, the baseline value will be taken as the average of the screening and randomization visit values. If one of the values is missing, the non-missing value will be used as baseline. The baseline value of safety parameters is defined as the last available value prior to the first dose of IMP. For the C-SSRS, the baseline value will be the worst assessment between the past 6 months evaluation at screening and the since last visit evaluation on Day 1. For participants randomized but not treated, the baseline value is defined as the last available value before randomization. Unless otherwise indicated, 2-sided p-values and 95% confidence intervals [CI(s)] will be provided for assessment of treatment differences.

To include all participants in analyses where baseline Gd-enhancing T1-hyperintense lesion presence or absence is used as a covariate, baseline status (present or absent) will be imputed for participants with a missing baseline count: if subsequent post-baseline counts for a participant are all 0, then impute baseline status as absent, otherwise, impute baseline status randomly based on the proportion of presence/absence at baseline in the overall population (Section 5.5). The random seed used for imputation will be the study number, ie, 16033 for imputation in EFC16033 and 16034 for imputation in EFC16034. Of note, the number of participants affected is minimal with no impact on results.

In general, continuous data will be summarized using the number of observations available, mean, standard deviation (SD), median, interquartile range (Q1, Q3), minimum, and maximum. Categorical and ordinal data will be summarized using the count and percentage of participants. Efficacy data will be analyzed in the ITT population, and safety data will be analyzed in the safety population, unless otherwise specified.

Unless otherwise specified, analyses will be performed by treatment group (and overall for baseline and demographics characteristics). For endpoints to be analyzed in pooled EFC16033+EFC16034 data [confirmed disability worsening (CDW) and confirmed disability improvement (CDI)], pooled data will be mentioned in the analyses. If no such specification is included, the analysis is within the individual study.

The primary analysis of each study and the analyses of pooled study data will be conducted after the completion and database lock of both studies.

Observation period

The observation periods for safety will be divided into 3 epochs:

- The **screening (pre-treatment) period** is defined as the period from the signed informed consent date up to first IMP administration.

- The **treatment period** (ie, treatment-emergent (TE) period) is defined as the period from the first IMP administration to the earliest of 1) last IMP administration + 10 days, 2) death date, or 3) last contact date.
- The **post-treatment period**, if applicable, is defined as the period from the last day of the treatment period to the participant's final study contact date.

The on-study period is defined as the time from the randomization until the end of the study, defined as the last scheduled visit for those participants who complete the study and the end-of-study date collected on the electronic case report form (eCRF) page "Completion of End of Study" for those who did not complete the study. If death is the end-of-study reason, date of death will be used.

Contingency measures for a regional or national emergency that is declared by a governmental agency

For participants in Ukraine, Russia, and Belarus who are impacted by national events and likely lost to follow-up (LFU) as of 24 February 2022, their data will be handled as follows:

- For the primary endpoint of annualized adjudicated relapse rate, all observed events, if any, up to the last contact date for the participant will be included in the analysis, and the observation duration will be from randomization to the last contact date. No imputation will be performed for the unobserved events that may happen after being LFU.
- For time to event secondary and tertiary endpoints, participants will be treated as prematurely discontinuing the study and will be censored at the date of last corresponding assessment (eg, last EDSS assessment).
- For all other secondary and tertiary endpoints as well as all safety endpoints, all observed data up to the last contact date will be included for analyses.

3.2 PRIMARY ENDPOINT(S) ANALYSIS

3.2.1 Definition of endpoint(s)

The primary endpoint is the ARR (this refers to the adjudicated ARR herein, unless otherwise specified) during the study period assessed by confirmed protocol-defined adjudicated relapses.

3.2.2 Main analytical approach

The purpose of the primary analysis of ARR is to assess the efficacy of tolebrutinib in an ITT setting. The null hypothesis for the primary efficacy endpoint of ARR is that there is no treatment difference between tolebrutinib and teriflunomide, and the alternative is that there is a between-treatment difference.

In this primary approach, off-treatment events of participants who prematurely discontinue study intervention will be included for analysis as per the ITT principle. Participants who permanently discontinue study intervention will be asked and encouraged to return to the clinic for

all remaining study visits. In this case, all events during the planned treatment period will be included and the observation duration will be from randomization to the EOS. If a participant withdraws from the study prior to the common end of study (CEOS), all observed events up to the last contact date will be included in the analysis, and the observation duration will be from randomization to the last contact date. No imputation will be performed for the unobserved events that may happen after study discontinuation. This estimand compares the rate of relapse for the participants randomized to tolebrutinib versus teriflunomide, regardless of what treatment participants actually received. It assesses the benefit of the treatment policy or strategy relative to teriflunomide. The study will be declared positive if the null hypothesis for ARR for tolebrutinib versus teriflunomide is rejected at the 2-sided 0.05 significance level.

ARR will be analyzed using a negative binomial regression model with robust variance estimation (1). The robust variance can be estimated by specifying the participant identifier in the repeated statement using SAS PROC GENMOD (version 9.4 or later). The model will include the total number of adjudicated relapses per participant occurring during the observation period as the response variable, with treatment group, gadolinium (Gd)-enhancing T1 lesions at baseline (presence, absence), EDSS strata (<4, ≥4) and geographic region (US, non-US) as covariates. Log transformed observation duration will be the offset variable.

The estimated ARR for each treatment group and corresponding 2-sided 95% confidence interval will be derived from the negative binomial model where the adjusted mean value within treatment group is based on a sample size dependent weight within each categorical covariate (ie, weight by overall proportion of participants in each stratum or categorical factor, also called “population weight”). The relative reduction in ARR with tolebrutinib compared to teriflunomide, its 2-sided 95% confidence interval and p-value will be provided.

The absolute difference in ARR between tolebrutinib and teriflunomide, two-sided 95% confidence interval and associated p-value derived using the delta method will also be provided as supportive information.

The unadjusted ARR (total number of adjudicated relapses during the study period divided by the total participant years on study) will also be presented by treatment group. Mean cumulative function plot will be provided for descriptive purposes.

3.2.3 Sensitivity analysis

The following sensitivity analyses will be performed to assess the robustness of the conclusion of the main analytical approach:

- ARR based on all Investigator-reported relapses, regardless of adjudication status, based on the ITT population.
- ARR for the first 24-month study period (from randomization; fixed 24-month duration) based on the participants completing at least 24 months on study, regardless of treatment status. All adjudicated relapses occurring from randomization through Month 24, including data collected after premature treatment discontinuation if applicable, will be used for analysis.

- ARR on-treatment analysis based on the randomized and exposed participants. This model will include adjudicated relapses occurring during treatment (first administration of IMP to last administration inclusive) as the response variable and the log transformed duration of treatment as the offset variable. Data collected after premature treatment discontinuation will not be included in this analysis.

In each case, a negative binomial model with the same set of covariates as specified in the primary analysis will be used.

3.2.4 Supplementary analyses

The time to first adjudicated relapse for tolebrutinib versus teriflunomide will be compared via a log-rank test stratified by EDSS strata (<4, ≥4) and geographic region (US, non-US). Participants without an adjudicated relapse event during the study will be censored at their last contact date. Kaplan-Meier (KM) plots of the cumulative incidence rate will be provided by treatment group. The proportion of participants with events at given time points (eg, Month 12, 18, 24, 30, …) will be calculated using the KM estimates. The hazard ratio (95% CI) will be estimated by a Cox proportional hazards model with terms for treatment, Gd-enhancing T1 lesions at baseline (presence, absence), EDSS strata (<4, ≥4), and geographic region (US, non-US).

3.2.5 Subgroup analyses

To assess the homogeneity of the treatment effect, analyses will be performed on the primary endpoint (ARR) across the following subgroups (categories with too few participants for the model to converge will be combined with another category; for factors with only two categories, no subgroup analysis will be performed if fewer than 5 participants in one category):

- Age (≤40, >40 years)
- Sex (Male, Female)
- Race (White, Black or African American, Asian, Other)
- EDSS strata (<4, ≥4)
- Gd-enhancing T1 lesions at baseline (presence, absence)
- Geographic region (US, non-US)
- Region (Eastern Europe [EU], Western EU, North America, rest of the world [ROW])
- Highly active disease (HAD) at baseline (Yes, No)
- Prior disease modifying therapy (DMT) use (Yes, No)
- Relapses in the one year prior to screening (<2, ≥2)
- Time since symptom onset in years (<5, 5 to <10, >=10)

Participants with HAD are defined as having 1 relapse in the previous year AND one of the following: at least 1 Gd enhancing T1-hyperintense lesion; or 9 or more T2-hyperintense lesions at baseline for participants who were already treated with DMT (any treatment with DMTs would be considered, if documented, during the prior year) or who had 2 or more relapses in the previous year, whether treated with DMTs or not.

