NOVARTIS

Clinical Development

OMB157/Ofatumumab®

COMB157G2301/COMB157G2302
NCT02792218/NCT02792231

A randomized, double-blind, double-dummy, parallel-group study comparing the efficacy and safety of ofatumumab versus teriflunomide in patients with relapsing multiple sclerosis

Statistical Analysis Plan (SAP)

Authors: [redacted]

Document type: SAP Documentation

Document status: Final (Addendum 1)

Release date: 17-Oct-2019

Number of pages: 86

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
15-Jul-2016	Prior to FPFV	Creation of final version	N/A - First version	NA
10-Jan-2019	Prior to DBL	Protocol amendments (1 and 2)	Amendment 1	Sections impacted by protocol amendments 1 or 2: - Section 1.2: Study objectives and endpoints - Section 2.2.1: Subgroup of interest - Section 2.5.3: Multiplicity adjustment - Section 2.6: Analysis of key secondary objectives - Section 2.7: Analysis of secondary objectives - Section 2.11: Patient reported outcomes [redacted] - Section 5.3.3: CTCAE grades definition - Section 3: Sample size calculation Section 2.8.6: Suicidality evaluation (to be consistent with latest MAP on eCSSRS data) Section 2.8.1: AEs (to add analyses to be used in legal requirements of ClinincalTrials.gov) Section 5.6: Rule of exclusion criteria of analysis sets (to add relevant PD that is updated in study PD document) Clarifications and minor corrections throughout the document

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
May 2019	Prior to DBL	Align with submission planning and update for corrections and clarity	Amendment 2	Section 2.1: General definition (Updated the CSR name to be consistent with name planned for submission) Section 2.4.1: Time at risk for AE/SAE (Applied the analysis data cutoff date) Section 2.4.2.3: Injection premedication (Updated to include medications either start or end on the injection date for completeness) Sections 2.6.2.1 and 2.6.2.2: Analysis for disability worsening and improvement (Updated model specifications for clarity and accuracy) Section 2.6.2.5: NfL (Added subgroup analysis on brian volume change to align with NfL pre-submission meeting briefing book submitted to FDA on Feb 8th 2019) Section 2.8.1: AEs (Added summary of AEs causing study drug interruptions as required by CSR) Section 2.8.5: Vital signs (minor correction to be consistent with Section 2.1.2.3) Section 2.8.7: Safety data after study drug discontinuation (Clarification that data within 100 days safety cutoff will be included in both “on-treatment” summary as well as in this section) 

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
July 2019	Prior to DBL	Update region definition used in statistical model based on blinded dry-run data and for corrections and clarity and align with TFL shells	Amendment 3	Section 2.1: General Definition (Region to be adjusted in the statistical model was updated based on statistical criteria as expected) Section 2.4.2.3: Injection related premedication (Updated the by-injection summary to replace injection 10+ with injection 10) Section 2.6.2.1: Analysis of key secondary endpoints 3mCDW and 6mCDW (Updated to present only month 18 and month 24 KM estimates) Section 2.6.2.2: Analysis of key secondary endpoint 6mCDW (same update as above) Section 2.6.2.5: Analysis of key secondary endpoint NfL (Updated the supportive analysis to include NfL baseline category as main effect which was missed in last version) Section 2.8.1.1.1: Injection reaction related AEs (Updated the by-injection summary to replace injection 10+ with injection 10; this will make trend in figure more interpretable) Section 5.5.2: Appendix on key secondary analysis (Removed the test related to proportional hazard assumptions but kept the visual check through log-log plots) Other minor corrections or clarifications made in the document;
Oct 2019	After DBL	Update to correct B-cell depletion analysis	Addendum 1	Section 2.8.3.4: Other special lab results [Updated to 1) correct the B-cell depletion definition from "<LLN or <baseline" to "<LLN"; 2) correct the data cutoff for B-cell depletion summaries from general safety cutoff to on-treatment period cutoff. Both updates are made in order to provide more accurate information on B-cell depletion from a scientific point of view: 1) the original depletion definition

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
				included "or <baseline" condition which can't represent a real depletion by itself but by chance that 50% of patients could have post-baseline values < Basline values; this specific definition of depletion therefore would artificially inflate the proportion of patients in comparator arm that have depletion and mask the differences in depletion between ofatumumab and comparator, leading to a potentially non-conservative estimate of the difference; the corrected definition "<LLN" captures the real depletion; 2) The original safety cutoff (i.e., last dose date +100) would mix repletion and depletion to some degree because fast B-cell repletion is expected after study drug discontinuation; the corrected cutoff using on-treatment period will concentrate on B-cell depletion as a result of study drug intake.]

Table of contents

List of abbreviations

AE	Adverse Event
ADA	Anti-drug-antibody
ALP	Alkaline Phosphate
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
ARBA	Annualized Rate of Brain Atrophy
ARR	Annualized Relapse Rate
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Classification
BIL	Bilirubin
BMI	Body Mass Index
BVL	Brain Volume Loss
BSSR	Blinded Sample Size Re-estimation
C-CASA	Columbia Classification Algorithm for Suicide Assessment
CDI	Confirmed Disability Improvement
CDW	Confirmed Disability Worsening
CNS	Central Nervous System
CRF	Case Report/Record Form (paper or electronic)
eCRF	Electronic Case Report Form
CRO	Contract Research Organization
CSPD	Clinical Summary Preparation Document
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
(e)CSSRS	(electronic) Columbia Suicide Severity Rating Scale
[redacted]	[redacted]
DMC	Data Monitoring Committee
DMT	Disease Modifying Treatment
EAST	Software for trial designs; name is derived from the benefit of “**Ea**rly **st**opping” a trial due to futility interim analysis
ECG	Electrocardiogram
(e)EDSS	(electronic) Expanded Disability Status Scale
EOS	End of Study
EOT	End of Treatment
[redacted]	[redacted]
FAS	Full Analysis Set
FDA	Food Drug Association
FS	Functional System/Functional score
FU	Follow up
Gd	Gadolinium
GdE	Gadolinium Enhancing
GGT	Gamma-Glutamyltransferase

9HPT	9-Hole Peg Test
Ig	Immunoglobulin
INR	International Normalized Ratio
ITT	Intention to treat
iv	Intravenous
IRR	Injection related reaction
IRT	Interactive Response Technology
KM	Kaplan-Meier
LDD	Last Dose Date
LFT	Liver Function Test
LLN	Lower Limit of Normal
M	Month
MedDRA	Medical Dictionary for Regulatory Activities
MH	Medical History
MRI	Magnetic Resonance Image
MS	Multiple Sclerosis
MSIS	Multiple Sclerosis Impact Scale
NB	Negative Binomial
NfL	Neurofilament light chain
NEDA	No Evidence of Disease Activity
od	once a day
PBVC	Percent Brain Volume Change
PD	Pharmacodynamics
PDS	Programming Datasets Specification
PG	Pharmacogenetics
PH	Proportional Hazard
PK	Pharmacokinetic
PML	Progressive Multifocal Leukoencephalopathy
po	oral(ly)
PPS	Per-Protocol Set
PRO	Patient Reported Outcome
PT	Preferred Term
q4, q12	Every 4, every 12
RMS	Relapsing MS
RRMS	Relapsing-Remitting MS
SAE	Serious Adverse Event
SAF	Safety Set
SAS	Statistical Analysis Software
sc	Subcutaneous
SDMT	Symbol Digit Modality Test
SPMS	Secondary progressive MS
SAP	Statistical Analysis Plan
SMQ	Standardized MedDRA Query

SOC	System Organ Class
TEAE	Treatment Emergent AE
TFLs	Tables, Figures, Listings
T25FW	Timed 25-foot Walk
TBIL	Total Bilirubin
ULN	Upper Limit of Normal
VAS	Visual Analog Scale
[redacted]	[redacted]
WHO	World Health Organization

1 Introduction

The purpose of this Statistical Analysis Plan (SAP) is to describe the implementation of the statistical analysis planned in the protocol for studies COMB157G2301 (G2301) and COMB157G2302 (G2302), which are of identical design conducted in parallel.

There will be three reports resulting from this SAP: (1) The clinical study report (CSR) of the G2301 study, (2) the CSR of the G2302 study, and (3) a combined report containing side-by-side presentation of demography and baseline characteristics between the two ASCLEPIOS studies, and key-secondary analyses on the pooled data from the two studies as defined in the clinical study protocols. Specifically the combined report will contain the analysis results concerning disability worsening and disability improvement from the combined data of the two studies. These three reports will be generated twice and referred to as 1) registration CSR and 2) final CSR. The registration CSR will be completed at the time of submission for new drug application. The final CSR will be completed after the final database lock.

This document is consistent with the current study protocols (version 02).

The analyses that will be performed for the blinded interim data reviews are out of the scope in this SAP and will be documented in a separate analysis plan. [redacted] These analyses will be documented in separate analysis plans or modeling reports and are out of scope for the present SAP.

1.1 Study design

Study G2301 is a randomized, double-blind, active comparator-controlled, parallel-group, multi-center study with variable treatment period in approximately 900 patients with relapsing multiple sclerosis (MS). The treatment duration for an individual patient will be variable based on when the End of Study (EOS) criteria are met. The maximal duration for an individual patient will be 30 months (2.5 years). Eligible patients will be randomized to receive either ofatumumab 20 mg sc injections once every 4 weeks (after initial loading regimen of three doses/week in the first 14 days) or teriflunomide 14 mg orally once daily via Interactive Response Technology (IRT). The randomization will be stratified by geographical region and by MS subtype (RRMS, SPMS).

A second study of identical design (G2302) will be conducted in parallel. Both studies have the same primary objective (reduction of annualized relapse rate (ARR)) and key secondary objectives, [redacted]. Both studies will be conducted globally, but centers/sites can only participate in one study to ensure independence between the two studies. Key secondary hypotheses with high sample size requirements, i.e. those related to disability worsening or disability improvement, will be tested on the basis of the combined data from the two studies (meta-analysis). Multiplicity adjustments are defined in Section 2.5.3. Poolability of the two studies will be assumed on the basis of the identical design, as well as the simultaneous and global conduct of the two studies. For information only, the heterogeneity of the treatment effect between the two studies will be tested for disability-related outcomes in the meta-analysis.

There is no unblinded efficacy interim analysis planned. Prior to the completion of enrolment, a blinded data review will be performed for the two studies to re-assess sample size assumptions for the ARR for each study separately and for 3-month confirmed disability worsening (3mCDW) for pooled studies. Based on this blinded review, the number of randomized patients to be enrolled may be increased to a maximum number of 1250 patients in each study.

The EOS will be reached for both studies when **all** of the following conditions are met simultaneously (in blinded data review):

1) Each study has collected sufficient data to provide 90% power for the primary ARR endpoint.
2) Across the two studies, sufficient 3mCDW events have been observed to provide 90% power for the key secondary 3mCDW endpoint.
3) Across the two studies, sufficient 6-month confirmed disability worsening (6mCDW) events have been observed to provide 80% power for the key secondary 6mCDW endpoint.

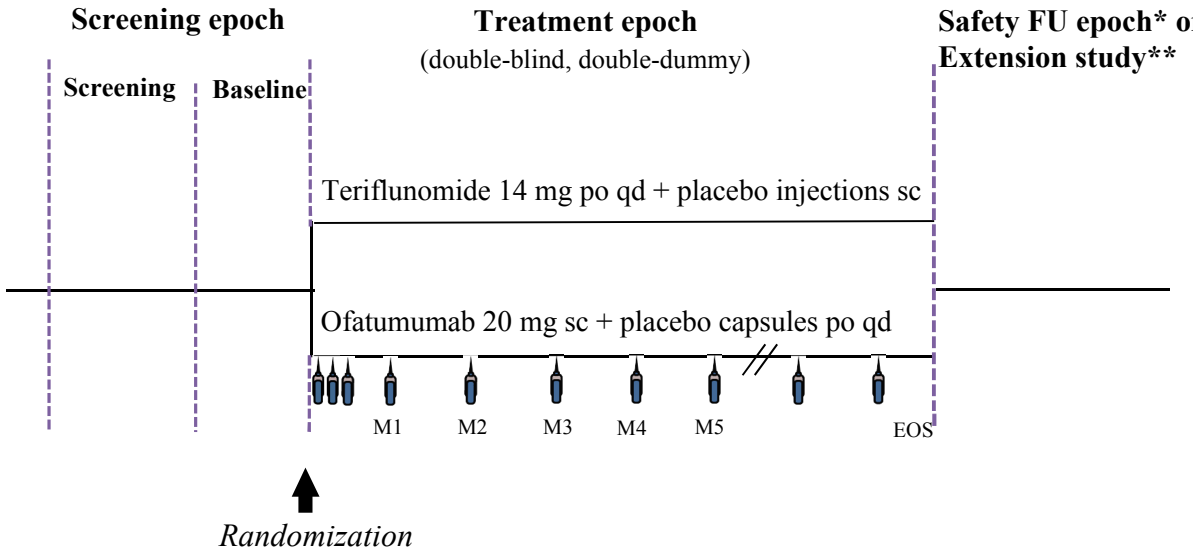
Details of the statistical analysis plan with respect to blinded data reviews are provided in the BSSR SAP.

The core study consists of three epochs: **Screening epoch** (including baseline), **Treatment epoch** (double-blind) and the post-treatment **Safety Follow up epoch** (Figure 1-1). Patients who complete the double-blind Treatment epoch (on study drug) may be eligible to enter an open-label ofatumumab Extension study that is planned (under a separate protocol).

All patients will have an EOS visit at the end of the Treatment epoch. Patients who prematurely discontinue double-blind study medication and agree to continue to follow the assessment schedule of the Treatment epoch (recommended) will have their end of treatment assessments (EOT) at the time of study medication discontinuation.

Patients who complete the Treatment epoch (on study drug), but do not enter the Extension study will be followed up for safety in the post-treatment Safety Follow-up (Safety FU) for a minimum of 9 months (or longer if indicated). The duration of the Safety FU epoch for an individual patient will depend on how long the patient has been off study drug at the time he/she enters the Safety FU epoch. For example, a patient who completed the Treatment epoch on study drug and had EOS (and will not enter the extension study), will be followed for at least 9 months in the Safety FU epoch. However, a patient who discontinued study drug earlier and had 4 months of follow up after study drug discontinuation in the Treatment epoch may only need to be followed for an additional 5 months in the Safety FU epoch (for a total of 9 months). Continued follow up beyond 9 months will be required for patients who have not repleted their B cells (to LLN or baseline values) or in whom teriflunomide plasma levels are still above 0.02 mg/L at 9 months, unless they have already started treatment with another MS disease modifying treatment (DMT).

Figure 1-1 Study design



*Patients who complete the Treatment Epoch and do not enter the planned Extension study or who prematurely discontinue study drug and do not agree to complete the study Treatment epoch or have less than 9 months of follow up after study drug discontinuation, will enter the Safety FU epoch. **Extension study will be conducted under separate protocol. M=month, EOS=End of Study, FU=Follow-Up, po qd= orally once a day

1.2 Study objectives and endpoints

1.2.1 Primary objective

Demonstrate that ofatumumab 20 mg sc once every 4 weeks (q4w) is superior to teriflunomide 14 mg po once daily in reducing the frequency of confirmed relapses as evaluated by the ARR in patients with relapsing MS.

1.2.2 Key secondary objectives

All disability related key secondary objectives will be addressed in the combined data (meta-analysis) from this study and the second study of identical design. All other objectives will be addressed based on the data from this study alone.

The key secondary objectives are to evaluate if ofatumumab 20 mg sc q4w is superior to teriflunomide 14 mg po once daily on the following efficacy measures:

1. Time to disability worsening as measured by 3-month confirmed worsening (3mCDW) on The Expanded Disability Status Scale (EDSS)

2. Time to disability worsening as measured by 6-month confirmed disability worsening (6mCDW) on EDSS

3. Time to disability improvement as measured by 6-month confirmed improvement (6mCDI) on EDSS

4. Number of T1 Gd-enhancing lesions per MRI scan

5. Number of new or enlarging T2 lesion on MRI per year (annualized T2 lesion rate)

6. Neurofilament light chain (NfL) concentration in serum

7. Rate of brain volume loss (BVL) based on assessments of percentage brain volume change from baseline

1.2.3 Other secondary objectives

Evaluate if ofatumumab 20 mg sc q4w is superior to teriflunomide 14 mg po once daily on the following efficacy measures:

- Time to first relapse
- Annualized relapse rates > 8 weeks after the onset of treatment
- Risk of a 3mCDW > 8 weeks after the onset of treatment
- Risk of a 6mCDW > 8 weeks after the onset of treatment
- Time to a 6-month confirmed cognitive decline (6mCCD), defined as a 4-point worsening on Symbol Digit Modalities Test (SDMT)
- Time to 6mCDW or 6mCCD, whichever is reached first
- Change in cognitive performance relative to baseline as measured by the SDMT
- Time to 6-month confirmed worsening of at least 20% in the timed 25-foot walk test (T25FW)
- Time to 6-month confirmed worsening of at least 20% in the 9-hole peg test (9HPT)
- Time to 6mCDI sustained until the EOS as measured by EDSS
- Number of new or enlarging T2 lesions between Month 12 and EOS
- Change in T2 lesion volume relative to baseline
- Proportion of patients with no evidence of disease activity (NEDA-4; defined in Section 2.7.1) at year 1 and 2
- Physical and psychological impact of MS disease as measured by the MSIS-29

In the subgroup of newly diagnosed, treatment-naïve patients, evaluate if:

- High NfL (above median) concentration at baseline is predictive of higher disease activity post baseline
- Patients with a high NfL (above median) concentration at baseline benefit from a stronger relative treatment effect of ofatumumab vs teriflunomide
- The safety profile of ofatumumab vs terifluomide is comparable in patients with high NfL (above median) concentration at baseline

Evaluate the safety and tolerability of ofatumumab 20 mg sc q4w compared to teriflunomide 14 mg po once daily

Evaluate the pharmacokinetics (PK) of ofatumumab

2 Statistical methods

2.1 Data analysis general information

Novartis statistical and programming team will be performing the CSR analysis as planned in this document unless otherwise specified. The Statistical Analysis System (SAS) 9.4 and/or R 15.2.1 or higher versions will be used.

Statistical safety analyses for the independent Data Monitoring Committee (DMC) will be conducted by a CRO. The process is described in the DMC charter.

Unless otherwise stated, summary tables/figures/listings will be on all patients in the respective analysis sets. Categorical data will be presented as frequencies and percentages. For continuous data, mean, standard deviations, median, 25th and 75th percentiles (optional), minimum and maximum will be presented.

The registration CSR analysis cut-off date will depend on the EOS date. All data up to this date will be included in the registration CSR. A complete analysis of the post-treatment safety follow-up will be provided in the final CSR when all patients have completed Safety FU epoch.

For efficacy analyses on the full-analysis set (FAS), all available data until the end of treatment epoch date (i.e., excluding the data collected in the safety follow-up epoch) will be considered and no other cut-offs will be applied.

For efficacy analyses on the per-protocol set (PPS), only on-treatment (as defined in Section 2.1.1) data will be considered, that is, for patients randomized to receive ofatumumab, data obtained 30 days after the last injection date will be excluded from the analyses; for patients

randomized to receive teriflunomide, data obtained after the last dose date of randomized study medication will be excluded from the analyses.

For safety analyses on the safety set, only data up to and including the safety cutoff of 100 days (5 x 20 days takes the long half-life of the comparator-drug into account) after last administration of study drug will be considered. Therefore, observations obtained more than 100 days after last administration of study drug will be excluded from the analyses. Nevertheless, all serious adverse events (SAE) and all deaths, regardless of the safety cut-off will be summarized.

Statistical models will include adjustments for regions by pooling centers to regions. The definition of region is intended to correspond to that used for the stratification of the randomization (as defined in Section 5.8). However, the definition of region may be modified if that is indicated based on statistical criteria (e.g., non-convergence). For statistical analysis where region is adjusted in the statistical models, regions "Asia Pacific" and "Latin America" will be combined with region "Other" due to small number of patients in these 2 regions.

Presentation of p-values: p-values from statistical tests will be presented with 3 decimal places, or as <.001 where applicable. Statistically significant p-values will be flagged with an asterisk. In general, this is when p-values ≤0.05. For key-secondary disability-related hypotheses, two asterisks will be used if p-value ≤0.04875 (=2*[0.025-0.025*0.025]) as these hypotheses will be tested at alpha minus alpha-squared (this condition will be footnoted on the relevant outputs).

All data (collected or derived) will be listed appropriately.

2.1.1 General definitions

Below summarize some general definitions to be used in the rest of the document.

Table 2-1 general definitions

Study treatment/Study drug	Both the investigational drug (ofatumumab) and the active comparator (teriflunomide) will be referred as study treatment or study drug.
Actual treatment	Actual treatment patients received. Ideally, actual treatment should be the same as randomized treatment. It could be different in case study drug is mis-dispensed. If patients received both investigational drug and active comparator accidentally during the study, their actual treatment will be the treatment to which they were exposed longer in duration. For safety analysis, patients will be analyzed according to the actual treatment received.
Date of first administration of study drug/first dose date	The first dose date of active study drug administration. For determining this date, matching placebo dummy treatment records will be excluded.
Date of last administration of study drug/last dose date	The last dose date of active study drug administration. For determining this date, matching placebo dummy treatment records will be excluded.

