Official Title: A Randomized, Double-Blind, Double-Dummy, Parallel-Group Study to Evaluate the Efficacy And Safety of Ocrelizumab in Comparison to Interferon Beta-1a (Rebif®) in Patients With Relapsing Multiple Sclerosis

NCT Number: NCT01412333

Document Date: SAP Version 5: 20 March 2023

STATISTICAL ANALYSIS PLAN

TITLE:	**A RANDOMIZED, DOUBLE-BLIND, DOUBLE-DUMMY, PARALLEL-GROUP STUDY TO EVALUATE THE EFFICACY AND SAFETY OF OCRELIZUMAB IN COMPARISON TO INTERFERON BETA-1A (REBIF®) IN PATIENTS WITH RELAPSING MULTIPLE SCLEROSIS**
PROTOCOL NUMBERS:	WA21092 and WA21093
STUDY DRUG:	Ocrelizumab (RO4964913)
VERSION NUMBER:	5
IND NUMBER:	100,593
EUDRACT NUMBERS:	2010-020337-99 and 2010-020315-36
SPONSOR:	F. Hoffmann-La Roche Ltd
PLAN PREPARED BY:	[redacted], Ph.D. [redacted], M.Sc.
DATE FINAL:	8 March 2011
DATES AMENDED:	Version 2: 14 January 2014 Version 3: 3 September 2014 Version 4: 14 May 2015 Version 5: See electronic date stamp on the last page of the document.

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE

January 2014

Study WA21093 number for pooling was not included in the first version. Several exploratory and sensitivity analyses have been removed in order to focus on the analyses to be performed for the regulatory submission. The hierarchy of the secondary efficacy endpoints has been modified. Efficacy analyses in some subgroups have been added for EU regulatory purposes. The intent-to-treat population includes all patients randomized to adhere to the standard definition.

August 2014

This Statistical Analysis Plan (SAP) was amended to implement EMA Scientific Advice and to increase statistical rigor (for example, replacing ranked ANCOVA analysis with Mixed-Effect Model Repeated Measures analysis (MMRM) in relevant analyses). This SAP has also been updated to align with the amended study protocols. In both the protocols and this SAP, the secondary and exploratory efficacy endpoints have been extensively modified to reflect the latest clinical and scientific thinking in the field of MS, and also to increase rigor (for example, the proportion of relapse-free patients was changed from a secondary to an exploratory efficacy endpoint, since it is closely related to the primary efficacy endpoint, hence adds only limited value to our understanding of efficacy).

More details have been added to specify the calculation of the baseline Expanded Disability Status Scale (EDSS) value. The average value of the screening EDSS score and the value at the baseline visit used as reference for the analyses of increase or decrease EDSS score (i.e., the Confirmed Disability Progression, Confirmed Disability Improvement (CDI), and mean change of EDSS over time). However, for the stratification, the value is rounded up to the next EDSS score (e.g., 3.75 is rounded up to 4. and 3.25 is rounded up to 3.5), in order to present a valid score for the baseline characteristics and randomization scheme. The definition of the confirmed disability progression has been slightly modified to take into account the existence of a baseline EDSS of 5.75, with no impact to the identification of the progression. The definition of the Confirmed Disability Improvement has been added. In order to simplify the definition of the safety population and adhere to the standard, the safety population now includes all patients who received any study drug.

This Statistical Analysis Plan covers analyses performed at the level of each individual study. The scope of this statistical analysis plan includes:

- All data collected in the 96-week comparative treatment phase
- All data collected in the Open-Label Extension (OLE) phase up to the date of clinical cutoff applied for the analysis of the 96-week comparative treatment phase

- All data collected in the Safety Follow-Up (SFU) phase up to the date of clinical cutoff applied for the analysis of the 96-week comparative treatment phase

Note that substudies conducted at certain selected sites, such as the substudy collecting Optical Coherence Tomography (OCT) scans, are not within the scope of this statistical analysis plan. These analyses will be planned, executed, and reported separately.

Analyses performed upon data from both studies combined are described in a separate analysis plan (i.e., the "Pooled SAP").

It is planned that this statistical analysis plan will be reviewed, to include the SAS code for primary and secondary analysis per FDA request, once more before the first of the two study database locks for the primary analysis, to ensure it remains fit for purpose and provides sufficient detail.

Protocol synopsis and schedule of assessment have been removed from the SAP and are available in the protocols.

Additional minor changes have been made to improve clarity and consistency.

May 2015

Updates made at this time include the following major aspects. The rationale for the updates is to provide full and complete information on these important aspects prior to database lock:

- Subgroup analyses for EU planning purposes have been removed; these are now covered in the pooled SAP.
- Detailed updates have been made to the safety analysis section.
- SAS code has been included for statistical models for primary and secondary analyses.
- Derivation of protocol-defined relapses has been added.
- Reasons for patient exclusion from the per-protocol population have been added.
- Sensitivity analyses have been added to describe alternative analyses when negative binomial models do not fit appropriately
- Methods have been added to handle MSFC values outside of normal ranges

Additional minor changes have been made to improve clarity and consistency.

Jan 2022

This SAP is updated to describe the modification to the pre-specified analyses during the double-blind phase, as well as the analyses to be performed on data from the open-label extension (OLE) period. The OLE differed from the double-blind phase of the study with respect to the objectives and the data collected. As a result, the statistical analysis plan required modification. Moreover, several modifications were made based on the

availability of the data. A list of these modifications can be found below. A table is also added to display the status of each reporting event.

Updated Analysis:

Some analyses that were pre-specified for the double blind period were modified after unblinding and are summarized below:

- The pre-specified exploratory analyses of T2 hyperintense lesion volume, the timed 25-foot walk (T25-FW) and nine-hole peg test (9-HPT) utilizing the change from baseline (see SAP version 4 May 2015 Sections 4.4.3.2, 4.4.3.7 and 4.4.3.8 respectively) did not meet assumptions of normality and so were adapted post-hoc to use a logarithmic transformation of the scores, hence the ratio of post-baseline to baseline scores are reported.

The pre-specified exploratory analysis of ARR for severe relapses (SAP version 4 May 2015 Section 2.2.3) was not conducted due to the low number of severe relapses observed during the study (two severe relapses in the IFN group and zero in the OCR group).

Additional Exploratory Analysis:

- ARR by year
- Annualized change in Non-enhancing T1 lesion volume radius
- Time to Onset of Confirmed Disability Progression for at least 48 Weeks
- Time to 24-week confirmed Expanded Disability Status Scale score$\geq$6.0 (time to requiring a walking-aid)
- Time to 48-week confirmed Expanded Disability Status Scale score$\geq$6.0 (time to requiring a walking-aid)
- Time to Onset of Composite Confirmed Disability Progression (EDSS or T25-FW or 9-HPT) for at least 12 Weeks
- Time to Onset of Composite Confirmed Disability Progression (EDSS or T25-FW or 9-HPT) for at least 24 Weeks

The analyses of EuroQoL EQ-5D questionnaire collected at baseline, Week 48 and Week 96 during the double-blind phase and yearly during the OLE phase were not conducted at the time of the Primary Analysis and will be reported descriptively in the final CSRs.

Protocol Updates:

Due to various protocol amendments, SAP is updated to reflect changes or updates made in the protocol amendments. Below are some of such changes;

- OLE treatment phase is extended till 31 Dec 2022 to allow additional long-term efficacy and safety data. Study can be ended by 31 Dec 2022 and an option is introduced for all ongoing participants of OPERA studies to enroll into a new open-label extension study (MN43964) prior to or following the closure of Study WA21092/WA21093, latest by End of 2022.
- PK/HAHA collection has been stopped since Dec 2019.
- Ocrelizumab must be suspended in the event of an active TB infection or if a female patient is pregnant or breastfeeding in OLE period. Ocrelizumab infusions may be restarted at the discretion of the Investigator and based on individual benefit-risk assessments, but only upon resolution of the active TB infection or after completion of pregnancy and breastfeeding.
- Patients starting other DMT or commercial ocrelizumab will discontinue from the study completely and will not enter safety follow-up.

Additional minor changes have been made to improve clarity and consistency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. BACKGROUND

Studies WA21092 and WA21093 are two Phase III studies with identical study design. This Statistical Analysis Plan (SAP) describes the analyses that will be performed for each study independently. The analyses will be conducted in exactly the same way for both studies; therefore, only one SAP has been written to be applied to both studies. The analysis of some secondary efficacy endpoints will necessitate the pooling of data from both studies in order to have sufficient statistical power to detect relevant treatment differences. Analyses of pooled data are described separately in the "Pooled SAP."

Study WA21493 is an ongoing double-blind, double-dummy, parallel-group, Phase II study of the efficacy and safety of ocrelizumab compared with interferon beta-1a intramuscular (IM) (Avonex) in patients with relapsing remitting multiple sclerosis (RRMS). The study was unblinded in November 2009 and has an open-label phase that is ongoing. After 24 weeks of treatment, both doses of ocrelizumab (600 mg and 2000 mg) demonstrated a strong effect with a highly statistically significant reduction versus placebo in signs of paraclinical disease activity as measured by the number of gadolinium (Gd)–enhancing brain lesions in magnetic resonance imaging (MRI), the primary endpoint of the study, at Weeks 12, 16, 20, and 24. Compared with placebo, this represents a relative reduction in the number of lesions of 89% and 96%, for the 600-mg and 2000-mg ocrelizumab groups respectively. In addition, a statistically significant reduction in the annualized relapse rate (ARR) was observed with both ocrelizumab doses compared with placebo. Moreover, both ocrelizumab doses were superior to the active comparator, Avonex, for both the MRI primary endpoint and the ARR.

Although Study WA21493 (which enrolled approximately 50 patients per arm) demonstrated significant treatment effects in both MRI and ARR endpoints at Week 24 versus Avonex, the sample size required to demonstrate a statistically significant difference in disability progression between ocrelizumab and an active comparator is estimated to be much greater (i.e., approximately 800 patients per arm). To demonstrate significant treatment benefit of ocrelizumab in disability progression, the Sponsor plans to pool data from two Phase III studies (WA21092 and WA21093), which both include approximately 400 patients per treatment arm (ocrelizumab or interferon beta-1a 44 μg subcutaneous [SC] [Rebif]), have identical inclusion and exclusion criteria, and were implemented concurrently.

2. WA21092 AND WA21093 STUDY DESIGN

The primary objective of Studies WA21092 and WA21093 is to assess whether the efficacy of ocrelizumab 600 mg (given as dual infusions of 300 mg on Days 1 and 15 of the first 24-week Dose and as a single infusion of 600 mg on Day 1 of each 24-week Dose thereafter) intravenously every 24 weeks is superior to interferon beta-1a 44 μg SC as measured by the annualized protocol-defined relapse rate (see Section 4.4.1 for the

efinition of protocol-defined relapse) by 2 years (96 weeks) in patients with relapsing forms of multiple sclerosis (RMS).

The studies consist of the following periods:

Screening:

Consenting patients entered a screening period to be evaluated for eligibility. The screening period lasted approximately 2 weeks, but it may have been prolonged for up to 8 weeks for relevant clinical, administrative, or operational reasons.

Treatment Period:

Double-blind, double-dummy, 96-week comparative treatment period

Eligible patients were randomized via an Interactive Voice and Web Response System (IxRS) into one of two treatment groups: ocrelizumab 600 mg regimen (Group A) or interferon beta-1a 44 μg SC (Group B).

Open-Label Extension Phase

Patients who complete the 96-week treatment period may become eligible for the open-label extension (OLE) phase of the study.

All patients will continue their treatment in the open-label ocrelizumab phase until 31 December 2022, as per the protocol. All patients will discontinue ongoing open-label extension phase and move into a new extension trial (OLERO) on or before 31 December 2022.

Safety Follow-Up Period:

Patients who discontinue treatment for any reason during the following periods will be entered into the Safety Follow-Up Period:

- During or after completion of the 96-week, double-blind, double-dummy treatment period
- During the OLE Phase Screening Period
- During the OLE Phase
- Patients who choose not to enter the OLE Phase or are not eligible for the OLE Phase after completing the 96-week, double-blind, double-dummy treatment period

Patients who start treatment with other DMT or commercial ocrelizumab will discontinue from the study completely and will not enter or continue in the Safety Follow-Up Period.

All patients ongoing in SFU at study completion should move to OLERO by 31 December 2022.

Table 1 and Table 2 give an overview of the study design and dosing regimen during the double-blind double-dummy treatment period and OLE, respectively.

Table 1 Overview of Dosing Regimen in the Double-Blind, Double-Dummy Treatment Period

Study Medication	Double-Blind, Double-Dummy Treatment Period [a, b]				
	1st Dose [c] (Weeks 1–24)		2nd Dose [c] (Weeks 24–48)	3rd Dose [c] (Weeks 48–72)	4th Dose [c] (Weeks 72–96)
	Day 1	Day 15	Week 24	Week 48	Week 72
A: Ocrelizumab 600 mg regimen	300 mg IV	300 mg IV	600 mg IV	600 mg IV	600 mg IV
B: Interferon beta-1a 44 μg SC regimen [d]	Interferon beta-1a 44 μg SC 3 times per week	→	→	→	→

IV = intravenous; SC = subcutaneous.

Note: 100 mg of methylprednisolone IV will be administered in both treatment arms prior to each infusion of ocrelizumab/ocrelizumab placebo.

[a] The double-blind, double-dummy, treatment period consists of 96 weeks of treatment (four Doses).

[b] The first Dose consists of two 300-mg ocrelizumab IV infusions separated by 14 days. Doses 2–4 consist of a single IV infusion of 600-mg ocrelizumab.

[c] Prior to each infusion, a clinical evaluation will be performed to ensure that the patient remains eligible for treatment.

[d] Refer to the protocol for detailed interferon beta-1a 44 μg SC dosing regimen.

Table 2 Overview of Dosing Regimen in the OLE Phase Screening Period and the OLE Phase

Study Medication	OLE Phase Screening Period	OLE Phase [a]				
		5th Dose [b, c,d]		6th Dose [b, c, d]	7th Dose [b, c, d]	Nth Dose [b, c, d]
		Day 1 Infusion	Day 15 Infusion			
Ocrelizumab 600 mg regimen	NA [e]	300 mg IV	300 mg IV	600 mg IV	600 mg IV	600 mg IV
Interferon beta-1a 44 μg SC regimen	Interferon beta-1a 44 μg SC 3 times per week[f]	NA [g]	NA [g]	NA [g]	NA [g]	NA [g]

IV = intravenous; N = "nth" Dose; NA = not applicable; OLE = open-label extension; SC = subcutaneous.

Note: 100 mg of methylprednisolone IV will be administered prior to each infusion of ocrelizumab.

To keep the blinding, all patients will receive the first dose of ocrelizumab in the open-label phase in a split form, even if they had been treated with ocrelizumab in the previous blinded phase

[a] The OLE Phase can terminate at any moment.

[b] The assessments requested for N represent the typical schedule of assessments.

[c] Prior to each infusion, a clinical evaluation will be performed to ensure that the patient remains eligible for treatment.

[d] The first Dose of the OLE Phase consists of two 300-mg ocrelizumab IV Infusions that are to be separated by 14 days. Dose 6 onward consists of a single IV infusion of 600 mg ocrelizumab.

[e] During the OLE Phase Screening Period there will be no administration of ocrelizumab.

[f] Refer to the protocol for detailed interferon beta-1a 44 μg SC dosing regimen.

[g] During the OLE, there will be no administration of interferon beta-1a 44 μg SC; patients previously assigned to interferon beta-1a 44 μg SC receive ocrelizumab.

2.1 PROTOCOL SYNOPSIS

The Protocol Synopsis and the Schedule of Assessments are included in the protocols.

2.2 OUTCOME MEASURES

2.2.1 Primary Efficacy Outcome Measure

The primary efficacy endpoint is the annualized protocol-defined relapse rate at 2 years (96-weeks).