Treatment by subgroup interaction and its p-value will be derived from a negative binomial model. This model will include the total number of adjudicated relapses occurring during the observation period as the response variable with treatment group, Gd-enhancing T1 lesions at baseline (presence, absence), EDSS strata (<4, ≥4), geographic region (US, non-US), subgroup (if different than the aforementioned covariates), and treatment by subgroup interaction as covariates. Log transformed observation duration will be the offset variable. If quantitative treatment by subgroup interaction is detected with nominal p-value <0.1 for any subgroup factor, a further investigation will be performed to evaluate possible qualitative interaction. The ARR (95% CI) within each subgroup separately will be provided from similar negative binomial models excluding the subgroup and treatment by subgroup interaction terms. Forest plots of relative risks and corresponding 95% CIs, and forest plots of risk differences and corresponding 95% CIs comparing tolebrutinib to teriflunomide within each subgroup will be provided.

3.3 SECONDARY ENDPOINT(S) ANALYSIS

The secondary endpoints detailed in this section are for assessing efficacy of tolebrutinib compared to teriflunomide on disability progression, lesion counts and volumes obtained via magnetic resonance imaging (MRI), cognitive performance, and quality of life.

Other secondary endpoint analyses are defined in Section 3.6.2 (AE, SAE), Section 3.6.3.1 (laboratory abnormalities), Section 3.7.1 (PK) and Section 3.7.2 (PD).

3.3.1 Key secondary endpoint(s)

3.3.1.1 Definition of endpoint(s)

Time to onset of 6-month CDW is defined as the time from randomization to the onset of a confirmed, sustained increase from baseline in EDSS score (of ≥1.5 points when the baseline score is 0, of ≥1.0 point when the baseline score is 0.5 to ≤5.5, of ≥0.5 points when the baseline EDSS score is >5.5) over at least 6 months that is not attributable to another etiology (eg, fever, concurrent illness, or concomitant medication).

- The initial onset increase from baseline EDSS score can be from a scheduled or unscheduled assessment.
- All intermediate EDSS scores (EDSS scores obtained after onset of disability and before the confirmatory assessment), if any, must maintain at least the minimum increase.
- Confirmatory EDSS assessment must be obtained at least 6 months after onset, at a routine quarterly visit, at least 30 days after any confirmed clinical relapse, and not be associated with an ongoing relapse.

Note: 6-month confirmation must be ≥154 days after onset. Given that visits are scheduled every 12 weeks with a time window of ±7 days for conducting EDSS assessment, the minimum per protocol allowed time between 2 EDSS assessments scheduled 24 weeks apart is 154 days (6 months=24 weeks=168 days minus a 7-day window for each of the 2 visits, ie, 168-14=154).

3.3.1.2 Main analytical approach

The primary analysis of the key secondary endpoint, time to onset of 6-month CDW, will be performed on the pooled ITT population from studies EFC16033 and EFC16034. The null hypothesis is the same as for the primary endpoint analysis, ie, no treatment difference between tolebrutinib and teriflunomide versus the alternative of a between-treatment difference.

In this primary ITT analysis:

- For participants who complete the study without an initial onset of disability worsening, the participant will be censored at the date of last EDSS assessment.
- For participants who have an initial onset of disability worsening but complete the study at the common study end without 3-month confirmation, the participant will be censored at the date of last EDSS assessment.
- For participants who complete the study, meet 3-month CDW criteria, and continue to meet criteria for EDSS disability worsening through the final study assessment, but do not reach 6-month confirmation, the 6-month CDW event status of the participant will be determined by an imputation approach. Since in this setting the partial missing data can reasonably be assumed to be missing at random, this approach leverages the partial information and follows the ITT principle. The details of this imputation approach are provided in Section 3.3.1.3.

If a participant dies due to MS, it will be considered a confirmed disability worsening regardless of the baseline EDSS or the change in EDSS. The time to onset will be calculated as (date of EDSS assessment at a tentative onset of the event – date of randomization + 1) or (date of death – date of randomization + 1) if a tentative onset date does not exist. Death for other reasons than MS will not be considered a disability worsening.

The time to onset of 6-month CDW for tolebrutinib versus teriflunomide will be compared via a log-rank test stratified by EDSS strata (<4, ≥4), geographic region (US, non-US), and study (EFC16033, EFC16034). KM plots of the cumulative incidence rate will be provided by treatment group to depict the course of onset of 6-month CDW over time. The proportion of participants with events at given time points (eg, Month 12, 18, 24, 30, …) will be calculated using the KM estimates. The hazard ratio (95% CI) will be estimated by a Cox proportional hazards model with robust variance estimation (1). The covariates are treatment group, Gd-enhancing T1 lesions at baseline (presence, absence), EDSS strata (<4, ≥4), geographic region (US, non-US), and study (EFC16033, EFC16034). An expanded model with a treatment-by-study interaction will be fit to the data, and between-study heterogeneity will be tested (type-3); the corresponding p-value will be provided.

3.3.1.3 Multiple imputation for missingness due to CEOS

For participants who met 3-month CDW criteria, and continued to meet criteria for EDSS disability worsening through the final study assessment, but do not reach 6-month confirmation, the event status (ie, confirmed or not confirmed) of the participant will be determined by a logistic regression model as follows:

$$Y_i \sim Bernoulli(p_{ij}), i = 1, \ldots, n_j \; ; \; j = 1,2$$

$$logit(p_{ij}) = ln\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{3j}x_{3ij}$$

where Y_{ij} is the indicator for event status of 6-month CDW for participant i from group j; 1 = 6-month CDW, 0 = no 6-month CDW, p_{ij} is the probability that participant i from group j has a 6-month CDW event, x_{1ij} is the EDSS strata (<4, ≥4), x_{2ij} is the geographic region (US, non-US), x_{3ij} is the Gd-enhancing T1 lesion status at baseline (absence, presence), β_{kj} are the model parameters ($k = 0,1,2,3; j = 1,2$). The logistic model parameters β_{kj} will be estimated from participants who do not have missing value due to CEOS. Separate logistic models will be estimated within each treatment arm. Then, these estimates will be used to calculate the predicted probability for participants who meet 3-month CDW criteria, and continue to meet criteria for EDSS disability worsening through the final study assessment, and missingness due to CEOS. Finally, the missing value for those participants with an initial onset of disability worsening and missingness due to CEOS will be imputed via a randomly generated value from a Bernoulli distribution with p_{ij}.

Using the above methodology, 1000 complete data sets will be generated with imputed missing data via PROC MI in SAS. Each of the data sets will be analyzed using the Cox proportional hazards model as described above. Log-rank test and KM plot will also be provided based on 1000 complete data sets based on the method in (2). The SAS MIANALYZE procedure will be used to generate statistical inferences by combining results from the 1000 analyses using Rubin's formula.

Sample SAS code is provided below:

```
PROC MI data=DATAIN out=DATAOUT nimpute=1000;
  by  Treatment_group;
  var EDSS_strata geographic_region Gd_T1;
  fcs logistic(Confirmation = EDSS_strata geographic_region Gd_T1/ link=glogit);
run;

PROC MIANALYZE data= DATAOUT;
   modeleffects Estimate;
   stderr StdErr;
run;
```

DATAIN only includes participants who meet the imputation criteria mentioned above. Details of multiple imputation can be found in SAS support document (3).

In addition to the analysis using pooled data, time to onset of 6-month CDW will also be analyzed within each study separately using the same statistical methods as for the primary analysis, but without study included in the model and without any treatment-by-study interaction analysis.

3.3.1.4 Sensitivity analysis (pooled data only)

To assess robustness of the primary analysis, a sensitivity analysis will be performed based on the observed confirmed 6-month CDW where missing or incomplete 6-month CDW will be treated censored at their last EDSS assessment. The analysis methods will be the same as those for the primary analysis of 6-month CDW but without imputation.