Study Day 1 or Day 1	The date of first administration of study drug/first dose date.
Study Day	All other study days will be labeled relative to Day 1. For events with dates on or after Day 1, study day for the event is calculated as (event start date – first dose date + 1). For events with dates before Day 1, study day for the event is calculated as (event start date – first dose date). Day 0 will not be used.
Duration of an event	Duration of an event is calculated as (event end date – event start date +1).
1 month	30 days; to be used in defining 3-month or 6-month confirmed disability worsening or improvement.
4 weeks	28 days; to be used in determining target days of scheduled visits. It is based on the scheduled injection frequency for ofatumumab (during the maintenance phase).
Day post-study drug discontinuation	Day post-study drug discontinuation for a particular event is calculated as (event start date – study drug discontinuation date).
Baseline	Baseline is the last assessment obtained prior to the first administration of study drug. No visit windows will be needed for the identification of baseline assessment. For pulse and blood pressure vital sign values, the baseline is the average of the non-missing values of the 3 measurements taken on the last visit prior to the first administration of study drug. The pre-injection assessment values on Day 1 vital sign CRF page will not be used for baseline derivations.
On-treatment period	For patients randomized to receive ofatumumab, on-treatment period includes days from the first injection date until 30 days after the last injection date; for patients randomized to receive teriflunomide, on-treatment period includes days from the first dose date until the last dose date. This definition considers patients on ofatumumab are scheduled to inject every 28 +/- 3 days. The on-treatment definition applies to efficacy analyses only. For calculation of compliance to study drug administration, similar definition of on-treatment period except based on actual treatment assignment will be used.
Safety cutoff (off-treatment)	Safety cutoff: Unless explicitly otherwise stated (e.g. SAEs and deaths), data up to and including the safety cutoff of 100 days after permanent study drug discontinuation will be

	included in the analysis and data beyond this time point for a given patient will be excluded from the safety analysis. The safety cutoff of 100 days (5x20 days) takes the long half-life of both investigational and comparator-drug into account. The safety cutoff applies to safety analyses only.
Nominal visits	Nominal visits are defined as all scheduled visits as per the clinical study protocol including the EOS and EOT visits. The definition of nominal visit excludes unscheduled visits. Only vital signs data collected on Day 1, Day 7, Day 14 and Month 1 protocol scheduled visit and ECG data collected on EOT and EOS visits will be summarized by nominal visit.
End of Study (EOS)	EOS, used in the context of individual patients, refers to EOS visit. EOS, used in the context of the entire study, refers to completion of treatment epoch for all patients.
End of Treatment (EOT)	EOT refers to EOT visit. Only patients who prematurely discontinue study drug but agree to continue to follow the schedule of assessments in the treatment epoch will have EOT visit.
End of treatment epoch date	This date is the date of discontinuation/study phase completion as recorded in the Study Phase Completion CRF page.
Last assessment on drug	It is the last assessment with non-missing value taken before or on the date of last administration of study drug. No visit windows will be needed for the identification of the last assessment on drug evaluation.

Table 2-2 Definition of time in key analyses

Time in study (ARR)	Time in study for ARR will be calculated as (end of treatment epoch date – first dose date+1)/365.25. The time in study by patient will be used as an offset variable to adjust for the various length patients have been observed and at risk of a confirmed MS relapse in the study.
Time from screening scan (MRI: key secondary analysis related)	**T2 lesions**: The time from screening scan will be calculated as (date of last scheduled MRI scan with a non-missing value for the number of new or enlarging T2 lesions during the treatment epoch – date of screening scan +1)/365.25. **Brain volume change**: Time of MRI assessment from screening scan will be calculated as (date of the scheduled MRI scan – date the screening scan +1)/365.25.

Time at risk for AE	**Time at risk for AE** is defined as the number of days spent in the study, from first to last administration of study drug, plus the safety data cut-off of 100 days. Time at risk for AE will be used for tables reporting AEs (including both SAE and non-SAE). If the corresponding last day of time at risk for AE is after the analysis cut-off date, then it will be truncated by the analysis cut-off date as defined in Section 2.1.
Time at risk for SAE	**Time at risk for SAE** is defined as the number of days spent in the study from the day of first administration of study drug to the end of study date (including safety follow up epoch). Time at risk for SAE will be used only for tables reporting SAEs. If a patient is still in safety follow up epoch at the time of reporting, time at risk for SAE is the number of days from the day of first administration of study drug to the analysis cut-off date as defined in Section 2.1.

2.1.2 Visit windows

2.1.2.1 Visit windows for treatment epoch

Visit-windows will be used for both efficacy and safety data summaries by visit. Visit windows define a time period "around" the targeted visit date as defined in the evaluation schedule of the clinical study protocol. Visit-windows are non-overlapping, and defined without gaps between consecutive visit windows. The width of visit windows may vary over the course of the study period.

Baseline assessments are defined in Section 2.1.1 and do not require a visit window.

The purpose of visit windows is to analyze data based on the actual study days (rather than "nominal" visits). E.g., if a patient's Month 1 visit is delayed; it is possible that the Month 1 data be re-aligned to visit-window Month 2 and be summarized under Month 2.

- For **efficacy analyses** (including PK concentration by visit analysis) all nominal visits (i.e. excluding unscheduled visits) will be mapped into one of the defined visit-windows. Note: for the derivation of disability worsening or improvement all visits (scheduled and unscheduled) need to be considered before the worsening or improvement can be confirmed (see Section 2.6.1.1). Similarly, for No Evidence of Disease Activity (NEDA-4), all visits (scheduled and unscheduled) need to be considered (Section 2.7.1).

-For **safety analyses** all visits (scheduled and unscheduled) will be mapped to visit windows. Safety data from unscheduled visits may be reported separately if applicable.

It is possible that more than one assessment of a patient fall into a particular visit-window. Section 2.1.2.3 deals with the statistical approaches to handle multiple visits in a given visit-window.

Tables displaying summary statistics "by visit" will also use the term *visit-window* as column header; this is to remind the reviewer that multiple assessments of a patient might be summarized. Below tables provide visit-windows definitions for applicable parameters.

Table 2-3 Visit-windows for EDSS*/T25FW

Visit-window	Start day	Target Day	End day
Week 12	1	84	126
Week 24	127	168	210
Week 36	211	252	294
Week 48	295	336	378
Week 60	379	420	462
Week 72	463	504	546
Week 84	547	588	630
Week 96	631	672	714
Week 108	715	756	798
Week 120	799	840	881

* For EDSS, assessment on Day 1 will be excluded as baseline EDSS can occur on Day 1 per protocol.

Table 2-4 Visit-windows for 9HPT/SDMT/MSIS-29/[redacted]

Visit-window	Start day	Target Day	End day
Week 24	1	168	252
Week 48	253	336	420
Week 72	421	504	588
Week 96	589	672	756
Week 120	757	840	923

Table 2-5 Visit-windows for MRI/[redacted]

Visit-window	Start day	Target Day	End day
Week 48	1	336	504
Week 96	505	672	839

Table 2-6 Visit-windows for routine laboratory values

Visit-window	Start day	Target Day	End day
Week 4	1	28	56
Week 12	57	84	126
Week 24	127	168	210
Week 36	211	252	294
Week 48	295	336	378
Week 60	379	420	462
Week 72	463	504	546

Week 84	547	588	630
Week 96	631	672	714
Week 108	715	756	798
Week 120	799	840	881

Table 2-7 Visit-windows for vital signs*

Visit-window	Start day	Target Day	End day
Week 12	1	84	126
Week 24	127	168	210
Week 36	211	252	294
Week 48	295	336	378
Week 60	379	420	462
Week 72	463	504	546
Week 84	547	588	630
Week 96	631	672	714
Week 108	715	756	798
Week 120	799	840	881

*Data collected from Day 1, Day 7, Day 14 and Month 1 protocol scheduled visit will not be mapped to the visit windows due to different data collection on those visits.

Table 2-8 Visit-windows for B-cell counts

Visit-window	Start day	Target Day	End day
Week 1	1	7	10
Week 2	11	14	21
Week 4	22	28	56
Week 12	57	84	126
Week 24	127	168	210
Week 36	211	252	294
Week 48	295	336	378
Week 60	379	420	462
Week 72	463	504	546
Week 84	547	588	630
Week 96	631	672	714
Week 108	715	756	798
Week 120	799	840	881

Table 2-9 Visit-windows for ADA data

Visit-window	Start day	Target Day	End day
Week 4	1	28	98
Week 24	99	168	252

Week 48	253	336	504
Week 96	505	672	839

Table 2-10 Visit-windows for biomarker data

Visit-window	Start day	Target Day	End day
Week 12	1	84	210
Week 48	211	336	504
Week 96	505	672	839

Table 2-11 Visit-windows for PK data

Visit-window	Start day	Target Day	End day
Week 4	1	28	56
Week 12	57	84	126
Week 24	127	168	252
Week 48	253	336	504
Week 96	505	672	839

2.1.2.2 Visit windows after study drug discontinuation

For summaries of data collected after study drug discontinuation, data from both treatment epoch and safety follow-up epoch will be considered. All reporting will be done based on visit windows defined relative to the last administration of study drug.

The visit window definitions are provided in Table 2-12 where the Start day and End day are relative to the date of last administration of study drug. For the "Last assessment on drug", the last assessment with non-missing value taken before or on the date of last administration of study drug will be summarized (no visit window applies). For the "Week 12 after LDD" visit-window, assessments taken at least 1 day after but no more than 126 days after the date of last administration of study drug will be considered. LDD stands for last dose date and will be footnoted in applicable outputs.

Table 2-12 Visit-windows after study drug discontinuation

Visit-window	Start day	Target Day	End day
Last assessment on drug	NA	NA	NA (see above or section 2.1.1)
Week 12 after LDD	2	84	126
Week 24 after LDD	127	168	210
Week 36 after LDD	211	252	294
Week 48 after LDD	295	336	378
Week 60 after LDD	379	420	462

Week 72 after LDD	463	504	546
Week 84 after LDD	547	588	630
Week 96 after LDD	631	672	714
Week 108 after LDD	715	756	798
Week 120 after LDD	799	840	881

2.1.2.3 Multiple assessments within visit windows

It is possible that multiple assessments of a patient fall into the same visit-window (e.g. due to unscheduled visits). All results (scheduled and unscheduled) will be displayed in listings, but only one value (observed or derived) will be selected for summary statistics by visit-window.

For **quantitative variables**, the assessment closest to the target day will be selected. If more than one assessment is at the same distance to the target day, the later one will be selected. For tables displaying the worst case scenario, such as shift tables or notable abnormalities, all assessments within a visit window will be used to identify the worst (e.g. the maximum or the minimum depending on parameter). Where applicable it will be defined for each parameter what the worst case is.

For **qualitative variables**, the worst record is selected; it is noted that in the relevant data subsection, worst case is always well defined.

2.2 Analysis sets

All screened patients (SCR): The SCR set comprises all patients who were screened.

Full analysis set (FAS): The FAS comprises all randomized patients with assigned treatments. Patients will be analyzed according to the randomized treatment assignment following the intention-to-treat (ITT) principle, even if they actually received no or a different treatment.

The FAS will be used for the summary of demography and baseline characteristics as well as for all efficacy analyses.

Per-protocol set (PPS): The PPS is a subset of FAS, consists of all randomized patients who take at least one dose of study medication and have no major protocol deviations that could confound the interpretation of analyses conducted on the FAS. Major protocol deviations will be determined according to the pre-defined protocol deviation criteria before treatment unblinding (e.g. non-compliance for a large proportion of the time in study). For analyses performed on the PPS, only efficacy data assessed during the on treatment period (as defined in Section 2.1.1) will be included.

The PPS will be used for the supportive analyses of the primary efficacy variable and selected key secondary variables.

Safety set (SAF): The SAF set includes all patients who received at least one dose of study medication. Patients will be analyzed according to the actual treatment received. If patients received both investigational drug and active comparator accidentally during the study, they

will be analyzed according to the treatment to which they were exposed longer in duration. The SAF will be used for all safety analyses.

2.2.1 Subgroup of interest

Only protocol specified subgroup analyses will be covered in the CSR. Subgroup analyses on the pooled study data for submission purpose are out of the scope of this SAP but will be specified in the relevant submission plan document (e.g., Clinical Summary Preparation Document (CSPD)). As subgroup analyses do not control for either type I or type II error, they should be considered as hypothesis generating in nature with the purpose of examining potential inconsistencies of a treatment among the many subgroups being examined. All confidence intervals and p-values for subgroup analysis will be presented without multiplicity adjustments.

Subgroups for efficacy analyses:

- NfL level (high:>median, low:<=median), where the median is defined based on all patients in the FAS of the combined data from both ASCLEPIOS studies.
- NfL levels (<Q1; >=Q1 but <Q2; >=Q2 but < Q3; >=Q3; where Q1, Q2 and Q3 are first, second and third quartiles of baseline neurofilament levels in all patients from studies G2301 and G2302 combined)
- Vitamin D levels (<Q1; >=Q1 but <Q2; >=Q2 but < Q3; >=Q3; where Q1, Q2 and Q3 are first, second and third quartiles of baseline vitamin D levels in all patients from studies G2301 and G2302 combined)
- Newly diagnosed (within 3 years prior to the screening visit) and treatment-naïve (no prior MS DMT) patients

These subgroups will be used for selected exploratory analyses as described in Section 2.12.

Subgroups for safety analyses:

- Age at baseline (<18, 18-30, 31-40, 41-55, >55)
- Gender (male, female)
- Gender x Age: 10 subgroups defined by age and gender: female & age <18; female & age of 18-30; female & age of 31-40; female & age of 41-55; female & age >55; male & <18; male & age of 18-30; male & age of 31-40; male & age of 41-55; male & age >55;
- Gender x body weight (by quartile): 8 subgroups defined by gender and baseline body weight: female & weight <Q1; female & weight >=Q1 but <Q2; female & weight >=Q2 but < Q3; female & weight >=Q3; male & weight <Q1; male & weight >=Q1 but <Q2; male & weight >=Q2 but < Q3; male & weight >=Q3; where Q1, Q2 and Q3 are first, second and third quartiles of baseline weight in all patients from studies G2301 and G2302 combined
- Race (White, Asian, Black or African Americian, Other)
    - White: those who selected “Caucasian” as race on demography CRF
    - Asian: those who selected “Asian” as race on demography CRF
    - Black or African American: those who selected “Black” as race on demography CRF

- Other: those who selected “Native American” or “Pacific Islander” or “Unknown” or “Other” as race on demography CRF

- *Patients who accidentally received both investigational drug and active comparator during the study
- **Patients in SAF who had data reported after date of last administration of study drug
- **Patients in SAF who had at least one eCSSRS performed after the date of last administration of study drug
- **Patients in SAF who had B-cell count data after the date of last administration of study drug
- **Patients in SAF who had teriflunomide data

*This subgroup will be used for exploratory analysis described in Section 2.8.1. **These subgroups will be used for corresponding safety follow up data analysis described in Section 2.8.7. Rest of the subgroups will be used for patient-year summaries described in Section 2.4.1.

2.3 Patient disposition, demographics and other baseline characteristics

2.3.1 Patient disposition

The number and percentage of patients who were screened but did not continue into the treatment period will be presented, along with the reason for discontinuation. Data collected on the screening phase disposition CRF page will be used to summarize this information. The summary will be on the SCR set.

The number and percentage of patients who completed the study on treatment or prematurely discontinued study drug will be presented, along with the primary reason for discontinuation. Data collected on the end of study treatment CRF page will be used to summarize this information. The summary will be on the FAS.

The number and percentage of patients who completed the study (i.e., treatment epoch) or prematurely discontinued study prior to the end of the treatment epoch will be presented, along with the primary reason for discontinuation. Data collected on the study phase completion CRF page will be used to summarize this information. The summary will be on the FAS.

Time to study drug discontinuations and time to study discontinuations by treatment group will be presented using Kaplan-Meier estimates in survival curves and in summary tables, along with the p-value from the log-rank test. Time to study drug discontinuation in (days) will be derived as [last dose date – first dose date + 1]. Time to study discontinuation (in days) will be derived as [date of discontinuation/study phase completion” for the treatment epoch – first dose date + 1]. The event flag for analysis of time to study drug discontinuations is set to 1 for patients who answered “No” to question “did the subject complete study treatment” as recorded on End of Study Treatment CRF page or 0 otherwise. The event flag for analysis of time to study discontinuations is set to 1 for patients whose “subject status” is not “completed” as recorded on Study Phase Completion CRF page or 0 otherwise. These analyses will be on the FAS.

The number and percentage of patients in each analysis set will be presented by treatment group. Protocol deviations will be summarized by deviation categories and treatment for the FAS. In addition, protocol deviations that led to exclusion from the analysis set will be summarized by deviation category, deviation terms and treatment groups for the FAS.

Patients exclusion from PPS will be listed for all patients with reasons for exclusion (i.e. including both protocol and non-protocol deviations).

2.3.2 Background and demographic characteristics

Background characteristics include recipient demographic characteristics (gender, race and ethnicity collected on the Demography CRF), age, height, body weight, BMI and employment status at baseline.

Age will be calculated from date of first administration of study drug and date of birth. Derived baseline height, body weight and BMI will be presented. These variables will be summarized by treatment group for the FAS using frequency distributions (for categorical variables) and descriptive statistics of mean, standard deviation, minimum, median and maximum (for continuous variables).

Demography will be presented by study but also side-by-side for the two studies in the combined report to evaluate poolability of the two ASCLEPIOS trials.

2.3.3 MS baseline disease characteristics

MS baseline characteristics, MS disease history and MS medication history will be summarized by treatment group for the FAS in the individual study reports but also side-by-side for the two studies in the combined report to evaluate poolability of the two ASCLEPIOS trials.

MS baseline characteristics include baseline EDSS, 9HPT, T25FW, SDMT, and key MRI parameters (e.g., number of Gd-enhancing T1 lesions, T2 lesion volume, and normalized brain volume).

MS disease history includes duration of MS since diagnosis (years), duration of MS since first symptom (years), number of relapses in the last 12 months prior to screening, number of relapses in the 12 to 24 months prior to screening, type of MS at study entry (i.e., RRMS or SPMS), time since onset of SPMS, and time since onset of most recent relapse (months) prior to screening.

Duration of MS since diagnosis (years) will be derived [(first dose date – MS diagnosis start date + 1)/365.25]; duration of MS since first symptom (years) will be derived as [(first dose date – first MS symptom date +1)/365.25]; time since onset of SPMS (years) will be derived as [(first dose date – conversion to SPMS date +1)/365.25]; and time since onset of most recent relapse (months) will be derived as [(first dose date – most recent relapse onset date + 1)/(365.25/12)]. In these calculations, partial dates if any will be imputed according to the rules specified in Section 5.1.3.3.

MS medication history of previous disease-modifying drugs (coded by WHO drug dictionary) will be summarized by preferred term (PT) and treatment group. The number and proportion of treatment-naïve patients (i.e., patients who have not been treated with any disease- modifying drug before study enrolment) will also be presented.

2.3.4 Medical history

Medical history will be summarized by treatment group for the FAS. Any condition entered on the Medical History (MH) CRF will be coded using the MedDRA dictionary. The medical history will be summarized by primary system organ class (SOC), preferred term (PT) and treatment group.

The MH conditions captured on the eCRF "Protocol solicited medical history or medical history possibly contributing to liver dysfunction" will be tabulated in the regular MH table for level 2 drugs (because the solicited events are captured for all patients) and shown and flagged in the regular MH listing. There is no separate deliverable required.

2.3.5 Smoking history and alcohol history

Data collected on smoking history and alcohol history CRF pages will be listed as appropriate.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

2.4.1 Exposure to study treatment / compliance / time at risk

Duration of exposure to study drug will be derived as follows:

- For patients whose actual treatment (Section 2.2 Safety set) is teriflunomide, duration of exposure will be calculated as (last dose date – first dose date +1- Σ [number of days with temporary study drug interruption]), which is the number of days between the first and the last day of study drug administration, excluding the number of days with temporary study drug interruption (as patients are scheduled to take the oral study medication once daily).
- For patients whose actual treatment is ofatumumab, duration of exposure will be calculated as (last injection date – first injection date + 31 – Σ [$(j+1)^{th}$ injection date – j^{th} injection date -31]), where j and j+1 refer to consecutive injections with injection dates more than 31 days apart (as patients are scheduled to take the subcutaneous injections every 28 +/- 3 days).

Duration of exposure to study drug will be summarized descriptively on SAF set by treatment group and duration category (i.e., ≥ 1 week, ≥ 2 weeks, ≥ 3 weeks, ≥ 4 weeks, ≥ 8 weeks, ≥ 12 weeks, ≥ 24 weeks, ≥ 36 weeks, ≥48 weeks, ≥ 60 weeks, ≥ 72 weeks, ≥ 84 weeks, ≥ 96 weeks, ≥ 108 weeks, ≥ 120 weeks, ≥ 132 weeks, ≥ 144 weeks, ≥ 156 weeks, ≥ 168 weeks). Descriptive statistics of duration in days will also be provided by treatment group.

For each treatment group, the number of patient-years is calculated as (the sum of the number of days of exposure for all patients in the group)/365.25 and will be summarized by age and gender, by gender and weight, as well as by race. Cutoffs for age subgroups and weight subgroups are defined in Section 2.2.1.

Time at risk is the censoring time used for estimating exposure adjusted incidence rate of SAEs or AEs in patients who did not experience the event of interest. For patients who have experienced the event of interest, the actual date of the SAE, or AE onset will be used. Details

about the calculation of exposure adjusted incidence rate of adverse events are described in Section 5.7.2. Time at risk also corresponds to the time period used for adverse event reporting.

- **Time at risk for AE** is defined as the number of days spent in the study, from first to last administration of study drug, plus the safety data cut-off of 100 days. Time at risk for AE will be used for tables reporting AEs (including both SAE and non-SAE). If the corresponding last day of time at risk for AE is after the analysis cut-off date, then it will be truncated by the analysis cut-off date as defined in Section 2.1.
- **Time at risk for SAE** is defined as the number of days spent in the study from the day of first administration of study drug to the end of study date (including safety follow up epoch). Time at risk for SAE will be used only for tables reporting SAEs. If a patient is still in safety follow up epoch at the time of reporting, time at risk for SAE is the number of days from the day of first administration of study drug to the analysis cut-off date as defined in Section 2.1.

Time at risk for AE and time at risk for SAE will be summarized in a similar way to duration of exposure to study drug.

Compliance to the study drug administration schedule will be calculated as duration of exposure to study drug in (days)/duration of on-treatment period (as defined in Section 2.1.1) in (days) ×100%. This rule means that compliance will be measured during the time interval the patient took study medication: premature discontinuation from study drug will not be considered non-compliance. Compliance to study drug administration will be summarized descriptively on SAF by treatment group. In addition, compliance will be summarized with cumulative number and percentage of patients in each compliance category (*i.e., ≥ 20%, ≥ 30%, ≥ 40%, ≥50%, ≥60%, ≥70%, ≥80%, ≥90%, ≥95%, ≥98%, =100%*).

2.4.2 Prior, concomitant and post therapies

Analyses described in this section will be performed on the SAF.

2.4.2.1 Concomitant medication

Records on the Prior and Concomitant Medications CRF page will be coded using the WHO drug dictionary. All medications will be classified as prior, concomitant or post study drug discontinuation medication as follows:

- Prior medications are defined as drugs taken and stopped prior to first dose of study medication.
- Concomitant medications are defined as drugs taken at least once between first dose and last dose of study medication (including those which were started prior to first dose and continued into the treatment period).
- Post-study drug discontinuation medications will be drugs started after the discontinuation of randomized study medication.

Medications will be categorized into one (and only one) of above classes based on recorded or imputed start and end dates. When incomplete or missing, dates will be imputed according to Novartis standards (details will be given in programming datasets specifications (PDS) document). If both start date and end date are completely missing and medication was not

collected on the “Previous MS Disease Modifying Treatment” page, medication will be classified into concomitant medication category.