Secondary Efficacy Outcome Measures

The secondary efficacy endpoints are as follows:

- The time to onset of confirmed disability progression (CDP) for at least 12-weeks, with the initial event of neurological worsening occurring during the 96-week, double-blind, double-dummy treatment period
- The total number of T1 Gd-enhancing lesions as detected by brain MRI at Weeks 24, 48, and 96
- The total number of new and/or enlarging T2 hyperintense lesions as detected by brain MRI at Weeks 24, 48, and 96
- The proportion of patients who have confirmed disability improvement for at least 12 weeks, with the initial event of neurological improvement occurring during the 96-week, double-blind, double-dummy treatment period
- The time to onset of confirmed disability progression for at least 24 weeks, with the initial event of neurological worsening occurring during the 96-week, double-blind, double-dummy treatment period
- The total number of new T1−hypo-intense lesions (chronic black holes) at Weeks 24, 48, and 96
- The change in Multiple Sclerosis Functional Composite Scale (MSFCS) score from baseline to Week 96
- The percentage change in brain volume as detected by brain MRI from Week 24 to Week 96
- The change in SF-36 Physical Component Summary (PCS) Score from baseline to Week 96
- The proportion of patients who have no evidence of disease activity (NEDA) by Week 96

2.2.2 Exploratory Efficacy Outcome Measures

The exploratory efficacy endpoints are as follows:

- In addition to below specified exploratory outcomes, specific primary, secondary and exploratory efficacy analysis will be repeated combined and separately for double blind and open label period
- The proportion of relapse-free patients by Week 96

- The percentage change in total T2 hyperintense lesion volume as detected by brain MRI from baseline to Week 96
- The ARR, based on all clinical relapses at the end of the 96-week comparative treatment period (protocol-defined relapses are a subset of all clinical relapses)
- The ARR of relapses requiring IV steroids therapy
- The percentage change in brain volume as detected by brain MRI from baseline to Week 96
- The change in timed 25-foot walk from baseline to Week 96
- The change in 9-hole peg test from baseline to Week 96
- The change in fatigue, as measured by the Modified Fatigue Impact Scale (MFIS) total score from baseline to Week 96
- The change from baseline in patient-reported depressive symptoms, as measured by the Center for Epidemiologic Studies Depression Scale (CES-D), from baseline to Week 96
- The change in Karnofsky Performance Status Scale from baseline to Week 96
- The percentage change in cortical gray matter volume from baseline to Week 96
- The percentage change in white matter volume from baseline to Week 96
- The proportion of patients who have confirmed disability improvement for at least 24 weeks, with the initial event of neurological improvement occurring during the 96-week, double-blind, double-dummy treatment period
- The proportion of patients who have disability improvement sustained for at least 12 weeks and sustained until the end of the 96-week, double-blind, double-dummy treatment period, with the initial event of neurological improvement occurring during the 96-week, double-blind, double-dummy treatment period
- The duration of the confirmed disability improvement
- The proportion of patients who at Week 96 have improved, stable, or worsened disability, compared to baseline
- The change in Expanded Disability Status Scale (EDSS) Score from baseline to Week 96
- The change in Quality of Life, as measured by the Short Form 36 Version 2 (SF-36 v2) Mental Component Summary Score from baseline to Week 96
- The change in EQ-5D-3L Index and visual analogue scale (VAS) from baseline to Week 48 and Week 96

2.2.3 Pharmacokinetic Analysis: Objectives and Outcome Measures

The pharmacokinetic (PK) endpoints are as follows:

- Develop a population PK model to describe the pharmacokinetics of ocrelizumab in patients with MS and to estimate inter- and intra-patient variability
- Determine individual post hoc estimates to derive PK exposure measures

- Explore and quantify the potential influence of covariates contributing to the interpatient variability in PK parameters
- Explore the relationship between ocrelizumab exposure and selected efficacy endpoints (e.g., annualized relapse rate)
- Explore the relationship between ocrelizumab exposure and appropriate safety parameters

2.2.4 Pharmacodynamic Outcome Measure

The pharmacodynamics (PD) endpoint is the CD19 count, as a PD marker.

2.2.5 Safety Outcome Measures

Safety will be assessed through regular neurological and physical examinations, vital signs, ECGs, and the occurrence of adverse events. In addition, the following will be examined:

- Non-MS pathology at all available MRI scans
- Columbia-Suicide Severity Rating Scale (C-SSRS)
- Standard hematology, chemistry, and urinalysis assessments
- Circulating B-cell total and subsets, T cells, natural killer cells, and other leukocytes
- Plasma immunoglobulins
- Anti-drug antibodies (ADAs), also called human anti-human antibodies (HAHAs)
- Antibody titers for mumps, rubella, varicella, and Streptococcus pneumonia
- MS relapses classified as serious

2.3 DETERMINATION OF SAMPLE SIZE

The sample size for this study was estimated on the basis of data from previous RRMS trials, with use of two-sided tests with an experiment-wise alpha of 0.05. The ARR at 96 weeks in patients receiving ocrelizumab is predicted to be 0.165 (SD of approximately 0.60), compared with 0.33 (SD of approximately 0.80) in patients receiving the control treatment (interferon beta-1a 44 μg SC); this represents a relative reduction of approximately 50% with ocrelizumab treatment compared with the active comparator. For the ARR, a t-test was used to determine the sample size between ocrelizumab and interferon beta-1a 44 μg SC. The sample size of 400 patients per treatment group provides 84% power, maintaining the type I error rate of 0.05 and with the assumption of a dropout rate of approximately 20% (with the assumption that relative reduction among patient dropout is 25%).

For confirmed disability progression, a two group test of equal exponential survival with exponential dropout was used to determine the sample size. Assuming the 2-year confirmed disability progression rate is 18% for the interferon beta-1a 44 μg SC arm and 12.6% for the ocrelizumab arm (this represents a relative reduction of 30% on ocrelizumab compared to the active comparator), and assuming a dropout rate of 20% over 2 years approximately, the sample size of 400 per arm will provide 80% power,

maintaining the type I error rate of 0.05 based on the pooled analysis of two RMS studies (800 patients treated with ocrelizumab 600 mg and 800 patients treated with interferon beta-1a 44 μg SC).

2.4 ANALYSIS TIMING

After the 96-week visit of the last randomized patient, approximately 12-weeks may be needed to allow the confirmation of the latest event of the 12-week confirmed disability progression. Therefore, the clinical cutoff date will be approximately 12-weeks after the last patient's 96-week visit when the status is clarified for each patient. Database lock and unblinding of the Sponsor will occur several weeks after the clinical cutoff to clarify all outstanding queries.

The sites and EDSS raters will remain blind until approximately 24-weeks after the 96-week visit of the last patient randomized, to allow the confirmation of the last 24-week confirmed disability progression, in case an updated analysis of this endpoint is requested at a later point.

The following reporting events are done / planned for this study:

Report	Clinical cutoff date	Status
Primary CSR	Opera 1: 10 April 2015 Opera 2: 12 May 2015	Completed
Pooled safety analyses for safety monitoring and Publication of updated interim results	Yearly data cut has been performed since February 2017 to conduct interim analyses used for publication purposes	Completed until 2021 and planned for subsequent years until the end of the study
Interim CSR (Time to milestone analysis)	03 January 2020	Completed
Final CSR	After the last patient last visit of the open label extension phase or Safety Follow-up Phase (31 Decemeber 2022)	Planned

CSR = clinical study report

3. STUDY CONDUCT

3.1 RANDOMIZATION

Patients were randomized into two groups (Group A or Group B) in a 1:1 ratio. An independent IxRS provider conducted randomization (with use of blocked randomization with a block size of 4) and holds the treatment assignment code. Patients were stratified by geographical region (United States vs. rest of world [ROW]) and baseline EDSS score (<4.0 and ≥ 4.0). (Note: If both screening EDSS and randomization EDSS scores were collected, the average of both values will be considered as baseline EDSS score for all analyses and for the stratification factor, unless otherwise specified.)

The patient randomization list was generated by the IxRS with use of a pre-defined randomization specification. The randomization list will not be available to the study center, study monitors, project statisticians, or the Sponsor's project team. Unblinding of individual patients should not occur except in the case of emergency situations. Any request from the investigator for information about the treatment administered to study patients for another purpose must be discussed with the Sponsor. Unblinding will be performed by means of the IxRS. In accordance with regulatory reporting requirements, the Sponsor will unblind the identity of the study medication for all unexpected serious adverse events that are considered by the investigator to be related to study drug per safety reference document(s), (e.g., Investigator's Brochure, Core Data Sheet, and Summary of Product Characteristics). Details of patients who are unblinded during the study will be included in the Clinical Study Report.

3.2 INDEPENDENT REVIEW FACILITY

MRI scans will be read by a centralized reading center. The centralized reading center is blinded to the treatment assignment, and the reading is performed in the absence of clinical information. Further details on scanning acquisition sequences; methods, handling, and transmission of the scans; certification of site MRI radiologist/technicians; and the procedures for the blinded analysis of the scans at the central reading center are described in a separate MRI technical manual.

3.3 EXPANDED DISABILITY STATUS SCALE CLEANING PROCESS

EDSS assessments are performed by a qualified examining investigator (a person other than the treating investigator), and the results are entered into an electronic device, and transferred to a central database. All EDSS results will then be checked in accordance with the Standard Operating Procedure (SOP) entitled "EDSS Assessment Check for the Roche Trials WA25046, WA21092, and WA21093" (see Appendix 1).

3.4 DATA MONITORING

An independent Data Monitoring Committee (iDMC) is reviewing cumulative data from the studies at approximately 4-month intervals until the primary analysis is completed. After all patients have completed the 96-week comparative treatment period, the iDMC will review cumulative data less frequently, but it is anticipated at least once per year. The iDMC is reviewing both efficacy and safety data; however, the studies will not be stopped for efficacy reasons. No iDMC review is conducted post unblinding of the study.

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATIONS

One patient population will be defined for the purpose of the safety analysis, and two patient populations will be defined for the efficacy analysis. All efficacy analyses will be performed using the intent-to-treat (ITT) population. The per-protocol population will be used for the primary analysis of the primary endpoint (ARR at 96 weeks) and the primary

analysis of the first secondary endpoint (time to onset of confirmed disability progression at a minimum of 12 weeks with the initial event of neurological worsening occurring during the 96-week, double-blind, double-dummy treatment period) in order to evaluate the influence of major protocol violators and as a sensitivity check to the ITT analysis.

4.1.1 Intent-to-Treat Population

All randomized patients will be included in the ITT population. Patients who prematurely withdrew from the study for any reason and for whom an assessment was not performed for whatever reason will still be included in the ITT analysis. Patients who received an incorrect therapy from that which was intended will be summarized according to their randomized treatment.

4.1.2 Per-Protocol Population

The per-protocol population will include all patients in the ITT population who adhere to the protocol. Patients may be excluded if they significantly violated the inclusion or exclusion criteria or deviated from the study plan. Specific reasons for warranting exclusion will be agreed to and documented in the Data Analysis Plan on the basis of the final version of the protocol prior to unblinding of the treatment groups. Only those patients with violations that are deemed to potentially affect the efficacy of study treatment will be excluded from the per-protocol population. Patients who received an incorrect therapy from that which was intended will be excluded from the per-protocol population.

The following patients will also be excluded from the per-protocol population:

- Diagnosis of primary progressive MS
- Disease duration from the onset of MS symptoms of more than 10 years in patients with an EDSS $\leq$2.0 at screening
- Known presence of other neurological disorders that may mimic MS
- Pregnancy or lactation
- Any concomitant disease that may require chronic treatment with systemic corticosteroids or immunosuppressants during the course of the study
- History or currently active primary or secondary immunodeficiency
- Treatment with any investigational agent within 24 weeks of screening or five half-lives of the investigational drug (whichever is longer); or treatment with any experimental procedures for MS
- Contraindication to interferon beta-1a 44 μg SC or incompatibility with interferon beta-1a 44 μg SC use
- Previous treatment with B-cell targeted therapies
- Any previous treatment with alemtuzumab (Campath), anti-CD4, cladribine, mitoxantrone, daclizumab, teriflunomide, laquinimod, total body irradiation, or bone marrow transplantation

- Treatment with cyclophosphamide, azathioprine, mycophenolate mofetil (MMF), cyclosporine, methotrexate (MTX), or natalizumab within 24 months prior to screening
- Treatment with fingolimod (FTY720, Gilenya) or other S1P receptor modulator (i.e., BAF312) or with BG12, within 24 weeks prior to screening
- Treatment with IV immunoglobulin within 12 weeks prior to baseline
- No diagnosis of relapsing MS, in accordance with the revised McDonald criteria (2010)
- Less than two documented clinical attacks within the last 2 years prior to screening and no clinical attack in the year prior to screening; or at least one clinical attack within 30 days prior to screening
- Neurological status instability within 30 days prior to both screening and baseline.
- Received no dose of ocrelizumab / ocrelizumab placebo and interferon beta-1a 44 μg SC / interferon beta-1a 44 μg SC placebo
- Received ocrelizumab + interferon beta-1a 44 μg SC placebo / ocrelizumab placebo + interferon beta-1a 44 μg SC verum but was not randomized
- Received ocrelizumab + interferon beta-1a 44 μg SC placebo / ocrelizumab placebo + interferon beta-1a 44 μg SC verum other than the group to which the patient was randomized at any point during the study
- Received study medication that has been mishandled (e.g., incorrect storage temperature) and was not approved subsequently for use

4.1.3 Pharmacokinetic Evaluable Population

The PK population will include all patients in the ocrelizumab treatment arm who had at least one measurable concentration value.

4.1.4 Safety Population

The Safety Population will include all patients who received any study drug. Randomized patients who receive an incorrect therapy from that which was intended will be summarized in the group according to the therapy actually received. Patients who are not randomized but who receive study drug will be included in the safety population and summarized according to the therapy actually received. Patients who received more than one study therapy will be summarized in the ocrelizumab group. Patients who received only ocrelizumab placebo but not interferon beta-1a 44 μg SC will be included in the interferon beta-1a44 μg SC−arm group as a conservative approach.

4.1.5 OLE Intent-to-Treat Population

All patients receiving any study drug during the OLE Phase will be included in the OLE ITT population. Patients will be summarized according to their randomized treatment during the double blind treatment period

4.1.6 OLE Safety Population

Similar to the safety population, all patients receiving any study drug during the OLE Phase will be included in the OLE safety population. Patients will be summarized according to the therapy received during the double blind treatment period, consistent with the safety population.

4.2 ANALYSIS OF STUDY CONDUCT

All data up to the point of the clinical cutoff date for each study will be included to evaluate study conduct. This will include complete data from the 96-week, double-blind, double-dummy treatment period and all available data from the OLE and SFU phases at the time of cutoff for the primary analysis. Data from Study WA21092 that relate to the period after the Study WA21092 clinical cutoff date but before the Study WA21093 clinical cutoff date will not be included in the Study WA21092 database that will be used in the primary analysis.

The following analyses will be performed to evaluate the study conduct:

- Summary of protocol violations
- Summaries of ITT, per-protocol-defined, and safety populations, including numbers of patients in each population, and reasons for exclusion from the per-protocol population
- Summary of patient disposition, including the number of Doses received, the number of patients entering into the SFU phase, and the number of patients entering into the OLE phase
- Summary and Kaplan-Meier plots of the following:
  - Time to discontinuation of study treatment during the 96-week, double-blind, double-dummy treatment period
  - Time to discontinuation from the study (available data from all study phases to be included)

After 96 weeks of double-blind phase, all eligible patients from both treatment groups, including patients from IFN group start receiving ocrelizumab treatment in the open-label extension phase. OLE allows additional data being collected for long-term safety and efficacy assessments. Patients who withdraw from treatment during OLE may start safety follow-up.

Analysis to evaluate study conduct as mentioned above will be repeated for the OLE phase as appropriate.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

For continuous variables, the mean, median, SD, minimum, and maximum will be calculated. For categorical variables, number and percentage in each category will be displayed. The units/categories to be used are indicated within the brackets and separated by commas.

Except where stated, all assessments of treatment group comparability will utilize the date of the baseline visit (not the date of the screening visit) as the reference point in time.

Summaries will be presented for ITT and safety populations.

4.3.1 Demography

- Age (years): summary statistics calculated will include mean, median, SD, minimum, and maximum, percentage, and number in each category (18–39, ≥40, also age categories <18, 18–65, >65, used in the Development Safety Update Report [DSUR]) will be displayed (age at randomization will be used)
- Sex: the number and percentage of male and female patients
- Race: the number and percentage of White; Black or African American; Asian (Indian Subcontinent, Other than Indian subcontinent), American Indian or Alaskan native; Native Hawaiian or other Pacific Islander; and Other
- Stratification Factor: (Geographical region) the number and percentage of United States and Rest of World patients
- Sub-region: the number and percentage of patients from European Union/Switzerland/Norway, Latin America, Non-EU/Israel/Africa, and USA/Canada/Australia
- Ethnicity: the number and percentage of Hispanic or Latino and not Hispanic or Latino
- Weight (kg): summary statistics as described for age
- Body Mass Index (BMI): summary statistics as described for age

4.3.2 Baseline Disease Characteristics

- Stratification Factor (Baseline EDSS): the number and percentage of patients in each EDSS category (<4.0, and ≥4.0); EDSS mean, median, SD, minimum, and maximum will be summarized. The baseline EDSS is defined as the average EDSS score from the screening and baseline visit EDSS scores (only values recorded up to and including the date of randomization will be used). For the analysis of mean change of EDSS, disability progression, and disability improvement, the exact value of the average EDSS score will be used. However, for the stratification factor, and for the randomization algorithm, the average will be rounded up to the next EDSS score, in order to assign a valid score (e.g., round 3.75 to 4.0, round 5.25 to 5.5). If one of the EDSS scores from the screening or baseline visit is missing, the other will be used for baseline EDSS. If EDSS score is missing at both the screening and baseline visit, the patient should not be randomized. Unless otherwise specified, this definition of baseline EDSS is used throughout.
- Number of relapses in the past year: the number and percentage of patients in each category (0, 1, 2, 3, and ≥4); a relapse is considered to happen in the past year if: date of randomization − date of relapse is ≤365 days

- Number of relapses in the past 2 years: the number and percentage of patients in each category (0, 1, 2, 3, and ≥ 4); a relapse is considered to happen in the past 2 years if: date of randomization $-$ the date of relapse is $\leq (365 \cdot 2)$ days
- Time since last onset of MS relapse prior to randomization (year): the number and percentage of patients in each category (>6 months prior to randomization, within the last 6 months from randomization); summary statistics calculated will include mean, median, SD, minimum, and maximum (calculated in years, i.e., divided by 365.25). Time is calculated as: date of randomization $-$ the date of last relapse
- Baseline Functional Systems Scores (FSSs) for each category (pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, cerebral [or mental]; rated 0–5 or 0–6 depending on the domain of FSS) and for ambulation (categorical; rated 0–12)
- Baseline MSFCS score, raw results for each component: summary statistics calculated will include mean, median, SD, minimum, and maximum

4.3.3 Multiple Sclerosis Disease History

- Duration since MS symptom onset (years): summary statistics calculated will include mean, median, SD, minimum, and maximum (calculated in years; i.e., divide by 365.25)
- Duration since MS diagnosis (years): summary statistics calculated will include mean, median, SD, minimum, and maximum (calculated in years; i.e., divide by 365.25)
- Duration since MS symptom/diagnosis onset will be calculated up to the randomization date. If the month of symptom/diagnosis onset date is missing, the month of January will be used. If the day of symptom/diagnosis onset date is missing, the first (1st) of the month will be used

4.3.4 Non-MS Disease History

- Did patient have a history of any non-MS diseases?: number and percentage of patients, (yes or no)
- Each non-MS disease: number and percentage of patients with a history of each disease

4.3.5 Prior Treatments for Multiple Sclerosis

- Did patient receive any prior treatment for MS?: number and percentage of patients, (yes or no)
- Did patient receive any prior treatment for MS with any interferon or glatiramer acetate?: number and percentage of patients, (yes or no)
- Each prior treatment for MS: number and percentage of patients receiving each treatment

4.3.6 Baseline Magnetic Resonance Imaging Data

- Number of Gd-enhancing T1 lesions at baseline: the number and percentage of patients in each category (0, 1, 2, 3, and ≥ 4); summary statistics calculated will include mean, median, SD, minimum, and maximum
- Volume of T2 lesions at baseline: summary statistics calculated will include mean, median, SD, minimum, and maximum
- Number of T2 lesions at baseline: summary statistics calculated will include mean, median, SD, minimum, and maximum. Also the number and percentage of patients in each category (0–5, 6–9, and >9) will be presented.
- Normalized brain volume at baseline: summary statistics calculated will include mean, median, SD, minimum, and maximum
- Number of T1 hypo intense lesion (black holes) count at baseline: summary statistics calculated will include mean, median, SD, minimum, and maximum

Past MRI Data (recorded at Screening):

- Number of previous Gd-enhancing T1 lesions: the number and percentage of patients in each category (i.e., 0, 1, >1, and not evaluable)
- Number of previous T2 lesions: the number and percentage of patients in each category (0–5, 6–9, and >9)

4.3.7 OLE Baseline

Demography and baseline characteristics analysis will be repeated and updated for the OLE period.