3.3.1.5 Subgroup analyses (pooled data only)

To assess the homogeneity of the treatment effect, analyses will be performed on the key secondary endpoint for the same subgroups as specified in Section 3.2.5. There will be no imputation for incomplete/missing data for these subgroup analyses.

Treatment by subgroup interaction and its p-value will be derived from a Cox proportional hazards model with terms for treatment, Gd-enhancing T1 lesions at baseline (presence, absence), EDSS strata (<4, ≥4), geographic region (US, non-US), study (EFC16033, EFC16034), subgroup (if different than the aforementioned covariates) and treatment by subgroup interaction as covariates. The HR (95% CI) comparing tolebrutinib to teriflunomide within each subgroup separately (Cox models without subgroup and treatment by subgroup interaction terms) will be provided and displayed in forest plots.

3.3.2 Supportive secondary endpoint(s)

3.3.2.1 Definition of endpoint(s)

Other secondary endpoints are listed in Table 1.

SDMT is used to assess divided attention, visual scanning, tracking and motor speed. The number of correct substitutions and number of items completed within a 90 second interval (max = 110) are recorded. The baseline mean and standard deviation of all participants are used to create the Z-score for SDMT number of correct substitutions: Z = {(SDMT score – baseline mean SDMT)/baseline SD SDMT)}. Z-scores will be used for analysis.

California verbal learning test-II (CVLT-II) is a verbal learning and memory test. It consists of recall and recognition of a list of 16 words. For each assessment, 5 trials are completed. The total correct, total intrusions, and total repetitions across the 5 trials combined are recorded and converted to a standardized score by sex within each of 7 age groups (4). Total Correct Recall Trials 1–5 is scaled to a normalized T–score metric, which has a mean of 50 and standard deviation of 10, with higher values indicating better performance. Total intrusions and total

repetitions are normed on a linear Z-score metric, with a mean of 0, standard deviation of 1.0, and increment values of plus or minus 0.5. Higher positive Z-scores on these measures reflect greater deficits. Standardized scores will be used for analysis.

MSQoL-54 is a multidimensional health-related quality of life measure that combines both generic and MS-specific items into a single instrument (5). The recall time, when relevant, is for the last 4 weeks. There are 2 derived summary scores, physical health composite and mental health composite, and 12 subscale scores: physical function, role limitations-physical, role limitations-emotional, pain, emotional well-being, energy, health perceptions, social function, cognitive function, health distress, overall quality of life, and sexual function, and 2 single-item measures: satisfaction with sexual function and change in health. The total number of items in each subscale is listed as the divisor for each subtotal. Due to missing data, the divisor can be less than that if not every item within a given scale has been answered. All scores are transformed linearly to a common 0 (worst) to 100 (best) score range.

3.3.2.2 Main analytical approach

Time to onset of 3-month CDW will be analyzed using the same statistical methods as for the primary analysis of time to onset of 6-month CDW. Primary analyses will be in the pooled EFC16033 and EFC16034 ITT population. Analysis within each study separately will also be provided.

Note: 3-month confirmation must be ≥70 days after onset. Given that visits are scheduled every 12 weeks with a time window of ± 7 days for conducting EDSS assessment, the minimum per protocol allowed time between 2 EDSS assessments scheduled 12 weeks apart is 70 days (3 months=12 weeks=84 days minus a 7-day window for each of the 2 visits, ie, 84-14=70).

Time to onset of 6-month CDI will be analyzed using similar statistical methods as for the primary analysis of time to onset of 6-month CDW, but without imputation. Participants with 6-month CDI onset without confirmation will be treated censored at the last EDSS assessment. Primary analyses will be in the pooled EFC16033 and EFC16034 ITT population. Analysis within each study separately will also be provided.

For categorical efficacy endpoints with count data (annualized rate of new/enlarging T2-hyperintense lesions and average number of new gadolinium- (Gd-) enhancing T1-hyperintense lesions per scan), similar negative binomial regression analyses as for the primary analysis of ARR will be performed in the ITT population. The response variable in this case will be the total count of the respective type of lesion across all scheduled MRI scans during the study period and the offset variable will be the log transformed study duration for T2 lesions and number of MRI scans for T1 lesions. Randomization stratification factors will remain as covariates in the model. For analysis of new Gd-enhancing T1 lesions, Gd-enhancing T1 lesions at baseline (presence, absence) will remain as a covariate, but for analysis of new/enlarging T2 lesion count, this covariate will be replaced by baseline T2 lesion count. The average number of lesions per scan (95% CI) for each treatment group will be estimated from the model along with the relative risk (95% CI, p-value) for tolebrutinib compared to teriflunomide.

Percent change in brain volume loss at the EOS compared to Month 6 will be analyzed using a mixed-effect model with repeated measures (MMRM) approach in the ITT population for estimating the treatment effect with appropriate transformation if necessary (eg, log-transform). Estimates derived from the MMRM approach will be weighted by overall proportion of participants in each stratum or categorical factor. The model will include percent change from Month 6 values at scheduled visits up to and including the end of study visit as the response variable, and treatment, EDSS strata (<4, ≥4), geographic region (US, non-US), visit, treatment by-visit interaction, Month 6 brain volume, and Month 6 brain volume-by-visit interaction as covariates; have an unstructured variance-covariance matrix and be fitted using restricted maximum likelihood estimation. If this model fails to converge, the following variance-covariance structures will be tested in this order:

- Heterogeneous Toeplitz (TOEPH)
- Heterogeneous first-order autoregressive (ARH(1))
- Heterogeneous compound symmetry (CSH)
- Toeplitz (TOEP)
- AR(1)
- CS

The first (co)variance structure yielding convergence will be used as the primary analysis. Comparisons between treatment arms will be made using least squares mean contrasts at the EOS with denominator degrees of freedom estimated using the Kenward-Roger approximation (6). When a variance-covariance structure other than unstructured is used, the denominator degrees of freedom will be estimated using the between within method (DDFM=BW in SAS PROC MIXED) (7).

Example SAS code for the MMRM is:

```
ods output Diffs = out1
        LSMeans = out2;
proc mixed data = DATAIN method = reml;
  class TRTPN VISIT USUBJID;
  model PCHG = EDSS REGION TRTPN VISIT TRTPN*VISIT M6 M6*VISIT /
       s ddfm=kr;
  repeated VISIT / type = un subject = USUBJID  r;
  lsmeans TRTPN TRTPN*VISIT / OM cl diff;
run;
```

where TRTPN = treatment group, VISIT refers to the protocol scheduled visits for the endpoint up to and include the EOS, USUBJID = unique participant ID, PCHG is the percent change from Month 6 in brain volume, M6 is the Month 6 brain volume, EDSS refers to the EDSS strata, and REGION refers to the geographic region strata. Type specifies the variance-covariance structure and can take on a value of UN, TOEPH, ARH(1), CSH, TOEP, AR(1) or CS as described above.

Change from baseline in SDMT number of correct substitutions, CVLT-II total correct and MSQoL-54 physical and mental health composite scores at EOS will be analyzed using

an analogous MMRM approach in the ITT population with the change from baseline values for the respective endpoint at scheduled visits up to and including the end of study visit as response variables, and treatment, EDSS strata (<4, ≥4), geographic region (US, non-US), visit, treatment by-visit interaction, baseline value for the endpoint being assessed and baseline value by-visit interaction as covariates.

For each endpoint, difference in least squares means, the corresponding 95% CI, and p-value will be provided for the comparison of tolebrutinib versus teriflunomide. Plots of least squares means (LSM) [±standard error (SE)] over time will be provided. Participants who discontinue study treatment before the common study end will be asked and encouraged to return for all remaining study visits and the additional off-treatment values measured up through the EOS visit will be included in the primary analysis. For participants who withdraw from the study before the CEOS, values will be missing after study discontinuation and no imputation will be performed. In order to assess the impact of deviation from the normality assumption, a sensitivity analysis using rank ANCOVA with the same covariates as for the MMRM analysis except visit and related interaction terms, will be performed for these continuous supportive secondary endpoints to provide the p-value for the comparison between the treatment groups. No other sensitivity, supplementary, or subgroup analyses will be performed for these supportive secondary endpoints.

3.4 TERTIARY/EXPLORATORY ENDPOINT(S) ANALYSIS

3.4.1 Definition of endpoint(s)

Tertiary/exploratory endpoints are listed in Table 1.