Medications in each of these 3 categories will be summarized separately by treatment group, ATC code and preferred term. ATC level 1 and level 3 (e.g., M [Musculo-skeletal system], M01A [anti-inflammatory and anti-rheumatic products, non steroids], etc.) will be used.

Data collected from the Previous MS disease modifying treatment pages or as "Injection related premedication" in the concomitant medication pages will not be included in this summary.

2.4.2.2 Surgical and medical procedures

Records on the surgical and medical procedures CRF page will be coded using the MedDRA dictionary. All procedures will be classified as prior, concomitant or post-study drug discontinuation procedure, in the same way as done for concomitant medications. Surgical and medical procedures in each of these 3 categories will be summarized separately by system organ class, preferred term and treatment group.

Imputation rules for start and end dates will follow the same rule as for the concomitant medications.

2.4.2.3 Injection related premedication

Injection related premedication will be identified by subcategory “Injection related premedication” in concomitant medication data set. Injection related premedication will also be summarized separately for each injection up to injection 10 and cumulatively for all injections.

For injection 1 summary, the injection related premedication with either start date or end date on the same day as the first injection date will be included and summarized for each of the three protocol specified types and for each combination of the specified types (type 1+ type 2, type 1+ type 3, type 2+ type 3, type 1+ type 2+ type 3). The three protocol specified types are steroids (type 1), antihistamines (type 2) , and antipyretics/analgesics (type 3) (corresponding to “Acetaminophen” as specified in the table shell for this summary). The steroids (type 1) will be identified by category “Steroid”. The antihistamines (type 2) will be identified by ATC level 3 “antihistamines for systemic use”. The antipyretics/analgesics (type 3) will be identified by ATC level 3 “other analgesics and antipyretics” and “anti-inflammatory and anti-rheumatic products, non-steroids”. In summaries for each combination of the specified types, the number and proportion of patients who took both (or all 3) types of injection related premedication at specified injection will be provided.
Rest injection specific summaries or cumulatively summaries will be reported similarly.

2.4.2.4 Previous MS disease modifying treatment

Data collected from the Previous MS disease modifying treatment pages will be summarized separately.

2.5 Analysis of the primary objective

2.5.1 Primary endpoint

The primary endpoint is the annualized relapse rate (ARR), which is defined as the number of confirmed MS relapses in a year. In the primary analysis, the ARR is estimated based on the FAS which follows the intent-to-treat principle in a negative binomial (NB) model by using individual relapse count as the response variable with natural log of time in study in years as an offset variable.

Two variables are required for the calculation of the ARR (excluding covariates):

- The cumulative number of confirmed MS relapses by patient is the response variable in the negative binomial model. All confirmed relapses with a start date on or after the date of first administration of study drug and prior to or on the end of treatment epoch date will be included in the analysis. Additional details are provided in Section 5.1.3.4.
    - The definition of a confirmed MS relapse is one accompanied by a clinically relevant change in the EDSS assessment, i.e. an increase of at least 0.5 points on the EDSS (total) score, or an increase of at least 1 point on at least two Functional scores (FSs), or an increase of at least 2 points on at least one FS, excluding changes involving bowel/bladder or cerebral FS, compared to the last EDSS assessment taken in the absence of (confirmed or unconfirmed) relapse and prior to the current relapse. EDSS obtained on the date as indicated on the Summary of MS Relapse eCRF page will be used. If such EDSS assessment is missing or not meeting the criteria to confirm the relapse, all other EDSS assessments taken within 30 days from the relapse start date (i.e., EDSS assessment date – relapse start date <=30) and before the relapse end date (EDSS assessment date < relapse end date) will be checked. If at least one of such available EDSS assessments meets the criteria, the relapse is a confirmed relapse. Otherwise, the relapse is considered an unconfirmed relapse.
- The time in study by patient will be used as an offset variable to adjust for the various length patients have been observed and at-risk of a confirmed MS relapse in the study. Time in study for ARR will be calculated as (end of treatment epoch date – first dose date+1)/365.25.

2.5.2 Statistical hypothesis, model, and method of analysis

The null hypothesis is that there is no difference in the ARR between ofatumumab 20mg sc once monthly and teriflunomide 14 mg po once daily in reducing the frequency of confirmed MS relapses as measured by ARR. The alternative hypothesis is that there is a difference between the two treatment groups.

Superiority of ofatumumab 20mg sc over teriflunomide 14 mg po will be concluded if the observed ARR on ofatumumab 20mg sc is lower than on teriflunomide 14 mg po and if the null hypothesis can be rejected at the two-sided significance level of 0.05.

The null hypothesis will be tested based on the FAS, using a negative binomial regression model with log-link, treatment and region as factors, number of relapses in previous year, baseline

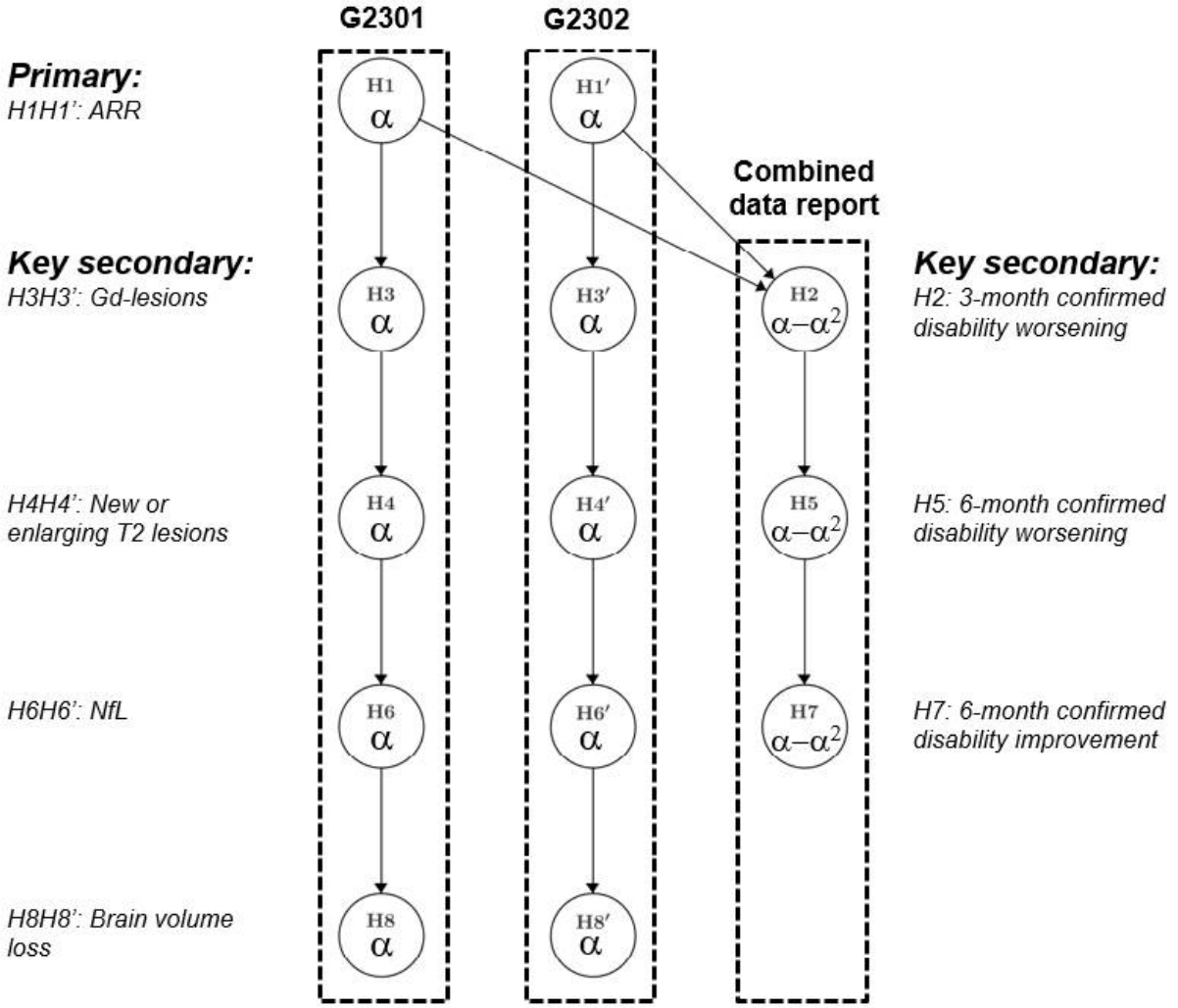
EDSS, baseline number of T1 Gd-enhancing lesions and the patient's age at baseline as covariates. In the analysis, the response variable is the number of confirmed relapses observed from each patient and the patient's time in study (natural log of time in years) is used as an offset variable to adjust for the varying lengths of patient's time in the study. The adjusted ARR (i.e., model-based estimate adjusted for covariates) for each treatment and the corresponding 95% confidence interval, and ARR ratio (also expressed as percentage reduction relative to control group) along with the 95% confidence interval for the ARR ratio and the corresponding p-value will be obtained.

The definition of Region is intended to correspond to that used for the stratification of the randomization. However, the definition of Region may be modified if that is indicated based on statistical criteria (e.g. non-convergence of models).

2.5.3 Multiplicity adjustment

The planned submission consists of two studies of identical design (G2301 and G2302), each with multiple endpoints. In order to control the type-I error rate ("false positive rate") at the level of the individual studies, and at the level of the submission as a whole, the testing strategy illustrated in Figure 2-1 will be implemented.

Figure 2-1 Multiple testing procedure



Testing procedure and type-I-error control in the planned ofatumumab submission which consists of studies G2301 and G2302 (both with identical design). Hypotheses can only be tested in sequential order as indicated by the arrows. The number associated with each hypothesis (α, or α-α^2) indicates the significance level at which that hypothesis can be tested. If the null-hypothesis for the primary objective (ARR) can be rejected within a study, MRI- and NfL-related hypotheses will be tested in sequential order within that study as long as all proceeding hypotheses can successfully be rejected. Disability-related hypotheses will only be tested in the combined data of the two studies, if the primary null-hypotheses can be rejected in both studies first. At the study-level, the type-I error rate (one-sided) is controlled at ≤0.025. In the submission, the type-I error rate is controlled at ≤0.000625 (=0.025^2) for the primary hypothesis and at ≤0.025 when considering all endpoints.

The primary hypothesis (ARR) and all MRI- and NfL-related key-secondary hypotheses will be tested in hierarchical order within study (Figure 2-1). The testing procedure starts with the statistical test of the primary null-hypothesis (ARR) and continues to lower ranking hypotheses as long as the proceeding null hypotheses can all be rejected in favor of ofatumumab in a two-

sided statistical test with a p-value ≤0.05. This testing procedure controls the type-I error rate to ≤0.05 within study.

If both studies independently reject the primary null-hypothesis (ARR) in favor of ofatumumab in a two-sided statistical test with p-value ≤0.05, disability-endpoints will be addressed in the combined data of G2301 and G2302, at the submission level. Disability endpoints will be tested in hierarchical order as indicated by arrows in (Figure 2-1). The testing procedure continues to the next lower ranking disability-hypothesis as long as the previous null-hypothesis can be rejected in favor of ofatumumab in a two-sided statistical test with a p-value ≤0.04875 (=2*[0.025-0.025^2]).

Provided the primary hypothesis can be rejected in both studies, disability-related endpoints can be tested regardless of the outcome of MRI- and NfL-related endpoints, and vice-versa.

Under the global null-hypothesis (i.e. no difference between ofatumumab and teriflunomide), the testing procedure controls the type-I error rate (one-sided) at the study-level to ≤0.025, and at the submission level to ≤0.000625(=0.025^2). Considering all possible configurations of true and false positive null hypotheses, the type-I error control at the level of the submission is ≤0.000625 for the primary objective, and ≤0.025 for all hypotheses.

The type-I error is controlled by the testing procedure. All confidence intervals and p-values in the study report will be presented without adjustments.

2.5.4 Handling of missing values/censoring/discontinuations

The primary NB model with an offset for the time in study adjusts for missing information (drop-out) under the assumption of non-informative drop-out, information is missing at random, and constant relapse rate over time. According to the protocol, patients who discontinue study treatment should remain in the study and follow the assessment schedule. The primary analysis will use all available data up to the end of treatment epoch date, irrespective of on or off study treatment. In addition, a sensitivity analysis will be conducted to allow for the possibility that relapse rates may be non-constant over time (i.e. higher during the onset-of-action of both drugs for a period of 8 weeks).

2.5.5 Supportive analyses

The primary analysis will be repeated to analyze all reported MS relapses (confirmed or unconfirmed).

The primary analysis will also be repeated using the per-protocol set to provide an analysis of on-treatment data from patients who have no major protocol violations (refer to Section 5.6 for details of how per-protocol set flag is derived for each patient). Only relapses with a start date during the on-treatment period (as defined in Section 2.1.1) will be included and natural log(time on study drug in years) rather than natural log(time in study in years) will be used as the offset variable in the negative binomial model.

To estimate relapse rates and the treatment effect between ofatumumab 20mg sc once monthly and teriflunomide 14 mg po once daily during the initial “onset of action” period of 8 week (≤56 days=8*7 days), and relapse rates and the treatment effect thereafter (>56 days; long-term efficacy) a sensitivity analysis will be conducted on the FAS. This analysis will be implemented

as a piecewise negative binomial model assuming different event rates and ARR-ratios before and after week 8, but constant dispersion and covariates adjustment (as in the primary analysis model). Details of the statistical model and implementation of the model are provided in Section 5.5.3.1.

For each time period (≤56 days, and >56 days), the adjusted ARR (i.e., model-based estimate adjusted for covariates) and the corresponding 95% confidence intervals will be provided by treatment, together an ARR ratio between ofatumumab 20mg sc once monthly and teriflunomide 14 mg po once daily (also expressed as percentage reduction relative to control group) with corresponding 95% confidence intervals and p-values.

Additionally, the time-to-first relapse will be analyzed in a Cox proportional hazards model on the FAS. In comparison with the primary analysis in a negative binomial model, the Cox proportional hazards model does not assume constant relapse rates (but rather it assumes proportional hazards). The Cox proportional hazards model will be specified with treatment, region, number of relapses in previous year, baseline EDSS, baseline number of T1 Gd-enhancing lesions and the patient's age at baseline as covariates. For patients with at least one relapse, the time to first relapse will be calculated as (start date of first relapse - date of first administration of study drug+1). For patients without relapses, they are censored with censored time as time in study (as defined in Section 2.5.1).

2.6 Analysis of the key secondary objectives

2.6.1 Key secondary endpoints

2.6.1.1 Disability worsening (3-month or 6-month confirmed)

A 3-month confirmed disability worsening (3mCDW) is defined as an increase from baseline in EDSS sustained for at least 3 months. Analogously, a 6-month confirmed disability worsening (6mCDW) is defined as an increase from baseline in EDSS sustained for at least 6 months. This means that after a scheduled or unscheduled visit at which the patient fulfills the disability worsening criterion as defined in Table 2-13, all EDSS assessments (scheduled or unscheduled) need to also fulfill the worsening criteria until the worsening ("the event") can be confirmed at the first scheduled visit that occurs in the absence of (confirmed or unconfirmed) relapse 3 months/90 days (or 6 months/166 days) after the onset of the worsening, or later.

In the time to event (3mCDW or 6mCDW) analysis, time will be calculated as (the date of EDSS assessment at onset of the event – date of first administration of study drug+1) for patients with the events. Censoring occurs in those patients who did not experience an event in the study (during the treatment epoch), this includes patients who had a "tentative" disability worsening that could not be confirmed due to an early discontinuation or any another reason. The censoring time is calculated as (the date of last EDSS assessment during the treatment epoch – date of first administration of study drug+1). Additional details are provided in footnote of Table 2-13.

Table 2-13 Criterion for disability worsening based on change in EDSS score

Total EDSS at baseline*	"Disability worsening" criterion
0	≥ +1.5

0.5 to 5	≥ +1
≥5.5	≥ +0.5

A 3-month confirmed disability worsening can have an onset at any scheduled or unscheduled visit if the disability worsening criterion is met. A disability worsening can only be confirmed at a scheduled visit in the absence of (confirmed or unconfirmed) relapse if, over a period of 3 months (≥90 days=3*30) time interval, all assessments meet the worsening criterion.

A 6-month confirmed disability worsening can have an onset at any scheduled or unscheduled visit if the disability worsening criterion is met. A disability worsening event can only be confirmed at a scheduled visit in the absence of (confirmed or unconfirmed) relapse if, over a period of 6 months (≥166 days=6*30-14 [visit window]) time interval, all assessments meet the worsening criterion.

If a patients dies due to MS (EDSS=10 at any time), it will be considered a confirmed disability worsening regardless of the baseline EDSS or the change in EDSS. The time will be calculated as (date of EDSS assessment at a tentative onset of the event – date of first administration of study drug+1) or (date of death – date of first administration of study drug + 1) if a tentative onset date does not exist. Note: Death for other reasons than MS (i.e. not EDSS=10) will not be considered a disability worsening.

*Baseline EDSS is defined as the last EDSS assessment prior to the first dose of study medication (protocol inclusion criterion is EDSS 0-5.5)

2.6.1.2 Disability improvement (6-month confirmed)

A 6-month confirmed disability improvement (6mCDI) is defined as a decrease from baseline EDSS sustained for at least 6 months. This means that after a scheduled or unscheduled visit at which the patient fulfills the disability improvement criterion as defined in Table 2-14, all EDSS assessments (scheduled or unscheduled) need to also fulfill the improvement criteria until the improvement ("the event") can be confirmed at the first scheduled visit that occurs 6 months/166 days after the onset of the improvement, or later. In the time to event (6mCDI) analysis, time will be calculated as (the date of EDSS assessment at onset of the event – date of first administration of study drug+1) for patients with the events. Censoring occurs in those patients who did not experience a 6mCDI event in the study (during the treatment epoch), this includes patients who had a "tentative" disability improvement that could not be confirmed due to an early discontinuation or any another reason. The censoring time is calculated as (the date of last EDSS assessment during the treatment epoch – date of first administration of study drug+1).

Table 2-14 Criterion for disability improvement based on change in EDSS score

Total EDSS at baseline*	"Disability improvement" criterion
0 to 1.5	No improvement possible
≥2 to 6	≤ -1
≥6.5 to 9.5**	≤ -0.5

6mCDI: A disability improvement can have an onset at any scheduled or unscheduled visit if the disability improvement criterion is met. A disability improvement can only be confirmed at a scheduled visit if, over a period of 6 months (≥166 days=6*30-14) time interval, all assessments meet the improvement criterion. A 6mCDI sustained until the end of study is defined as a 6mCDI after which all EDSS assessments meet the disability improvement criterion through End of Study.

*Baseline EDSS is defined as the last EDSS assessment prior to the first dose of study medication

**protocol inclusion criterion is EDSS 0-5.5

2.6.1.3 Number of Gd-enhancing lesions per scan

Number of Gd-enhancing lesions will be obtained from each MRI scan per protocol assessment schedule. To estimate the number of Gd-enhancing lesions per scan, below variables will be derived:

- The total number of Gd-enhancing lesions during the treatment epoch will be derived by taking the sum of numbers of Gd-enhancing lesions from all scheduled MRI scans during the treatment epoch. MRI scans taken within 30 days after the termination of steroid therapy will not be included in analysis of Gd lesion related endpoint.
- The number of MRI scans will be derived by counting the number of scheduled MRI scans with non-missing values for the number of Gd-enhancing lesions during the treatment epoch.

2.6.1.4 Annualized rate of new or enlarging T2 lesions

The number of new or enlarging T2 lesions as compared to the baseline MRI scan will be obtained from each MRI scan per protocol assessment schedule. To estimate the annualized rate of new or enlarging T2 lesions, below variables will be derived:

- The total number of new or enlarging T2 lesions during the treatment epoch will be derived by taking the number of new or enlarging T2 lesions from the last scheduled MRI scan with a non-missing value during the treatment epoch.
- The time (in years) from screening scan will be calculated as (date of last scheduled MRI scan with a non-missing value for the number of new or enlarging T2 lesions during the treatment epoch – date of screening scan +1)/365.25.

2.6.1.5 Neurofilaments light chain (NfL)

The NfL concentration in serum will be collected per protocol assessment schedule. To estimate the geometric mean of NfL, below variable will be derived:

- The log-transformed value of NfL at each time point will be derived by taking the natural logarithm of the NfL concentraction values.

2.6.1.6 Brain volume loss

The normalized brain volume will be obtained at baseline MRI scan and percent change from baseline in brain volume will be obtained from each MRI scan per protocol assessment schedule. To estimate the annual rate of percent change from baseline in brain volume, all available scheduled MRI assessments taken during the treatment epoch will be used and the below variable will be derived for each assessment.

- Time of MRI assessment from screening scan will be calculated as (date of the scheduled MRI scan – date the screening scan +1)/365.25.

2.6.2 Statistical hypothesis, model, and method of analysis

2.6.2.1 Disability worsening (3-month or 6-month confirmed)

Hypothesis and Analysis of disability worsening (3mCDW, 6mCDW)

The hypothesis and the analysis methods will be identical for 3mCDW and 6mCDW. For brevity the hypothesis and analysis methods are only specified in full for 3mCDW.

The null hypothesis is that there is no difference in the time to 3mCDW between ofatumumab 20mg sc and teriflunomide 14 mg po

- Superiority of ofatumumab 20mg sc over teriflunomide 14 mg po will be concluded if there is a reduction in risk (estimated hazard ratio from Cox-model<1) in patients treated with ofatumumab compared with teriflunomide and the observed p-value for the between-treatment comparison is less than the two-sided significance level of 0.04875 (=2*[0.025-0.025^2]). The multiplicity adjustment explained in Section 2.5.3 applies.

Confirmatory analysis for disability-related endpoints: The confirmatory analysis of time to 3mCDW will be done in a meta-analysis based on the combined FAS populations from G2301 and G2302.

The null hypothesis will be tested using a stratified Cox proportional hazards model with study as stratum, treatment, and region as factors and baseline EDSS as a continuous covariate. For confirmatory purposes the hazard ratio between ofatumumab and teriflunomide will be esitmated with 95% confidence interval and p-value from the combined data from both studies in this meta-analysis. In addition, an expanded model with a treatment-by-study interaction will be fit to the data and between-study heterogeneity will be tested as the type-3 test of the treatment-by-study interaction; the corresponding p-value will be provided. For information only, the hazard ratio for each study will be estimated by study from the same expanded model with corresponding 95% confidence intervals and p-values.

Supportive analysis: The confirmatory analysis of time to 3mCDW will also be done in a meta-analysis based on the combined per-protocol sets from G2301 and G2302.