EDSS at OLE baseline is defined as the latest EDSS score prior to or on the start of the OLE period.

For all other analysis, OLE Baseline score is the latest score prior to or on the OLE first dose, collected at Week 96 and/or OLE Week 0, unless specified specifically. Patients who have none of these scores available will be excluded from the analyses requiring OLE baseline score.

4.4 EFFICACY ANALYSIS

All statistical hypotheses for the primary and secondary endpoints and treatment comparisons will be tested at the 5% significance level ($\alpha = 0.05$) against two-sided alternatives.

For all assessments, the baseline value will be used as the last non-missing value on or before the date of the first infusion of study medication. See Section 4.3.2 for baseline EDSS and Section 4.4.2.4 for time-to-confirmed sustained disability progression.

All primary and secondary efficacy endpoints will be analyzed stratified by geographical region (United States vs. ROW) and baseline EDSS (<4.0 vs. ≥ 4.0).

For all analyses during OLE and combined periods of double-blind and OLE, a pooled efficacy analysis of WA21092 and WA21093 will be performed.

4.4.1 Primary Efficacy Endpoint, Annualized Protocol-Defined Relapse Rate by 2 Years (96 weeks)

Significance Level

The null hypothesis will be tested at the $\alpha = 0.05$ level (two-sided test).

- H_0 (null hypothesis): there is no statistically significant difference between the ocrelizumab group and interferon beta-1a 44 μg SC group in the annualized protocol-defined relapse rate at 2 years.
- H_1 (alternative hypothesis): there is a statistically significant difference between the ocrelizumab group and interferon beta-1a 44 μg SC group in the annualized protocol-defined relapse rate at 2 years.

Annualized protocol-defined relapse rate at 2 years between the ocrelizumab dose groups and interferon beta-1a 44 μg SC group will be compared using the negative binomial model, adjusting for geographical region (United States vs. ROW) and baseline EDSS score (<4.0 vs. ≥ 4.0). If the test result for comparing the 600-mg ocrelizumab group and the interferon beta-1a 44 μg SC group is statistically significant at $\alpha < 0.05$ level (two-sided test), it will be concluded that the 600-mg ocrelizumab group demonstrated a superior effect of reducing the annualized protocol-defined relapse rate when compared with the interferon beta-1a 44 μg SC group.

Protocol-Defined Relapse

A protocol-defined relapse is defined as the occurrence of new or worsening neurological symptoms attributable to MS and immediately preceded by a relatively stable or improving neurological state of least 30 days. Symptoms must persist for >24 hours and should not be attributable to confounding clinical factors (e.g., fever, infection, injury, or adverse reactions to concomitant medications). The new or worsening neurological symptoms must be accompanied by objective neurological worsening consistent with an increase of at least half a step on the EDSS score, or 2 points on one of the appropriate FSS, or 1 point on two or more of the appropriate FSS. The change must affect the selected FSS (i.e., Pyramidal, Gait, Cerebellar, Brainstem, Sensory, or Visual). Episodic spasms, sexual dysfunction, fatigue, mood change, or bladder or bowel urgency or incontinence will not suffice to establish the diagnosis of a relapse.

Information related to a protocol-defined relapse will be captured on a clinical relapse event case report form (CRF) page. If the following criteria are satisfied, the clinical relapse will qualify as a protocol-defined relapse:

- Check-box of "Did symptoms persist for >24 hours and were not being attributable…" on the Clinical MS relapse event is checked.
- EDSS scores from a visit occurring on or after the onset date of relapse needs to be at least half a step from the previous EDSS score; OR for FSS domains involved in

the relapse event (indicated on the Clinical MS Relapse page), an increase of at least 2 points on one appropriate FSS domain or at least 1 point on two or more appropriate FSS domains.

- There is no protocol-defined relapse within 30 days before the start date of the clinical relapse.

Protocol defined relapse will be derived following the steps below:

1. Clinical relapse is reported on eCRF.
2. "Did symptoms persist for $>$24 hours and were not being attributable…" is checked 'Yes' on the Clinical MS relapse event form.
3. Check if the first EDSS assessment at a visit (unscheduled or scheduled) on or after the onset date of the relapse is increased by $\geq$0.5 steps from the previous EDSS; OR SELECTED FSS domains relevant to the relapse event (pyramidal, ambulation, cerebellar, brainstem, sensory, or converted visual) are increased by $\geq$2 points on one domain or $\geq$1 point on two or more domains. When deriving this step do the following:

– Take the last EDSS/FSS score before each clinical relapse onset date

– Take the first EDSS/FSS score on or after each clinical relapse onset date

– Calculate the difference between the two scores

– Select clinical relapses where there is an increase of $\geq$0.5 in EDSS OR $\geq$2 on one appropriate FSS domain OR $\geq$1 on two or more appropriate FSS domains

4. For each relapse that satisfies the 3 criteria above, check if the following relapses are within 30 days (i.e., the onset dates are $\leq$30 days apart). If they are within 30 days, then the later relapses are not protocol-defined relapses.

Analysis Methods

The ITT Population analysis will be presented.

The total number of protocol-defined relapses for each patient will be counted. The exposure time will be calculated as follows:

Exposure Time = (Early treatment discontinued date or date of Week 96 visit) − date of study Day 1 + 1)/365.25

All available data during the 96-week treatment period will be used for the analysis. For patients who discontinue the treatment early, only data that are collected before the early treatment discontinuation will be used in this analysis (with the exception of subsequent data used to confirm disability progression post-discontinuation of study treatment).

The annualized protocol-defined relapse rate at Week 96 will be analyzed using a negative binomial model, fitted in SAS using the GENMOD procedure. The model will include, for each patient, the total number of protocol-defined relapses with onset between randomization date and early treatment discontinued date/date of Week 96 as response variable and treatment group, baseline EDSS score ($<$4.0 vs. $\geq$4.0), and

geographical region (United States vs. ROW) as covariates. In order to account for different study treatment exposure durations among patients, log-transformed exposure time will be included in the model as an "offset" variable for appropriate computation of relapse rate. The rate ratio and the associated two-sided 95% confidence interval will be provided to compare the ocrelizumab and interferon beta-1a 44 μg SC groups. The estimated relapse rate and its two-sided 95% confidence intervals will be provided for each treatment group.

Sample SAS code can be found below (SAS code is regarded as "draft" until fully validated at the analysis stage):

```
    proc genmod data=ARR;
        class ARMCD REGION BEDSSCAT;
        model N_PDR = ARMCD REGION BEDSSCAT / offset=EXPLOG              link=log
dist=negbin type3;
        lsmeans ARMCD / exp cl;
        ods output lsmeans=lsm;
        ods output ParameterEstimates=est;
    run;
```

The unadjusted annualized protocol-defined relapse rate at Week 96 will also be presented for each treatment group. This is defined for each treatment group as the total number of protocol-defined relapses for all patients in the treatment group, divided by the total number of patient-years of exposure to that study treatment.

Sensitivity and Robustness Checks

The primary analysis described above will be repeated for the per-protocol population as a sensitivity analysis.

The primary analysis described above will be repeated for the ITT population but will exclude patients who received no study medication, as a sensitivity analysis.

As a sensitivity analysis, the primary analysis negative binomial model described above will be adjusted by the following additional covariates: number of relapses occurring within the 2 years prior to study entry, baseline presence of Gd lesions (present or absent), prior MS treatment, and age (<40, ≥ 40).

A Poisson model with the same covariates as the primary analysis negative binomial model will be fitted.

Three further sensitivity analyses will be undertaken:

- The annualized protocol-defined relapse rate at Week 96 will be calculated including all protocol-defined relapses occurring during the 96-week, double-blind, double-dummy phase or the SFU Phase, up to 96 weeks after randomization. In this analysis, the exposure time will be calculated as follows:

  Exposure Time = ((Date of Week 96 visit or date of last visit up to 96 weeks after randomization) − date of study Day 1 + 1)/365.25

- A sensitivity analysis for the annualized protocol-defined relapse rate at Week 96 using multiple imputations will be performed to explore the potential influence of informative dropouts on the results of the primary efficacy analyses. The multiple imputations will be used to impute the events for patients who discontinued treatment early during the 96-week, double-blind, double-dummy phase without any protocol-defined relapse during the 30 days prior to discontinuation of study treatment. Multiple imputation inference involves three distinct phases:

1. The missing data are filled in "m" times to generate "m" complete data sets. Instead of filling in a single value for each missing value, multiple imputation replaces each missing value with a set of "m" plausible values that represent the uncertainty about the right value to impute.
2. The "m" complete data sets are analyzed using standard statistical analyses.
3. The results from the "m" complete data sets are combined to produce inferential results.

In this analysis, 50% of the patients who discontinued treatment early during the 96-week, double-blind, double-dummy phase without any protocol-defined relapse during the 30 days prior to discontinuation of study treatment will be randomly assigned to have an event at the date of treatment discontinuation; the other 50% of these patients will be censored at the date of treatment discontinuation. A total of 1000 (m = 1000) imputed datasets will be produced.

The primary model described in the "Analysis Methods" section above will be applied to each of the imputed datasets. Each imputed dataset will produce an estimate of the difference between ocrelizumab and interferon beta-1a 44 μg SC. The multiple imputation estimator of the difference between ocrelizumab and interferon beta-1a 44 μg SC is the average of the individual 1000 estimators. The variance of the estimator is the combination of the between- and within-imputation variability (Carpenter and Kenward 2007).

- The annualized protocol-defined relapse rate at Week 96 will be calculated such that patients who discontinued treatment early during the 96-week, double-blind, double-dummy phase without any protocol-defined relapse during the 30 days prior to discontinuation of study treatment will be counted as having had a relapse on the date of treatment discontinuation. This approach is similar to above, but here the described approach is applied to 100% of such patients (hence no imputation), rather than randomly to 50% of such patients (with multiple imputation).

OLE Phase

The annualized protocol-defined relapse rate as defined above will be calculated for every year from Study Day 1 and from the start of the OLE period (Year 3).

Each yearly interval will be defined as the follows,

- Year 1: Treatment start date up to -1 day before Week 48; if WEEK 48 is missing then take the minimum of treatment start date+48*7 or treatment end date as the last day
- Year 2 (only exists if treatment end date is strictly after the last day of Year 1): Year 1 end date+1 up to either OLE start date if it is available, or the treatment end date
- Year 3 (only exists if OLE start date is not missing and OLE end date is strictly after the last day of Year 2): Year 2 end date+1 up to 1 day before OLE Week 48; if OLE Week 48 is missing then take the minimum of (OLE start date+48*7 or OLE end date or last known alive date as the last day).
- Year 4: Similar to Year 3 but from OLE Week 48+1 up to 1 day before OLE Week 96.
- The subsequent years will be similar to Year 4 definition with addition of 48 weeks for yearly interval.

So, ARR will be calculated and presented at every 48 weeks interval.

A GEE Poisson model will be performed, with repeated measurements for each patient, the yearly total number of protocol-defined relapses as response variable and treatment group, baseline EDSS score (< 4.0 vs. ≥ 4.0), geographical region (United States vs. ROW), year number, and the interaction between treatment group and year number as covariates.

Summary table should contain;
- ARR for each treatment arm and rate ratio within each year
- Rate ratio between consecutive years for each treatment arm
- Rate ratio between Year 1 ocrelizumab arm and OLE Year 1 (i.e., Year 3) interferon beta-1a44 μg SC arm and corresponding p-value
- Rate ratio between Year 2 ocrelizumab arm and OLE Year 2 (i.e., Year 4) interferon beta-1a44 μg SC arm and corresponding p-value

p-values will be reported for rate ratio within each year and between consecutive years for comparison between treatments.

4.4.2 Secondary Efficacy Endpoints

4.4.2.1 Rationale for Hierarchy of Secondary Efficacy Endpoints

The hierarchical order of secondary efficacy endpoints is shown below, with endpoints listed in descending order of importance. The rationale for this hierarchical order of secondary endpoints is based primarily on clinical meaningfulness (i.e., those endpoints that are clinically more meaningful are listed higher in the hierarchy). In situations where endpoints have similar clinical relevance, those endpoints with a greater chance of achieving a statistically significant treatment difference are listed higher in the hierarchy. Established, rather than novel, endpoints are generally given higher priority within the hierarchy.

For patients, reduction in disability progression is a highly meaningful clinical outcome, reflecting the degree to which they are able to maintain independence and quality of life.

The following is list of all secondary endpoints, also called key endpoints:

- The time to onset of confirmed disability progression for at least 12 weeks, with the initial event of neurological worsening occurring during the 96-week, double-blind, double-dummy treatment period is listed as the first secondary efficacy endpoint on the basis of its clinical meaningfulness.
- The second and third secondary efficacy endpoints listed in the hierarchy are MRI parameters (T1 Gd-enhancing lesions and new and/or enlarging T2 hyperintense lesions) that historically have shown to correlate well with the clinical course.
- The fourth secondary efficacy endpoint, disability improvement, is a clinical endpoint that, if achieved, would have even greater impact on the quality of life of patients than reduction in disability progression. However, it represents a novel endpoint. Therefore, this endpoint is placed lower within the hierarchy than the above endpoints.
- The fifth secondary efficacy endpoint is the time to onset of confirmed disability progression for at least 24 weeks, with the initial event of neurological worsening occurring during the 96-week, double-blind, double-dummy treatment period, but the modality being tested here (EDSS) is identical to that of the first secondary endpoint. Given that the number of 24-week CDP events will be lower than the number of 12 week CDP events, there is likely a lower chance of achieving a statistically significant treatment difference, and hence, the 24 week CDP is listed lower than the 12 week CDP. It is, however, listed relatively higher in the hierarchy as the fifth endpoint, given the greater clinical meaningfulness of disability progression than the remaining endpoints.
- The sixth and eighth secondary endpoints, chronic black holes and brain volume, are indicators of brain tissue loss. The possible correlation of such brain-atrophy measures and long-term clinical outcomes has been suggested in other studies (Barkhof et al. 2009). Therefore, because they are imaging modalities and the correlation to clinical outcomes is weaker than the imaging endpoints that are listed higher in the hierarchy, these endpoints are listed lower in the hierarchy than the clinical endpoints.
- The EDSS does not adequately assess upper limb function and cognitive impairment; therefore, the MSFC scale is used to address this gap in clinical disability assessment. However, since the MSFC scale is not as widely used in other studies and not as well-accepted as a validated endpoint, it is listed seventh in the hierarchy.
- The SF-36 is one of the most widely used and validated instruments for measuring quality of life in patients with MS. It is, however, listed as ninth in the hierarchy below all the other clinical measures because quality-of-life instruments in general are not as widely used in confirmatory clinical studies compared with the other clinical measures.

- The tenth and final secondary endpoint is a composite measure encompassing the absence of disease activity based on clinical and MRI parameters (i.e., consisting of the primary endpoint and secondary endpoints 1, 2, and 3 above). It is a highly meaningful endpoint for patients but is listed as last in the hierarchy because it is a combination of previous secondary endpoints.

4.4.2.2 Statistical Testing Strategy for Secondary Efficacy Endpoints

Secondary efficacy endpoints will be tested in the sequence presented in the hierarchical order listed in Figure 1, all at the $\alpha = 0.05$ level.

All p-values will be reported and will be interpreted as either confirmatory or non-confirmatory (i.e., descriptive only). The circumstances for confirmatory and for non-confirmatory interpretation of p-values are described below, and the sequence of the confirmatory tests is illustrated in Figure 1.

The first secondary endpoint in the hierarchy is the time to onset of confirmed disability progression for at least 12 weeks with the initial event of neurological worsening occurring during the 96-week, double-blind, double-dummy treatment period, and the fifth secondary endpoint is the same, except progression is defined as sustained for at least 24 weeks during the 96-week comparative treatment period. The fourth secondary endpoint is the proportion of patients who have confirmed disability improvement for at least 12 weeks with the initial event of neurological improvement occurring during the 96-week, double-blind, double-dummy treatment. For these three endpoints, only when data from both Studies WA21092 and WA21093 are combined will there be sufficient statistical power to detect relevant treatment differences. For the other secondary efficacy endpoints and the primary efficacy endpoint (annualized protocol-defined relapse rate by 2 years), there is sufficient statistical power within each study to detect relevant treatment differences, without needing to combine data from the two studies.