The T25-FW is a quantitative mobility and leg function performance test based on a timed 25-foot walk; time in seconds to walk 25 feet (lower bound: 2.2 sec; upper bound: 180 sec). For each assessment, 2 trials are performed. The score is the average of the times from the 2 trial (or single score if only 1 completed). The baseline mean and standard deviation of all participants are used to create Z-scores: Z = -{(Average 25-FW - Baseline Mean T25-FW) /Baseline SD T25-FW}. The negative of the Z-score is used so that improvement is a positive number and worsening a negative number. For a participant who could not complete the 25-FW due to a “physical limitation”, the maximum time (180 sec) will be used. Missing values for other reasons will not be imputed. In case of values below (above) the lower (upper) bound, the value will be set to the respective bound.

The 9-HPT is a brief, standardized, quantitative test of upper extremity function (lower bound: 10 sec; upper bound: 300 sec). Two consecutive trials with the dominant hand are immediately followed by two consecutive trials with the non-dominant hand. The total time to complete the task is recorded. To transform the data so that higher Z-scores correspond to an improved outcome and lower Z-scores correspond to deterioration, the inverse of the test values are taken prior to computing the Z-score. The baseline mean and standard deviation of all participants are used to create Z-scores: Z = {(1/trialarm,left + 1/trialarm,right) / 2.0 – Baseline mean (1/9-HPT)} / Baseline SD(1/9-HPT). The two trials for each hand are averaged, converted to the reciprocals of the mean times for each hand and then the two reciprocals are averaged. Where 1/trialarm,left = (1/average of the two times of the left arm trials); if one is missing then use the one available. For

a participant who could not complete the 9-HPT due to a "physical limitation", the maximum time (300 sec) will be used. Investigators will be reminded that missing data should be avoided. However, in case of missing data for other reason than inability to complete the test, missing data will be treated as follows:

- In case only 1 trial was done for a hand, the average of the two trials will be replaced by the time for the available trial.
- In case 9-HPT was assessed for only one hand, the overall 9-HPT will not be calculated and considered missing.

Missing values for other reasons will not be imputed. In case of values below (above) the lower (upper) bound, the value will be set to the respective bound.

MSFC-3 is a multidimensional, three-component performance scale to assess the degree of impairment in MS patients. Here, the MSFC-3 score is defined as the average of the T25-FW, 9-HPT and SDMT Z-scores.

EQ-5D-5L is a standardized health-related quality of life questionnaire that provides a simple, generic measure of health for clinical and economic appraisal. It consists of 2 pages, the EQ-5D descriptive system and the EQ visual analogue scale (VAS). The EQ-5D descriptive system comprises 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. EQ-5D-5L health status will be converted into a single index value by using the value sets based on UK population and according to the crosswalk algorithm developed by B. van Hout (8). The index score is anchored at 1 (full health) with 0 (negative values) corresponding to states being as bad or worse than dead. The EQ VAS records the respondent's self-rated health on a vertical VAS ranging from 0 (worst imaginable health state) at the bottom to 100 (best imaginable health state) at the top. Overall, higher scores indicate high health utility.

3.4.2 Main analytical approach

Summary statistics including mean, standard deviation, median, interquartile range (Q1, Q3), minimum, and maximum will be provided for the following continuous data endpoints at scheduled visits over the study period including the EOS. Only non-missing data will be included. For change from baseline, only participants with baseline and post-baseline data will be summarized.

- EDSS score and change from baseline
- Number of Gd-enhancing T1 lesions and change from baseline
- Number of new Gd-enhancing T1 lesions
- Number of new/enlarging T2 lesions per 6-month interval
- T2 lesion volume and change from baseline
- Cumulative number of new T1-hypointense lesions

- T1-hypointense lesion volume and change from baseline.
- Number of phase rim lesions and change from baseline
- Normalized brain volume at screening and percent change in brain volume at post baseline time points
- Number of slowly evolving lesions at EOS compared to screening
- Volume of slowly evolving lesions at EOS compared to screening
- SDMT number of correct substitutions Z-score and change from baseline
- SDMT number of completed items and change from baseline
- CVLT-II standardized scores for total number correct and change from baseline
- CVLT-II standardized scores for total number of intrusions and change from baseline
- CVLT-II standardized scores for total number of repetitions and change from baseline
- MSFC-3 and change from baseline
- T25-FW raw score and Z-score for leg function/ambulation and change from baseline
- 9-HPT raw score and Z-score for arm/hand function and change from baseline
- MSQoL-54 score and change from baseline for the 2 derived composite scores, 12 subscale scores and 2 single item scores
- EQ-5D-5L health state utility value score and change from baseline
- EQ VAS score and change from baseline

Plots of mean (±SD) and/or mean change (±SD) and/or median over time will also be provided for better visualization.

No evidence of disease activity-3 (NEDA-3) is defined as absence of all of the following:

- 6-month CDW
- Active MRI lesions (both new or enlarged T2-hyperintense lesions and Gd-enhancing T1-hyperintense lesions)
- MS relapses

The proportion of adjudicated relapse-free participants at the EOS and the proportion of participants with NEDA-3 at Months 18, 24, and the EOS will be analyzed in the ITT population via a stratified Cochran-Mantel-Haenszel (CMH) test, stratified by EDSS strata (<4, ≥ 4) and geographical region (US, non-US). The relative ratio for tolebrutinib versus teriflunomide, 95% confidence interval, and p-value will be provided.

Time-to-event endpoints confirmed over 3 and 6 months (20% worsening in 9-HPT, 20% worsening in T25-FW, 4-point decrease in SDMT) will be analyzed using similar statistical methods as for the primary analysis of time to onset of 6-month CDW (key secondary endpoint) in the pooled EFC16033 and EFC16034 ITT population, but without multiple imputation for

incomplete/missing data (ie, censoring at last assessment). Analysis within each study separately will also be provided.

Change from baseline in total volume of T2 hyperintense lesions (M18, M24, and EOS), MSFC-3 Z-score (M12, M18, M24, and EOS), and EQ-5D-5L index and VAS scores (by scheduled visit) will be analyzed in the ITT population using a similar MMRM approach as described for the continuous supportive secondary endpoints in Section 3.3.2. Analysis of T2 lesion volume will be performed with appropriate transformation if necessary (eg, log-transform). In order to assess the impact of deviation from the normality assumption, a sensitivity analysis using rank ANCOVA with the same covariates as for the MMRM analysis except visit and related interaction terms, will be performed for these continuous tertiary endpoints to provide the p-value for the comparison between the treatment groups.