Kaplan-Meier curves (and/or cumulative incidence plots) will be provided by treatment for the combined study populations to present the time-dependent cumulative probability of patients reaching 3mCDW. Kaplan-Meier curves (and/or cumulative incidence plots) and KM estimates by treatment will also be provided by study.

By-treatment Kaplan-Meier (KM) estimates (and/or 1-KM estimates) will be calculated for the combined study data at month 18 and month 24, with 95% confidence intervals. Similar estimates will be provided by study.

The log-rank test stratified by study will be performed for the combined study data as a supportive analysis. The log-rank test for each study data will also be performed for information only.

To estimate the hazard ratio between ofatumumab 20 mg sc once monthly and teriflunomide 14 mg po once daily during the initial “onset of action” period of 8 weeks (≤56 days=8*7 days), and thereafter (>56 days; long-term efficacy) a sensitivity analysis will be conducted. The Cox-

proportional hazards model will be specified as above, but in addition a time-dependent indicator variable (0 if record corresponds to the first 8 weeks, 1 otherwise) and an indicator-by-treatment interaction will be included in the model. Details of the statistical model and implementation of the model are provided in Section 5.5.3.2. From this model separate hazard ratios with 95% confidence intervals and p-values will be estimated for the first 8 weeks of treatment and the time period thereafter.

A sensitivity analysis will be performed similarly to the confirmatory analysis but patients who prematurely discontinued study with reason being "Lack of efficacy" as recorded in study phase completion CRF page and all patients who died during the treatment epoch (regardless whether due to MS or not) are considered having the disability worsening (3mCDW or 6mCDW) and time will be calculated based on a tentative onset date if it exists or on the study discontinuation date or death date as applicable (i.e., date of EDSS assessment at a tentative onset/EOS date/date of death- date of first administration of study drug+1).

An additional sensitivity analysis will be performed in which patients who prematurely discontinued study due to "Lack of efficacy" or died (due to any reason) during the treatment epoch will be considered having the disability worsening (3mCDW or 6mCDW) if patients were randomized to the ofatumumab treatment group.

2.6.2.2 Disability improvement (6-month confirmed)

Hypothesis and Analysis of disability improvement (6mCDI)

The null hypothesis is that there is no difference in the time to 6mCDI between ofatumumab 20mg sc and teriflunomide 14 mg po

- Superiority of ofatumumab 20mg sc over teriflunomide 14 mg po will be concluded if there is an improved change of a 6mCDI (estimated hazard ratio from Cox-model>1) in patients treated with ofatumumab compared with teriflunomide and the observed p-value for the between-treatment comparison is less than the two-sided significance level of 0.04875 (=2*[0.025-0.025^2]). The multiplicity adjustment explained in Section 2.5.3 applies.

Confirmatory analysis for disability-related endpoints: The confirmatory analysis of time-to-6mCDI will be done in a meta-analysis based on the combined FAS populations from G2301 and G2302.

The null hypothesis will be tested using a stratified Cox proportional hazards model with study as stratum, treatment, and region as factors and baseline EDSS as a continuous covariate. For confirmatory purposes the hazard ratio between ofatumumab and teriflunomide will be estimated with 95% confidence interval and p-value from the combined data from both studies in this meta-analysis. In addition, an expanded model with a treatment-by-study interaction will be fit to the data and between-study heterogeneity will be tested as the type-3 test of the treatment-by-study interaction; the corresponding p-value will be provided. For information only, the hazard ratio for each study will be estimated by study from the same expanded model with corresponding 95% confidence intervals and p-values.

Supportive analysis: Kaplan-Meier curves (and/or cumulative incidence plots) and by-treatment Kaplan-Meier (KM) estimates (and/or 1-KM estimates) will be calculated for the

combined study data at month 18 and month 24, with 95% confidence intervals. Similar estimates will be provided by study.

The log-rank test stratified by study will be performed for the combined study data as a supportive analysis. The log-rank test for each study data will also be performed for information only.

2.6.2.3 Number of Gd-enhancing lesions per scan

The null hypothesis is that there is no difference in the number of Gd-lesion per scan between ofatumumab 20mg sc and teriflunomide 14 mg po

- Superiority of ofatumumab 20mg sc over teriflunomide 14 mg po will be concluded if there are fewer Gd-lesions per scan (estimated rate ratio from a negative binomial model<1) in patients treated with ofatumumab compared with teriflunomide and the observed p-value for the between-treatment comparison is less than the two-sided significance level of 0.05. The multiplicity adjustment explained in Section 2.5.3 applies.

Confirmatory analysis of the number of Gd-lesions per scan: The confirmatory analysis of the number of Gd-lesions per MRI-scan will be done based on the FAS using a negative binomial regression model with log-link. The total number of Gd-enhancing T1 lesions (derived as in Section 2.6.1.3) will be used as the response variable, and the natural log of the number of MRI-scans (derived as in Section 2.6.1.3) will serve as the offset variable to adjust for the different number of MRI-scans between patients related to the flexible follow-up time in this study. The model will include treatment and region (factors), and age, and number of Gd-enhancing T1 lesions at baseline as continuous covariates.

The number of Gd-lesions per scan will be estimated by treatment with 95% confidence interval. The between treatment effect will be calculated as rate-ratio with 95% confidence interval and p-value. In addition the relative reduction in the number of Gd-lesion per scan will be computed as one (1) minus the rate-ratio and expressed as a percentage.

Supportive analysis: The number of Gd-lesions will be summarized descriptively by visit-window and treatment group.

2.6.2.4 Annualized rate of new or enlarging T2 lesions

The null hypothesis is that there is no difference in the number of new or enlarging T2 lesions between ofatumumab 20mg sc and teriflunomide 14 mg po

- Superiority of ofatumumab 20mg sc over teriflunomide 14 mg po will be concluded if there are fewer new or enlarging T2 lesions (estimated rate ratio from a negative binomial model<1) in patients treated with ofatumumab compared with teriflunomide and the observed p-value for the between-treatment comparison is less than the two-sided significance level of 0.05. The multiplicity adjustment explained in Section 2.5.3 applies.

Confirmatory analysis of the Annualized rate of new or enlarging T2 lesions: The confirmatory analysis of the annualized rate of new or enlarging T2 lesions will be done based on the FAS using a negative binomial regression model with log-link. The number of new or enlarging T2 lesions on the last available MRI scan relative to baseline (derived as in Section

2.6.1.4) will be used as the response variable, and the natural log of the time (in years) of the MRI assessment from the baseline scan (as defined in Section 2.6.1.4) will serve as the offset variable to adjust for the various lengths of follow-up times between patients in this study. The model will include treatment and region (factors), and age, and baseline volume of T2 lesions as continuous covariates.

The number of new or enlarging T2 lesions per year will be estimated by treatment with 95% confidence interval. The between treatment effect will be calculated as rate-ratio with 95% confidence interval and p-value. In addition the relative reduction in the number new or enlarging T2 lesions per year will be computed as one (1) minus the rate-ratio and expressed as a percentage.

Supportive analysis: The number of new or enlarging T2 lesions collected at each visit will be summarized descriptively by treatment group.

The number of new or enlarging T2 lesions are a cumulative measure of disease activity. Since the number of new or enlarging T2 lesion is assessed relative to the MRI scan collected at screening (i.e. well before the start of study medication intake), it is expected that this analysis underestimates the true treatment effect. Therefore, a supportive analysis will be performed based on the new or enlarging T2 lesions between the Month 12 scan and the End of Study scan.

2.6.2.5 Neurofilament light chain (NfL)

The null hypothesis is that there is no difference in NfL between ofatumumab 20 mg sc and teriflunomide 14 mg po by month 3.

- Superiority of ofatumumab 20 mg sc over teriflunomide 14 mg po will be concluded if NfL levels are already lower at month 3 in patients treated with ofatumumab compared with teriflunomide and the observed p-value for the between-treatment comparison is less than the two-sided significance level of 0.05. The multiplicity adjustment explained in Section 2.5.3 applies.

- **Confirmatory analysis of NfL:** The NfL concentration (geometric mean concentration) will be estimated by treatment and time point with 95% confidence intervals using a repeated measures model on the basis of all evaluable log-transformed NfL values (the natural log will be used). The response variable is the log-transformed values of the NfL level as the NfL level is expected to follow log-normal distribution. The exponentiated treatment differences are the geometric mean ratios (GMR) which will be reported by visit window with 95% confidence intervals and p-values. The statistical hypothesis test will be based on the treatment contrast and p-value obtained at month 3. An unstructured covariance matrix will be used. The model will include treatment, region and visit window as factors, and age, number of Gd-enhancing lesions at baseline, baseline T2 lesion volume and the log-transformed NfL baseline concentration as continuous adjustments. The model will also include a treatment-by-timepoint interaction to allow the treatment effect to vary over time. The treatment effect of ofatumumab versus teriflunomide will be visualized in a line plot with 95% confidence intervals.

Supportive analysis: If ofatumumab significantly reduces NfL as compared with teriflunomide, the following additional analyses will be performed. In all patients (FAS), and additionally in the subgroup of newly diagnosed (within 3 years prior to the screening visit) and treatment-

naïve (no prior MS DMT) patients, the hypotheses will be tested that patients with a high (>median) NfL concentration at baseline will experience more new or enlarging T2 lesions, a higher ARR, higher brain volume loss and tend to have a worse disease course on EDSS during the study compared with patients with low (≤median) NfL concentration at baseline. The median will be estimated from all patients in the combined ASCLEPIOS I and ASCLEPIOS II patients and then applied in all analyses.

- The T2 lesion rate will be estimated using a negative binomial model similar to the one specified in Section 2.6.2.4. In addition, the model will include NfL baseline category and a treatment-by-NfL baseline category interaction. The T2 lesion rate will be estimated by treatment within baseline NfL category. The treatment effect will be expressed as a lesion rate ratio with 95% confidence interval, for low and high baseline NfL category, respectively. From the same model the high vs low NfL contrasts will be estimated with 95% confidence intervals. The statistical hypothesis test with regards to the prognostic value of NfL for new lesion formation will be based on the high vs low baseline NfL category contrast in the teriflunomide group.
- The ARR will be estimated using a negative binomial model similar to the primary model specified in Section 2.5.2. In addition, the model will include NfL baseline category and a treatment-by-NfL baseline category interaction. The ARR will be estimated by treatment within baseline NfL category. The treatment effect will be expressed as an ARR-ratio with 95% confidence interval, for low and high baseline NfL category, respectively. From the same model the high vs low NfL contrasts will be estimated with 95% confidence intervals. The statistical hypothesis test with regards to the prognostic value of NfL for on-study relapses will be based on the high vs low baseline NfL category contrast in the teriflunomide group.
- The disease course based on EDSS will be analyzed as time to 3mCDW, 6mCDW or 6mCDI using Kaplan-Meier curves and a Cox proportional hazard model based on the combined data from COMB157G2301 and COMB157G2302. The model will include study as stratum, treatment, region and baseline NfL category as factors and baseline EDSS as a continuous covariate. In addition the model will include a treatment-by-NfL interaction term. The treatment effect will be expressed as hazard ratio with 95% confidence interval, for low and high baseline NfL category, respectively. The statistical hypothesis test of the prognostic value of NfL for disability changes will be based on the main effect of NfL category. It is acknowledged that the study will not be powered to show an effect on disability outcomes in subgroups. Kaplan-Meier curves will be produced with by treatment and NfL high or low category.
- The brain volume loss rate will be estimated using a random coefficient model similar to the one specified in Section 2.6.2.6. In addition, the model will include NfL baseline category and a treatment by time by NfL baseline category interaction as well as the 3 corresponding 2-way interactions. The annual rate of brain volume change will be estimated by treatment within baseline NfL category. The treatment effect will be expressed as an rate-ratio with 95% confidence interval, for low and high baseline NfL category, respectively. From the same model the high vs low NfL contrasts will be estimated with 95% confidence intervals. The statistical hypothesis test with regards to the prognostic value of

NfL for on-study brain volume loss will be based on the high vs low baseline NfL category contrast in the teriflunomide group.

- The benefit/risk profile of ofatumumab in the subgroup of newly diagnosed, treatment-naïve patients with high NfL will be compared to the overall trial results based on the above described efficacy analyses. The safety profile in the subgroup of newly diagnosed, treatment-naïve patients with high NfL will be compared to the overall trial population by summarizing adverse events by treatment side-by-side for the subgroup and the overall trial by SOC and PT as well as by predefined adverse events as defined in the latest version of case retrieval sheet (eCRS) at the time of analysis implementation (i.e., study database lock).

2.6.2.6 Brain volume loss

The null hypothesis is that there is no difference in brain volume change between ofatumumab 20mg sc and teriflunomide 14 mg po

- Superiority of ofatumumab 20mg sc over teriflunomide 14 mg po will be concluded if there is less brain volume loss (positive difference in slope from the random coefficient model) in patients treated with ofatumumab compared with teriflunomide and the observed p-value for the between-treatment comparison is less than the two-sided significance level of 0.05. The multiplicity adjustment explained in Section 2.5.3 applies.

Confirmatory analysis of brain volume loss: The percentage change from baseline in brain volume will be estimated on the basis of all scheduled scans up to the last available MRI scan that evaluated percentage brain volume change relative to baseline. A random coefficients model will be used as main analysis for this endpoint in order to adjust for the various length of follow-up time in this study. Brain volume change was approximately linear over time and approximately normally distributed in 3 independent studies of the fingolimod phase 3 program. The random coefficients model will include treatment, region as fixed effects (factors), and time, number of Gd-enhancing lesions at baseline, baseline T2 volume, and normalized brain volume at baseline as continuous covariates and treatment by time interaction. Time as a continuous covariate allows for estimation of different slopes and intercepts among treatment groups. The model will also contain random terms to account for deviations about the population slope and intercept. The statistical test will address the question whether there is a difference in the slope of brain volume change by treatment group. The annual rate of percent change from baseline in brain volume is approximates as the population slope within the treatment group. Since the random coefficients model allows for the estimation of different slopes and intercepts among the treatment groups, differences between treatments at any time can be summarized using this model. Model estimates of percentage brain volume change at Months 12 and 24 will be provided by treatment with corresponding 95% confidence interval. Treatment differences will be estimated at the same time points with confidence interval and a p-value for the test of a difference in slope.

Supportive analysis: Additionally, brain volume change and the annualized rate of brain atrophy (ARBA) will be summarized by visit. Calculation of ARBA is provided in Section 2.7.1. Cross-sectional comparisons at each visit will be made using ANCOVA models with ARBA as the response variable, treatment and region as factors, number of Gd-enhancing lesions at

baseline, baseline T2 volume, and normalized brain volume at baseline as continuous covariate. Line plots of percentage brain volume change over time by treatment will be provided.

In addition, the percentage change from baseline in brain volume will be estimated on the basis of all scheduled scans up to the last available MRI scan that evaluated percentage brain volume change relative to baseline in a repeated measures model. The unstructured covariance matrix will be used. The model will include treatment and scanning time point as factors (Month 12 and Month 24), region, number of Gd-enhancing lesions at baseline, baseline T2 volume and normalized brain volume at baseline as covariates. For the purpose of this model, visit windows will be used to map all available MRI assessments to Months 12 (Week 48) and 24 (Week 96) as defined in Table 2.5. The comparison of treatment differences in mean at Month 12 and 24 will be made using this model.

Percentage brain volume change is cumulative measure of disease activity. Since the percentage brain volume change is assessed relative to the MRI scan collected at screening (i.e. well before the start of study medication intake), it is expected that these analyses tend to underestimate the true treatment difference between study drugs.

2.6.3 Handling of missing values/censoring/discontinuations

2.6.3.1 Disability related endpoints (3mCDW, 6mCDW and 6mCDI)

Patients discontinuing study treatment should remain in the study and continue to follow the assessment schedule. The primary analysis of the time to 3-month/6-month confirmed disability worsening or the time to 6-month confirmed improvement uses all the available data up to the end of treatment epoch date, irrespective of on or off study treatment.

All patients in FAS will be included in the primary analysis. Patients who do not reach the endpoint by the end of treatment epoch date will be censored at date of last EDSS assessment during the treatment epoch.A per-protocol analysis of time to 3mCDW and time to 6mCDW will be performed to analyze on-treatment data (on-treatment period as defined in Section 2.1.1) from patients who have no major protocol deviations.

2.6.3.2 MRI endpoints

The primary analysis will use all available data up to the end of treatment epoch date, irrespective of on or off study treatment.

For the number of Gd-enhancing lesions and the number of new or enlarging T2 lesions, the primary NB model with an offset for the total number of scans or for the time from screening scan respectively adjusts for missing information (drop-out) under the assumption of non-informative drop-out, information is missing at random, and constant intensity (#lesion/time) of lesion formation over time.

For the percent change from baseline in brain volume, the primary analysis of random coefficient model will include longitudinal data from all scheduled post-baseline MRI scans as response. Under the assumption of linearity of the percentage brain volume change over time, the model corrects for missing values.

2.7 Analysis of secondary efficacy objective(s)

2.7.1 Secondary endpoints

Other secondary efficacy endpoints include:

- Time to first confirmed relapse: For patients with at least one confirmed relapse, the time to first confirmed relapse will be calculated as (start date of first confirmed relapse - date of first administration of study drug+1). For patients without confirmed relapses, they are censored with censored time as time in study (as defined in Section 2.5.1).
- ARR > 8 weeks after the onset of treatment (estimated from sensitivity analysis of primary endpoint) (details as described in Section 5.5.3.1).
- Risk of a 3mCDW > 8 weeks after the onset of treatment (estimated from sensitivity analysis of corresponding key secondary endpoint) (details as described in Section 5.5.3.2).
- Risk of a 6mCDW > 8 weeks after the onset of treatment (estimated from sensitivity analysis of corresponding key secondary endpoint) (details as described in Section 5.5.3.2).
- Time to 6-month confirmed cognitive decline on SDMT (6mCCD): similar to definition of 6mCDW (Section 2.6.1.1) but replacing the worsening criteria by at least an increase by 4 points from baseline in SDMT oral score. Patients who do not reach the endpoint by the end of treatment epoch date will be censored at date of last SDMT assessment during the treatment epoch.
- Time to 6mCDW or 6mCCD: defined as either a 6-month confirmed disability worsening, or a 6-month confirmed cognitive worsening, whichever is first. If a patient has both events achieved, time variables for this composite endpoint will be the same as the event occurred earlier. If 6mCDW is achieved but 6mCCD is not, the time variables for this composite endpoint will be the same as the time to 6mCDW event. If 6mCCD is achieved but 6mCDW is not, the time variables for this composite endpoint will be the same as the time to 6mCCD event. If neither events achieved, the patient is censored. The time variable will take the larger value in the censoring times corresponding to these 2 individual events.
- SDMT oral score by visit-window
- Time to 6-month confirmed worsening of at least 20% in the timed 25-foot walk test (T25FW): similar to definition of 6mCDW (Section 2.6.1.1) but replacing the worsening criteria by at least 20% increase from baseline in T25FW score. Patients who do not reach the endpoint by the end of treatment epoch date will be censored at date of last T25FW assessment during the treatment epoch.
- Time to 6-month confirmed worsening of at least 20% in the 9-Hole Peg Test (9HPT): similar to definition of 6mCDW (Section 2.6.1.1) but replacing the worsening criteria by at least 20% increase from baseline in 9HPT score in at least one hand. Patients who do not reach the endpoint by the end of treatment epoch date will be censored at date of last 9HPT assessment during the treatment epoch.

- 6mCDI sustained until EOS: same as 6mCDI except that the rest EDSS scores (scheduled or unscheduled) after the confirmation visit of the 6mCDI during the treatment epoch have to meet the disability improvement criteria as well.
- Disability status during the study (classified as improved, stable, or worsened): patients will be classified as improved if 6mCDI sustained until EOS was met (regardless whether 6mCDW status was met at an earlier time point); or classified as stable if 6mCDW was not met & 6mCDI sustained until EOS was not met; or classified as worsened if 6mCDW was met & 6mCDI sustained until EOS was not met.
- MS relapse characteristics include severity, recovery, hospitalization, and steroid treatment. The severity of MS relapses will be derived according to the criteria in Table 2-15. All relapses (confirmed or unconfirmed) will be considered. For confirmed relapses, the EDSS used to confirm the relapse (as specified in Section 2.5.2) will be used to determine the severity. For unconfirmed relapses, all EDSS assessments taken within 30 days from the relapse start date and before the relapse end date will be checked among which the most severe case will be selected to determine the severity. If no such EDSS assessment is available, the severity will be set to "missing EDSS". If none of such EDSS assessments meet the mild, moderate or severe conditions, the severity will be set to "no worsening in EDSS".

Table 2-15 Severity of MS relapse

Mild relapse	Moderate relapse	Severe relapse
EDSS increase of 0.5 point	EDSS increase of 1 or 2 points	Exceeding Moderate criteria
or	or	or
1 point FS change in one to three systems	2-point FS change in one or two systems	Exceeding Moderate criteria
	or	or
	1-point change in four or more systems	Exceeding Moderate criteria
Definition is based on the EDSS obtained to confirm the relapse as compared to the last EDSS (scheduled or unscheduled) taken in the absence of (confirmed or unconfirmed) relapse and prior to the current relapse. EDSS refers to total score; FS refers to functional score; all of the 7 functional scores are considered in this derivation.		