As a consequence of this, for analyses of secondary efficacy endpoints at the individual study level:

- For the first secondary efficacy endpoint (the time to onset of confirmed disability progression for at least 12 weeks), the study-level p-value will be interpreted as non-confirmatory, due to inadequate statistical power at the study level to detect relevant treatment differences.
- The second secondary efficacy endpoint (total number of T1 Gd-enhancing lesions at Weeks 24, 48, and 96) will be tested in a confirmatory manner if and only if, in the analysis of both studies combined, the first secondary efficacy endpoint reaches a significance level of 0.05 (i.e., pooled analysis $p \leq 0.05$). If, in the analysis of the combined studies, the first secondary efficacy endpoint pooled analysis $p > 0.05$, then the second and subsequent secondary efficacy endpoint p-values within the hierarchy will be interpreted as non-confirmatory.
- The third secondary efficacy endpoint (total number of new and/or enlarging T2 hyperintense lesions at Weeks 24, 48, and 96) will be tested in a confirmatory manner if and only if the second secondary efficacy endpoint (total number of

T1 Gd-enhancing lesions at Weeks 24, 48, and 96) reaches a significance level of 0.05 (i.e., $p \leq 0.05$). If the second secondary efficacy endpoint $p > 0.05$, then the third and subsequent secondary efficacy endpoint p-values within the hierarchy will be interpreted as non-confirmatory.

- For the fourth (proportion of patients who have confirmed disability improvement for at least 12 weeks) and fifth (time to onset of confirmed disability progression for at least 24 weeks) secondary efficacy endpoints, the study-level p-value will be interpreted as non-confirmatory, due to inadequate statistical power at the study level to detect relevant treatment differences.
- The sixth secondary efficacy endpoint (total number of new T1-hypo-intense lesions (chronic black holes) at Weeks 24, 48, and 96) will be tested in a confirmatory manner if and only if, in the analysis of the combined studies, the fifth secondary efficacy endpoint (time to onset of confirmed disability progression for at least 12 weeks) reaches a significance level of 0.05 (i.e., pooled analysis $p \geq 05$). If, in the analysis of both studies combined, the fifth secondary efficacy endpoint pooled analysis $p > 0.05$, then the sixth and subsequent secondary efficacy endpoint p-values within the hierarchy will be interpreted as non-confirmatory.
- The seventh (and subsequent) secondary efficacy endpoint will be tested in a confirmatory manner if and only if the sixth (or immediately previous) secondary efficacy endpoint reaches a significance level of 0.05 (i.e., $p \leq 0.05$). If the sixth (or immediately previous) secondary efficacy endpoint $p > 0.05$, then the seventh (or current) and all subsequent secondary efficacy endpoint p-values within the hierarchy will be interpreted as non-confirmatory.

Furthermore, for analyses of secondary efficacy endpoints where data from both studies are combined (so there is sufficient statistical power for all primary and secondary efficacy endpoint comparisons):

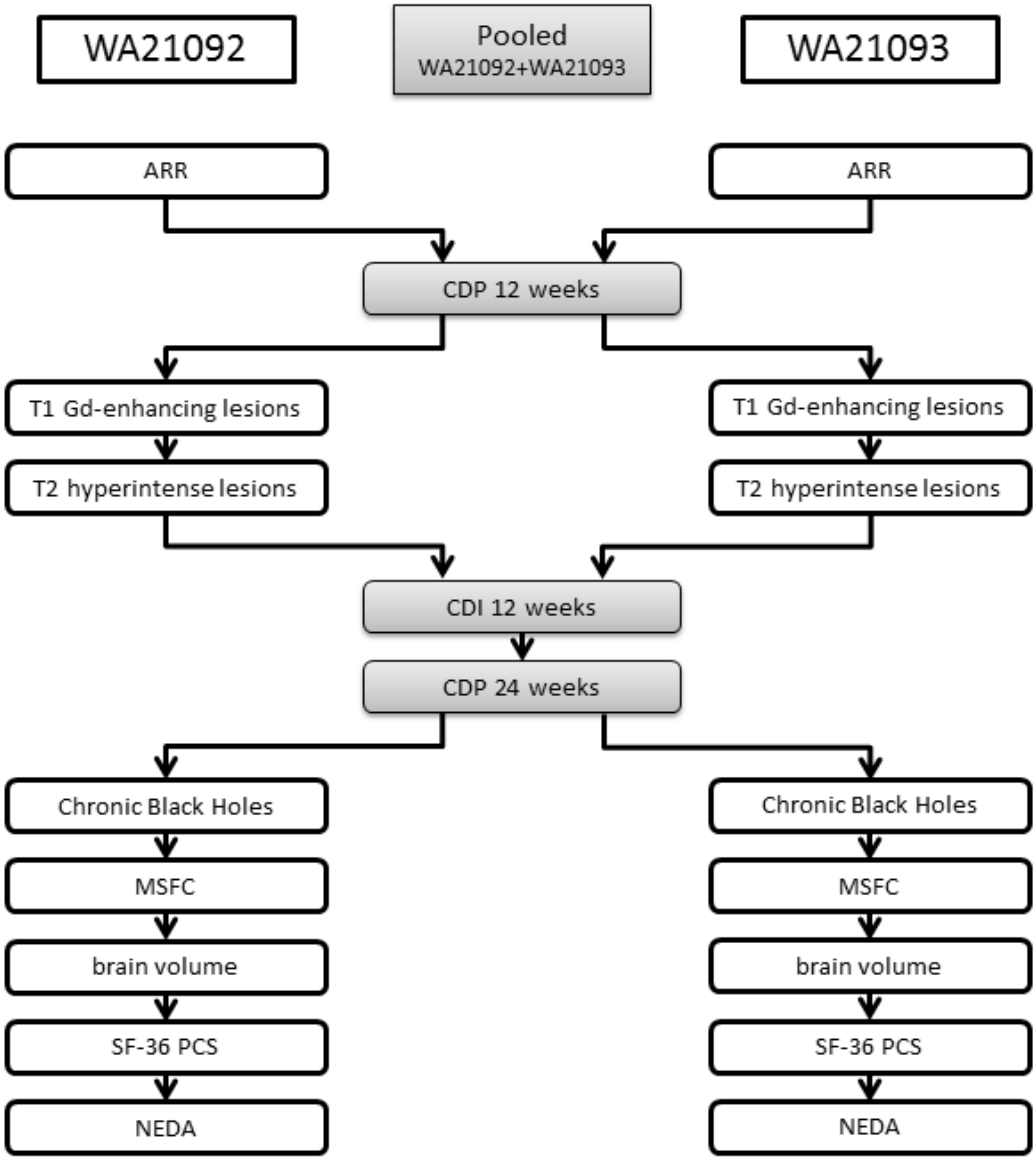
- The first secondary efficacy endpoint (the time to onset of confirmed disability progression for at least 12 weeks) will be tested in a confirmatory manner if and only if the primary efficacy endpoint (annualized protocol-defined relapse rate by 2 years), reaches a significance level of 0.05 (i.e., pooled analysis $p \leq 0.05$). If the primary efficacy endpoint pooled analysis $p > 0.05$, then all secondary efficacy endpoint pooled analysis p-values within the hierarchy will be interpreted as non-confirmatory.
- The second (and subsequent) secondary efficacy endpoint will be tested in a confirmatory manner if and only if the first (or immediately previous) secondary efficacy endpoint reaches a significance level of 0.05 (i.e., pooled analysis $p \leq 0.05$). If the first (or immediately previous) secondary efficacy endpoint pooled analysis $p > 0.05$, then the second (or current) and all subsequent secondary efficacy endpoint pooled analysis p-values within the hierarchy will be interpreted as non-confirmatory.

The sequence of the confirmatory testing is illustrated in Figure 1 below.

The arrows on the left illustrate the hierarchy of the statistical tests of Study WA21092. The arrows on the right illustrate the hierarchy of the statistical tests of Study WA21093. The boxes in the middle highlight the analyses performed in the pooled data set of WA21092 and WA21093.

For example, for Study WA21092, the primary endpoint ARR should be positive in both studies ($p \leq 0.05$). After the poolability of both Studies WA21092 and WA21093 has been observed (see more details in the Pooled SAP), the first secondary endpoint is tested in the pooled dataset WA21092 and WA21093. Then, the analysis of T1 Gd-enhancing lesions is performed for Study WA21092. If this is statistically significant ($p \leq 0.05$), the T2 hyperintense lesions are analyzed for Study WA21092. If this is statistically significant ($p \leq 0.05$), the analysis of confirmed disability improvement is done in the pooled data of Studies WA21092 and WA21093. If this is statistically significant ($p \leq 0.05$), then the analysis of Confirmed Disability Progression for at least 24 weeks is performed in the pooled data. If this is statistically significant ($p \leq 0.05$), the Chronic Black Holes endpoint is analyzed in Study WA21092. Then, the sequence of tests continues in the order of MSFC, brain volume, SF-36 PCS, and NEDA for Study WA21092.

Figure 1 Hierarchical Order of Key Efficacy Endpoints



ARR = annualized relapse rate; CDI = Confirmed Disability Improvement; CDP = confirmed disability progression; Gd = gadolinium; MSFC = Multiple Sclerosis Functional Composite Scale; NEDA = no evidence of disease activity; SF-36 PCS = short form 36 Physical Component Summary.

4.4.2.3 Time to Onset of Confirmed Disability Progression for at Least 12 Weeks with the Initial Event of Neurological Worsening Occurring during the 96-Week, Double-Blind, Double-Dummy Treatment Period

Disability progression as measured by EDSS will be assessed in all patients by the independent examining investigator at screening and baseline visit, and every 12 weeks throughout the study until the end of the double-blind, double-dummy treatment period at Week 96, after which disability progression is assessed at Weeks 12 and 24 in the OLE

and then every 24 weeks thereafter. Additional EDSS assessments for individual patients may be requested between visits (i.e., during an MS relapse).

Section 2.4 provides additional information on the timing of this analysis.

The EDSS is based on a standard neurological examination; the seven categories of the EDSS representing functional systems (Pyramidal, Cerebellar, Brainstem, Sensory, Bowel and Bladder, Visual, plus "Other") are rated and scored (collectively, FSS). Each score of the FSS is an ordinal clinical rating scale ranging from 0 to 5 or 6. These ratings are then used in conjunction with observations and information concerning ambulation and use of assistive devices (which will also be scored) to determine the EDSS score. The EDSS is a disability scale that ranges in 0.5-point steps from 0 (normal) to 10.0 (death).

Because of the variability of EDSS, a single measurement at the baseline visit might not provide a sufficiently reliable baseline EDSS. Therefore, using an average score from two separate measurements may improve the reliability of baseline EDSS and the reliability of confirmed disability progression seen in the study. Inclusion criteria require neurological stability for 30 days prior to both screening and baseline visits. Generally, it takes 2 weeks for screening. There should be little change in EDSS except for the measurement variability during the screening period, assuming patients are clinically neurologically stable during this period.

In the derivation of time-to-onset confirmed-disability progression, baseline EDSS is defined as the average of the EDSS scores at the screening and baseline visits, without rounding.

Progression is defined a ≥ 1–point increase in EDSS score from a baseline EDSS score of 0.0–5.5 inclusive, and a 0.5-increase from a baseline EDSS score higher than 5.5.

For example, for a patient with a baseline EDSS score of 5.25 or 5.5, the progression is defined as an EDSS score of at least 6.5. For a patient with a baseline EDSS score of 5.75 or 6.0, the progression is defined as an EDSS score of at least 6.5.

The inclusion criterion of EDSS (0–5.5) only applies to screening EDSS. Hence, it is still possible for a patient's baseline EDSS score (derived from both screening and Day 1 EDSS results) to be >5.5. Confirmation requires the sustained change in EDSS for at least 12 weeks from the initial progression. Initial progression can happen at any visit during the 96-week, double-blind, double-dummy treatment period.

Confirmation of disability progression must occur at the regularly scheduled visit that is at least 12 weeks (84 days) after initial progression. If a patient has a missing EDSS at the scheduled visit occurring at least 84 days after an initial progression or the scheduled visit occurs several days before the 84-day window after an initial progression

(e.g., the visit window is ± 4 days), confirmation of the disability progression must be on the basis of the assessment at the next scheduled visit. There may be EDSS assessments at unscheduled or scheduled visits that are < 84 days after the initial progression that are between the initial progression visit and the confirmation visit. Disability progression cannot be confirmed unless the EDSS values meet the minimum change required for progression. The non-confirmatory EDSS assessments (if any) between the initial disability progression and the confirmation of disability progression should be at least as high as the minimum change required for progression. All initial disability progression events up to Week 96 with corresponding confirmation visits at a subsequent scheduled visit will be taken into account for the statistical analysis, irrespective of whether the confirmation visit occurred during the treatment phase or in the SFU phase or OLE phase after study-drug discontinuation. Thus, patients who prematurely discontinue study drug treatment should remain in the SFU phase of the study, and every effort should be made to obtain a follow-up EDSS status at the next scheduled visit. Patients who, according to the above definition, did not have onset of confirmed disability progression by the Week−96 visit, by the time of the early discontinuation of treatment, or are lost to follow up will be censored at the date of the last EDSS assessment during the 96-week, double-blind, double-dummy treatment period.

Data from the two studies (with respect to the ocrelizumab group vs. the interferon beta-1a 44 μg SC group) will be pooled for analysis of this endpoint at the $\alpha = 0.05$ level. Only data collected during the 96-week treatment period will be used for this analysis, although confirmation of disability progression during the 96-week, double-blind, double-dummy treatment period may occur during a subsequent study phase (i.e., during the SFU phase or the OLE phase).

Time to confirmed disability progression in the ocrelizumab group and in the interferon beta-1a 44 μg SC group will be compared using a two-sided log-rank test stratifying by geographical region (United States vs. ROW), and baseline EDSS (< 4.0 vs. ≥ 4.0). The proportion of patients with confirmed disability progression will be estimated using Kaplan-Meier methodology. A by-treatment Kaplan-Meier plot will be presented for this endpoint. The overall hazard ratio will be estimated using a stratified Cox regression model with treatment group as covariate and the same stratification factors used in the stratified log-rank test above.

Sample SAS code can be found below (SAS code is regarded as “draft” until fully validated at the analysis stage):

```
* Log-rank test;
    proc lifetest data=CDP12;
        time CDP12TIME*CNSR(1);
        strata REGION BEDSSCAT / group=ARMCD;
        ods output HomTests =_logrank;
    run;

* Cox regression model;
proc phreg data=CDP12;
```

```
    class ARMCD / descending;
    model CDP12TIME*CNSR(1) = ARMCD / rl;
    strata REGION BEDSSCAT;
    ods output ParameterEstimates=est;
run;
```

The analysis described above will be repeated for the per-protocol population as a sensitivity analysis.

Additionally, the following sensitivity analyses will be performed:

- Use of two different methods for handling of missing data:

  Method 1: with use of the same kind of multiple imputation approach described at the end of Section 4.4.1, patients with an initial disability progression during the 96-week, double-blind, double-dummy treatment period who discontinue the treatment early and do not have a subsequent visit with EDSS measurement will be randomly imputed as an event of confirmed disability progression at the date of initial disability progression with a probability of 50%; the other 50% of these patients will be censored at the date of initial disability progression. A total of 1000 (m = 1000) imputed datasets will be produced.

  Method 2: Patients with an initial disability progression during the 96-week, double-blind, double-dummy treatment period who discontinue the treatment early and do not have a subsequent visit with EDSS measurement will be imputed as an event of confirmed disability progression at the date of initial disability progression.

- Adjustment for some additional baseline factors: Sensitivity analysis will use the same tests above but will, in addition, adjust for the number of relapses occurring within a 2-year period prior to study entry, baseline presence of Gd lesions (present or absent), prior MS treatment, and age (<40 vs. ≥ 40 years).

4.4.2.3.1 OLE Phase

CDP12 analysis cannot be repeated for the OLE phase since EDSS assessments are performed only at every 24 weeks post OLE Week 24.

4.4.2.4 Total Number of T1 Gadolinium–Enhanced Lesions as Detected by Brain Magnetic Resonance Imaging at Weeks 24, 48, and 96

The total number of T1 Gd–enhanced lesions will be calculated as the sum of the individual number of T1 Gd–enhanced lesions at Weeks 24, 48, and 96. Data from other unscheduled assessments will not be included in this summary or analysis.

The same approach to analysis will be applied here as described in Section 4.4.1, “Analysis Methods” section, with the following exceptions:

- In order to account for patients receiving varying numbers of brain MRI scans during the study, the log-transformed number of brain MRI scans received will be included in the model as an “offset” variable for appropriate computation. This approach is preferable to imputation of missing values using the average of non-missing

observations, because only three timepoints would give a high variability in the imputed average values. Additionally, the "offset" variable approach aligns more closely with approaches to analysis used a) in regulatory submissions for other MS therapies, b) for the primary endpoint, and c) in this program's Phase II study.

- Baseline covariates here will be as follows: baseline Gd lesion (present or not), treatment group, baseline EDSS score (<4.0 vs. ≥ 4.0), and geographical region (United States vs. ROW).