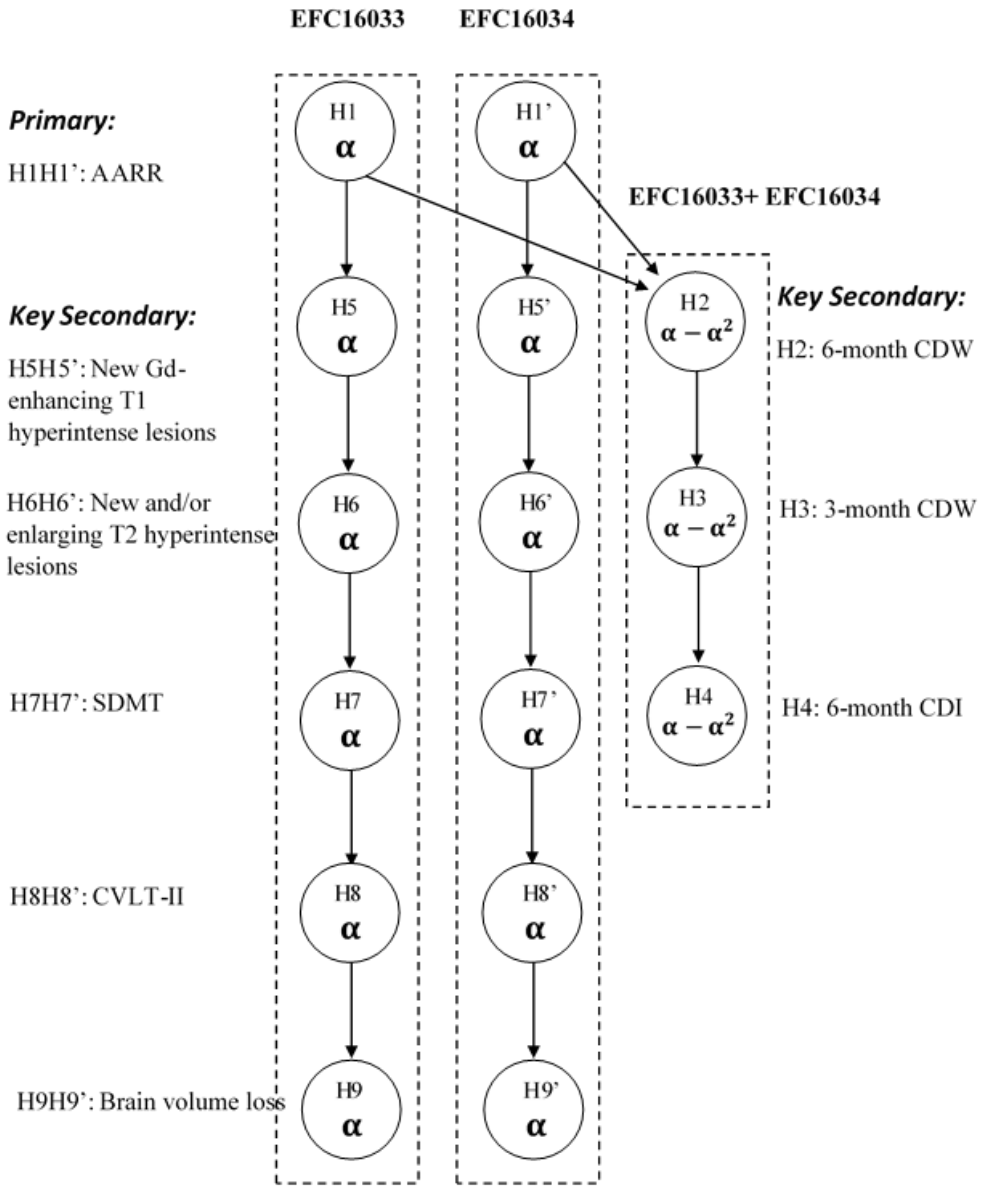
The participant/lesion time courses for magnetization transfer ratio (MTR) recovery in Gd-enhancing lesions detected at Months 6 and 12 and normalized T1 intensity evolution in slowly evolving lesions will be listed.

Missing data will not be imputed in the analysis of tertiary endpoints, nor will other sensitivity or subgroup analyses be conducted.

3.5 MULTIPLICITY ISSUES

To strongly control the type 1 error rate at the level of the individual studies and at the level of the pooled studies, the testing strategy in (9) illustrated in Figure 1 will be implemented. This allows for sequential, hierarchical testing of selected endpoints within each study given the null hypothesis for the primary endpoint (ARR) of the study is rejected at 2-sided 0.05. Additionally, given the null hypothesis for the primary endpoint (ARR) is rejected in both studies (each at 2-sided 0.05), it allows for sequential, hierarchical testing of selected disability endpoints in the pooled EFC16033 and EFC16034 study population. Hypotheses can only be tested in sequential order as indicated by the arrows. A specific hypothesis can only be tested if all proceeding null hypotheses in the sequence have been rejected. The number associated with each hypothesis (α and α-α^2) indicates the significance level at which the hypothesis can be tested. The testing is performed at 2-sided α=0.05 within each study, but the corresponding efficacy claim can be made only if both studies demonstrate statistical significance for the associated endpoint. The family-wise type I error rate for each study is controlled at 2-sided α=0.05. The type I error rate for the primary endpoint (H1, H1') and secondary endpoints (H5, H5'; H6, H6'; H7, H7'; H8, H8'; H9, H9') is controlled at 2-sided α^2 (0.05^2). For all secondary endpoints, including those based on the pooled data (H2, H2'), (H3, H3') and (H4, H4'), the family-wise type I error is controlled at 2-sided 0.05.

Figure 1 - Testing procedure and type 1 error control



3.6 SAFETY ANALYSES

All safety analyses will be performed for each study on the safety population as defined in Section 2, unless otherwise specified, using the following common rules:

- The analysis of the safety variables will be descriptive, and no testing is planned. The summary of safety results will be presented by treatment group.
- Safety data in participants who do not belong to the safety population (eg, exposed but not randomized) will be provided separately.

3.6.1 Extent of exposure

The extent of IMP exposure will be assessed by the duration of IMP exposure and summarized within the safety population.

Duration of IMP exposure

Duration of IMP exposure is defined as last IMP administration date – first IMP administration date + 1 day, regardless of unplanned intermittent discontinuations. If the date of the last dose of IMP is missing, the duration of IMP will be left as missing.

Duration of IMP exposure will be summarized quantitatively and categorically:

- >0 and <24 weeks
- ≥24 and <48 weeks
- ≥48 and <72 weeks
- ≥72 and <96 weeks
- … (24-week intervals)

Additionally, the cumulative duration of treatment exposure (expressed in participant-years) will be provided.

Treatment compliance

A given administration will be considered non-compliant if the participant did not take the planned dose of treatment as required by the protocol, ie, once daily oral dose. No imputation will be made for participants with missing or incomplete data.

Percentage of compliance for a participant will be defined as the number of days that the participant was compliant divided by the total number of days of planned administration during the treatment period.

Treatment compliance will be summarized quantitatively and categorically: <80%, ≥80%.

3.6.2 Adverse events

General common rules for adverse events

All adverse events (AE(s)) will be graded according to National Cancer Institute Common Terminology for Adverse Events (NCI-CTCAE version 5.0, 27 Nov 2017) and coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using the Medical Dictionary for Regulatory Activities (MedDRA) version currently in effect at Sanofi at the time of database lock.

The AEs will be analyzed in the following 3 categories:

- Pre-treatment AEs: AEs that developed, worsened, or became serious during the pre-treatment period.
- Treatment-emergent adverse events (TEAE(s)): AEs that developed, worsened, or became serious during the treatment period
- Post-treatment AEs: AEs that developed, worsened, or became serious during the post-treatment period

Similarly, if occurring, deaths will be analyzed in the pre-treatment, treatment-emergent, and post-treatment period.

The primary focus of AE reporting will be on TEAEs. Pre-treatment and post-treatment AEs will be described separately.

An AE with incomplete or missing date/time of onset (occurrence, worsening, or becoming serious) will be classified as a TEAE unless there is definitive information to determine it is a pre-treatment or a post-treatment AE.

If the assessment of the relationship to IMP is missing for an AE, this AE will be assumed as related to the IMP. If the grade is missing for 1 of multiple treatment-emergent occurrences of a specific AE for a participant, the grade will be imputed with the maximal grade of the other occurrences. If the grade is missing for all the occurrences, the grade will be left as missing.

Multiple occurrences of the same event in the same participant will be counted only once in the tables within a treatment period, using the maximum (worst) grade by treatment period.

The AE tables will be sorted as indicated in Table 4.

Table 4 - Sorting of AE tables

AE presentation	Sorting rules
SOC and PT	By the internationally agreed SOC order and decreasing frequency of PTs[a]
PT	By decreasing frequency of PTs[a]
AESI category and PT	By AESI category (protocol order) and decreasing frequency of PTs[a]

a Sorting will be based on the tolebrutinib treatment group incidence; alphabetic order in case of equal frequency.

Analysis of all adverse events

The overview of TEAEs will be generated presenting the number (%) of participants with:

- Any TEAE
- Any treatment emergent SAE
- A TEAE leading to death
- Any TEAE leading to permanent discontinuation of IMP

- Any treatment emergent AESI
- Any TEAE considered by the investigator as related to IMP

An additional overview summary including the number and rate of events will be provided.

The AE summaries of Table 5 will be generated with number (%) of participants experiencing at least one event.

Listings of SAEs from randomized participants, AEs leading to treatment/study discontinuation and severe AEs will be provided.