[Panitch et al 2002]

- ARR time-based and ARR patient based: ARRs using a "time-based approach" are calculated by taking the total number of relapses observed for all patients within a treatment group divided by the total number of days in study of all patients within the treatment group and multiplied by 365.25 days. Also a "patient-based approach" is presented, where individual ARRs are computed and summarized over patients within a treatment group. For above ARR calculations, one analysis will consider confirmed relapses only. Another analysis will consider both confirmed and unconfirmed relapses together. T2 lesion volume and its change from baseline MRI scans by visit-window

- Number of new or enlarging T2 lesions on MRI between Month 12 and EOS: The number of new T2 lesions on the EOS scan relative to the Month 12 scan will be collected; The time (in years) from the Month 12 scan to the EOS MRI scan will be calculated as (date of EOS MRI scan – date of Month 12 MRI scan +1)/365.25.
- Number of new or enlarging T2 lesions on MRI at yearly visits relative to screening scan: The number of new T2 lesions at yearly visit scans relative to the screening scan will be collected; The time (in years) from the screening scan to the yearly MRI scan will be calculated as (date of yearly MRI scan – date of screening MRI scan +1)/365.25.
- Annualized rate of brain atrophy (ARBA) by visit-window: ARBA describes the "averaged annual percentage change" in brain volume. It is designed to adjust for differences in the time spent between two scans that are compared by standardizing the percentage change in brain volume (PBVC) to 1 year. The logic of interest rates applies. ARBA = $[(PBVC/100+1)^{(365.25/days)}-1]*100$, where "PBVC" represents the percentage brain volume change obtained between 2 scans and "days" stands for the number of days between the two scans that are being compared. Days are calculated as post-baseline MRI scan date-screening MRI scan date +1.
- Proportion of patients free of clinical and MRI disease activity (No evidence of disease activity; NEDA-4): NEDA-4 is defined as no 3mCDW, no confirmed MS relapse, no new or enlarging T2 lesions on any MRI scan (scheduled or unscheduled) compared to baseline, and brain volume change >-0.4%/year on all MRI scans (scheduled and unscheduled; brain volume as measured by ARBA). The main analysis will consider only those patients who were followed-up to the assessment time point in the analysis (e.g. only patients with ≥12 months of follow-up in the 12-month assessment of disease freedom, etc.). Intermediate missing values (e.g. due to missing MRI assessments) will be considered not free of disease activity. For the derivation of NEDA-4, data up to day 365.25 will be considered in the year 1 analysis and data up to day 730.5 will be considered in the year 2 analysis.
- EDSS total score by visit-window
- The 25-foot Timed Walking Test (T25FW) score by visit-window: two trials will be performed at each scheduled visit. The average scores of the two trials per visit will be calculated as $T25FW_{score} = \frac{1}{2}(T25FW_1 + T25FW_2)$. In case of missing data, the average of the non-missing values will be taken.
- The 9-Hole Peg Test (9HPT) score by visit-window: four trials (two of each arm) will be performed at each scheduled visit. The 9HPT score per visit will be calculated for the dominant and the non-dominant hands separately and then for the average of both hands. In case of missing data, the average of the non-missing values will be taken.
  - $9HPT_{dominant\ hand} = (9HPT_{dominant\ hand,\ trial\ 1} + 9HPT_{dominant\ hand,\ trial\ 2})/2$
  - $9HPT_{non\text{-}dominant\ hand} = (9HPT_{non\text{-}dominant\ hand,\ trial\ 1} + 9HPT_{non\text{-}dominant\ hand,\ trial\ 2})/2$
  - $9HPT_{total} = (9HP_{non\text{-}dominant\ hand} + 9HP_{dominant\ hand})/2$.
- NfL related secondary hypotheses are specified in Section 2.6.2.5 as 'supportive analysis' for the key-secondary endpoint NfL.

2.7.2 Statistical model and method of analysis

Methods of analyses for efficacy endpoints defined in Section 2.7.1 are described below.

- The analysis of "Time of first relapse" and "Annualized relapse rates > 8 weeks after the onset of treatment" is described in (Section 2.5.5) as sensitivity analysis to the primary analysis.
- The analysis of the "Risk of a 3mCDW > 8 weeks after the onset of treatment" and "Risk of a 6mCDW > 8 weeks after the onset of treatment" is described in (Section 2.6.2.1) as a supportive analysis to the key-secondary 3mCDW and the 6mCDW, respectively.
- Time to 6mCCD will be analyzed in the combined FAS from studies G2301 and G2302, in a Cox proportional hazards model with study as stratum, treatment, and region as factors and baseline SDMT-result as a continuous covariate. By study analysis will be performed for information only.
- Time to 6mCDW or 6mCCD (composite endpoint) will be analyzed in the combined FAS from studies G2301 and G2302, in a Cox proportional hazards model with study as stratum, treatment, and region as factors and baseline EDSS and baseline SDMT scores as continuous covariates. By study analysis will be performed for information only.
- SDMT scores and change from baseline will be summarized by visit-window and treatment in the combined FAS from studies G2301 and G2302. For the change from baseline values, a repeated measures mixed effects analysis adjusted for study, treatment, region, and SDMT baseline scores will be performed.
- Time to 6-month confirmed worsening of at least 20% in the timed 25-foot walk test (T25FW) will be analyzed in the combined FAS from studies G2301 and G2302, in a Cox proportional hazards model with study as stratum, treatment, and region as factors and the baseline T25FW-result as a continuous covariate. By study analysis will be performed for information only.
- Time to 6-month confirmed worsening of at least 20% in the 9-hole peg test (9HPT) will be analyzed in the combined FAS from studies G2301 and G2302, in a Cox proportional hazards model with study as stratum, treatment, and region as factors and baseline 9HPT-result as a continuous covariate. By study analysis will be performed for information only.
- The 6mCDI sustained until EOS will be analyzed in the same way as for the 6mCDI (as described in Section 2.6.2.2)
- The disability status during the study will be summarized descriptively by treatment group (i.e., number and percent of patients improved, stable or worsened will be provided) and compared between treatment groups via a Chi-square test.
- For MS relapse characteristics, summary statistics will be presented. The proportion of patients hospitalized for relapse and the proportion of patients with severe relapses will be compared between treatment groups using a Chi-square test.
- For the presentation of descriptive statistics, ARR time based and ARR patient based will be reported by treatment group.

- Change in T2 lesion volume on MRI will be summarized by visit-window. Cross-sectional analyzes will be done by using a rank ANCOVA with treatment and region (factors), and baseline T2 volume as a continuous covariate.
- Number of new or enlarging T2 lesions on MRI between EOS and Month 12 will be analyzed in a negative binomial regression model, and presented similarly to that used for the annualized rate of new or enlarging T2 lesions. The number of new T2 lesions on the EOS scan relative to the Month 12 scan will be used as the response variable. The natural log of the time (in years) from the Month 12 scan to the EOS MRI scan will serve as the offset variable.
- Number of new or enlarging T2 lesions on MRI at yearly visits relative to baseline will be analyzed in a negative binomial regression model and presented similarly to that used for the annualized rate of new or enlarging T2 lesions. The number of new T2 lesions on the yearly MRI scan relative to the screening scan will be used as the response variable. The natural log of the time (in years) from the screening scan to the yearly MRI scan will serve as the offset variable. Due to the variable length of patient study participation, this analysis will only be performed for visits where there are at least a total of 100 patients with non-missing data on this endpoint. In case of non-convergence, region will be removed from the regression model.
- The analysis of ARBA is described in (Section 2.6.2.6) as supportive analysis to the key secondary endpoint of brain volume change.
- Proportion of patients free of clinical and MRI disease activity (No evidence of disease activity; NEDA-4) will be analyzed cross-sectionally at year 1 and year 2 in a logistic regression model with treatment and region as factor, and age, baseline EDSS, and number of Gd-lesions at baseline as covariates. A sensitivity analysis will be conducted considering all patients who discontinued from the study prior to the assessment timepoint, or have missing MRI assessments prior to the assessment timepoint, as not free of disease activity.
- EDSS, T25FW, 9HPT: The scores and their change from baseline values will be summarized by visit-window and treatment in the combined FAS from studies G2301 and G2302. For the change from baseline values, a repeated measures mixed effects analysis adjusted for study, treatment, region, and corresponding baseline values will be performed.
- NfL related secondary hypotheses are specified in Section 2.6.2.5 as ‘supportive analysis’ for the key-secondary endpoint NfL.

2.7.3 Handling of missing values/censoring/discontinuations

As a general rule, missing data will not be imputed in any secondary endpoint analyses. All the available data up to the end of treatment epoch date, irrespective of on or off study treatment will be used as applicable. Details of handling discontinuations or censoring are provided in the relevant endpoints or analysis of endpoints sections (Section 2.7.1 and/or Section 2.7.2).

2.8 Safety analyses

Safety analyses will be conducted using the safety (SAF) set. Patients will be grouped by the actual treatment received. If patients received both investigational drug and active comparator

accidentally during the study, they will be analyzed according to the treatment to which they were exposed longer in duration. Unless explicitly stated otherwise, only data up to and including the safety cut-off of 100 days after last administration of study drug will be included in the analysis and data beyond this time point for a given patient will be excluded from the safety analysis. The safety cut-off of 100 days (5 x 20 days) takes the long half-life of both the investigational and comparator-drug into account.

The assessment of safety will be primarily based on the frequency of adverse events (including death and non-fatal serious adverse events). Additional safety assessments include laboratory tests, physical examination (including examination of skin), vital sign measures, ECG evaluations and assessment of suicidality. Clinically significant findings in these additional safety assessments will be reported as adverse events and analyzed as such. In addition all safety assessments will be summarized or listed as appropriate.

2.8.1 Adverse events (AEs)

An adverse event (AE) is any untoward medical occurrence (i.e., any unfavorable and unintended sign [including abnormal laboratory findings], symptom or disease) in a patient or clinical investigation of a patient after providing written informed consent for participation in the study. That means that a patient can report AEs before having started study medication. For reporting purposes, the main focus will be on treatment emergent adverse event (TEAE), defined as any adverse event which started on or after the day of first dose of study medication.

Except for serious TEAES and death, only TEAEs up to and including safety cut-off of 100 days after last administration of study drug will be included in the analyses. All serious TEAES and death will be included, regardless of safety cut-off.

AEs will be reported by primary system organ class (SOC) and preferred term according to the Medical Dictionary for Regulatory Activities (MedDRA). The MedDRA version used for reporting the study will be described in a footnote.

The number and percentage of patients reporting any TEAEs (referred to as incidence of any TEAEs later) will be summarized by primary SOC, preferred term and treatment. Separate summaries will be provided for serious TEAEs, drug related TEAEs, TEAEs leading to permanent discontinuation of study drug, TEAEs causing study drug interruption, and most common TEAEs (≥ 2% in any of the treatment groups). Additionally, incidence of any TEAEs will also be summarized by SOC, preferred term, and maximum CTCAE grade. Missing CTCAE grade will not be imputed..

Odds ratios (investigational drug vs. control) with 95% confidence intervals will be presented along with the summary of any TEAEs and the summary of serious TEAEs. Details in calculations of odds ratios and their 95% confidence intervals are provided in Section 5.7.1. Incidence of any TEAEs will be plotted by SOC, along with the odds ratio (and corresponding 95% confidence interval) for TEAEs in each SOC. Odds ratios provide appropriate inferences for events that are likely to occur at certain times (e.g., injection related reactions) while no constant risk over time is expected.

Given the flexible duration of patients follow-up in the study, any TEAEs and serious TEAEs will also be summarized by reporting exposure-adjusted incidence rates (assuming a Poisson-process for adverse events), by primary SOC, preferred term and treatment. The analysis will

take into account the time-at-risk as defined in Section 2.4.1 for each patient. The incidence rate ratios (investigational drug vs. control) and corresponding 95% confidence intervals will be presented along with this summary. Details in calculations of incidence rates and their 95% confidence intervals as well as incidence rate ratios and their 95% confidence intervals are provided in Section 5.7.2. The exposure adjusted incidence rates provide appropriate inferences for events that are likely to occur at any time during the observation period with constant risk over time (e.g., infection or cancer).

If a patient reported more than one adverse event within the same primary system organ class, the patient will be counted only once with the maximum CTCAE grade at the system organ class level, where applicable.

All AEs will be presented in listings.

For the subgroup of patients who accidentally received both investigational drug and active comparator during the study, a listing of adverse events along with drug administration information will be provided.

2.8.1.1 Selected AEs of special interest / grouping of AEs

Selected tables will be produced for Adverse Events of Special Interest (AESI) (i.e., risks) defined in the latest version of case retrieval sheet (eCRS) at the time of analysis implementation (i.e., study database lock). Specifically, incidence of TEAEs that fulfill the risk search terms as defined in eCRS will be summarized by risk name, preferred term and treatment with odds ratios and 95% confidence intervals presented. The exposure-adjusted incidence rate of TEAEs that fulfill the risk search terms as defined in eCRS will also be summarized by risk name, preferred term and treatment with incidence rate ratios and 95% confidence intervals presented. Similarly, separate summaries will be provided for serious TEAEs that fulfill the risk search terms as defined in eCRS. Additionally, incidence of any TEAEs that fulfill the search terms as defined in eCRS will also be summarized by risk name, preferred term, and maximum CTCAE grade.

2.8.1.1.1 Injection reaction related AEs

Incidence of injection site reaction AEs and injection systemic reaction AEs as collected in the relevant CRF pages will be reported as part of the AE summary tables as two preferred terms respectively. For summaries of injection systemic reaction AEs specified in this section, , reaction/symptom start date and time will be compared with the injection date and time. Only reactions/symptoms within 24 hours after injections will be included (i.e., time to onset of reaction <= 24 hours). The time to onset of reaction will be derived as (reaction start date/time – injection date/time) and rounded to the closest integer in hours.

Symptoms listed in the injection site reaction or injection systemic reaction CRF pages will be summarized by providing the number and percentage of patients with each of the symptoms and pre-specified grouping of symptoms as well as overall. These summaries will be provided for each injection up to injection 10 and cumulatively for all injections.

For the injection site reaction, no grouping of symptoms will be specified. For the injection systemic reaction, symptoms will be grouped under 6 categories as defined below and the

number and percentage of patients with at least one symptom reported under the category will be provided for each category.

- Skin/mucosal tissue symptoms: Rash, Urticaria, Pruritus general, Flushing
- Respiratory compromise: Dyspnea, Bronchospasm, Chest discomfort, Cough
- Related to change in vital signs: Hypotension, Hypertension, Dizziness, Tachycardia
- Gastrointestinal symptoms: Nausea, Vomiting, Abdominal pain, Diarrhea
- Musculoskeletal/connective tissue symptoms: Arthralgia, Myalgia, Back pain
- Other manifestations: Fever, Headache, Chills, Asthenia, Fatigue, and Other

For the injection systemic reaction, time to onset of the first symptoms after each injection will be summarized by cumulative hour intervals (i.e., 0 to 1, 0 to 2, 0 to 3, etc. until 0 to 24 hours). These summaries will be provided for each injection up to injection 10 and cumulatively for all injections.

The proportion of patients with injection site reactions and the proportion of patients with injection systemic reactions will be plotted by treatment group against the injection sequence numbers (injections 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10).

For the first 3 injections, additional summaries will be provided by category of injection related premedication. Specifically, 9 categories are defined as 1) no injection related premedication, 2) type 1, 3) type 2, 4) type 3, 5) type 1+ type 2, 6) type 1+ type 3, 7) type 2+ type 3, 8) type 1+ type 2+ type 3, 9) Any injection related premedication. The types 1 to 3 refer to protocol specified three types of injection related premedication as stated in Section 2.4.2.3.

A separate listing for injection site reaction AEs and injection systemic reaction AEs will be provided where all reported injection related reactions (IRR) (regardless of timing) will be listed along with the corresponding injection sequence number and time to onset from the most recent injection as well as from Day 1. The listing will be sorted by injection of first occurrence of symptoms. Specifically, under the first injection, patients whose first IRR was associated with the first injection will be listed; within a patient all IRRs to all injections will be listed chronologically. Under the second injection, patients whose first IRR was associated with the second injection will be listed; similarly, within a patient all IRRs to all injections will be listed chronologically. The listing will continue with subsequent injections until all patients with IRRs are included in the listing.

2.8.1.1.2 Liver safety related AEs

Standardized MedDRA Queries (SMQs) are groupings of terms from one or more MedDRA SOCs that relate to a defined medical condition or area of interest. They are intended to aid in case identification.

SMQ Table: The number and percentage of patients experiencing adverse events categorized under the SMQ module drug-related hepatic disorders – comprehensive search (SMQ code 20000006, broad search) and its SMQ sub-modules will be reported.

SMQ-PT Table: The more detailed SMQ-PT table including the respective preferred term frequencies falling under the SMQ drug-related hepatic disorders – comprehensive search (SMQ code 20000006, broad search) and its SMQ sub-modules will be provided.

2.8.1.1.3 Other analyses

For the legal requirements of ClinicalTrials.gov and EudraCT, two required tables on non-serious treatment emergent adverse events with an incidence greater than 5% and on treatment emergent serious adverse events and SAE suspected to be related to study treatment will be provided by system organ class and preferred term on the safety set population.

If for a same patient, several consecutive AEs (irrespective of study treatment causality, seriousness and severity) occurred with the same SOC and PT:

•a single occurrence will be counted if there is ≤ 1 day gap between the end date of the preceding AE and the start date of the consecutive AE

•more than one occurrence will be counted if there is > 1 day gap between the end date of the preceding AE and the start date of the consecutive AE

For occurrence, the presence of at least one SAE / SAE suspected to be related to study treatment / non SAE has to be checked in a block e.g., among AE's in a ≤ 1 day gap block, if at least one SAE is occurring, then one occurrence is calculated for that SAE.

The number of deaths resulting from SAEs suspected to be related to study treatment and SAEs irrespective of study treatment relationship will be provided by SOC and PT.

2.8.2 Deaths

Death, if meaningful number of cases reported (i.e., 5 or more cases), will be summarized by providing the number and percentage of patients by treatment group. All deaths as recorded in the final database (i.e., up to database lock) will be included.

2.8.3 Laboratory data

Data summaries will be provided in SI units. The summary of laboratory evaluations will be presented for three groups of laboratory tests: Hematology, Chemistry and Urinalysis. On presenting summary statistics, laboratory data will be grouped and displayed in an alphabetical order within the Hematology and Chemistry groups and subgroups. Refer to Section 5.3.1 for subgroup definitions.

Descriptive summary statistics (mean, median, standard deviation, Min and Max) of the change from baseline in the laboratory result to each study visit-window by treatment group will be presented. Change from baseline will only be summarized for patients with both baseline and post baseline values and will be calculated as:

change from baseline = post baseline value – baseline value

In addition, shift tables will be provided for all parameters to compare a patient's baseline laboratory evaluation relative to the post-baseline values. For the shift tables, the grade level based on CTC grade (as defined by CTCAE as listed in Section 5.3.3) will be used to evaluate whether a particular laboratory test value was Grade 0, 1, 2, 3, or 4 relative to whether or not

the baseline value was Grade 0, 1, 2, 3 or 4. These summaries will be presented by laboratory test and treatment group. Shift table will also be provided for urinalysis results to compare baseline to post-baseline extreme values (negative, +, 2+, 3+, or 4+).

The number and percentage of patients with new or worsening laboratory abnormalities based on CTC grade (as defined by CTCAE as listed in Section 5.3.3) in each visit-window and at any time post baseline will be presented. Patients with specific laboratory abnormalities (defined by CTC grade 3 and 4) will be listed.

Number of patients with newly occurring liver enzymes abnormalities will be summarized. Newly occurring liver enzymes abnormalities are defined in Section 5.3.2.

Liver function tests will also be presented graphically as matrix plots of each of the parameters (ALT, AST, TBIL, ALP) maximum post-baseline/ULN normalized.

For the shift tables, abnormalities based on CTC grades tables and liver frequency distribution tables, all applicable post-baseline values will be checked against the respective criteria and the rules for handling multiple laboratory assessments within visit windows will not be applied.

For continuous variables databased as <lower limit, these will be imputed as being half of the lower limit.

All above summaries include only data up to and including safety cut off.

2.8.3.1 Liver events

Detailed information will be captured for patients with a reported liver event. For most studies patient listings may be appropriate because only a few liver events are expected, if any. However, respective tables with descriptive statistics may be provided if the number of liver events allows a meaningful interpretation, specifically if the number of liver events are greater than or equal to 10.

2.8.3.2 Renal events

Detailed information will be captured for patients with a reported renal event. The overall frequency and percentage of patients with confirmed renal events will be summarized by treatment group. Answers to “Event first identified through”, “Renal event associated signs and symptoms”, and “Any alternative diagnoses” (collected on diagnosis, follow up and overview pages) will also be summarized descriptively by treatment group. All data collected on the standard renal events CRF pages will be listed appropriately.

2.8.3.3 Serology at screening

Collected data of laboratory serology to determine patient’s eligibility includes parameters with respect to hepatitis and HIV viruses. Parameters with respect to hepatitis include 1) anti-hepatitis A virus IgM; 2) hepatitis B surface antigen and anti-hepatitis B core antigen IgM/IgG; 3) anti-hepatitis C virus antibody; 4) anti-hepatitis E virus IgM.

The result of each serology parameter is a categorical value. The number and percentage of patients with values in each of the categories will be provided by treatment group for all above mentioned parameters. Serology test should be performed once at screening. For patients with

multiple assessments, the latest result will be used in the summary. Additional serology data with numerical values may be collected but will not be reported.

2.8.3.4 Other special lab results

Other non-routine laboratory data include pregnancy test results, B-cell counts, teriflunomide plasma levels, total IgG, total IgM, [redacted]. All data will be listed appropriately.

The B-cell counts, total IgG, total IgM, neurofilaments [redacted] will be summarized using descriptive statistics by treatment group and visit-window.

In addition, number and percentage of patients with B-cells < the lower limit of normal (LLN) value (i.e., B-cell depleted) will be presented by treatment group and visit-window. Graphic presentation of B-cell counts distribution will also be provided by treatment group and visit-window. All B-cell summaries defined in this section will include data up to the end of the on-treatment period (i.e., last dose date + 30). This is to provide more accurate information about B-cell depletion because fast B-cell repletion is expected after study drug discontinuation and applying the safety cutoff (i.e., last dose date +100) would mix repletion and depletion to some degree.

Number and percentage of patients meeting the notable low level criteria in IgG or IgM at least once will be provided by treatment group. A notably low IgG level is defined as a level that is 20% below the LLN and a notably low IgM level is defined as a level 10% below the LLN.

2.8.4 ECG data

ECG data will be collected at baseline visit, EOT visit and EOS visit. Clinically significant findings from ECG evaluations will be reported as AEs and included in the analysis of AEs. ECG parameters include max heart rate, mean PR duration, mean QT duration, mean QRS duration, and QT corrected using Fridericia's correction formula (all as collected on the ECG CRF). Descriptive statistics of each ECG parameter will be provided by treatment group for baseline and for the 2 nominal visits (i.e., EOT and EOS visits).

The number and percentage of patients meeting the criteria defined in Table 2-16 will be provided for each criterion by treatment group for baseline and for the 2 nominal visits.

Table 2-16 Criteria for relevant ECG absolute or change from baseline values

Absolute values criteria:	Changes from baseline criteria:
Heart rate: HR <40 or HR > 120 beats/min Pulse rate: PR <110 or PR >200 msec QRS complex: < 70 msec or > 120 msec QTcF < 350 or > 450 msec (males) QTcF <360 or > 460 msec (females)	QRS complex : increase > 25% compared to baseline QTcF > 500 msec and QTcF increase > 60 msec

2.8.5 Vital signs

Vital sign measurements include sitting systolic and diastolic blood pressures, sitting pulse, body temperature, height and body weight.

Three sitting measurements of blood pressure (SBP and DBP) and pulse will be taken at each vital sign assessment.

For post-baseline assessments, the blood pressure and pulse values will be the average of the non-missing values of the 3 measurements. If more than one blood pressure/pulse assessment (scheduled or unscheduled) exists in a particular visit-window (as defined in Section 2.1.2.1), derivation should follow the rules as defined in Section 2.1.2.3. Derivation of baselines for blood pressure and pulse are provided in Section 2.1.1.