Sample SAS code can be found below (SAS code is regarded as "draft" until fully validated at the analysis stage):

```
proc genmod data=MRIT1;
    class ARMCD REGION BEDSSCAT BGDLESFL;
    model N_MRIT1 = ARMCD REGION BEDSSCAT BGDLESFL /
offset=MRILOG link=log dist=negbin type3;
    lsmeans ARMCD / exp cl;
    ods output lsmeans=lsm;
    ods output ParameterEstimates=est;
run;
```

If the model fails to converge due to high number of zero T1 lesion counts, a logistic regression model will be performed on Gd lesion (present or not) adjusted for the same baseline variables. Patients withdrawn from treatment and having no T1 lesions: those withdrawn because of lack of efficacy or death are considered as having T1 lesions; otherwise, it will be considered a missing observation:

```
proc logistic data=MRIT1;
    class ARMCD REGION BEDSSCAT BGDLESFL;
    model N_MRIT1_CAT = ARMCD REGION BEDSSCAT BGDLESFL;
    oddsratio ARMCD /cl=wald;
    ods output OddsRatiosWald = oddsr;
run;
```

4.4.2.4.1 OLE Phase

This analysis will be modified to include data collected during the OLE period. This analysis will be an exploratory analysis and will be performed for each visit from baseline up to the last mature visit. The last mature visit is defined as the visit where the majority of patients have completed prior to the data cutoff date. Of note, MRI is only collected once a year during OLE.

Adjusted rate and adjusted rate ratio based on the negative binomial models may not be estimable due to low number of lesions. Hence, unadjusted rate (total number of lesions divided by number of scans), unadjusted rate ratio and corresponding confidence intervals based on the exact poisson test will be presented.

4.4.2.5 Total Number of New and/or Enlarging T2 Hyperintense Lesions as Detected by Brain Magnetic Resonance Imaging at Weeks 24, 48, and 96

The total number of new and/or enlarging T2 lesions will be calculated as the sum of the individual number of new and/or enlarging lesions at Weeks 24, 48, and 96. Data from other unscheduled assessments will not be included in this summary or analysis.

- The same approach to analysis will be applied here as described in Section 4.4.2.4, with the following exception.
- Baseline covariates here will be as follows: baseline T2 lesion count, treatment group, baseline EDSS score (<4.0 vs. ≥4.0) and geographical region (United States vs. ROW).

Sample SAS code can be found below (SAS code is regarded as "draft" until fully validated at the analysis stage):

```
proc genmod data=MRIT2;
    class ARMCD REGION BEDSSCAT;
    model N_MRIT2 = ARMCD REGION BEDSSCAT BT2LES /
offset=MRILOG link=log dist=negbin type3;
    lsmeans ARMCD / exp cl;
    ods output lsmeans=lsm;
    ods output ParameterEstimates=est;
run;
```

If the model fails to converge due to a high number of zero T2 lesion counts, a logistic regression model will be performed on new/enlarging T2 lesion (present or not) adjusted for the same baseline variables. Patients withdrawn from treatment and having no T2 lesions: those withdrawn due to lack of efficacy or death are considered as having T2 lesions; otherwise, it will be considered a missing observation.

4.4.2.5.1 OLE Phase

This analysis will be modified to include data collected during the OLE period. This analysis will be an exploratory analysis and will be performed for each visit from baseline up to the last mature visit. The last mature visit is defined as the visit where the majority of patients have completed prior to the data cutoff date. Of note, MRI is only collected once a year during OLE.

Adjusted rate and adjusted rate ratio based on the negative binomial models may not be estimable due to low number of lesions. Hence, unadjusted rate (total number of lesions divided by number of scans), unadjusted rate ratio and corresponding confidence intervals based on the exact poisson test will be presented.

4.4.2.6 Proportion of Patients Who Have Confirmed Disability Improvement Sustained for at Least 12 Weeks with the Initial Event of Neurological Improvement Occurring during the 96-Week, Double-Blind, Double-Dummy Treatment Period

This endpoint will be analyzed only for the subgroup of patients with a baseline EDSS score≥2.0. Exactly the same approach to data derivation will be used for disability improvement as for disability progression (refer to Section 4.4.2.3, although note that here the endpoint is a binary improved/not improved variable, rather than a time-to-event endpoint). In particular, the same approach to the timing of the confirmation of disability improvement will be applied as for disability progression. The baseline EDSS is the average of the screening EDSS score and the score of the baseline visit, without rounding. For patients with a baseline EDSS score≥2 and≤5.5, disability improvement is defined as a reduction in EDSS score≥1.0 compared to baseline EDSS score. For patients with a baseline EDSS score>5.5, disability improvement is defined as a

reduction in EDSS score of ≥0.5. All patients without disability improvement will be counted as not improved, independent of follow-up time

This analysis requires both Studies WA21092 and WA21093 to be combined to have sufficient statistical power to detect relevant treatment differences. In the study-level analysis, the p-value for this treatment comparison will be interpreted as non-confirmatory.

The proportions in treatment groups will be compared using the Cochran-Mantel-Haenszel (CMH) χ^2 test stratified by geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs. ≥4.0).

Sample SAS code can be found below (SAS code is regarded as "draft" until fully validated at the analysis stage):

```
proc freq data=CDI12;
    tables REGION*BEDSSCAT*ARMCD*CDI12 / cmh;
    ods output CMH=pval;
    ods output CommonRelRisks=risks;
run;
```

4.4.2.6.1 OLE Phase

CDI12 analysis won't be repeated for the OLE phase since EDSS is only collected every 24 weeks post OLE Week 24.

4.4.2.7 Time to Onset of Confirmed Disability Progression for at Least 24 Weeks with the Initial Event of Neurological Worsening Occurring during the 96-Week, Double-Blind, Double-Dummy Treatment Period

This is the same as the 12-week confirmed disability progression, except that the confirmation of disability progression must occur at the regularly scheduled visit that is ≥24 weeks (≥161 days) after initial disability progression.

All initial disability progression events up to Week 96 with corresponding confirmation visits at the next schedule visit will be taken into account for the statistical analysis. The same analysis principles as described in Section 4.4.2.3 will be applied to the 24-week disability endpoint. The sensitivity analyses will be performed for the 24-week disability endpoint in the same way as for the 12-week endpoint.

Sample SAS code can be found below (SAS code is regarded as "draft" until fully validated at the analysis stage):

```
* Log-rank test;
    proc lifetest data=CDP24;
        time CDP24TIME*CNSR(1);
        strata REGION BEDSSCAT / group=ARMCD;
        ods output HomTests =_logrank;
    run;

* Cox regression model;
    proc phreg data=CDP24;
        class ARMCD / descending;
```

```
    model CDP24TIME*CNSR(1) = ARMCD / rl;
    strata REGION BEDSSCAT;
    ods output ParameterEstimates=est;
run;
```

4.4.2.7.1 OLE Phase

This analysis will be repeated to include EDSS assessments performed during the OLE period for the combined DB+OLE period, with the same baseline as for the analysis during DB.

Analysis containing only data from OLE will be performed considering the OLE baseline.

Moreover, the difference of survival curves will be analyzed every 24 weeks. Greenwood formula will be used to compute the corresponding 95% CI and p-value.

4.4.2.8 Total Number of New T1-Hypo-Intense Lesions (Chronic Black Holes) at Weeks 24, 48, and 96

The total number of new T1-hypo-intense lesions (chronic black holes) will be calculated as the sum of the individual number of new lesions at Weeks 24, 48, and 96. Data from other unscheduled assessments will not be included in this summary or analysis.

The same approach to analysis will be applied here as described in Section 4.4.2.4, with the following exception:

- Baseline covariates here will be as follows: treatment group, baseline T1-hypo-intense lesions count, baseline EDSS score (<4.0 vs. ≥ 4.0) and geographical region (United States vs. ROW).

Sample SAS code can be found below (SAS code is regarded as “draft” until fully validated at the analysis stage):

```
proc genmod data=MRICBH;
    class ARMCD REGION BEDSSCAT;
    model N_MRICBH = ARMCD REGION BEDSSCAT BT1LES /
offset=MRILOG link=log dist=negbin type3;
    lsmeans ARMCD / exp cl;
    ods output lsmeans=lsm;
    ods output ParameterEstimates=est;
run;
```

If the model fails to converge due to high number of zero chronic black hole counts, a logistic regression model will be performed on chronic black hole (present or not), adjusted for the same baseline variables. Patients withdrawn from treatment and having no chronic black holes: those withdrawn due to lack of efficacy or death are considered as having chronic black holes; otherwise, it will be considered a missing observation.

Note that the protocol and previous versions of this SAP did not make it clear that only new T1 hypo-intense lesions would be considered.

4.4.2.8.1 OLE Phase

This analysis will be modified to include data collected during the OLE period. This analysis will be an exploratory analysis and will be performed for each visit from baseline

up to the last mature visit. The last mature visit is defined as the visit where the majority of patients have completed prior to the data cutoff date. Of note, MRI is only collected once a year during OLE.

Adjusted rate and adjusted rate ratio based on the negative binomial models may not be estimable due to low number of lesions. Hence, unadjusted rate (total number of lesions divided by number of scans), unadjusted rate ratio and corresponding confidence intervals based on the exact poisson test will be presented.

4.4.2.9 Change in Multiple Sclerosis Functional Composite Scale Score from Baseline to Week 96

There are three primary measures in a MSFC score: 1) Timed 25-Foot walk; 2) 9-Hole Peg Test (9-HPT); and 3) Paced Auditory Serial Addition Test (PASAT-3 version). The MSFCS is based on the concept that scores for these three dimensions (arm, leg, and cognitive function) are combined to create a single score (the MSFC) that can be used to detect change over time in a group of patients with MS. Since the three primary measures differ in what they actually measure (time for the 9-HPT and 25-Foot Timed Walk, but number of correct answers for the PASAT-3), a common metric for these variables has been used to create a composite score for the three different measures. Z-score was selected for this purpose. The results from each of these three tests are transformed into Z-scores and averaged to yield a composite score for each patient at each timepoint.

$$\text{MSFC Score} = \{Z_{arm}, \text{average} + Z_{leg}, \text{average} + Z_{cognitive}\} / 3.0$$

Where Zxxx = Z-score

The following are MSFC components:

- Trial 1, Timed 25-Foot Walk
- Trial 2, Timed 25-Foot Walk
- Trial 1, Dominant Hand, 9-HPT
- Trial 2, Dominant Hand, 9-HPT
- Trial 1, Non-Dominant Hand, 9-HPT
- Trial 2, Non-Dominant Hand, 9-HPT
- PASAT-3

There are three derived results for three tests: 1) the average (reciprocal) scores from the four trials on the 9-HPT (the two trials for each hand are averaged, converted to the reciprocals of the mean times for each hand and then the two reciprocals are averaged); 2) the average scores of two 25-Foot Timed Walk trials; and 3) the number correct from the PASAT-3. The Z-score will be created for these three derived results to yield the composite score by the following formula.

Formula for Creating the MSFC Score to Compare Groups within a Study (The Preferred Method)

MSFC Score = { {(Average $(_{1/9\text{-HPT}})$ − Baseline Mean $(_{1/9\text{-HPT}})$)/ Baseline Std Dev $(_{1/9\ \text{HPT}})$}
+ {(Average $_{25\text{-Foot Walk}}$ − Baseline Mean $_{25\text{-Foot Walk}}$)/Baseline Std-Dev $_{25\text{-Foot Walk}}$}
+{(PASAT-3 − Baseline Mean $_{\text{PASAT-3}}$) /Baseline Std Dev $_{\text{PASAT-3}}$} } / 3.0

Note: "Average (1/9-HPT)" is the average of the inverse (reciprocal) for the mean time of the two trials on the right hand and reciprocal of the mean time of the two left-hand trials from the test patient; Baseline Mean (1/9-HPT) and Std Dev (1/9-HPT) are the baseline values from each patient in all study groups combined at the baseline assessment; "Average 25-Foot Walk" is the mean time from the two trials of the 25-foot timed walk; and we take the negative value of the Z-score to make the direction of change the same as the other components. Similarly, the Baseline Mean and Std Dev 25-Foot Walk are of all Baseline Groups combined; and "PASAT-3" is the score from the test patient, and the Baseline Mean PASAT-3 and Std Dev PASAT-3 of the combined baseline assessments. These are discussed below.

Note that changing the signs of the Z-scores for the 25-Foot Walk and using inverse value of 9-HPT to ensure that a negative value implies worsening, and a positive value implies improvement. More details can be found in the MSFC Manual.

- Missing data will be imputed as follows:
- For results within a given test:

  25-Foot Timed Walk trials: If the test results are not available due to a "physical limitation," impute the maximum possible value for the scale. If the test results are not available but it is NOT due to a "physical limitation," use the result from the other trial to impute the missing value.

  9-HPT: If the test results are not available due to a "physical limitation," impute the maximum possible value for the scale. If the test results are not available but it is NOT due to a "physical limitation," impute the missing test score using the result from the other trial of the same hand. If the result from the other trial of the same hand is also missing, use the average score from the other hand (or available score from one trial if the result from one trial only is available).

- For the composite score:

  All available data will be used for analysis regardless of the timing of any protocol-defined relapse.

Outliers will be applied with the following rules:

- For 9-HPT, the lower bound is set to be 10 seconds (Oxford Grice et al. 2003), and the upper bound is set to be 300 seconds, according to the MSFC Manual.
- For 25-FWT, the lower bound is set to be 2.2 seconds (Bohannon 1997), and upper bound is set to be 180 seconds, according to the MSFC Manual.

- For PASAT, the lower and upper bounds are set to be 0 and 60 according to the MSFC Manual.
- For values outside the lower and upper bounds, they will be treated as missing, and the imputation rule applied as defined above.

The MSFC score will be summarized at baseline and at Weeks 12, 24, 36, 48, 60, 72, 84, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The change from baseline to Weeks 12, 24, 36, 48, 60, 72, 84, and 96 in MSFC score will be computed and summarized using descriptive statistics for each treatment group.

For the assessment of differences in the mean change from baseline in MSFC score up to Week 96, a Mixed-Effect Model Repeated Measures analysis (MMRM) incorporating post-randomization data collected up to 96 weeks of treatment will be used to assess all data collected over time with consideration of the variance−covariance matrix of the repeated measures. This method allows for a general unstructured variance−covariance matrix and will include data from patients with incomplete data from some scheduled timepoints.

The model will be implemented in SAS using the MIXED procedure and will include the change from baseline in MSFC score as the dependent variable. The fixed effects in the model will include independent variables of randomized treatment, visit (nominal post $\leq$baseline visits as per the Schedule of Assessments) and treatment-by-visit interaction, along with the following baseline covariates: baseline MSFC score, baseline MSFC score by visit interaction, geographical region (United States vs. ROW), and baseline EDSS score ($<$4.0 vs$\geq$ 4.0). Visit will be treated as a repeated variable within a patient. Patient, treatment, and visit will be treated as factor variables. An unstructured variance−covariance structure will be applied to model the within-patient errors. The model will be fitted using the Restricted Maximum Likelihood method (REML). Denominator degrees of freedom will be estimated using Satterthwaite's approximation.

To estimate the difference between the ocrelizumab and interferon beta-1a 44 μg SC groups in mean change from baseline to Week 96, a treatment-by-visit interaction contrast will be constructed (i.e., the treatment group contrast at Week 96). On the basis of this analysis, least square means, standard errors, and the 95% confidence interval for the treatment difference will be reported.

Graphical presentations for least square means and 95% confidence intervals will be used to illustrate trends over time.

Sample SAS code can be found below (SAS code is regarded as "draft" until fully validated at the analysis stage):

```
proc mixed data=MSFCS method=REML;
    class REGION BEDSSCAT ARMCD VISIT USUBJID;
    model change  =  BFCS  REGION  BEDSSCAT  ARMCD  VISIT
```

```
   BFCS*VISIT  ARMCD*VISIT / ddfm= satterthwaite;
    repeated VISIT / type=un subject = USUBJID;
    lsmeans ARMCD*VISIT / pdiff cl;
    ods output lsmeans = lsm;  * contains the adjusted means;
    ods output diffs = dif;  * contains treatment differences;
run;
```

4.4.2.9.1 OLE Phase

This analysis won't be repeated for the OLE phase as this data is not being collected during the OLE phase.

4.4.2.10 Percentage Change in Brain Volume as Detected by Brain Magnetic Resonance Imaging Scan from Week 24 to Week 96

Brain volume is recorded as an absolute "normalized" value at the baseline visit then recorded at subsequent visits as a percentage change relative to the absolute value at the baseline visit. Therefore, brain volume at Week 24 will be calculated as the brain volume at the baseline visit multiplied by 1+([percentage change in brain volume from baseline visit to Week 24]/100). This value will be used to determine the percentage change in brain volume at Weeks 48 and 96 relative to Week 24.

Percentage change in brain volume relative to Week 24, as detected by brain MRI scan, will be computed and summarized at Weeks 48 and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis.

For the assessment of differences in the mean percentage change in brain volume on MRI scans from Week 24 to Week 96, an MMRM analysis will be undertaken using the same approach as described in Section 4.4.2.9, except for the following:

- The response variable will be the percentage change from Week 24 to subsequent visits (Weeks 48 and 96). Baseline brain volume data will not feature in this analysis.
- Baseline covariates here will be as follows: brain volume at Week 24, geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs ≥ 4.0)
- To estimate the difference between the ocrelizumab and interferon beta-1a 44 μg SC groups in mean change from Week 24 to Week 96, a treatment-by-visit interaction contrast will be constructed (i.e., the treatment group contrast at Week 96)

Sample SAS code can be found below (SAS code is regarded as "draft" until fully validated at the analysis stage):

```
proc mixed data=MRINBV method=REML;
    class REGION BEDSSCAT ARMCD VISIT USUBJID;
    model change  =  W24NBV  REGION  BEDSSCAT  ARMCD  VISIT
   W24NBV*VISIT  ARMCD*VISIT / ddfm= satterthwaite;
    repeated VISIT / type=un subject = USUBJID;
    lsmeans ARMCD*VISIT / pdiff cl;
    ods output lsmeans = lsm;  * contains the adjusted means;
    ods output diffs = dif;  * contains treatment differences;
run;
```

The choice of Week 24 as “baseline” in this analysis, rather than the baseline visit, is a widely accepted approach to establish a stable baseline and avoid the effect of pseudoatrophy commonly reported for immunomodulatory therapies.