Table 5 - Analyses of adverse events

Type of AE	MedDRA levels
All TEAEs	Primary SOC and PT
	PT (incidence ≥ 2%)
Severe (grade ≥3) TEAEs	Primary SOC and PT
	PT
TEAEs related to IMP as per Investigator's judgment	Primary SOC and PT
	PT
Treatment-emergent SAEs	Primary SOC and PT
	PT
Treatment emergent SAEs related to IMP as per Investigator's judgment	Primary SOC and PT
Treatment emergent AESIs	AESI category and PT
TEAEs leading to permanent treatment discontinuation	Primary SOC and PT
	PT
TEAEs leading to death (death as an outcome of the AE as reported by the Investigator in the AE page)	Primary SOC and PT
Pre-treatment AEs	Overview[a]
	Primary SOC and PT
Post-treatment AEs	Overview[a]
	Primary SOC and PT

a Will include the following AE categories: any AEs, any severe (grade ≥ 3) AE, any serious AEs, any AEs leading to death

Analysis of deaths

In addition to the analyses of deaths included in Table 5, the number (%) of participants in the following categories will be provided:

- Deaths during the treatment-emergent and post-treatment periods
- Deaths in non-randomized participants or randomized but not treated participants

Analysis of adverse events of special interest (AESIs)

Adverse events of special interest (AESIs) are as described in Table 6. Number (%) of participants experiencing at least one event will be provided. Tables will be sorted as indicated in Table 4. A listing of AESIs will also be provided.

Table 6 - Selections for AESIs

AESIs	Selection
Pregnancy of a female participant or female partner of a male participant	Dedicated CRF page (AECAT="PREGNANCY DATA")
Symptomatic overdose (serious or non-serious) with IMP	Dedicated CRF page (AECAT="OVERDOSE DATA"); must be symptomatic and related to IMP
Increase of ALT > 3xULN (confirmed)	Dedicated CRF page (AECAT="ALT INCREASE DATA") and AESI marked "Y"
ECG observation of atrial fibrillation/atrial flutter	CMQ30003 for selection and AESI check box marked
Severe infection (NCI CTCAE grade 3 or above), that may or may not meet seriousness criteria (eg, a grade 3 opportunistic infection)	SOC of Infections and Infestations, grade 3 or above and AESI marked "Y"
Moderate or severe hemorrhagic events (NCI CTCAE grade 2 or above), including, but not limited to, symptomatic bleeding, bleeding in a critical area or organ such as the CNS, or intraocular bleeding	SMQNAME='Haemorrhage terms (excl laboratory terms)" for selection, grade 2 or above, and AESI marked "Y"
Thrombocytopenia, platelet count <75 x 10^9/L	PT="Thrombocytopenia" and AESI marked "Y"

3.6.3 Additional safety assessments

3.6.3.1 Laboratory variables, vital signs and electrocardiograms (ECGs)

The following laboratory variables, vital signs, and electrocardiogram (ECG) variables will be analyzed. They will be converted into standard international units [and conventional unit, if applicable].

- Hematology:
  - Red blood cells and platelets: red blood cell count, platelet count, hemoglobin, hematocrit
  - White blood cells: white blood cell count, neutrophils, lymphocytes, monocytes, basophils, eosinophils
- Clinical chemistry:
  - Metabolism: glucose, total protein, albumin, creatine phosphokinase, lipase
  - Electrolytes: sodium, potassium, chloride, calcium, bicarbonate

  - Renal function: creatinine, creatinine clearance, blood urea nitrogen. Creatinine clearance will be derived with the equation of Cockcroft and Gault using weight assessed at the same visit as creatinine.
  - Liver function: alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, total and direct bilirubin
- Urinalysis for quantitative analysis: pH, specific gravity
- Vital signs: heart rate, systolic and diastolic blood pressure, weight, temperature
- ECG variables: heart rate, PR, QRS, QT, and corrected QTcF (according to Fridericia)

For laboratory data reported as <xx or >yy, a numeric value of xx/2 or yy, respectively, will be used for quantitative analysis.

Quantitative analyses

For all laboratory variables, vital signs, and ECG variables above, descriptive statistics for results and changes from baseline will be provided for each analysis visit window during the treatment period. These analyses will be performed using central measurements only for laboratory variables.

For each laboratory parameter, vital sign parameter and ECG parameter, mean changes from baseline with the corresponding standard error will be plotted over time.

Analyses according to PCSA

Potentially clinically significant abnormality (PCSA) analyses will be performed based on the PCSA list currently in effect at Sanofi at the time of the database lock. For parameters for which no PCSA criteria are defined, similar analyses will be done using the normal range.

Analyses according to PCSA (normal range) will be performed based on the worst value during the treatment-emergent period, using all measurements (either local or central, either scheduled, nonscheduled or repeated).

For laboratory variables, vital signs, and ECG variables above, the incidence of participants with at least one PCSA during the treatment-emergent period will be summarized regardless of the baseline level and according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA (normal range) criterion or criteria

Additional analyses for drug-induced liver injury

The following additional analyses will be performed for drug-induced liver injury:

- Time to onset of the initial ALT elevation (>3 x ULN, >5 x ULN, >8 x ULN, >10 x ULN), the initial AST elevation (>3 x ULN, >5 x ULN, >8 x ULN, >10 x ULN), and the initial

ALT or AST elevation >3 x ULN with total bilirubin elevation >2 x ULN will be presented by treatment group with KM curves.

- A graph of the distribution of peak values of ALT versus peak values of total bilirubin during the treatment-emergent period will be provided.
- The normalization (to ≤1 x ULN or return to baseline) of elevated liver function tests of ALT will be summarized by categories of elevation (3 x ULN, 5 x ULN, 10 x ULN, 20 x ULN for ALT), with the following categories of normalization: never normalized, normalized before study intervention discontinuation, and normalized after study intervention discontinuation. Note that a participant will be counted only under the maximum elevation category.

3.6.3.2 Analysis of suicidality assessment

The number (%) of participants with suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent based on the C-SSRS during treatment will be summarized. A shift table for baseline versus during treatment responses will be provided according to the categories of no suicidal ideation or behavior, suicidal ideation and suicidal behavior.

3.7 OTHER ANALYSES

3.7.1 PK analyses (EFC16033 only)

Tolebrutinib and M2 metabolite PK individual parameters (at least C_{max}, t_{max}, AUC_{0-24}) and corresponding descriptive statistics (mean, geometric mean, median, standard deviation, coefficient of variation, minimum, and maximum) will be determined by POP PK analysis by PKDM department using sampling times on Month 6, 9 and 12.

Tolebrutinib and M2 metabolite PK parameters will be reported in a standalone report outside of the CSR.

3.7.2 Pharmacodynamic (PD) analyses

Blood samples for lymphocyte phenotyping (EFC16033 only) will be collected during the study for a subset of participants randomized (approximately 400). Flow cytometry will be used for evaluation of change (and % change) from baseline in lymphocyte count and phenotype subsets. Additional pharmacodynamics evaluations will include assessment of neurofilament light chain (NfL) levels in plasma, Chi3L1, and Ig levels in serum.

These parameters will be summarized by actual treatment group on the safety population using the following descriptive statistics: mean, geometric mean, median, standard deviation, coefficient of variation, minimum, and maximum.

3.8 INTERIM ANALYSES

No formal efficacy interim analysis is planned.

4 SAMPLE SIZE DETERMINATION

While the primary endpoint for each study is ARR, the studies have been sized and are event driven based on the secondary endpoint of 6M-CDW.

Approximately 1200 people will be screened to achieve 900 (±10%) participants per study randomly assigned to the study intervention: approximately half to each treatment group, tolebrutinib and teriflunomide. This study sample size provides greater than 90% power to detect a 45% relative reduction of ARR (primary endpoint) with tolebrutinib compared to teriflunomide, based on a negative binomial distribution for the number of relapses with the following assumptions:

- ARR of 0.29 for teriflunomide
- 1:1 randomization
- variance inflation factor (defined as variance divided by mean) of 1.5 for each arm
- 20% study discontinuation by Month 24

These calculations are based on 2-sided $\alpha = 0.05$. For the primary ITT analysis, study duration will be variable (approximately 18 to 36 months) given the event-driven trial design based on the key secondary efficacy endpoint (confirmed 6-month CDW). All randomized participants are expected to have at least an 18-month study duration, with more than 60% (approximately 550 participants) having at least 24 months. In this subset of participants, the study will also have 90% power to detect a 45% reduction for the analysis of 24-month ARR (analysis of ARR during a fixed 24-month period from randomization).