Height will be collected at screening visit only and will be summarized in the baseline characteristic summary only.

Analyses of vital sign measurements (excluding data collected on Day 1, Day 7, Day 14 and Month 1 protocol scheduled visit) using descriptive summary statistics (mean, median, standard deviation, min, max) for the change from baseline for each post-baseline visit-window will be performed. These descriptive summaries will be presented by vital sign parameter and treatment group. Change from baseline will only be summarized for patients with both baseline and post-baseline values and will be calculated as:

change from baseline = post-baseline value – baseline value

The number and percentage of patients with clinically notable vital signs will be presented. For clinical notable vital signs values, refer to Section 5.4.

For vital signs data collected on Day 1, Day 7, Day 14 and Month 1 protocol scheduled visit, pre and post-injection vital signs data including temperature, pulse rate and blood pressure are collected. Change from pre-injection to post-injection in these 3 parameters will be summarized by nominal visit.

All above summaries include only data up to and including safety cut off.

2.8.6 Suicidality evaluations

The Columbia Suicide Severity Rating Scale (C-SSRS) questionnaire will be administered electronically (eC-SSRS), via an interactive voice response (IVR) system. However, as per internal Novartis guidelines, a CRF page ‘Supplemental Data for Suicidal Ideation and Behavior Categories’ is also used as an unplanned/unscheduled visit for those cases when the subject did not conduct the phone interview because, e.g. the subject was hospitalized and unable to conduct the interview or the subject refused to conduct the interview/withdrew from the study and external information on suicidal ideation and behavior (SIB) is required. In such cases, SIB information will still be collected from external parties (spouse, caregiver, nurse, investigator, etc.) through means of this CRF page. Further, for deaths due to suicide, the site should fill out the Supplemental Data CRF (“completed suicide” tickbox) to ensure accurate reporting of such cases. When reporting the SIB data, data from both sources (IVR eC-SSRS and Supplemental CRF) will be used in a pooled manner and no distinction will be made between the two.

Table 2-17 Standard SIB events as categorized by C-SSRS

Category number	C-SSRS category
Suicidal Ideation	
1	Wish to be dead
2	Non-specific active suicidal thoughts
3	Active suicidal ideation with any methods (not plan) without intent to act
4	Active suicidal ideation with some intent to act, without specific plan
5	Active suicidal ideation with specific plan and intent
Suicidal behavior	
6	Preparatory acts or behavior
7	Aborted attempt
8	Interrupted attempt
9	Actual attempt
10	Completed suicide
Self-injurious behavior, without suicidal intent	
11	Non-suicidal self-injurious behavior

Definition of 'all prior history' and 'recent history'

SIB will be collected once before randomization, either at screening or at baseline visits. SIB assessments obtained before the first dose of study drug will be defined by two components: *all prior history and recent history.*

All prior history will be defined as the SIB results obtained from the *lifetime* assessment, and *recent history* will be defined as the SIB results obtained from the *pre-defined period (i.e., past 6 months or 24 months)* assessment of the screening or baseline visit. In case multiple assessments collected before first dose of study drug, results should be derived as the worst case value per the reporting timeframe. Worst case is defined by an answer 'yes' to the SIB category.

Data summaries

SIB data will be summarized for the Safety set. The number and percentage of subjects with suicidal ideation, suicidal behavior and self-injurious behavior without suicidal intent will be presented by analysis-period (recent history, all prior history and any time post-baseline) and treatment group. The following 14 events will be included in the summary table:

- Each of the 11 categories listed in Table 2-17, separately
- Any suicidal ideation or behavior (a 'yes' answer to at least one of the 10 suicidal ideation and behavior questions in analysis-period of interest)
- Any suicidal ideation (answered 'yes' to at least one of the 5 suicidal ideation questions in analysis-period of interest)
- Any suicidal behavior (answered 'yes' to at least one of the 5 suicidal behavior questions in analysis-period of interest)

In addition, the number and percentage of subjects with the following post-baseline events will be presented (Nilsson et al. 2013):

Note: in these definitions, a category of 0 is assigned to a subject without suicidal ideation (i.e. a 'no' answer to all suicidal ideation categories)

- Worsening suicidal ideation compared to recent history: An increase in the maximum suicidal ideation category at any time post-baseline from the maximum suicidal ideation category during pre-treatment recent history.
- Worsening serious suicidal ideation compared to recent history: An increase in the maximum suicidal ideation category to 4 or 5 at any time post-baseline from not having serious suicidal ideation (categories of 0-3) during pre-treatment recent history.
- Emergence of serious suicidal ideation compared to recent history: A maximum suicidal ideation category of 4 or 5 at any time post-baseline from no suicidal ideation (category 0) during pre-treatment recent history.
- Improvement in suicidal ideation at last on-treatment measurement compared to recent history: A decrease in suicidal ideation score at last on-treatment measurement from pre-treatment recent history.
- Emergence of suicidal behavior compared to all prior history: The occurrence of suicidal behavior (Categories 6-10) at any time post-baseline from not having suicidal behavior (Categories 6-10) prior to treatment (includes "lifetime" and any other assessments prior to treatment taken prior to treatment).

For those analyses, each subject can only be counted once for each event. However, a subject can be counted in several different events.

Suicidal ideation and behavior data will be listed. Detailed answers to C-SSRS items will be listed separately for subjects with any suicidal ideation at any time post-baseline (i.e. a 'yes' answer to at least one of the five suicidal ideation questions at any time post-baseline) and for a subject with any suicidal behavior at any time post-baseline (i.e. a 'yes' answer to at least one of the five suicidal behavior questions at any time post-baseline).

2.8.7 Safety evaluation during the Safety follow-up

Safety data collected after last administration of study drug includes adverse events, vital signs, routine laboratory parameters, laboratory assessments to measure total IgG, total IgM, B-cell repletion and residual teriflunomide PK levels, and eCSSRS. No safety cutoff date will be applied in the analyses defined in this section. Safety data within the safety cutoff date but after last administration of study drug will also be included.

The number and percentage of patients with at least one TEAEs that started after the date of last administration of study drug will be reported by SOC, preferred term, and treatment group. The subset of patients in SAF who had data reported after the date of last administration of study drug will be included in the analysis and the number of patients in this subset will be the denominator in calculating the percentage.

The vital signs data collected after study drug discontinuation will be summarized as summary statistics of the vital sign measures and of the change from baseline in the vital sign measures by treatment group for each visit-window after study drug discontinuation (as defined in Section 2.1.2.2). The SAF will be used for these analyses.

The B-cell counts and its change from baseline values will be summarized by treatment group for each visit-window after study drug discontinuation. In addition, the change from last assessment on study drug in B-cell count will also be summarized similarly. Graphic presentation of B-cell counts distribution will also be provided by treatment group for each visit-window after study drug discontinuation. Additionally, number and percentage of patients with B-cells within 10%, 10 to 20%, 20 to 50%, 50% to 80%, 80 to 100% of baseline, or with B-cells < the lower limit of normal (LLN), or with B-cell repletion will be presented by treatment group for each visit-window after study drug discontinuation. B-cell repletion is defined as B-cell counts having returned to their baseline value or to LLN (i.e., >=100% of baseline or >=LLN). The analysis will be limited to the subset of patients in SAF who had B-cell count data after the date of last administration of study drug; the number of patients in this subset will be the denominator in calculating the percentage.

Patients with teriflunomide plasma data will be listed as only small percentage of patients are expected to have such data.

The eCSSRS data collected after study drug discontinuation will be summarized similarly as in Section 2.8.6. The subset of patients in SAF who had at least one eCSSRS performed after the date of last administration of study drug will be included in the analysis and the number of patients in this subset will be the denominator in calculating the percentage.

As stated in Section 2.1, the registration CSR will contain only partial data collected from the Safety FU epoch because the registration CSR cutoff date will depend on EOS date of the study. A complete analysis of the post-treatment safety follow-up will be provided in the final CSR when all patients have completed the Safety FU epoch.

2.9 Pharmacokinetic endpoints

[redacted]

2.9.2 Immunogenicity assessment

Samples will be analyzed for the presence of human anti-drug antibodies (ADA). The data will be summarized by visit and overall (i.e. proportion of patients with ADA) as an assessment of the immunogenicity potential of ofatumumab. A listing by patient will also be provided. All patients in FAS with non-missing values will be included in these analyses.

2.10 PD and PK/PD analyses and modelling

PK concentration data summarized by visit together with by-patient listings will be provided in the study report. All patients in FAS with non-missing values will be included in these analyses.

[redacted]

2.11 Patient-reported outcomes

For patient-reported outcomes, data from unscheduled visits will not be summarized but will be listed only. The FAS will be used for all analyses of patient reported outcomes.

2.11.1 MSIS-29

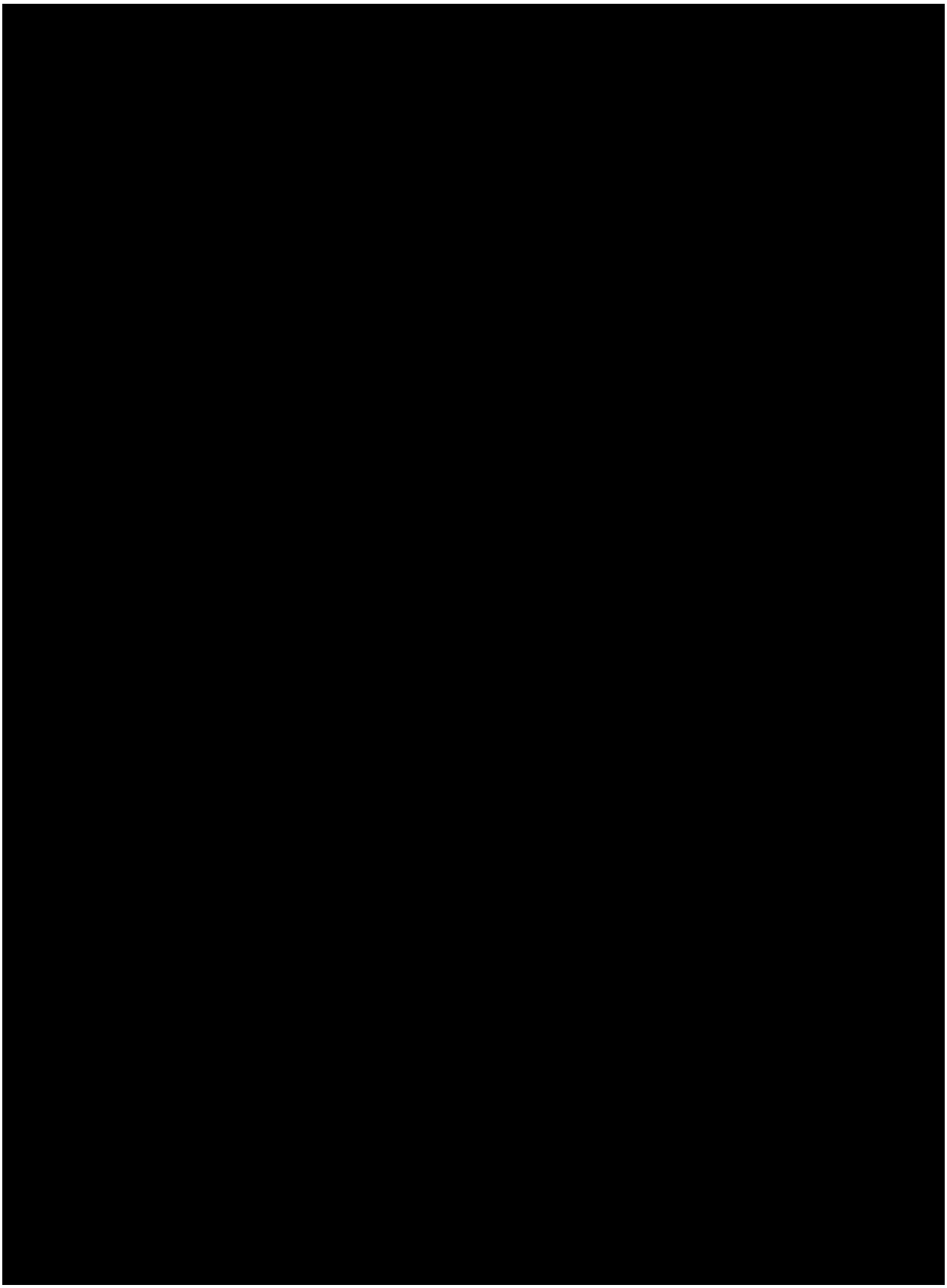
The Multiple Sclerosis Impact Scale version 2 (MSIS-29 v2) is a 29-item, self-administered questionnaire that includes two domains, physical and psychological. The questions in the scale ask the patient for their views about the impact of MS on their day-to-day life during the past two weeks. Each of the 29 items will be given a score between 1 (not at all) and 4 (extremely) (MSIS-29 v2 guidelines). The physical impact score is computed by summing up Items 1-20 (resulting in a score between 20 and 80). The psychological impact score is computed by summing up Items 21-29 (resulting in a score between 9 and 36).

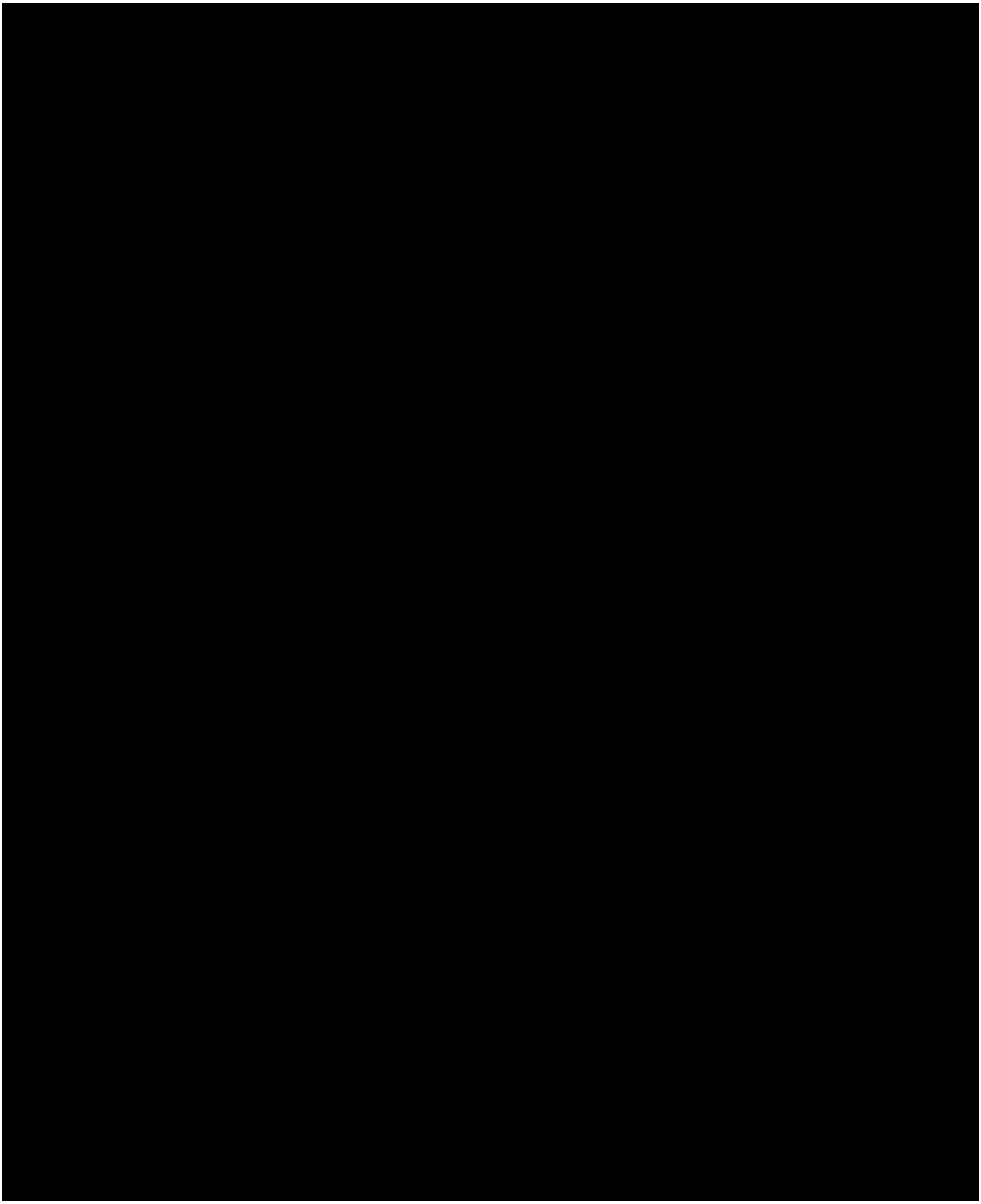
For ease of interpretation, these scores are transformed to a scale of 0-100 using the following formula: 100 x (observed score minus lowest possible score) / (highest possible score minus lowest possible score). The higher the score, the greater is the degree of disability.

In case of missing item scores, the following rule applies: if more than 50% (i.e., more than 10 for the physical scale and more than 4 for the psychological scale) of the item scores are missing, the scale score will not be calculated and set to missing. If no more than 50% of the item scores are missing, the scale score will be imputed as the average of the non-missing item scores multiplied by the number of items in the scale (20 for the physical scale and 9 for the psychological scale).

Both the MSIS-29 physical impact score and the psychological impact score and the corresponding changes from baseline will be summarized by visit-windows. The change from baseline at post-baseline visit-windows will be compared between treatment groups using a repeated measures mixed effects analysis adjusted for treatment, visit-windows (as class variable), region, and the corresponding baseline score.

[redacted]





2.14 Interim analysis

Two separate SAPs will be developed to document methodology details, one for analyses that will be performed for DMC (as mentioned in Section 2.14.2), the other for analyses that will be performed for blinded data reviews (as mentioned in Section 2.14.3).

2.14.1 Unblinding interim analysis

No unblinding interim efficacy analysis is planned for this study.

2.14.2 Analysis for Data Monitoring Committee (DMC)

Regular interim analyses on risk-benefit evaluation will be performed for the independent DMC by an independent team of statisticians and programmers who are not otherwise involved in the conduct of the study. The data review will be done by the DMC. The DMC will be independent of Novartis and the team involved in the study conduct.

2.14.3 Blinded data reviews

The purpose of the blinded data reviews is to ensure an adequate sample size and follow-up time to address the scientific objectives of the study without un-necessarily exposing patients by adapting design features such as sample size and follow-up duration as indicated by the accumulating data.

Blinded data reviews will be conducted in this study for:

(1) Blinded sample size re-estimation: relapse (and disability) related assumptions (primary endpoint) will be re-assessed based on a review of blinded data to adjust sample size if indicated by the accumulating data. This review will occur prior to the completion of enrolment.

(2) Declaration of EOS: data collected by the study on relapse rates and disability worsening events will be monitored. EOS will be declared only once when sufficient

information has been collected to address the study's primary and key secondary objectives.

Details of the statistical analysis plan with respect to blinded data reviews are provided in the BSSR SAP.

3 Sample size calculation

Sample size requirements for this study are primarily driven by the disability-related key-secondary endpoints. **A total of 900 patients will be randomized to study drug in a 1:1 ratio (450 per treatment arm)**. A second study of identical design with the same sample size will be conducted in parallel. Both studies are independently powered to address the primary endpoint (ARR) and all key-secondary MRI endpoints.

Key-secondary outcomes related to disability will be analyzed in the combined populations of both trials, provided the primary null-hypothesis can be rejected in both studies. Other secondary efficacy endpoints, as well as safety will be analyzed by study.

The total sample size of this study can be increased from 900 patients per study to a maximum of 1250 patients per study based on a blinded sample size re-estimation if either the blinded relapse rate or the event rate of 3mCDW is substantially lower than expected.

All sample size calculations were done in EAST 6, version 6.3, Cytel Inc.

Primary endpoint:

- **ARR:** The total sample size of 900 randomized patients for this trial is sufficient to achieve ≥90% power for the demonstration of superiority of ofatumumab over teriflunomide based on the primary endpoint (ARR).

Key-secondary endpoints:

The power mentioned for the key-secondary endpoints is conditional on the successful rejection of the null hypothesis for the primary endpoint (ARR), and the successful rejection of all key-secondary endpoints that are to be tested at a higher hierarchical level. Multiplicity adjustments and the testing procedure are defined in Section 2.5.3.

- **3mCDW:** A total sample size of 1800 patients across two studies of identical design is sufficient to provide ≥90% power and for the demonstration of superiority based on the 3mCDW.
- **6mCDW & 6mCDI:** A total sample size of 1800 patients across two studies of identical design is sufficient to provide ≥80% power for an analysis of 6mCDW and 6mCDI.
- **MRI endpoints:** A total sample size of 900 randomized patients for this trial is sufficient to provide ≥80% power for all MRI endpoints which are part of the testing procedure (Number of Gd-T1 lesions per scan, annualized rate of new or enlarging T2 lesions, annualized rate of BVL).
- **NfL**: A total sample size of 900 randomized patients for this trial is sufficient to provide a ≥ 90% power for an analysis of the NfL concentration in serum.

3.1.1 Sample size for the primary endpoint (ARR)

A negative binomial distribution of relapses is assumed for the primary analysis; this is a common assumption in MS. The demonstration of a relative reduction of the ARR in patients treated with ofatumumab (λ_{ofa}=0.168) compared with those treated with teriflunomide (λ_{ter}=0.28) by 40% (λ_{ofa} / λ_{ter} =0.6) with a **power of 90%** at a one-sided alpha-level of 0.025 in a study with 1.5 years follow-up and under the assumption of a dispersion parameter κ=0.82 requires a sample size of 322 completers per treatment arm (644 completers for the study). Allowing for 20% uninformative dropouts equally distributed across treatment arms, a total sample size of **805 randomized patients** is required for the study to demonstrate superiority of ofatumumab based on ARR. The same sample size, with the same assumptions would provide approximately 95% power for the demonstration of superiority of ofatumumab over teriflunomide at a one-sided alpha level of 0.000625 (=0.025^2) using the pooled data from two studies of identical design. The formula proposed by Keene et al. 2007 was used for the sample size calculation for the primary endpoint.

3.1.1.1 Justification of assumptions for the primary endpoint

In the pivotal studies, patients treated with teriflunomide 14 mg (the control treatment for this study) had an ARR=0.32 (Confavreux et al. 2014) and ARR=0.37 (O'Connor et al. 2011). In an additional phase 3 study of teriflunomide versus Interferon beta-1a, patients treated with teriflunomide 14 mg had an ARR=0.26 which was not significantly different from that observed in patients treated with Interferon beta-1a (ARR=0.22). For the purpose of this study an ARR= 0.28 is assumed for patients treated with teriflunomide 14 mg.