4.4.2.10.1 OLE Phase

This analysis will be extended to include data collected during the OLE period. Percentage change from Week 24, baseline and OLE baseline until the last mature visit will be analyzed.

4.4.2.11 Change in Quality of Life, as Measured by the Short Form 36 Version 2 Physical Component Summary Score from Baseline to Week 96

The Short Form 36 Version 2 (SF-36v2) is a 36-item, self-reported, generic measure of quality of life that has been widely used in multiple disease areas. It is composed of eight health domains: Physical Functioning (PF), Role-Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role-Emotional (RE), and Mental Health (ME). On the basis of these domain scores, the PCS Score and the Mental Component Summary (MCS) Score will be computed. The standard form of the instrument, with a 4-week recall, is being administered in the studies.

Scoring and rules for missing items will follow the SF-36v2 User’s Manual. In brief, scoring for each health domain scale involves (a) recoding item response values, (b) summing recoded response values for all items in a given scale to obtain the scale raw score, and (c) transforming the scale raw score to a 0−100 score. The PCS score is computed by (a) multiplying each health domain Z-score by a scale-specific physical factor score coefficient, (b) summing the resulting products, and (c) converting the product total to a T score. The MCS score is computed in the same manner, using scale-specific mental factor score coefficients.

The PCS scores will be summarized at baseline and Weeks 48 and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The change from baseline to Weeks 48 and 96 in PCS scores will be computed and summarized using descriptive statistics for each treatment group.

For the assessment of differences in the mean change in PCS score during the 96-week comparative treatment period from baseline up to Week 96, MMRM analyses will be undertaken using the same approach as described above in Section 4.4.2.9, except baseline covariates here will be as follows: Baseline PCS score, geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

Sample SAS code can be found below (SAS code is regarded as “draft” until fully validated at the analysis stage):

```
proc mixed data=PCSS method=REML;
    class REGION BEDSSCAT ARMCD VISIT USUBJID;
```

```
    model change  =  BPCS  REGION  BEDSSCAT  ARMCD  VISIT
   BPCS*VISIT  ARMCD*VISIT / ddfm= satterthwaite;
    repeated VISIT / type=un subject = USUBJID;
    lsmeans ARMCD*VISIT / pdiff cl;
    ods output lsmeans = lsm;  * contains the adjusted means;
    ods output diffs = dif;  * contains treatment differences;
run;
```

4.4.2.11.1 OLE Phase

Additional data from the OLE period will be descriptively analyzed.

4.4.2.12 Proportion of Patients Who Have No Evidence of Disease Activity by Week 96

This endpoint will be defined only for those patients with a baseline EDSS score $\geq$2.0. All available data during the 96-week treatment period will be used for the analysis. Patients who complete the 96-week treatment period will be considered as having evidence of disease activity if at least one protocol-defined relapse, a CDP event or at least one MRI scan showing MRI activity (defined as Gd-enhancing T1 lesions, or new or enlarging T2 lesions) was reported during the 96-week treatment period, otherwise the patient will be considered as having NEDA. Patients who discontinue treatment early with at least one event before early discontinuation will be considered as having evidence of disease activity.

Even if an event was not reported before early discontinuation, the patient will be considered as having evidence of disease activity if the reason for early discontinuation is lack of efficacy or death; otherwise, it will be considered a missing observation.

Only data collected during the 96-week treatment period will be used for this analysis, although confirmation of disability progression during the 96-week, double-blind, double-dummy treatment period may occur during a subsequent study phase (i.e., during the SFU phase or the OLE phase).

The proportion of patients with NEDA within treatment groups will be compared using the CMH χ^2 test stratified by geographical region (United States vs. ROW), and baseline EDSS score ($<$4.0 vs.$\geq$ 4.0).

Sample SAS code can be found below (SAS code is regarded as “draft” until fully validated at the analysis stage):

```
proc freq data=NEDA;
    tables REGION*BEDSSCAT*ARMCD*NEDA / cmh;
    ods output CMH=pval;
    ods output CommonRelRisks=risks;
run;
```

4.4.2.12.1 OLE Phase

Proportion of patients with NEDA between OLE start and OLE Week 96 will be presented by randomized arms in the OLE ITT population.

4.4.3 Exploratory Efficacy Endpoints

4.4.3.1 Proportion of Relapse-Free Patients by 96 Weeks

All available data during the 96-week treatment period will be used for the analysis. Patients who complete the 96-week treatment period will be considered as not relapse-free if at least one protocol-defined relapse was reported during the 96-week treatment period; otherwise, the patient will be considered as relapse-free. Patients who discontinued treatment early with at least one relapse before early discontinuation will be considered as not relapse-free. Even if the patient did not have a relapse before early discontinuation, if the reason for early discontinuation is lack of efficacy or death, the patient will be considered as not relapse-free; otherwise, it will be considered as a missing observation.

The proportions within treatment groups will be compared using CMH χ^2 test stratified by geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

4.4.3.1.1 OLE phase

Time to the first protocol defined relapse will also be analyzed during the DB and OLE period. Hazard ratio between randomized arms based on the Cox model, p-value based on the log-rank test and Kaplan-Meier estimates will be presented.

4.4.3.2 Percentage Change in Total T2 Lesion Volume as Detected by Brain Magnetic Resonance Imaging from Baseline to Week 96

The total volume of T2 lesions on MRI scans of the brain will be summarized at baseline and Weeks 24, 48, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The percentage change from baseline to Weeks 24, 48, and 96 in total volume of T2 lesions on MRI scans of the brain will be computed and summarized using descriptive statistics (e.g., mean, SD, median, minimum, maximum, and n) for each treatment group.

For the assessment of differences in the mean percentage change in total T2 lesion volume during the 96-week comparative treatment period from baseline up to Week 96, an MMRM analysis will be undertaken using the same approach as described in Section 4.4.2.9, except baseline covariates here will be as follows: Baseline T2 lesion volume, geographical region (United States vs. ROW), and baseline EDSS (<4.0 vs. ≥ 4.0).

This analysis has been modified as described in Section 4.4.3.23, due to the normality assumption for the percentage change in total T2 lesion volume being violated.

4.4.3.3 Annualized Relapse Rate, Based on All Clinical Relapses at the End of the 96-Week Comparative Treatment Period

All clinical relapses reported on the MS relapse page will be used for this analysis (protocol-defined relapse is a subset of clinical relapse).

The annualized clinical relapse rate by Week 96 will be analyzed using the same general approach as described in Section 4.4.1, "Analysis Methods" for protocol-defined relapse (primary efficacy endpoint). However, the sensitivity and robustness checks described in Section 4.4.1 will not be performed here.

4.4.3.4 Annualized Protocol-Defined Relapse Rate by 2 Years (96 weeks) for Relapses Requiring Intravenous Steroid Therapy

All protocol-defined relapses requiring treatment with IV steroids will be used for this analysis.

The annualized relapse rate by Week 96 will be analyzed using the same general approach as described in Section 4.4.1, "Analysis Methods" for protocol-defined relapse (primary efficacy endpoint). However, the sensitivity and robustness checks described in Section 4.4.1 will not be performed here.

4.4.3.5 Percentage Change in Brain Volume as Detected by Brain Magnetic Resonance Imaging from Baseline to Week 96

Percentage change in brain volume relative to the baseline visit, as detected by brain MRI scan, will be computed and summarized at Weeks 24, 48 and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis.

For the assessment of differences in the mean percentage change in brain volume on MRI scans during the 96-week comparative treatment period from baseline up to Week 96, an MMRM analysis will be undertaken using the same approach as described in Section 4.4.2.9, except baseline covariates here will be as follows: baseline brain volume, geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

4.4.3.5.1 OLE Phase

This analysis will be performed as described in Section 4.4.2.10.1.

4.4.3.6 Time to Onset of Confirmed Disability Progression for at Least 48 Weeks with the Initial Event of Neurological Worsening Occurring during the 96-Week, Double-Blind, Double-Dummy Treatment Period

This is the same as the 12-week confirmed disability progression, except that the confirmation of disability progression must occur at the regularly scheduled visit that is ≥ 48 weeks (≥ 329 days) after initial disability progression.

All initial disability progression events up to Week 96 with corresponding confirmation visits at the next schedule visit will be taken into account for the statistical analysis. The same analysis principles as described in Section 4.4.2.3 will be applied to the 48-week disability endpoint.

4.4.3.6.1 OLE Phase

This analysis will be repeated to include the EDSS assessments performed during the OLE period for the combined DB+OLE period, with the same baseline as for the analysis during DB. Analysis containing only data from OLE will be performed considering the OLE baseline.

Moreover, the difference of survival curves will be analyzed every 24 weeks. Greenwood formula will be used to compute the corresponding 95% CI and p-value.

4.4.3.7 Time to Confirmed EDSS≥6 for at least 24 weeks during 96-Week, Double-Blind, Double-Dummy Treatment Period

In the derivation of time-to-onset confirmed EDSS≥6, only patients with baseline≤5.5 will be considered. Baseline EDSS is defined as the average of the EDSS scores at the screening and baseline visits, without rounding. An increase in EDSS score to≥6.0 from a baseline EDSS≤5.5 during the 96-week, double-blind, double-dummy treatment period which sustains for at least 24 weeks will be considered as an event.

Data from the two studies will be pooled for analysis of this endpoint. Only events with onset during the 96-week treatment period will be used for this analysis, although confirmation of EDSS≥6 may occur during a subsequent study phase (i.e., during the SFU phase or the OLE phase).

Time to confirmed EDSS≥6 will be analyzed using a two-sided log-rank test stratifying by geographical region (United States vs. ROW), and baseline EDSS (<4.0 vs. ≥4.0). The proportion of patients with confirmed EDSS≥6 will be estimated using Kaplan-Meier methodology. A Kaplan-Meier plot will be presented for this endpoint by treatment groups. The overall hazard ratio will be estimated using a stratified Cox regression model with treatment group as covariate and stratified by the same stratification factors used in the stratified log-rank test above.

4.4.3.7.1 OLE Phase

This analysis will be repeated to include the EDSS assessments performed during the OLE period for the combined DB+OLE period, with the same baseline as for the analysis during DB. Analysis containing only data from OLE will be performed considering the OLE baseline.

Moreover, the difference of survival curves will be analyzed every 24 weeks. Greenwood formula will be used to compute the corresponding 95% CI and p-value.

4.4.3.8 Time to Confirmed EDSS≥6 for at least 48 weeks during 96-Week, Double-Blind, Double-Dummy Treatment Period

This is the same as the time to 24-week confirmed EDSS≥6, except that the confirmation must occur at the regularly scheduled visit that is≥48 weeks (≥329 days) after initial disability progression. The same analysis as described in Section 4.4.3.7 will be applied to the 48-week confirmed EDSS≥6 endpoint.

4.4.3.8.1 OLE Phase

This analysis will be repeated to include the EDSS assessments performed during the OLE period for the combined DB+OLE period, with the same baseline as for the analysis during DB. Analysis containing only data from OLE will be performed considering the OLE baseline.

Moreover, the difference of survival curves will be analyzed every 24 weeks. Greenwood formula will be used to compute the corresponding 95% CI and p-value.

4.4.3.9 Ratio in Timed 25-Foot Walk relative to Baseline at Week 96

The average time (for the 25-foot walk) from two trials will be used in this analysis. The imputation method for missing values in either of the two trials will be the same as described in Section 4.4.2.9.

All available data at Weeks 12, 24, 36, 48, 60, 72, 84 and 96 will be used for analysis, regardless of the timing of any protocol-defined relapses.

The timed 25-foot walk will be summarized at baseline and at Weeks 12, 24, 36, 48, 60, 72, 84, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The change from baseline to Weeks 12, 24, 36, 48, 60, 72, 84, and 96 in the timed 25-foot walk will be computed and summarized using descriptive statistics for each treatment group.

For the assessment of ratio in the timed 25-foot walk relative to baseline up to Week 96, MMRM analyses will be undertaken.

The model will be implemented in SAS using PROC MIXED and will include the log-transformed ratio of post-baseline to the baseline in timed 25-foot-walk at each visit as the dependent variable. The fixed effects in the model will include independent variables of randomized treatment, visit (nominal post-baseline visits as per the Schedule of Assessments), baseline-by-visit interaction and treatment-by-visit interaction, along with the following baseline covariates, log-transformed 25-foot-timed-walk at baseline visit, geographical region (U.S. vs. rest of world), and baseline EDSS score (<4.0 vs.≥4.0). Visit will be treated as a repeated variable within a patient. Patient, treatment and visit will be treated as factor variables. An unstructured variance−covariance structure will be applied to model the within-patient errors. The Restricted Maximum Likelihood method (REML) will be used for estimates of variance components. Denominator degrees of freedom will be estimated using Satterthwaite's approximation.

To estimate the difference between the ocrelizumab and interferon beta-1a 44 μg SC groups in ratio relative to baseline at Week 96, a treatment-by-visit interaction contrast will be constructed (i.e., the treatment group contrast at Week 96). On the basis of this analysis, back-transformed estimates of least square means, and the 95% confidence interval (CI) for the ratio of timed 25-foot walk at each visit relative to the baseline will be reported. Graphical presentations for least square means and 95% CIs will be used to illustrate trends over time.

Sample SAS code can be found below (SAS code is regarded as "draft" until fully validated at the analysis stage):

```
proc mixed data=T25FW method=REML;
        class REGION BEDSSCAT ARMCD VISIT USUBJID;
        model log(AVAL/BASE) = log(BASE) REGION BEDSSCAT
ARMCD VISIT log(BASE)*VISIT ARMCD*VISIT /
        ddfm= satterthwaite;
        repeated VISIT / type=un subject=USUBJID;
        lsmeans ARMCD*VISIT / pdiff cl;
        ods output lsmeans = lsm;  * contains the adjusted means;
        ods output diffs = dif;  * contains treatment differences;
    run;
```

4.4.3.9.1 OLE Phase

This analysis will be extended to include data collected during the OLE period. Ratio relative to baseline at all scheduled visits until the last mature visit during the combined DB+OLE period and ratio relative to OLE baseline at all scheduled visits until the last mature visit during the OLE period will be analyzed.

4.4.3.10 Ratio in 9-Hole Peg Test relative to Baseline at Week 96

The average score of the 9-hole peg test from four trials will be used in this analysis. The method for calculating the average score and the imputation method for missing values from any of the four trials will be the same as in Section 4.4.2.9.

All available data at Weeks 12, 24, 36, 48, 60, 72, 84, and 96 will be used for analysis regardless of the timing of any protocol-defined relapses.

The 9-hole peg test will be summarized at baseline and at Weeks 12, 24, 36, 48, 60, 72, 84, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The change from baseline to Weeks 12, 24, 36, 48, 60, 72, 84, and 96 in the 9-hole peg test will be computed and summarized using descriptive statistics for each treatment group.

For the assessment of ratio in the 9-hole peg test relative to baseline up to Week 96, MMRM analyses will be undertaken using the same approach as described in Section 4.4.3.9.

Baseline covariates here will be as follows: log-transformed baseline 9-hole peg test, geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

4.4.3.10.1 OLE Phase

This analysis will be extended to include data collected during the OLE period. Ratio relative to baseline at all scheduled visits until the last mature visit during the combined DB+OLE period and ratio relative to OLE baseline at all scheduled visits until the last mature visit during the OLE period will be analyzed.

4.4.3.11 Change in PASAT from Baseline to Week 96

All available data at Weeks 12, 24, 36, 48, 60, 72, 84 and 96 will be used for analysis, regardless of the timing of any protocol-defined relapses.

The PASAT will be summarized at baseline and at Weeks 12, 24, 36, 48, 60, 72, 84, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The change from baseline to Weeks 12, 24, 36, 48, 60, 72, 84, and 96 in the PASAT will be computed and summarized using descriptive statistics for each treatment group.

For the assessment of differences in the mean change in the PASAT during the 96-week comparative treatment period from baseline up to Week 96, MMRM analyses will be undertaken using the same approach as described in Section 4.4.2.9.

Baseline covariates here will be as follows: Baseline PASAT, geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

4.4.3.11.1 OLE Phase

This analysis won't be repeated for the OLE phase as this data is not being collected during the OLE period.

4.4.3.12 Change in Fatigue, as Measured by the Modified Fatigue Impact Scale Total Score from Baseline to Week 96

The Modified Fatigue Impact Scale (MFIS) measures fatigue effects in terms of physical, cognitive, and psychological impact. A higher score indicates greater impact of fatigue on a patient's activities.

- The Physical Subscale is the sum of items 4, 6, 7, 10, 13, 14, 17, 20, and 21 on the MFIS. Range is 0–36.
- The Cognitive Subscale is the sum of items 1, 2, 3, 5, 11, 12, 15, 16, 18, and 19. Range is 0–40.
- The Psychosocial Subscale is the sum of items 8 and 9. Range is 0–8.
- The Total MFIS score is the sum of all items. Range is 0–84.

Subscale scores will not be computed for patients with missing items on the subscale. Per the developer of the questionnaire, the Total MFIS score may be computed if up to two items are missing by first dividing the total score by the number of items answered and then multiplying by the number of total items in the scale. If more than two items are missing, the form will be set to missing.

The MFIS total score will be summarized at baseline and Weeks 48 and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The change from baseline to Weeks 48 and 96 in the MFIS total score will be computed and summarized using descriptive statistics for each treatment group. The MFIS subscale scores will only be listed.