Studies EFC16033 and EFC16034 pooled together will have a total sample size of at least 1800 participants. EFC16033 along with EFC16034 are planned as an event-driven trials to have 90% power, by pooling the two studies, for assessing the key secondary efficacy endpoint of time to onset of 6-month CDW. Each study will continue until approximately 162 events are projected to have occurred in the pooled data, to ensure approximately 90% power to detect a 40% risk reduction in 6-month CDW in participants treated with tolebrutinib compared to those treated with teriflunomide, based on an assumed 24-month event rate of 12% in the teriflunomide arm. Without pooling (ie, approximately 81 events per study), each study will have more than 60% power to detect a 40% risk reduction in 6-month CDW. These calculations are based on 2-sided testing and $\alpha = 0.05$.

Event rates in the tolebrutinib arm are assumed to be similar to those observed in the ocrelizumab arms of RMS trials (9) The forecast of CDW risk reduction by 40% is based on current knowledge of SAR442168, similar to that of ocrelizumab. The forecast of the event rates in the teriflunomide group is based on observations in previous teriflunomide trials, taking into account tendencies of event rate decrease in recent trials of other MS DMTs such as those that have been observed in beta-interferon groups in ocrelizumab pivotal trials (9, 10).

5 SUPPORTING DOCUMENTATION

5.1 APPENDIX 1 LIST OF ABBREVIATIONS

AE(s):	adverse event(s)
AESIs:	adverse events of special interest
ARR:	annualized relapse rate
ATC:	anatomic therapeutic category
CDI:	confirmed disability improvement
CDW:	confirmed disability worsening
CEOS:	common end of study
CI(s):	confidence interval(s)
CMH:	Cochran Mantel Haenszel
CVLT-II:	California verbal learning test-II
DB:	double-blind
DMT:	disease modifying therapy
eCRF:	electronic Case Report Form
EDSS:	Expanded Disability Status Scale
EOS:	end of study
EU:	Europe
Gd:	gadolinium
HAD:	highly active disease
HLGT:	high-level group term
HLT:	high-level term
IC:	informed consent
IMP:	investigational medicinal product
KM:	Kaplan-Meier
LFU:	lost to follow-up
LLT:	lower-level term
LSM:	least squares means
LTS:	long-term safety
MedDRA:	Medical Dictionary for Regulatory Activities
MMRM:	mixed-effect model with repeated measures
MTR:	magnetization transfer ratio
NCI-CTCAE:	National cancer institute common terminology for adverse events
PCSA:	potentially clinically significant abnormality
PD:	pharmacodynamic
PT:	preferred term
RMS:	relapsing forms of multiple sclerosis
ROW:	rest of the world
SAP:	statistical analysis plan
SD:	standard deviation
SE:	standard error
SOC:	system organ class

TE:	treatment emergent
TEAE(s):	treatment-emergent adverse event(s)
VAS:	visual analogue scale
WHO-DD:	World Health Organization-drug dictionary

5.2 APPENDIX 2 PARTICIPANT DISPOSITION

The number (%) of participants included in each of the analysis populations listed in Table 3 will be summarized.

Screened participants are those with a signed informed consent. Screen failures are defined as participants who consent to participate in the study but are not subsequently randomized. The number of screened participants will be summarized along with the number (%) of screen failures overall and by reason(s).

The number (%) of participants in the following categories will be provided:

- Exposed but not randomized, if applicable
- Randomized but not exposed, if applicable
- Randomized and exposed
  - Completed the study intervention period
  - Did not complete the study intervention period including main reason for intervention discontinuation and reason for intervention withdrawal by participant
- Completed the study period
- Did not complete the study period including main reason for study discontinuation
- Status at last contact (alive, dead)
- Entered LTS study

Reasons for permanent study intervention and study discontinuation, “adverse event”, and “other reasons” will be split as related versus not related to COVID-19, if applicable.

For all categories of participants (except for nonrandomized) percentages will be calculated using the number of randomized participants as the denominator.

In addition, the number (%) of participants screened, screened-failed, randomized, with permanent intervention discontinuation and with early study discontinuation will be provided by country and site. Listings of other reasons for intervention discontinuation, and participants who are exposed but not randomized will be provided.

Protocol deviations

Critical and major protocol deviations (automatic or manual) will be summarized in the randomized population as well as displayed separately as related versus not related to COVID-19 if applicable. A listing of critical and major protocol deviations will be provided.

5.3 APPENDIX 3 DEMOGRAPHICS AND BASELINE CHARACTERISTICS, PRIOR OR CONCOMITANT MEDICATIONS

Demographics, baseline characteristics, medical surgical history

The following demographic and baseline characteristics, smoking and alcohol history, disease characteristics at baseline, and medical and surgical history will be summarized using descriptive statistics in the randomized population.

Demographic and baseline characteristics

- age in years as a quantitative variable and in categories (≤40, >40; 18 to 55)
- sex (Male, Female)
- race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Multiple, Unknown, Not reported)
- ethnicity (Hispanic or Latino, Not Hispanic or Latino, Unknown, Not reported)
- region (US, non-US; North America, Western Europe, Eastern Europe, ROW)
- weight in kg as a quantitative variable
- BMI in kg/m^2 as a quantitative variable and in categories (<25, ≥25 to <30, ≥30)

Smoking and alcohol habits

- smoking history (Never, Current, Former)
- cessation prior to screening for Former smokers in months
- cigarettes per day for smokers
- frequency of alcohol drinking in the past 12 months (Never, Occasional, At least monthly, At least weekly, At least daily)
- # standard alcohol drinks on typical day when drinking (1 or 2, >2)

Disease characteristics at baseline

- prior DMT (0, 1, 2, >2)
- MS type (Relapsing remitting, Secondary progressive)
- time since symptom onset in years
- time since diagnosis in years as a quantitative variable and in categories (<1, ≥1 to 5, ≥5 to 10, ≥10)
- time since most recent relapse in months
- number of relapses within the past 1 year as a quantitative variable and in categories (0, 1, 2, ≥3)
- number of relapses within the past 2 years as a quantitative variable and in categories (0, 1, 2, ≥3)

- baseline EDSS score as a quantitative variable (average of screening and baseline scores when both available)
- time since most recent MRI prior to screening in months
- number of active Gd-enhancing lesions on last MRI prior to screening as a quantitative variable and in categories (0, 1, 2, ≥3)
- screening EDSS score (<4, ≥4).
- EDSS strata (<4, ≥4).
- Gd-enhanced T1 lesions at baseline as a quantitative variable and in categories (presence, absence).
- HAD at baseline (Yes, No).

Baseline safety and efficacy parameters (apart from those listed above) will be presented along with the safety and efficacy summaries.

Relevant medical (or surgical) history collected in the eCRF will be coded to a LLT, PT, HLT, HLGT, and associated primary SOC using the MedDRA version currently in effect at Sanofi at the time of database lock and will be summarized by primary SOC and HLT (internationally agreed SOC order and decreasing frequency of HLTs in the Overall group).

Prior or concomitant medications

All medications will be coded using the version of the World Health Organization-Drug Dictionary (WHO-DD) currently in effect at Sanofi at the time of database lock.

- Prior medications are those the participant used prior to first IMP intake. Prior medications can be discontinued before first administration or can be ongoing during treatment period.
- Concomitant medications are any interventions received by the participant concomitantly to IMP during the treatment period.
- Post-treatment medications are those the participant took in the period running from the end of the concomitant medications period up to the end of the study, if applicable.
- A given medication can be classified as a prior medication and/or as a concomitant medication and/or as post-treatment medication. If it cannot be determined whether a given medication was taken prior or concomitantly or post, it will be considered as prior, concomitant, and post-treatment medication.

Any prior DMT use will be summarized by the standardized medication name in the randomized population, sorted by decreasing frequency in the Overall group. Additionally, prior, concomitant, and post-treatment (if applicable) medications will be summarized for the randomized and exposed population by anatomic and therapeutic level. The summaries will be sorted by decreasing frequency of anatomic category (ATC) based on incidence in the tolebrutinib group. In case of equal frequency, alphabetical order will be used. Participants will be counted once in each ATC category (anatomic or therapeutic) linked to the medication.

5.4 APPENDIX 4 DATA HANDLING CONVENTIONS

Analysis windows for time points

The following analysis windows will decide how the scheduled and/or unscheduled visits will be used in the by-visit analyses of efficacy, safety, and PD variables.

A measurement (scheduled or unscheduled) will be used if it is available and measurement date is within the analysis window.