Ocrelizumab is an anti-CD20 monoclonal antibody with a similar mode of action as ofatumubab but is administered as infusion. In the ocrelizumb phase 3 trials an ARR of 0.156 (OPERA I) and 0.155 (OPERA II Hauser et al. 2015) were observed for patients treated with Ocrelizumab 600 mg, the corresponding interferon beta-1a controls had an ARR of 0.292 (OPERA I) and 0.290 (OPERA II); the corresponding relative reductions in ARR were 46% (OPERA I) and 47% (OPERA II). In the ocrelizumab phase 2 trial, patients treated with ocrelizumab 600 mg had an ARR=0.13, those treated with 1000 mg an ARR=0.17 (Kappos et al. 2011). Compared with the Interferon beta-1a control in the same trial (ARR=0.36), ocrelizumab 600 mg and 2000 mg reduced the ARR by 64% or 53%, respectively. Compared to placebo controls from the same trial (ARR=0.64), ocrelizumba 600 mg and 2000 mg reduced the ARR by 80% or 73%, respectively.

Ofatumumab 60 mg every 12 weeks administered subcutaneously showed in the phase 2 trial (unpublished data) approximately 34% relative reduction in relapse rate versus placebo over the first 24 weeks, but 60% in the second 12 weeks (week 12 to 24). The relative reduction of Gd-enhancing lesions of approximately 90% observed in patients treated with ofatumumab relative to those treated with placebo was similar to that observed with ocrelizumab and placebo, suggesting similarly strong anti-inflammatory potency of ofatumumab and ocrelizumab. Based on the combined findings, a relative reduction in the ARR of 40% is assumed between patients treated with ofatumumab compared to those treated with teriflunomide, which corresponds to an absolute ARR=0.168 for patients treated with ofatumumab, similar to that observed with ocrelizumab in phase 2 and 3.

The dispersion parameter of k=0.82 is assumed based on the values observed for relapse data in the fingolimod phase III program. In the pivotal teriflunomide phase 3 program 290/1086=27% of the patients discontinued over 104 weeks (O'Connor et al. 2011) in one trial, and 348/1196=30% over 130 weeks in another one (Confavreux et al. 2014). In OPERA I and OPERA II discontinuation rates were lower. For the purpose of the primary analysis we conservatively assume that 20% of the randomized patients will not contribute to the primary analysis at all. This is a conservative assumption because patients who discontinue prematurely from the study can contribute with partial data (relapse counts, exposure) to the primary analysis. In this information-based design patients will be followed until the end of the trial (see EOS criteria), rather than to be observed for a fixed time period. Based on the 3mCDW endpoint it is anticipated that the majority of patients will be exposed for 1 to 2 years. For ease of calculation a fixed follow-up time of 1.5 years is assumed for the primary endpoint. Of note, the power for the primary endpoint will depend on the actual follow-up. For instance, if all other assumptions were kept fixed but patients were followed for only 1 year instead of 1.5 years, the power in each cohort would drop to approximately 78%. If, on the other hand, patients were followed up for 2 years, the power for the primary endpoint would increase to approximately 95%.

3.1.2 Sample size the number of GdT1 lesions per scan

We assume a negative binomial distribution of the number of Gd-T1 lesions per scan. Assuming 0.26 Gd-T1 lesions per scan for teriflunomide and a 90% relative reduction with ofatumumab (i.e. 0.026 Gd-T1 lesions per scan for ofatumumab), and a dispersion parameter κ_{Gd}=5.3 (as observed in the combined FREEDOMS and FREEDOMS II studies for the same parameter, data on file), a total sample size of 210 patients with available MRI scans would be sufficient to provide **90% power** for the statistical test at a one-sided alpha 0.025. Allowing for 20% dropouts, a total sample size of **264 randomized patients** (132 per arm) is required for this endpoint.

3.1.3 Sample size for the annualized rate of new or enlarging T2 lesions

We assume a negative binomial distribution of the number of T2 lesions. Assuming 2.1 new or enlarging T2 lesions over 1.5 years for teriflunomide (1.4 new or enlarging T2 lesions per year with a mean follow-up of 1.5 years) and an 80% relative reduction with ofatumumab (i.e. 0.42 T2 lesions over 1.5 years; 0.28 lesions per year for ofatumumab), an average follow-up of 1.5 years, and a dispersion parameter κ_{Gd}=3.1 (as observed in the combined fingolimod Phase 3 placebo-controlled trials, data on file) for the same parameter), a total sample size of 66 patients with available MRI scans would be sufficient to provide **90% power** for the statistical test at a one-sided alpha 0.025. Allowing for 20% dropouts, a total sample size of **84 randomized patients** (42 per arm) is required for this endpoint.

3.1.4 Sample size for NfL

We assume a log-normal distribution for NfL concentrations. At month 3, geometric mean NfL concentrations of μ_1=25 pg/ml in teriflunomide and μ_2=17pg/ml in ofatumumab treated patients are assumed, leading to an assumed treatment difference on the log-scale of 0.386 pg/ml. We further assume a common standard deviation of 0.700 pg/ml. A total sample size of 139 randomized patients with available NfL assessments would be sufficient to provide 90% power

for the statistical test at a one-sided alpha 0.025. Allowing for 20% dropouts, a total sample size of 174 patients (87 per arm) is required for this endpoint.

3.1.5 Sample size for the annualized rate of brain volume loss (BVL)

We assume a normal distribution for the percentage brain volume change. Assuming a mean annualized rate of BVL of 0.45% on teriflunomide and 0.338% on ofatumubab (25% relative reduction) and a common standard deviation of 0.5%, a sample size of 621 patients with available MRI assessments would provide **80% power** at a one-sided alpha level of 0.025. Allowing from 20% dropouts, a total sample size of **778 randomized patients** are required for this endpoint (ca. 390 per arm).

3.1.6 Sample size for 3-month confirmed disability worsening (3mCDW)

For planning purposes only an exponential distribution of event times and proportional hazard is assumed. The cumulative event probability over two years is assumed as 15% vs 9.5% in patients treated with teriflunomide vs ofatumumab, respectively. The assumed cumulative event rates translate to a relative risk reduction of 38.6% (hazard ratio=0.614) in patients treated with ofatumumab compared with teriflunomide. The detection of a hazard ratio of 0.614 in a log-rank test with **90% power** at a one-sided alpha-level of 0.025-0.025[2] requires 178 qualifying events. Allowing for 20% dropouts over 2 years in both arms, a total sample size of **1774 randomized patients** is required. Assuming an accrual with 500 patients in the first 6 months, and 1500 patients per year thereafter, it is expected that accrual can be completed within approximately 18 months, and that the study duration is projected to last between 2.5 to 3 years.

For comparison to these assumptions, in the pivotal trials of teriflunomide, the cumulative probability of a 3-month confirmed disability worsening over 2 years in patients treated with teriflunomide 14 mg was 15.8% (Confavreux et al. 2014) and 20% (O'Connor et al. 2011), respectively. Furthermore, in combined ocrelizumab phase 3 studies (Hauser et al. 2015) the Kaplan-Meier estimates were 15.2 and 9.8 at year 2 for interferon beta-1a and ocrelizumab 600 mg, respectively; this corresponds to a hazard ratio 0.6 (40% relative risk reduction between ocrelizumab and interferon beta-1a).

3.1.7 Sample size for 6-month confirmed disability worsening (6mCDW)

The cumulative event probability over two years is assumed as 12% vs 7.548% in patients treated with teriflunomide vs ofatumumab, respectively. The assumed cumulative event rates translate to a relative risk reduction of 38.6% (hazard ratio=0.614) in patients treated with ofatumumab compared with teriflunomide. The detection of a hazard ratio of 0.614 in a log-rank test with **80% power** at a one-sided alpha-level of 0.025-0.025[2] requires 133 qualifying events. Allowing for 20% dropouts over 2 years in both arms, a total sample size of **1662 randomized patients** is required (831 per arm).

For comparison: In combined ocrelizumab Phase 3 studies (OPERA 1 and OPERA 2, Hauser et al. 2015) the Kaplan-Meier estimates were 12.0 and 7.4 at year 2 for interferon beta-1a and ocrelizumab 600 mg, respectively; this corresponds to a hazard ratio 0.6 (40% relative risk reduction between ocrelizumab and interferon beta-1a).

3.1.8 Sample size for 6-month confirmed disability improvement (6mCDI)

The cumulative event probability over two years is assumed as 12% vs 18.8% in patients treated with teriflunomide vs ofatumumab, respectively. The assumed cumulative event rates translate to a relative increase in chance by 63% (hazard ratio=1/0.614=0.163) in patients treated with ofatumumab compared with teriflunomide. The detection of a hazard ratio of 1/0.614 in a log-rank test with 80% power at a one-sided alpha-level of 0.025- 0.025^2 requires 133 qualifying events. Allowing for 20% dropouts over 2 years 1052 patients who contribute to the analysis are required; due to the definition of 6mCDI, patients with a baseline EDSS <2 cannot contribute to this analysis. Based on the combined fingolimod phase 3 trials (FREEDOMS, FREEDOMS II and TRANSFORMS, data on file) it is assumed that approximately 35% of all randomized patients will have baseline EDSS <2. Hence, **1620** (1052/0.65) **randomized patients** are required to provide **80% power.**

4 Change to protocol specified analyses

Changes to protocol specified analyses are in Section 2.6.2.5 NfL and Section 2.12 Biomarkers. All changes are listed below:

- Protocol (Section 9.5.1.2) specified analysis for NfL high vs. low subgroup on selected efficacy endpoints where brain volume was not included; the final analysis plan in NfL section includes the brain volume endpoint for this subgroup analysis.

[redacted]

All changes are made to be consistent with the NfL high vs. low subgroup analyses and/or to align with NfL pre-submission meeting briefing book submitted to FDA on Feb 8th 2019.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

Missing or partial dates are not allowed in completing the study drug administration CRF pages. The end date of study drug will be the last end date or last injection date for patients randomized to teriflunomide or ofatumumab, respectively.

5.1.2 AE date imputation

For the start and end dates of the adverse event records, when incomplete or missing, dates will be imputed according to Novartis standards (details will be given in programming datasets specifications (PDS) document).

5.1.3 Concomitant medication date imputation

For the start and end dates of the concomitant medication records, when incomplete or missing, dates will be imputed according to Novartis standards (details will be given in programming datasets specifications (PDS) document).

5.1.3.1 Prior therapies date imputation

Same as above.

5.1.3.2 Post therapies date imputation

Same as above.

5.1.3.3 Relapse date imputation

Missing or partial dates are not expected for the start and end dates of relapses on the MS relapses CRF pages. In case partial dates (unknown day with month and year available) exist in the final database, the following rules will apply:

- For start date, it will be imputed by the first day of the month or the first dose date if it occurs in the same month as the first dose date.
- For end date, it will be imputed by the last day of the month or truncated to have a duration of maximally 90 days.

5.1.3.4 Other imputations

For the calculation of duration or time since relevant history events as specified in Section 2.3.3 (MS disease baseline characteristics), partial dates will be imputed for the MS diagnosis start date, the first MS symptom date, the conversion to SPMS date, and the most recent relapse onset date via below imputation rules.

- If the year is missing or impossible (e.g. 12-Jan-1911), then the date will be imputed as missing.
- If the year is not missing and possible, but the month is impossible or missing (e.g. 17-XXX-2010), then the year will be kept and date will be imputed as July 1st (e.g., 1-July-2010).
- If the year and the month are not missing and possible, but the day is impossible or missing (e.g. 31-FEB-2009), then the year and month will be kept, and date will be imputed as 15th (e.g., 15-FEB-2009).
- The imputed dates should be prior to the screening visit date. That is, if imputed dates are on or after the screening visit date, the dates will be imputed to be one day before the screening visit date.

5.1.3.5 Data handling for relapses within 30 days of onset of previous relapses or relapses with duration beyond 90 days

According to protocol definition of MS relapses, the start date of a new relapse has to be at least 30 days after the start date of a previous relapse (i.e., start date of a new relapse – start date of a previous relapse >=30). If a relapse is recorded with a start date less than 30 days after the start date of a previous relapse, below data manipulation will be done to combine them into a single relapse by creating a new relapse record with the following information.

- Start date: take the earliest start date.
- End date: take the latest end date. If one of the end dates is missing, set it to missing.
- Date of EDSS intended to confirm the relapse:
    - Take the date of EDSS by which the relapse can be confirmed.
    - If more than one EDSS assessments meet the above criteria, take the date of EDSS by which the worst severity value is derived.
    - If no EDSS assessment meets the above criteria, take the earliest date of EDSS as captured on the relapse CRF page.
- Severity: take the value representing the worst case (severe>moderate>mild>missing)
- "Did the relapse affect daily activities?", "Hospitalization?", "Steroid used?", "Recovery status": for each of these characteristics, take the value representing the worst case (yes>no for first 3 questions; no>partial>complete recovery for last question).

According to CRF completion guidelines, maximum duration of a relapse is 90 days. If a relapse is recorded with a duration longer than 90 days, the end date will be truncated to have a duration of exactly 90 days. This applies also to the artificial record created by above procedure. Missing end date of relapse is not allowed. In the rare cases that missing end date exist in the final database, it will be imputed so that the duration of relapse is exactly 90 days.

5.2 AEs coding/grading

AEs are coded using the Medical dictionary for regulatory activities (MedDRA) terminology. The latest MedDRA version will be used and will be described in the footnote of relevant outputs.

AEs will be assessed according to the Common Terminology Criteria for Adverse Events (CTCAE).

The CTCAE represents a comprehensive grading system for reporting the acute and late effects of cancer treatments. CTCAE grading is by definition a 5-point scale generally corresponding to mild, moderate, severe, life threatening, and death. This grading system inherently places a value on the importance of an event, although there is not necessarily proportionality among grades (a grade 2 is not necessarily twice as bad as a grade 1). The CTCAE grade of 5 (death) is not used; rather, 'fatal' is collected as AE outcome and death information is also collected on a separate (e)CRF page.

5.3 Laboratory parameters derivations

For each patient, the estimated creatinine clearance values (without collecting urine) will be calculated using the Cockcroft-Gault formula (as specified in Table 5-1). In these calculations, the body weight is the last measurement collected on or before the day when the patient takes the laboratory test and age should also be calculated based on the time when the patient takes the laboratory test.

If the creatinine value is collected in the unit *µmol/L* (SI unit), it will be converted to *mg/dL* in order to use the formulas. The conversion is via the equation below:

- mg/dL = 88.4 µmol/L (e.g., creatinine = 2.0 mg/dL = 176.8 µmol/L).

Table 5-1 Creatinine clearance calculation

Variable	Formula
Creatinine clearance [mL/min] using Cockcroft-Gault formula (Cockcroft and Gault 1976)	= $(140 - A) \times W \ / (72 \times C) \times G$ Where *A* is age [years] *W* is body weight [kg] *C* is the serum concentration of creatinine [mg/dL] *G* is a constant: *G*=1 for males and *G*=0.85 for females.

The estimated creatinine clearance will be included as one of the laboratory parameters.

5.3.1 Laboratory test groups and subgroups

On presenting lab results, grouping parameters by family will ease the review. Table below shows a possible set of lab parameters and their corresponding classification.

Table 5-2 Laboratory tests

Order	Laboratory Group Subgroups	Tests [SI unit]
1	**Hematology**	
	Red Blood Cells	**Hematocrit [%]** **Hemoglobin [g/L]** **Platelets [10E9/L]** **Red cell count [10E12/L]**
	White Blood Cell Differential	**Absolute Basophils [10E9/L]** **Absolute Eosinophils [10E9/L]** **Absolute Lymphocytes [10E9/L]** **Absolute Monocytes [10E9/L]** **Absolute Neutrophils [10E9/L]** **Basophils [%]** **Eosinophils [%]** **Lymphocytes [%]** **Monocytes [%]** **Neutrophils [%]** **White Cell Count [10E9/L]**

Order	Laboratory Group Subgroups	Tests [SI unit]
	B-cells	**CD19+ B-cell counts[cells/µL]**
	Immunoglobulin	**Total IgG [g/L]** **Total IgM [g/L]**
2	**Chemistry**	
	Renal Function	**Creatinine [umol/L]** **Blood urea nitrogen [mmol/L]**
	Liver Function Tests	**ALT [U/L]** **Albumin [g/L]** **Alkaline Phosphatase [U/L]** **AST [U/L]** **Bilirubin (direct/conjugated) [umol/L]** **GGT [U/L]** **Total Bilirubin [umol/L]** **Total protein [g/L]**
	Other Enzymes	**Amylase [U/L]**
	Lipids	**Cholesterol HDL [mmol/L]** **Cholesterol LDL [mmol/L]** **Triglycerides [mmol/L]** **Total Cholesterol [mmol/L]**
	Other	**Random glucose [mmol/L]** **C-Reactive protein (CRP) [mg/L]**
	Electrolytes / Metabolism Tests	**Bicarbonate [mmol/L]** **Calcium [mmol/L]** **Chloride [mmol/L]** **Magnesium [mmol/L]** **Phosphate [mmol/L]** **Potassium [mmol/L]** **Sodium [mmol/L]**
3	**Urinalysis**	**Blood** **Glucose** **Specific gravity** **Albumin** **Protein**

5.3.2 Newly occurring liver enzymes abnormalities

Below lists the criteria for "events" of newly occurring liver enzymes abnormalities:

- ALT > 3, 5, 10, 20x ULN
- ALT or AST > 3, 5, 8, 10, 20x ULN
- ALT or AST > 3x ULN & TBIL > 1.5x ULN
- ALT or AST > 3x ULN & TBIL > 2x ULN
- ALP > 1.5, 2, 5x ULN
- TBIL > 1, 1.5, 2x ULN
- ALP > 3, 5x ULN & TBL > 2x ULN
- ALT or AST > 3x ULN & TBIL > 2x ULN & ALP ≤ 2x ULN
- ALT or AST > 3x ULN & (nausea or vomiting or fatigue or general malaise or abdominal pain or (rash and eosinophilia))

When a criterion contains multiple laboratory parameters (e.g., ALT > 3xULN & TBL > 2xULN), unless otherwise requested by the project clinical team/Brand Safety Leader (BSL), the criterion should be only considered to be met when the elevation in both parameters occurs on the same sample day (as evidenced by the same date that the lab samples were taken).

The "events" are defined in the Novartis safety guideline on hepatotoxicity (Novartis: Philippe Close 2011), section: Safety parameters for special liver event analyses.

5.3.2.1 Definition of characteristics for liver event overview summary

Characteristics: Liver events are categorized based on ALT and ALP measurements using the closest lab assessment +/- 7 days from the onset of the liver event into the following characteristics 'Hepatocellular' (ALT > 2xULN or ALT/ULN:ALP/ULN > 5), 'Cholestatic' (ALP > 2xULN or ALT/ULN:ALP/ULN ≤ 2), 'Mixed' (ALT > 2xULN and ALP > ULN and 2 < ALT/ULN:ALP/ULN ≤ 5), 'None' (if none of the three above qualifies), and 'Unknown' (in case of a missing ALT or ALP values). Note that the categories are not mutually exclusive. The definition is consistent with Novartis safety guideline on hepatotoxicity.

5.3.3 CTCAE grades for laboratory parameters

Table 5-3 CTCAE grades for laboratory parameters (CTCAE Version 5)

		Grade			
Abnormality	**Lab parameter**	**1**	**2**	**3**	**4**
Hematology					
Anemia	Hemoglobin (g/L)	<LLN - 100 g/L	<100 - 80 g/L	<80 g/L transfusion indicated	

		Grade			
Abnormality	**Lab parameter**	**1**	**2**	**3**	**4**
Platelet count decreased	Platelets (thrombocytes) (10^9/L)	<LLN-75.0 x 10^9/L	<75.0 - 50.0 x 10^9/L	<50.0 - 25.0 x 10^9/L	<25.0 x 10^9/L
White blood cell decreased	Leukocytes (WBCs) (10^9/L)	<LLN - 3.0 x 10^9/L	<3.0 - 2.0 x10^9/L	<2.0 - 1.0 x 10^9/L	<1.0 x 10^9/L
Neutrophil count decreased	Absolute neutrophil count (10^9/L)	<LLN - 1.5 x10^9/L	<1.5 - 1.0 x10^9/L	<1.0 - 0.5 x10^9/L	<0.5 x10^9/L
Lymphocyte count decreased	Absolute lymphocyte count (10^9/L)	<LLN x 0.8 - 10^9/L	<0.8 - 0.5 x 10^9/L	<0.5 - 0.2 x 10^9/L	<0.2 x 10^9/L
Lymphocyte count increased	Absolute lymphocyte count (10^9/L)		>4 - 20 x 10^9/L	> 20 x 10^9/L	
Chemistry					
Liver function					
Alanine aminotransferase increased	ALT (SGPT) (U/L)	>ULN - 3.0 x ULN if baseline was normal; >1.5 – 3.0 x baseline if baseline was abnormal	>3.0 - 5.0 x ULN if baseline was normal; >3.0 – 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 – 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal
Aspartate aminotransferase increased	AST (SGOT) (U/L)	>ULN - 3.0 x ULN if baseline was normal; 1.5 – 3.0 x baseline if baseline was abnormal	>3.0 - 5.0 x ULN if baseline was normal; >3.0 – 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 – 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal
Blood bilirubin increased	Bilirubin (µmol/L)	>ULN - 1.5 x ULN if baseline was normal; >1.0 – 1.5 x baseline if baseline was abnormal	>1.5 - 3.0 x ULN if baseline was normal; >1.5 – 3.0 x baseline if baseline was abnormal	>3.0 - 10.0 x ULN if baseline was normal; >3.0 – 10.0 x baseline if baseline was abnormal	>10.0 x ULN if baseline was normal; >10.0 x baseline if baseline was abnormal
GGT increased	Gamma-glutamyl transferase (GGT) (U/L)	>ULN - 2.5 x ULN if baseline was normal; 2.0 – 2.5 x baseline if baseline was abnormal	>2.5 - 5.0 x ULN if baseline was normal; >2.5 – 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 – 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal
Alkaline phosphatase increased	Alkaline Phosphatase	>ULN - 2.5 x ULN if baseline	>2.5 - 5.0 x ULN if baseline was	>5.0 - 20.0 x ULN if baseline was	>20.0 x ULN if baseline was

		Grade			
Abnormality	**Lab parameter**	**1**	**2**	**3**	**4**
	(U/L)	was normal; 2.0 – 2.5 x baseline if baseline was abnormal	normal; >2.5 – 5.0 x baseline if baseline was abnormal	normal; >5.0 – 20.0 x baseline if baseline was abnormal	normal; >20.0 x baseline if baseline was abnormal
Renal function Note: A semi-colon (;) indicates 'or' within the description of the grade.					
Creatinine increased*	Creatinine (µmol/L)	>ULN - 1.5 x ULN	>1.5 - 3.0 x baseline; >1.5 - 3.0x ULN	>3.0 baseline; >3.0 - 6.0 x ULN	>6.0 x ULN
Other enzymes					
Serum amylase increased	Amylase (U/L)	>ULN - 1.5 x ULN	>1.5 - 2.0 x ULN; >2.0 - 5.0 x ULN and asymptomatic	>2.0 - 5.0 x ULN with signs or symptoms; >5.0 x ULN and asymptomatic	>5.0 x ULNand with signs or symptoms
Serum amylase** increased	Amylase (U/L)	>ULN - 1.5 x ULN	>1.5 - 2.0 x ULN;	>2.0 - 5.0 x ULN	>5.0 x ULN
Lipids					
Cholesterol high	Cholesterol Total (mmol/L)	>ULN- 7.75 mmol/L	>7.75-10.34 mmol/L	>10.34-12.92 mmol/L	>12.92 mmol/L
Hypertriglyceridemia	Triglycerides (mmol/L)	1.71 - 3.42 mmol/L	>3.42 - 5.7 mmol/L	>5.7 - 11.4 mmol/L	>11.4 mmol/L

*Highest grade will be assigned if more than one grade criteria are met for an observed value of a patient.