For the assessment of differences in the mean change in MFIS total during the 96-week comparative treatment period from baseline up to Week 96, MMRM analyses will be undertaken using the same approach as described above in Section 4.4.2.9, except baseline covariates here will be as follows: baseline total MFIS score, geographical region, and baseline EDSS score ($<$4.0 vs. $\geq$4.0).

4.4.3.13 Change in Patient-Reported Depressive Symptoms, as Measured by the Center for Epidemiologic Studies Depression Scale, from Baseline to Week 96

The CES-D is a 20-item self-report questionnaire measuring depressive symptomology. It includes mood, well-being, somatic, and interpersonal domains.

Items are scored using the following table:

Table 3 Center for Epidemiologic Studies Depression Scale

CES-D Items	Rarely or none of the time	Some or a little of the time	Occasionally or a moderate amount of time	Most or all of the time
4, 8, 12, and 16	3	2	1	0
All other items	0	1	2	3

CES-D = Center for Epidemiologic Studies Depression Scale.

A total CES-D score is computed by summing the scored items. The total score ranges 0–60. If four or fewer items on a patient's questionnaire are missing, the computational mean is used for imputation. If more than four items are missing, the form is set to missing.

The total CES-D score will be summarized at baseline and Weeks 48 and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The change from baseline to Weeks 48 and 96 in total CES-D score will be computed and summarized using descriptive statistics for each treatment group.

For the assessment of differences in the mean change in total CES-D score during the 96-week comparative treatment period from baseline up to Week 96, an MMRM analysis will be undertaken using the same approach as described in Section 4.4.2.9, except baseline covariates here will be as follows: baseline total CES-D score, geographical region (United States vs. ROW), and baseline EDSS score ($<$4.0 vs. $\geq$ 4.0).

4.4.3.14 Change in Karnofsky Performance Status Score from Baseline to Week 96

The Karnofsky Performance Scale allows for the classification of patients according to their functional impairment. This scale is usually used to compare effectiveness of different therapies and to assess the prognosis in individual patients. The Karnofsky score ranges 100−0, where 100 is “perfect” health (normal, no complaints, or no evidence of disease) and 0 is death. Scores at 10-point intervals between 100 and 0 are assigned to represent health status.

The Karnofsky Performance Status Score will be summarized at baseline and Weeks 24, 48, 72, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The change from baseline to Weeks 24, 48, and 96 in Karnofsky Performance Status Score will be computed and summarized using descriptive statistics for each treatment group (i.e., frequency counts or summary statistics, as appropriate).

For the assessment of differences in the mean change in Karnofsky Performance Status Score during the 96-week comparative treatment period from baseline up to Week 96, an MMRM analysis will be undertaken using the same approach as described in Section 4.4.2.9, except baseline covariates here will be as follows: baseline Karnofsky Performance Status Score, geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

4.4.3.15 Percentage Change in Cortical Grey Matter Volume as Detected by Brain MRI from Baseline to Week 96

Cortical grey matter volume as assessed on MRI scans of the brain will be summarized at baseline and Weeks 24, 48, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The percentage change from baseline to Weeks 24, 48, and 96 in cortical grey matter volume on MRI scans of the brain will be computed and summarized using descriptive statistics for each treatment group.

For the assessment of differences in the mean percentage change in cortical grey matter volume on MRI scans during the 96-week comparative treatment period from baseline up to Week 96, an MMRM analysis will be undertaken using the same approach as described in Section 4.4.2.9, except baseline covariates here will be as follows: baseline cortical grey matter volume, geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

4.4.3.15.1 OLE Phase

This analysis will be extended to include data collected during the OLE period. Percentage change from Week 24, baseline and OLE baseline until the last mature visit will be analyzed.

4.4.3.16 Percentage Change in White Matter Volume as Detected by Brain MRI from Baseline to Week 96

White matter volume as assessed on MRI scans of the brain will be summarized at baseline and Weeks 24, 48, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The percentage change from baseline to Weeks 24, 48, and 96 in white matter volume on MRI scans of the brain will be computed and summarized using descriptive statistics for each treatment group.

For the assessment of differences in the mean percentage change in white matter volume on MRI scans during the 96-week comparative treatment period from baseline up to Week 96, an MMRM analysis will be undertaken using the same approach as described above in Section 4.4.2.9, except baseline covariates here will be as follows: baseline white matter volume, geographical region (United States vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

4.4.3.16.1 OLE Phase

This analysis will be extended to include data collected during the OLE period. Percentage change from Week 24, baseline and OLE baseline until the last mature visit will be analyzed.

4.4.3.17 Proportion of Patients who Have Disability Improvement Sustained for At Least 24 Weeks with the Initial Event of Neurological Improvement Occurring during the 96-Week, Double-Blind, Double-Dummy Treatment Period

This endpoint is the same as the 12-week confirmed disability improvement endpoint (refer to Section 4.4.2.6), but based on a 24-week period of disability improvement.

The same approach to analysis as described in Section 4.4.2.6 will be applied to the 24−week confirmed disability improvement endpoint.

4.4.3.17.1 OLE Phase

This analysis will be repeated to include the EDSS assessments performed during the OLE period for the combined DB+OLE period, with the same baseline as for the analysis during DB. Analysis containing only data from OLE will be performed considering the OLE baseline.

4.4.3.18 Proportion of Patients Who have Disability Improvement Sustained for at Least 12 Weeks and Sustained until the End of the 96-Week, Double-Blind, Double-Dummy Treatment Period, with the Initial Event of Neurological Improvement Occurring during the 96-Week, Double-Blind, Double-Dummy Treatment Period

This endpoint is similar to the 12-week confirmed disability improvement endpoint (see Section 4.4.2.6), but it is required that improvement is sustained until the end of the 96-

week double-blind double-dummy treatment period. Otherwise, patients will be classified as not improved.

The same approach to analysis as described in Section 4.4.2.6 will be applied here.

4.4.3.19 Duration of Confirmed Disability Improvement

This analysis will be performed for the patients with a Confirmed Disability Improvement for 12 weeks identified as in Section 4.4.2.6.

The duration of the Confirmed Disability Improvement is the time from the date of onset of the event, until the first visit where the EDSS score no longer fulfills the criteria for disability improvement. For the patients who have a sustained improvement until after the end of the double-blind treatment period, the duration of the confirmed disability improvement will be censored at the last EDSS score measured during the 96 week, double-blind, double-dummy treatment period. For patients with the onset of the CDI at the last visit during this period and confirmed in the open-label or safety follow-up phase, the duration of the CDI will be censored at Day 1. The Kaplan-Meier estimates of the duration of CDI will be calculated. No comparison between the randomized arms will be made, since this analysis is performed only for the patients with a CDI.

4.4.3.20 Proportion of Patients Who at Week 96 Have Improved, Stable or Worsened Disability, Compared with Baseline

Stable disability is defined as a difference in EDSS score of no greater than ± 0.5 comparing the Week 96 and baseline scores. Worsened disability is defined as an increase in EDSS score of > 0.5 at Week 96 compared to baseline. Improved disability is defined as a reduction in EDSS score of > 0.5 at Week 96 compared with baseline. The proportion of patients with improved, stable, or worsened disability compared with baseline in each treatment group will be reported. In addition, the proportion of patients with improved or stable disability (combining the proportion or patients improved and the proportion stable) in each treatment group will be reported.

Patients with missing EDSS data at baseline or Week 96 are excluded from this analysis. Interval EDSS scores are not taken into account in this analysis.

The proportions with improved, stable, and worsened disability within treatment groups will be analyzed using multinomial logistic regression with baseline covariates, geographical region (United States vs. ROW), and baseline EDSS score (< 4.0 vs. ≥ 4.0).

4.4.3.21 Change in EDSS Score from Baseline to Week 96

The EDSS scores will be summarized at baseline (without rounding) and Weeks 48 and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The change from baseline to Weeks 48 and 96 in EDSS scores will be computed and summarized using descriptive statistics for each treatment group.

For the assessment of differences in the mean change in EDSS score during the 96-week comparative treatment period from baseline up to Week 96, MMRM analyses will be undertaken using the same approach as described above in Section 4.4.2.9, except baseline covariates here will be as follows: geographical region (United States vs. ROW) and baseline EDSS (continuous without rounding).

Graphical presentations for least square means and 95% confidence intervals will be used to illustrate trends over time.

4.4.3.21.1 OLE Phase

This analysis will be extended to include data collected during the OLE period. Change from baseline at all scheduled visits until the last mature visit during the combined DB+OLE period and change from OLE baseline at all scheduled visits until the last mature visit during the OLE period will be analyzed.

4.4.3.22 Change in Quality of Life, as Measured by the Short Form 36 Version 2 Mental Component Summary Score from Baseline to Week 96

The derivation of the SF-36 MCS Score was described previously in Section 4.4.2.11.

The MCS Score will be summarized and analyzed in exactly the same way as the SF-36 PCS Score, as described in Section 4.4.2.11.

4.4.3.22.1 OLE Phase

Additional data from the OLE period will be descriptively analyzed.

4.4.3.23 Ratio in Total T2 Lesion Volume as Detected by Brain MRI Scans relative to Baseline at Week 96

The total volume of T2 lesions on MRI scans of the brain will be summarized at baseline and Weeks 24, 48, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The percentage change from baseline to Weeks 24, 48, and 96 in total volume of T2 lesions on MRI scans of the brain will be computed and summarized using descriptive statistics (e.g., mean, SD, median, minimum, maximum, and n) for each treatment group.

For the assessment of ratio in total T2 lesion volume relative to baseline up to Week 96, MMRM analyses will be undertaken using the same approach as described above in Section 4.4.3.9, except the baseline covariates here will be as follows: log-transformed baseline T2 lesion volume, geographical region (U.S. vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

All T2 lesion volumes below the lower limit of quantification of 0.009 cm^3 (3 ,voxels) will be set to this limit.

4.4.3.23.1 OLE Phase

This analysis will be extended to include data collected during the OLE period. Ratio relative to baseline at all scheduled visits until the last mature visit during the combined DB+OLE period and ratio relative to OLE baseline at all scheduled visits until the last mature visit during the OLE period will be analyzed.

4.4.3.24 Ratio in Total T1 Lesion Volume as Detected by Brain MRI Scans relative to Baseline at Week 96

The total volume of T1 lesions on MRI scans of the brain will be summarized at baseline and Weeks 24, 48, and 96 using descriptive statistics for each treatment group. Data from other unscheduled assessments will not be included in this summary or analysis. The percentage change from baseline to Weeks 24, 48, and 96 in total volume of T1 lesions on MRI scans of the brain will be computed and summarized using descriptive statistics (e.g., mean, SD, median, minimum, maximum, and n) for each treatment group.

For the assessment of ratio in total T1 lesion volume relative to baseline up to Week 96, MMRM analyses will be undertaken using the same approach as described above in Section 4.4.3.9, except the baseline covariates here will be as follows: log-transformed baseline T1 lesion volume, geographical region (U.S. vs. ROW), and baseline EDSS score (<4.0 vs. ≥ 4.0).

All T1 lesion volumes below the lower limit of quantification of 0.009 cm3 (3 ,voxels) will be set to this limit.

4.4.3.24.1 OLE Phase

This analysis will be extended to include data collected during the OLE period. Ratio relative to baseline at all scheduled visits until the last mature visit during the combined DB+OLE period and ratio relative to OLE baseline at all scheduled visits until the last mature visit during the OLE period will be analyzed.

4.4.3.25 Time to Onset of Composite Confirmed Disability Progression (cCDP) for at Least 12 Weeks and at Least 24 Weeks

Time to onset of cCDP over the double-blind treatment period, defined as an increase in EDSS that is sustained for at least 12/24 weeks (0.5 or 1, same criteria as for the secondary endpoint time to onset of 12/24-weeks CDP) or a 20 percent increase in T25-FW that is sustained for at least 12/24 weeks or a 20 percent increase in 9-HPT that is sustained for at least 12/24 weeks. These analyses will be performed for the double-blind period only.

4.4.3.26 Change in EuroQoL 5 Dimension 3 Levels (EQ-5D-3L) VAS and Index from Baseline to Week 48 and Week 96

The EQ-5D questionnaire comprises the following 5 dimensions, each describing a different aspect of health: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three levels: no problems (1), some problems

(2) and extreme problems (3). This part, called the EQ-5D descriptive system, provides a 5-dimensional description of health status.

A unique health state is defined by combining one level from each of the 5 dimensions. A total of 243 possible health states is defined in this way. Each state is referred to in terms of a 5-digit code. For example, state 11111 indicates no problems on any of the 5 dimensions, while state 11223 indicates no problems with mobility and self-care, some problems with performing usual activities, moderate pain or discomfort and extreme anxiety or depression. The EQ-5D-3L health state is then converted into a single summary number, the EQ-5D-3L index value, using validated translations provided by EuroQoL that can be found in the EuroQoL website (https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/valuation/). The EQ-5D-3L index is frequently used in economic evaluations: it represents societal preference values for the full set of EQ-5D-3L health states with the best state (perfect health) and death being assigned values of 1 and 0, respectively.

The United Kingdom value set of this questionnaire will be used for all participants regardless of their country.

The EQ-5D-3L descriptive system is followed by a VAS (EQ VAS), which records the respondent's self-rated health ranging from 0 (worst imaginable health state) to 100 (best imaginable health state).

The EQ-5D-3L index score and VAS score will be summarized at baseline and Weeks 48 and 96 using descriptive statistics for each treatment group. The change from baseline to Weeks 48 and 96 will be also presented.

4.4.3.26.1 OLE Phase

This analysis will be extended to include data collected during the OLE period. Change from baseline at all scheduled visits until the last mature visit during the combined DB+OLE period will be analyzed.

4.4.3.27 NfL

NfL data collected during the double-blind and open label phase will be summarized descriptively. Summary statistics of the absolute NfL values will be presented at all scheduled visits along with the percentage change from baseline. A corresponding plot will be produced. The analyses will be performed for the DB period only as well as for the combined DB+OLE period.

4.4.4 Subgroup Analyses

The primary and some secondary efficacy endpoints (12 and 24 week CDP) will be summarized and analyzed by the following subgroups:

- Demography: age ($\geq$40 vs. $<$40 years)
- Demography: sex (male vs. female)

- Demography: race (White vs. Other)
- Body weight: ≥ 75 kg versus < 75 kg
- Stratification factor: geographical region (United States vs. ROW)
- Stratification factor: baseline EDSS score (< 4.0 vs. ≥ 4.0)
- Previous lesions: baseline Gd-enhancing lesion (0 vs. > 0)

Estimates of treatment effect and associated 95% CIs will be presented in forest plots. Unadjusted p-values will also be presented for these analyses.

4.4.4.1 Efficacy Data from the Safety Follow-Up Phase

Exploratory analyses of efficacy data collected in the safety follow-up phase are not envisioned.

4.5 PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

4.5.1 Pharmacokinetic Analyses

Ocrelizumab serum concentration−time data will be analyzed using a population approach. The primary population PK parameters (clearances and volumes) for ocrelizumab will be estimated by means of a non-linear mixed effect model fitted to the PK data collected in studies WA21092, WA21093, and WA21493. Clearances with associated inter-patient variability may be characterized by a saturable and non-saturable clearance as well as an intercompartmental clearance depending on the final structural model. Volumes with associated inter-patient variability may be characterized by central and peripheral volumes depending on the final structural model. Exposure (area under the concentration−time curve [AUC]) to ocrelizumab will be estimated. The selection of other parameters will depend on the final PK model used for this analysis.

Patients who have measurable concentrations of ocrelizumab will be included in the PK analysis unless major protocol deviations or unavailability of information (e.g., exact blood sampling time) occurred that may interfere with PK evaluation. The influence of covariates, such as age, gender, weight, ADA, and baseline CD19+ cell count on these individual PK parameter estimates will be investigated.

Details will be described in a Modeling and Simulation Analysis Plan, and results will be reported separately from the main CSR.

4.5.2 Pharmacodynamic Analyses

The correlation between individual ocrelizumab exposure and selected safety and efficacy parameters will be explored in order to characterize the exposure and /or dose-response relationship. This may include but is not limited to CD19+ cell count, ARR, infusion-related reactions (IRRs), serious infections, and other safety or efficacy parameters if of interest. Any such analyses will be reported separately in the PK report.

4.6 SAFETY ANALYSES

Safety data include but are not restricted to adverse event data, laboratory data, previous and concomitant treatment data, infusion information including IRR data, withdrawal data, fatalities, ECG, vital signs, ADA data, and dosing information.

This SAP describes the analyses of all available safety data collected until the end of the trial. The safety data will be summarized descriptively for the DBP, OLE and SFU periods:

4.6.1 Exposure of Study Medication

This section deals with exposure in the context of safety analyses. In the context of efficacy analyses, exposure time is as defined above in Section 4.4.

Exposure of ocrelizumab/ocrelizumab placebo and interferon beta-1a 44 μg SC / interferon beta-1a 44 μg SC placebo over the course of the study will be summarized using descriptive statistics.

"Duration of Observation" and "Duration of Exposure" are used interchangeably within this SAP.

Definition of the Dose: A Dose of ocrelizumab can be given as one infusion or two infusions administered 2 weeks apart.

Patients will be considered to have received a Dose of treatment if at least part of one infusion of that Dose (either Day 1 or Day 15 for dual infusions) was given. If a Dose is completely missed instead of delayed, the next Dose number will be based on the number of previous doses received.

The duration of observation for a patient will be calculated as:

(Date of last contact* − date of first Dose) + 1

* Earliest of 1) date of CCOD for the reporting event, 2) date of last contact from the study completion page, or 3) date of death.

For the controlled treatment period, the CCOD will be variable for each patient at 96 weeks from treatment.

For patients who are receiving interferon beta-1a 44 μg SC and who switch to active ocrelizumab treatment, exposure of interferon beta-1a 44 μg SC will end on the study day prior to their first active Dose of ocrelizumab.