After applying these time windows, if multiple assessments are associated to the same time point, the closest from the targeted study day will be used. If the difference is a tie, the value after the targeted study day will be used. If multiple valid values exist within a same day, then the first value of the day will be selected.

If there is no measurement for a given parameter in an analysis window, data will be considered missing for the corresponding visit and similar quarterly windows will be applied if the treatment period extends beyond 36 months.

Table 7 - Analysis window definition of efficacy variables

Visits	Target Day	EDSS		MRI		T25FW, 9-HPT, SDMT, CVLT-II		MSQoL-54, EQ-5D-5L	
		Start Day	End Day	Start Day	End Day	Start Day	End Day	Start Day	End Day
Baseline	1	-40	1	-120	1	-40	1	-40	1
Month 3	84	2	126			2	126		
Month 6	168	127	210	2	252	127	210	2	252
Month 9	252	211	294			211	294		
Month 12	336	295	378	253	420	295	378	253	420
Month 15	420	379	462			379	462		
Month 18	504	463	546	421	588	463	546	421	588
Month 21	588	547	630			547	630		
Month 24	672	631	714	589	840	631	714	589	756
Month 27	756	715	798			715	798		
Month 30	840	799	882			799	882	757	924
Month 33	924	883	966			883	966		
Month 36	1008	967	1050	841	1176	967	1050	925	1092
EOS		Start Day: last visit window upper limit +1 End Day: EOS visit date Nominal visit (9000)							
FU for participants who completed treatment to EOS but do not enter LTS		≥(Date of EOS+1) Nominal visit (8010)							
pEOT		Start Day: date of discontinuation+1 End Day: start of the next scheduled visit window Map to the closest scheduled visit							

Table 8 - Analysis window definition of safety variables

AVISIT	Target Day	Hematology, biochemistry		Vital Signs, C-SSRS, B-HCG Test		Weight		ECG		Urinalysis		Coagulation	
		Start Day	End Day	Start Day	End Day	Start Day	End Day	Start Day	End Day	Start Day	End Day	Start Day	End Day
Screening	-1	-40	-1	-40	-1	-40	-1	-40	-1	-40	-1	-40	-1
Day 1	1	1	1	1	1	1	1						
Month 1	28	2	42										
Month 2	56	43	70										
Month 3	84	71	98	2	126	2	126	1	126	1	210		
Month 4	112	99	126										
Month 5	140	127	154										
Month 6	168	155	210	127	210	127	252	127	210				
Month 9	252	211	294	211	294			211	294				
Month 12	336	295	378	295	378	253	420	295	504	211	420		
Month 15	420	379	462	379	462								
Month 18	504	463	546	463	546	421	588			421	588		
Month 21	588	547	630	547	630								
Month 24	672	631	714	631	714	589	756	505	840	589	756		
Month 27	756	715	798	715	798								
Month 30	840	799	882	799	882	757	924			757	924		
Month 33	924	883	966	883	966								
Month 36	1008	967	1050	967	1050	925	1050	841	1050	925	1050		
EOS		Start Day: last visit window upper limit +1; End Day: EOS visit date; Nominal visit (9000)											
Liver function tests		Nominal visit at Month 0.5, 1.25, 1.5, 1.75, 2.25, 2.5, 2.75, 7, 8, 10, 11.											
FU for participants who completed treatment to EOS but do not enter LTS		>= Date of EOS+1 Nominal visit (8010)											
pEOT		Start Day: date of discontinuation+1; End Day: start of the next scheduled visit window Map to the closest scheduled visit											

Table 9 - Analysis window definition of PD variables

AVISIT	Target Day	Plasma samples (NfL), serum samples (Chi3L1)		Serum samples (Ig levels)		Lymphocyte phenotypes by flow cytometry		Blood sample for archiving	
		Start Day	End Day	Start Day	End Day	Start Day	End Day	Start Day	End Day
Screening	-1							-40	-1
Day 1	1	1	1	1	1	1	1		
Month 1	28								
Month 2	56								
Month 3	84	2	126						
Month 4	112								
Month 5	140								
Month 6	168	127	252	2	252				
Month 9	252								
Month 12	336	253	420	253	420				
Month 15	420								
Month 18	504	421	672	421	672				
Month 21	588								
Month 24	672								
Month 27	756								
Month 30	840	673	924	673	924				
Month 33	924								
Month 36	1008	925	1050	925	1050				
EOS		Start Day: last visit window upper limit +1; End Day: EOS visit date Nominal visit (9000)							
pEOT		Start Day: date of discontinuation+1; End Day: start of the next scheduled visit window Map to the closest scheduled visit							

Lymphocyte testing will be done for selected participant at EOS or pEOT whichever comes first.

Unscheduled visits

Unscheduled visit measurements of laboratory data, vital signs, and ECG will be used for computation of baseline, the last on-treatment value, analysis according to PCSAs, and the shift summaries for safety. They will also be included in the by-visit summaries (central lab only for laboratory data) if they are re-allocated to scheduled visits based on the analysis windows defined above.

5.5 APPENDIX 5 SAS CODE FOR IMPUTATION OF BASELINE GD-ENHANCING T1 LESION STATUS

```
/*select subjects with missing baseline GD-enhancing T1 lesion*/
 data pt_list;
   set _adsl_5;
   if gdt1bl='';
   keep usubjid gdt1bl;
 run;

 /*get all Nb of Gad Enhanc T1 Hyperintense Lesions for those subjects*/
 data gd;
    merge pt_list(in=a)
              add.admr;
   by usubjid;
   if a;
   if paramcd='T1GDLNUM' and avisitn>0 and anl01fl='Y';
   keep usubjid paramcd param aval avisit avisitn anl01fl;
 run;

/*get ITT population and subjects with non-missing baseline GD-enhancing T1 lesion*/
 data itt;
   set _adsl_5;
   where ittfl='Y' and gdt1bl^='';
 run;

proc freq data=itt ;
  table gdt1bl /missing out=pct;
run;

/*get observed rate of presence subjects*/
 proc sql;
   select (percent)/100 into:pct
   from pct
   where gdt1bl='presence';
 quit;

proc sql;
  create table gd1 as
  select usubjid, sum(aval) as sum
  from gd
  group by usubjid;
quit;

data ptlist;
```

```
  set gd1;
  if sum=0;
run;

/* randomly assign value to patients with missing IGDT1BL based on observed rate*/
data _adsl_6;
  merge _adsl_5
          ptlist(in=b);
  by usubjid;
  if ittfl="Y" then do;
     call streaminit(16034);/*specify random seed as study id, for study EFC16033 random seed
will be
                                  16033*/
    if gdt1bl ne '' then IGDT1BL=gdt1bl;
    else if b then IGDT1BL = 'absence';
    else do;
       p=RAND("Bernoulli", &pct.);
       if p=0 then IGDT1BL = 'absence';
       if p=1 then IGDT1BL = 'presence';
    end;
  end;
 run;
```

6 REFERENCES

1. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. J Am Stat Assoc. 1989;84(408):1074-8.

2. Zhao Y, Herring AH, Zhou H, Ali MW, Koch GG. A multiple imputation method for sensitivity analyses of time-to-event data with possibly informative censoring. J Biopharm Stat. 2014;24(2):229-53.

3. Yuan YC. Multiple imputation for missing data: Concepts and new development. Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference; 2000.

4. Delis DC, Kramer JH, Kaplan E, Ober BA. (2000). CVLT-II, California verbal learning test: Adult version: manual.

5. Vickrey BG, Hays RD, Harooni R, Myers LW, Ellison, GW. A Health-Related Quality of Life Measure for Multiple Sclerosis. Quality Life Res. 1995;4(3):187-206.

6. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997;53(3):983-97.

7. Kaifeng L, Devan V. Mehrotra. Specification of covariance structure in longitudinal data analysis for randomized clinical trials. Statistics in Medicine. 2010;29:474-88.

8. Van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15(5):708-15.

9. Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, et al. Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. N Engl J Med. 2017;376(3):221-34.

10. O'Connor P, Wolinsky JS, Confavreux C, Comi G, Kappos L, Olsson TP, et al. Randomized trial of oral teriflunomide for relapsing multiple sclerosis. N Engl J Med. 2011;365(14):1293-303.

Signature Page for VV-CLIN-0628257 v4.0
efc16033-efc16034-16-1-9-sap

Approve & eSign	[redacted]
Approve & eSign	[redacted]