**CTCAE 4.03 definition. For reporting in CSR, the grade definition per CTCAE 4.03 will be used for this parameter only.

5.4 Clinical notable vital signs

Clinically notable vital sign values to be used for this study are listed below.

Table 5-4 Vital signs clinically notable values

Vital Sign	Notable criteria
Pulse (beats/min, bpn)	>120 bpm Or <50 bpm
Systolic Blood Pressure (mmHg)	≥160 mmHg Or ≤90 mmHg

Vital Sign	Notable criteria
Diastolic Blood Pressure	≥100 mmHg Or ≤50 mmHg
Temperature (°C)	>38.3 °C (>101 °F)
Body weight (kg)	≥7% from baseline weight

5.5 Statistical models

5.5.1 Primary analysis

The null hypothesis is that there is no difference in the ARR between ofatumumab 20mg sc once monthly and teriflunomide 14 mg po once daily in reducing the frequency of confirmed MS relapses as measured by ARR. The alternative hypothesis is that there is a difference between the two treatment groups. H_0: $ARR_o / ARR_t = 1$ vs. H_a: $ARR_o / ARR_t < 1$. Where ARR_o and ARR_t are ARR of ofatumumab and ARR of teriflunomide respectively.

The null hypothesis will be tested using a negative binomial regression model with log-link, treatment and region as factors, number of relapses in previous year, baseline EDSS, baseline number of T1 Gd-enhancing lesions and the patient's age at baseline as covariates.

The SAS procedure GENMOD will be used to conduct the analysis. In GENMOD, the log of the dispersion parameter will be used (lognb) as an option in model statement. The natural log of time in year is used as an offset by specifying offset option in the model statement.

5.5.2 Key secondary analysis

For disability related key secondary endpoints, the Cox proportional hazard model will be the key secondary analysis to test H_0: $\lambda_o/\lambda_t = 1$ vs. H_a: $\lambda_o/\lambda_t < 1$. Where λ_o and λ_t are hazard of disability events in ofatumumab and teriflunomide treatment groups respectively. A stratified Cox proportional hazards model with study as stratum, treatment, and region as factors and baseline EDSS at baseline as a continuous covariate. The model will contain a treatment-by-study interaction.

The Cox proportional hazard (PH) model assumes that the effect of each covariate is the same at all points in time. To check the PH assumption for the treatment covariate, a plot of log-log survivor function vs. time for each treatment group will be used. If the hazards are proportional, the log-log survivor functions should be parallel. Approximate parallelism between the curves for the treatment groups will provide supportive evidence of the proportional hazards assumption.

If the assumption of proportional hazards is questionable based on graphical analyses, then the interpretation of the estimated hazard ratio will need to be done with caution.

The SAS procedure PHREG will be used to conduct the Cox PH analysis. The SAS procedure LIFETEST will be used to conduct the KM analysis.

For the key secondary endpoints related to Gd lesions and T2 lesions, mathematical details are similar to the ones for primary analysis described in Section 5.5.1 as the negative binomial models will be used.

For the key secondary endpoint related to brain volume change, a random coefficients model will be used to test H_0: $\beta_o - \beta_t = 0$ vs. H_a: $\beta_o < \beta_t$. Where β_o and β_t are annual rate of percent change from baseline in brain volume in ofatumumab and teriflunomide treatment groups respectively. The random coefficients model will include treatment, region as fixed effects (factors), and time, number of Gd-enhancing lesions at baseline, baseline T2 volume, and normalized brain volume at baseline as continuous covariates. Time as a continuous covariate allows for estimation of different slopes and intercepts among treatment groups. The model will also contain random terms to account for deviations about the population slope and intercept. The annual rate of percent change from baseline in brain volume is approximates as the population slope within the treatment group.

Percent brain volume change was approximately linear over time and approximately normally distributed in 3 independent studies of the fingolimod phase 3 program. Data collected from this study will be checked descriptively by plotting the population mean percent change from baseline against time before database lock in a blinded fashion (pooling all patients).

The SAS procedure MIXED will be used to conduct the analysis. Unstructured covariance will be assumed.

5.5.3 Important supportive analysis of primary or key secondary endpoints

5.5.3.1 Supportive analysis of primary endpoint:

To estimate the ARR and the treatment effect during the initial "onset of action" period of 8 week (≤56 days=8*7 days), and the ARR and the treatment effect thereafter (>56 days; long-term efficacy) a sensitivity analysis will be conducted on the FAS. The number of confirmed relapses for the i^{th} patient from treatment group j who spent t_i years in the study $N_{ij}(t_i)$ is assumed to follow a piecewise constant negative binomial distribution with different ARR before and after day 56 but with a constant dispersion k.

$$N_{ij}(t_i) = N_{ij}(t_i^1) + N_{ij}(t_i^2) \quad where \quad \begin{cases} N_{ij}(t_i^1) \sim NegBin(\lambda_j^1 * t_i^1, k), \quad t_i^1 \le 56 \\ N_{ij}(t_i^2) \sim NegBin(\lambda_j^2 * t_i^2, k), \quad t_i^2 = t_i - 56 \quad for \quad t_i > 56 \end{cases}$$

In the above notation, t_i^1 and t_i^2 are the times spent in the study before and after day 56 respectively. The total time in study $t_i = t_i^1 + t_i^2$.

5.5.3.1.1 Implementation in SAS

The piecewise constant negative binomial regression model can be re-written as a Poisson-Gamma mixture model which can be implemented in SAS via NLMIXED procedure. Specifically, let $N_{ij}(t_i)|U_i$ denote a Poisson distributed random variable where U_i is a Gamma distributed random variable with mean 1 and variance $1/\alpha=k$. The model assumed for the mean of $N_{ij}(t_i)|U_i$ is expressed by time period, before or after day 56 respectively:

$$\begin{cases} E(N_{ij}(t_i^1)\,|\,U_i) = \lambda_1^1 \cdot t_i^1 \cdot \exp(\theta^1 \cdot trt_j + x_i^T\gamma)\cdot U_i, & t_i^1 \le 56 \ days \\ E(N_{ij}(t_i^2)\,|\,U_i) = \lambda_1^2 \cdot t_i^2 \cdot \exp(\theta^2 \cdot trt_j + x_i^T\gamma)\cdot U_i, & t_i^2 = t_i - t_i^1 \ and \ t_i > 56 \ days \end{cases}$$

Where x_i^T denotes the vector of covariates for subject i, and trt_j is an indicator variable which is one for all patients assigned to ofatumumab and zero for those assigned to teriflunomid, $\exp(\gamma)$ quantifies the effect of the covariates. Thus, λ_1^1 and λ_1^2 represent the ARR for teriflunomid up to day 56 and thereafter, respectively. The treatment effect (ARR-ratio) is $\exp(\theta^1)$ during the first 56 days, and $\exp(\theta^2)$ thereafer. Hence, the ARR on ofatumumb is $\lambda_2^1 = \lambda_1^1 \exp(\theta^1)$ during the first 56 days, and $\lambda_2^2 = \lambda_1^2 \exp(\theta^2)$ thereafter.

In case of non-convergence of this model, all covariates will be dropped from the model.

As the annualized relapse rate is to be estimated, time will be divided by 365.25 (as in year scale) for implementation in SAS. The data set will be prepared in such a way that there will be 2 records, one for period ≤56 days and the other for period >56 days, for patients whose actual time in study is greater than 56 days. For period ≤56 days, the number of confirmed relapses with relapse starting date on or before day 56 will be calculated and the time in study variable will be set to 56/365.25. For period >56 days, the number of confirmed relapses with relapse starting date after day 56 will be calculated and the time in study variable will be set to (value of time in study overall-56 then divided by 365.25). The time period indicator variable will be populated for both records appropriately (e.g., 0 for period ≤56 days and 1 for period >56 days). For patients whose actual time in study variable is ≤56 days, the observed record (i.e., same as the overall number of confirmed relapses and time in study) will be populated with time period indicator as ≤56 days. Table 5-5 provides an example of the data set where other covariates are to be added accordingly.

Table 5-5 Key data structure for piecewise NB model

Patient	Treatment	Treatment code	Period code	Time in study (total)	Time in study (by period)	No. of confirmed relapses (total)	No. of confirmed relapses (by period)
1	OFA20	1	0	100/365.25	56/365.25	3	1
1	OFA20	1	1	100/365.25	44/365.25	3	2
2	TER14	0	0	45/365.26	45/365.26	2	2

Additional patient level variables will be derived to be used as covariates in this model.

Continuous covariates will be centered by the mean. For example, age will be centered by subtracting the mean of age as (age – mean(age)). The mean will be calculated based on all patients in the study.

Factors will be dummy coded. There are 3 factors in the model, i.e. treatment, period, and region. Treatment code and Period code as shown in Table 5-5 can be considered as dummy

coded variables as each factor has 2 levels only where the reference level will have a value of 0. Region will have 5 levels (as in Section 5.8) so that 4 dummy coded region variables will be derived. The reference region will be the "North America and AUS" region. The rest 4 regions will each have a dummy coded variable with value of 1 if patients are in this region or 0 if patients are not in this region. Patients in the reference region will have 0 values in all 4 region dummy coded variables.

5.5.3.2 Supportive analysis of key secondary endpoints (3mCDW and 6mCDW):

To estimate the hazard ratio between treatment groups during the initial "onset of action" period of 8 weeks (≤56 days=8*7 days), and thereafter (>56 days; long-term efficacy) a sensitivity analysis will be conducted for the 3mCDW and 6mCDW. A stratified Cox proportional hazards model simililar to that for the main analysis will be used with study and time period indicator as strata, treatment and region as factors and baseline EDSS as a continuous covariate. The model will also contain an interaction term of time period indicator by treatment.

The SAS procedure PHREG will be used to conduct the analysis.

The data set for Cox's PH model will include a time to event variable and event status variable (0 as censored) as usual. To fit the above mentioned stratified Cox's PH model, an artificial record will be created for patients whose actual time to event variable is >56 days. This artificial record will set the time to event variable equal to 56 days and event status variable as censored. A time period indicator variable will be populated for all records as 0 if the time to event variable ≤56 days and 1 if the time to event variable >56 days. Table 5-6 provides an example of the data set where other covariates are to be added accordingly. **Bolded** records are artificial records.

Table 5-6 Key data structure for stratified Cox's PH model

Patient	Treatment	Treatment code	Period code	Time to event	Event status
1	**OFA20**	**1**	**0**	**56**	**0**
1	OFA20	1	1	100	1
2	TER14	0	0	45	1
3	**OFA20**	**1**	**0**	**56**	**0**
3	OFA20	1	1	155	0

5.6 Rule of exclusion criteria of analysis sets

Patient classification in the analysis sets is entirely based on protocol deviation and non-protocol deviation criteria. Details are provided in Table 5-7 and Table 5-8.

Table 5-7 Protocol deviations that cause patients to be excluded

Deviation ID	Description of Deviation	Exclusion in Analysis sets
EXCL01	Patients with primary progressive MS or non-relapsing/no MRI activity-SPMS	Excluded from PPS
EXCL02	Patients meeting criteria for neuromyelitis optica	Excluded from PPS
EXCL03	Disease duration of more than 10 years in patients with EDSS score of 2 or less	Excluded from PPS
EXCL08	Patients with an active chronic disease (or stable but treated) of the immune system other than MS or with a known immunodeficiency syndrome	Excluded from PPS
EXCL19	Teriflunomide (if discontinued for reasons related to safety or lack of efficacy) at any time prior to randomization	Excluded from PPS
EXCL25B	Patients currently treated with leflunomide	Excluded from PPS
EXCL49	History of clinically significant CNS disease (e.g. stroke, traumatic brain or spinal injury, history or presence of myelopathy) or neurological disorders which may mimic MS	Excluded from PPS
EXCL52	History of hypersensitivity to any of the study drugs or excipients (including lactose intolerance) or to drugs of similar chemical classes	Excluded from PPS
INCL01	**Informed consent obtained = 'No' or missing**	Excluded from all analysis sets
INCL03	No diagnosis of MS as defined by 2010 revised McDonald criteria	Excluded from PPS
INCL06	Neurologically stable within 1 month prior to randomization	Excluded from PPS
WITH11	Patient diagnosed with PML and study treatment was not discontinued	Excluded from PPS
TRT01	Patient received incorrect study drug	Excluded from PPS
TRT02	Patient received damaged or expired study drug	Excluded from PPS
COMD01	Immunosuppressive/chemotherapeutic medications or procedures, including	Excluded from PPS

Deviation ID	Description of Deviation	Exclusion in Analysis sets
	cyclosporine, azathioprine, methotrexate, cyclophosphamide, mitoxantrone, lymphoid irradiation and hematopoietic stem cell transplantation while taking double-blind treatment	
COMD02	Monoclonal antibodies targeting the immune system, including but not limited to natalizumab, alemtuzumab, and B-cell depleting agents such as rituximab, ocrelizumab and obinutuzumab while taking double-blind treatment	Excluded from PPS
COMD03	Any other immunomodulatory or disease- modifying MS treatment, including but not limited to, fingolimod, interferon beta, glatiramer acetate, dimethyl fumarate or systemic corticosteroids (except when given for MS Relapse treatment) while taking double-blind treatment	Excluded from PPS
COMD04	Leflunomide while taking double-blind treatment	Excluded from PPS
COMD06	Daclizumab while taking double-blind treatment	Excluded from PPS
OTH01	EDSS rater acted as a treating physician or treating physician acted as EDSS rater	Excluded from PPS
OTH02	Not following per protocol blinding procedures such that the integrity of the study is compromised	Excluded from PPS

Table 5-8 Patient Classification

Analysis Set	PD ID that cause patients to be excluded	Non-PD criteria that cause patients to be excluded
SCR	NA	NA
FAS	INCL01	Not assigned a valid randomization number Mis-randomized patients who didn't take study drug
PPS	As listed in Table 5.7	Not in FAS;

Analysis Set	PD ID that cause patients to be excluded	Non-PD criteria that cause patients to be excluded
		Compliance to study drug administration < 80%
		EDSS baseline value >=7.0
SAF	INCL01	No double-blind study drug taken

5.7 Statistical method details for TEAEs

5.7.1 Odds ratios and 95% confidence interval

For an investigational drug group with n_1 patients at risk, independent from the control group with n_0 patients at risk, of whom x_1 and x_0 experience a certain event with probability π_1 and π_0 respectively, the odds ratio (OR) is estimated as

$\frac{p_1/(1-p_1)}{p_0/(1-p_0)}$ with $p_1 = x_1/n_1$ and $p_0 = x_0/n_0$, simplifying to $\frac{x_1 \times (n_0 - x_0)}{x_0 \times (n_1 - x_1)}$.

The OR and a conditional exact 95% confidence interval will be obtained by using the SAS procedure PROC FREQ with statement EXACT OR. The data set will include a patient level response variable indicating whether a patient has at least one AE of a specific type (1=yes, 0=no).

By default SAS sets the OR to missing if the odds for one of the treatments cannot be calculated (no infinity in SAS). This is problematic as infrequent but very serious adverse events that happen exclusively on active treatment can easily be missed if the OR for such highly imbalanced advents is set to missing and sorted last. To ensure that potentially important imbalances appear at the top of a table if it is sorted by OR, the following additional rules are applied (considering the numerator and the denominator of the OR that would lead to divisions by zero):

- NO CASE CLAUSE [0/0]: If ($x_1=0$ and $x_0=0$) or ($x_1=n_1$ and $x_0=n_0$), i.e. there are either no events in both treatment groups, or all patients in both treatment groups had the event, then the OR cannot be estimated. The OR can be left missing for the sorting, but should be displayed on the output as "N.E." indicating not estimable (if it has to be shown on the table, in this case it needs to be explained in the footnote).
- SAFETY CLAUSE[~/0]: If ($x_0=0$ and $x_1\neq 0$) or ($x_0\neq n_0$ and $x_1=n_1$), i.e. there are no events on control, but some on investigational drug, or everyone on investigational drug experienced the event, but not everyone on control, then OR will be imputed as 100000 for the sorting (so that it comes up top if sorted by the OR). On the output present the OR as ">100".
- EFFICACY CLAUSE[0/~]: If ($x_1=0$ and $x_0 \neq 0$) or ($x_1\neq n_1$ and $x_0 = n_0$), i.e. there are no events on investigational drug, but some on control, or, everyone on control experienced the event, but not everyone on investigational drug, then OR will be imputed as

1/100000 for the sorting (so that it comes prior to the missing ones). On the output present the OR as "<1/100".

5.7.2 Exposure adjusted incidence rate and 95% confidence interval

It will be assumed that for each of n patients in a clinical trial the time t_j (j=1,...,n) to the first incidence of a certain event is observed, or if the event was not experienced, the time is censored at the end of the observation period. The sequence of first occurrences of an event will be modeled to follow approximately a Poisson process θ with constant intensity λ. The rate parameter λ will be estimated by $\hat{\lambda}=\frac{D}{T}$, where $T=\sum_{j=1}^{n} t_j$ and D is the number of patients with at least one event. Conditionally on T, an exact 95% confidence interval for a Poisson variable with parameter θT and observed value D can be obtained based on (Garwood, 1936), from which an exact 95% confidence interval for $\hat{\lambda}=\frac{D}{T}$ will be derived as follows (Sahai, 1993; Ulm, 1990):

Lower confidence limit $L=\frac{0.5c_{\alpha/2,2D}}{T}$ for D>0, 0 otherwise,

Upper confidence limit $U=\frac{0.5c_{1-\alpha/2,2D+2}}{T}$

where $c_{\alpha,k}$ is the α^{th} quantile of the Chi-square distribution with k degrees of freedom.

The SAS procedure GENMOD will be used to fit a Poisson regression model with only treatment as factor, the log-link and natural logarithm of time as the offset variable. The incidence rate (and 95% confidence intervals) are obtained as the exponentiated estimates by treatment (e.g. obtained by statement LSMEANS with option EXP). The data set will include a patient level response variable indicating whether the patient experienced at least one event of a specific type (1=yes, 0=no) and a patient level time variable which is derived as follows:

- For patients who had at least one event of that type: time is calculated as (start day of the first event - date of first administration of study drug+1).
- For patients who did not have an event of that type: time at risk as defined in Section 2.4.1 will be used.

The time will be divided by 365.25 before taking the logarithm transformation as offset variable so that the estimated incidence rate corresponds to an annualized rate.

5.8 Regions for randomization stratification

Regions	G2301	G2302
Western Europe		
Austria		x
Belgium	x	x
Denmark	x	

Finland		x
France	x	x
Germany	x	x
Greece	x	
Italy	x	x
Netherlands	x	
Norway		x
Portugal	x	x
Spain	x	x
Sweden	x	
Switzerland	x	x
United Kingdom	x	x
Eastern Europe		
Bulgaria	x	x
Croatia	x	x
Czech Republic	x	x
Estonia	x	
Hungary	x	x
Latvia		x
Lithuania		x
Poland	x	x
Romania	x	x
Slovakia	x	x
Turkey	x	x
North America and AUS		
Australia	x	x
Canada	x	x
United States	x	x
Asia Pacific		
India	x	x
South Korea	x	
Taiwan		x
Thailand	x	
Latin America		
Argentina	x	x
Brazil		x

Chile	x	
Mexico	x	x
Peru		x
Others		
Egypt		x
Israel	x	
Kuwait	x	
Lebanon		x
Saudi Arabia	x	
Russia Federation	x	x
South Africa		x
United Arab Emirates		x

6 References

Cockcroft DW, Gault MJ (1976) Prediction of creatinine clearance from serum creatinine. Nephron; 16:31-41

Confavreux C, O'Connor P, Comi G, et al. (2014) Oral teriflunomide for patients with relapsing multiple sclerosis: a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Neurol: Mar;13(3):247-56

Garwood F. (1936) Fiducial Limits for the Poisson Distribution. Biometrika 28, 437-442.

Gassmann-Mayer et al. (2011) Clinical and statistical assessment of suicidal ideation and behavior in pharmaceutical trials, Clinical Pharmacology & Therapeutics 90 4, 554–560.

Hauser SL, Comi GC, Hartung HP et al. (2015) Efficacy and safety of ocrelizumab in relapsing multiple sclerosis - results of the interferon-beta-1a-controlled, double-blind, Phase III OPERA I and II studies. ECTRIMS Online Library. Hauser S. Oct 9, 2015; 116634

Kappos L, Li D, Calabresi PA et al. (2011) Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomized, placebo-controlled, mutlicentre trial. Lancet; 378: 1779-87

Keene ON, Jones MRK, Lane PW, Anderson J. (2007) Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. Pharmaceutical Statistics; 6: 89-97

O'Connor P, Wolinsky JS, Confavreux C et al. (2011) Randomized Trial of Oral Teriflunomide for Relapsing Multiple Sclerosis. N Engl J Med; 365:1293-303

Panitch H, Goodin DS, Francis G, et al. (2002) Randomized, comparative study of interferon beta-1a treatment regimens in MS. The EVIDENCE trial. Neurology. 2002;59:1496–506.

Novartis: Philippe Close (2011): *Hepatotoxicity: Clinical Development Safety Guideline.*

Novartis: COMB157G SAP for BSSR: statisticial analysis plan for blinded sample size re-estimation

Sahai H, Khurshid A (1993). Confidence intervals for the ratio of two Poisson means. The Mathematical Scientist, 18, 43–50.

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