The duration of observation, within a Dose is defined in a similar manner as:

(Day prior to first infusion in $n+1^{th}$ Dose* − date of first infusion in n^{th} Dose) + 1

*With the exception of the last Dose received by the patient where date of last contact is used as defined above.

Exposure to ocrelizumab and exposure to interferon beta-1a 44 μg SC during the controlled treatment phase will be summarized. For ocrelizumab, the treatment duration, number of infusions received, and the total cumulative dose (derived to mg) will be summarized by study treatment arm. An additional summary will be presented, summarizing at each infusion the percentage of patients who received $<80\%$ and $\geq 80\%$ of the planned infusion. The number and percentage of infused patients who were pre-treated with steroids will be summarized. For interferon beta-1a 44 μg SC, the number of Doses and dose intensity (based on the number of syringes received rather than the actual Dose) will be presented by study treatment arm.

4.6.2 Columbia-Suicide Severity Rating Scale

The number and percentage of patients with suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent at any time after the start of study treatment, based on the Columbia Suicide Severity Rating Scale (C-SSRS), will be presented. In addition, shift tables will be presented to show changes from baseline in C-SSRS categories. A listing of patients with suicidal ideation, suicidal behavior, or self-injurious behavior, based on the C-SSRS, will be presented.

4.6.3 Adverse Events

Adverse events will be defined as all adverse events including IRRs and serious MS relapses, but excluding non-serious MS relapses. Therefore, those adverse events recorded on the “adverse event” and “infusion-related reaction” CRF pages will be included.

For each adverse event recorded, the term entered by the investigator describing the event (the “reported term”) will be assigned a standardized term (the “preferred term” [PT]) based on the Medical Dictionary for Regulatory Activities (MedDRA) World Health Organization (WHO) dictionary of terms. All analyses of adverse event data will be performed using the PTs unless otherwise specified.

All adverse events will be mapped to PTs and super class terms. For all summary tables, the adverse events will be sorted by body system (in decreasing order of overall incidence) and then by PT (in decreasing order of overall incidence). Additionally, the most frequent adverse events ($\geq 5\%$ in any treatment arm) will be presented by PT.

Adverse events will be assigned to a Dose if the adverse event onset date is on or after the date of the first infusion of that Dose, but before the first infusion of the next treatment Dose. Adverse events that are reported during the safety follow-up period or B-cell monitoring period will be assigned to the last Dose received. Hence the last Dose for a patient may be of a variable length from 24 to 72 weeks or longer. Adverse events

that start prior to the first Dose and worsen during treatment (i.e. treatment emergent) will be assigned to the first treatment Dose.

All summaries and listings of adverse events will be based on the safety population unless otherwise stated. Summaries of adverse events will be generated by summarizing the incidence of treatment-emergent adverse events only. Treatment-emergent events are defined as those adverse events with observed or imputed onset date on or after the start date of study treatment. Only where the most extreme intensity is greater than the initial intensity will events with an onset date prior to the start of study treatment (and with an end date on or after the start of study treatment) be considered treatment emergent. An adverse event with a completely missing non-imputed start date will be assumed to be treatment emergent unless the adverse event has a complete non-imputed end date that is prior to start of study treatment.

For each treatment group, the incidence count for each adverse event PT will be defined as the number of patients reporting at least one treatment emergent occurrence of the event. The incidence rate will be calculated as the incidence count divided by the total number of patients in the population. Each table will also present the overall number of patients experiencing at least one adverse event and the total number of adverse events reported (multiple occurrences of the same adverse event in 1 patient will be counted only once).

The number of patients who experienced a related adverse event will be summarized by system organ class (SOC) and PT. Adverse events will be summarized by SOC and PT by intensity grade. Multiple occurrences of the same event within a patient will be counted once at the greatest intensity/highest grade for this PT. For adverse events leading to death, the most extreme intensity will be overwritten by Grade 5 (death). Any adverse events and the SOC overall rows of the summary table will count patients according to adverse events by intensity/grades.

All patient deaths, regardless of treatment received, will be listed.

Serious adverse events will be defined as all serious adverse events including serious MS relapses and serious IRRs. The number of patients who experienced a serious adverse event will be summarized by SOC and PT. Related serious adverse events will be summarized by SOC and PT. Additionally, the most frequent serious adverse events ($\geq$1%) will be presented by PT.

A separate listing with relapses considered as serious adverse events will be presented.

A patient may experience an adverse event that leads to the discontinuation of their study treatment. Discontinuation of study treatment for an adverse event may not necessarily lead to discontinuation from the study because patients can enter the safety follow-up periods of the protocol. Only adverse events that led to the discontinuation of

study treatment are of interest. Patients who withdraw early from the study because of adverse events will be summarized under disposition. Because of the double-dummy design, the study-drug regimen consists of more than one combination of two treatments (ocrelizumab and interferon beta-1a 44 μg SC). An analysis of adverse events leading to study drug discontinuation will not be performed for the individual treatments of the regimen within one study arm, because the investigators were instructed to discontinue both treatments. The number of patients who experienced an adverse event that led to discontinuation of study treatment will be summarized by SOC and PT.

The number of patients who experienced an adverse event that led to modification or interruption of study drug will be summarized by SOC and PT.

4.6.3.1 Selected Adverse Events

Infection

Infections will be defined from the adverse event data using the MedDRA SOC of "Infections and Infestations." During Primary CSR, adverse events that are reported as an infection by the investigator will also be included.

During Primary CSR, an infection has been defined as serious if the event is a serious adverse event or if the non-serious infection was treated with an IV anti-infective. For subsequent reporting, an infection has been defined as serious if the event is a serious adverse event.

The number of infections and serious infections per 100 patient-years will be calculated and summarized separately. The ninety-five percent CI of the incidence rates will be calculated using exact method based on Poisson distribution.

Summary of infections will be presented by SOC and PT, severity, and Dose. Similar summary tables will be also produced by serious infections.

Summary tables of infections and serious infections by pathogen information codes as well as listing of infections/serious infections with pathogen codes will be produced.

Infusion-Related Reactions

An IRR and its corresponding symptoms are collected on the dedicated eCRF.

The symptom(s) of an IRR and the IRR itself may be of different intensities. As other symptoms can be recorded as free-text on the eCRF page, symptoms will be coded in the MedDRA and summarized by PTs.

IRRs are categorized by the time of the event occurring (1) during the infusion, (2) after completion of the infusion while patient is in the clinic, and (3) within 24 hours of completion of the infusion and patient is not in the clinic. The number and percentage of patients with at least one IRR will be presented by infusion (patients with multiple events within an infusion will count only once). In addition, the total number of IRRs will be

summarized (multiple events will be counted). The total and percentage (based on the total number of patients with at least one IRR) will be summarized by most extreme intensity.

IRRs will be summarized by intensity and by infusion. For multiple events within a patient, the most extreme intensity will be taken.

Symptoms of IRR and symptoms of serious IRR will also be presented by infusion. This analysis will be produced for both the controlled treatment period and the OLE period.

The number of patients with at least one IRR, the total number of IRRs, and the number of IRRs will be summarized by the time of events.

Pregnancies

Any pregnancies that occur during the study will be listed.

4.6.4 Magnetic Resonance Imaging Data

Non-MS pathology that is reported by the local safety radiologist will be listed by treatment group.

4.6.5 Laboratory Data

General laboratory evaluation

All laboratory assessments will be summarized or listed when applicable.

Absolute value and change from baseline values at each visit will be summarized. The baseline value will be the last value prior to the first dose of study medication. When laboratory data are presented over time, these laboratory values will be time-windowed into a common visit structure. If multiple values of the same laboratory parameter occur within the same time window, the worst value for that parameter will be presented in the summary table.

A summary of the number and percentage of patients with abnormal laboratory outcomes will be produced by treatment group for each parameter.

In addition, for the liver laboratory parameters, the number and percentage of patients with an elevated post-baseline AST or ALT level during the 96-week controlled treatment period will be summarized by treatment.

CD19

The median CD19 cell count will be displayed graphically over time from the first infusion of study medication. The absolute CD19 count and change from baseline in the CD19 count will be summarized over time. At each timepoint, the number and percentage of patients whose CD19 counts have repleted will be presented. Repletion is defined as the CD19 count having returned to their baseline value or lower limit of normal ($\geq$80 cells/μL), whichever is lowest.

Immunoglobulins

The mean immunoglobulin levels (IgG, IgM, and total Ig) will be displayed graphically over time from the first infusion of study medication. The absolute values and change from baseline will be summarized over time. At each timepoint, the number and percentage of patients with immunoglobulin levels lower than the lower limit of normal will be presented.

Anti-Drug Antibodies

Anti-drug antibodies (ADAs), also called human anti-human antibodies (HAHAs), will be summarized over the 96-week controlled treatment period by treatment group. The baseline prevalence and post-baseline incidence of ADAs will be displayed. The number of patients with treatment-induced ADA will also be displayed. A table will be presented that summarizes ocrelizumab serum concentrations (μg/mL) at timepoints where ADA samples were collected and analyzed. A listing by treatment of anti-ocrelizumab antibody data will be presented for patients with at least one ADA sampled datum.

Antibody Titers

If data are available, antibody titers for mumps, rubella, varicella, and *Streptococcus pneumoniae* will be summarized for number and percentages of patients with positive level.

For Mumps, an assay method was updated due to bacterial contamination in tubes. And, results from both the assay methods, Diamedix Elisa and Diasorin Liaison XL to be presented for OLE Phase.

4.6.6 Vital Signs

Vital signs, physical examination, and ECG results will be presented in patient listings. Change from study baseline in vital signs at each visit will be summarized by treatment group. Changes in vital signs from baseline to post-infusion timepoints will also be summarized for each infusion.

4.6.7 Safety by Treatment and Period

To explore safety by treatment and period, all of the available safety data will be presented by treatment regimen and by period (i.e., All periods and OLE period separately, as appropriate).

The adverse event profile will be produced by treatment regimen for each period separately. The number of events per 100 patient-years will be presented to account for differing amounts of exposure.

Adverse events and serious adverse events will be presented by SOC and PTs for each period separately. Also, infection and serious infection will be presented by SOC and PTs for overall as well as each period separately.

4.6.8 Double-Blind Period

All safety data prior to the first active dose of ocrelizumab in OLE will be presented by the actual treatment received during the DB period, including the SFU data post-Week 96 until the start of OLE and SFU data post-Week 96 for patients who withdrew during the double-blind period.

4.6.9 OLE Period

All data collected after the first dose of OLE ocrelizumab will be presented by the actual treatment received during the DB period, including all SFU and B-cell monitoring data for patients who withdraw during the OLE. Some of the analyses performed for the double-blind treatment period will be repeated and updated for the safety endpoints collected during the OLE period. In general, AE, SAE profile, infections, serious infections, IRRs, conmeds and laboratory analysis such as CD19, Immunoglobulins will be presented.

Analyses pooling all safety data from the first dose of ocrelizumab (double blind and OLE periods) across multiple MS trials are described separately in a Global Safety Pooling SAP.

4.6.10 Type I Interferon-Neutralizing Antibody Data

Type I interferon-neutralizing antibody data will be summarized in listings and additional analysis will be performed as appropriate.

4.6.11 COVID-19 analyses:

In line with the guidance from the Sponsor and health authorities, several patient listings and summaries will be prepared in order to assess the impact of the COVID-19 pandemic on the study conduct and results. Summaries and listings will be prepared for major protocol deviations due to the pandemic, patients infected with COVID-19 and AEs & death associated with COVID-19. Further details will be provided in the COVID-19 CSR Appendix.

4.7 MISSING DATA

All methods for handling missing data and associated sensitivity analyses are described above, section by section, for each endpoint.

4.8 INTERIM ANALYSES

Updated efficacy analysis of additional data from OLE period are performed on a yearly basis after the CSR primary database lock. Regular safety updates are assessed in safety pooling analysis from multiple studies.

5. REFERENCES

Barkhof F, Calabresi PA, Miller DH, et al. Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. Nat Rev Neurol 2009; 5:256–266.

Bohannon RW. Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. Age Ageing 1997;26:15–19.

Carpenter JR, Kenward MG. Missing data in randomised controlled trials: A practical guide. 2007. Final report available at http://missingdata.lshtm.ac.uk/downloads/rm04_jh17_mk.pdf.

Oxford Grice KA, Vogel KA, Le V, et al. Adult Norms for a Commercially Available Nine Hole Peg Test for Finger Dexterity. Am J Occup Ther 2003;57:570–3.

Appendix 1 Standard Operating Procedure "EDSS Assessment Check for the Roche Trials WA25046, WA21092, and WA21093"

EDSS Assessment Check SD02

Neurologische Klinik und Poliklinik

Standard Operating Procedure (SOP)

EDSS assessment check for the Roche trials WA25046, WA21092 and WA21093 SD02			
Attachments Curriculum vitae of the USB Experts Contract of the USB Experts in copy Conflicts of interests of USB Experts			
Version: 1.0			
Target group: Roche, Neurostatus Systems GmbH, USB EDSS Expert Team			
Function	Name, Company	Date	Signature
Author	[redacted], USB		
Approved	Prof. [redacted], [redacted] Roche		
Approved	Prof. [redacted], USB [redacted]		
Approved	[redacted] Neurostatus Systems GmbH		
Approved	[redacted], USB		

TABLE OF CONTENTS

1. AIM

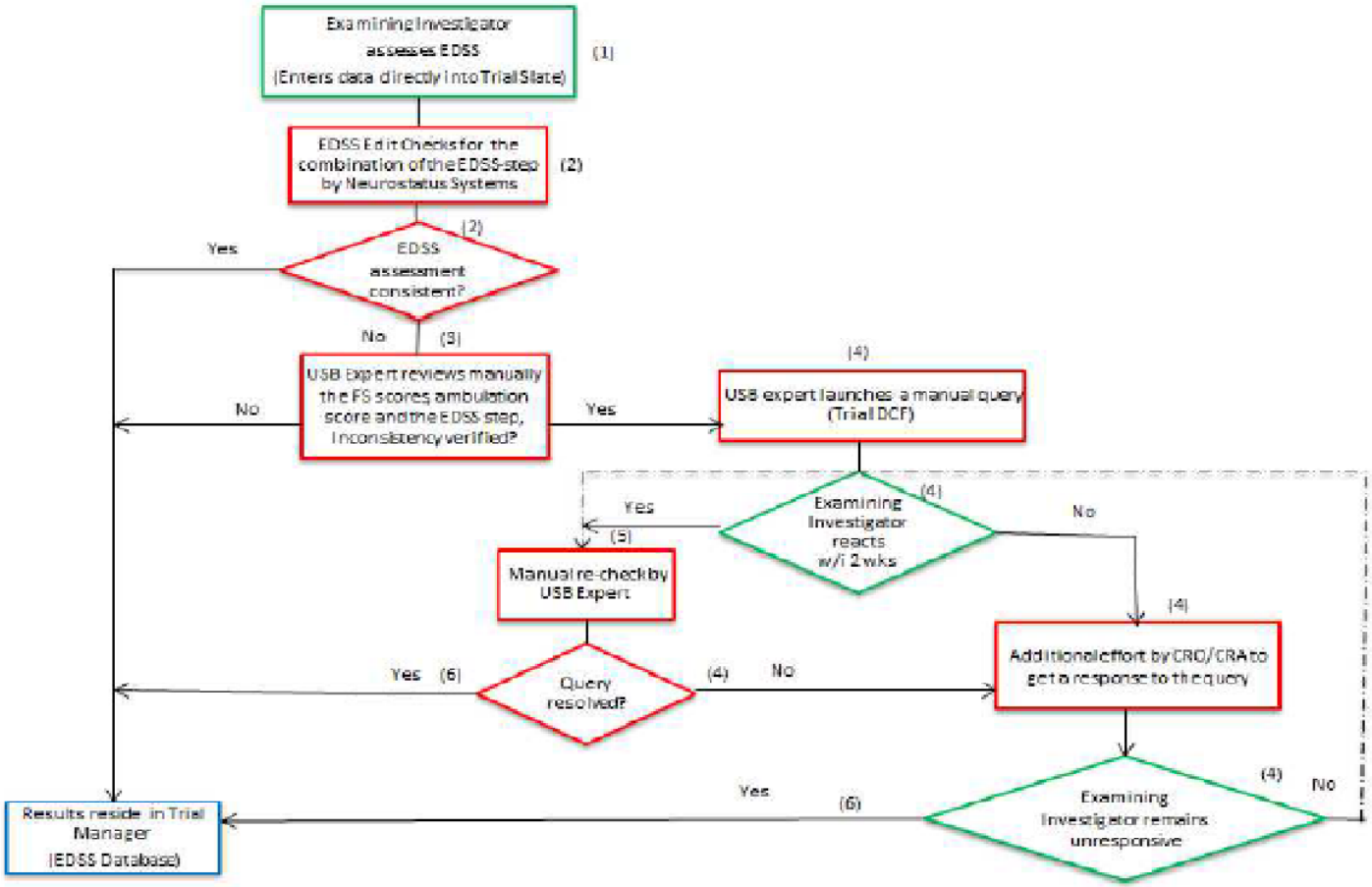
The aim of this SOP is to describe the process and the roles and responsibilities for the review of the Expanded Disability Status Scale (EDSS) assessments. In addition this SOP further outlines how to detect and handle inconsistent EDSS assessments and John Kurtzke's Functional Systems scores (FSS) in the Roche pivotal Multiple Sclerosis trials WA25046, WA21092 and WA21093 according to ICH guidelines.

2. BACKGROUND

Currently, John Kurtzke's Functional Systems and the EDSS are the most widely accepted clinical outcome measures for the evaluation of neurological impairment and disability in Multiple Sclerosis (MS) clinical trials. The determination of the EDSS step is primarily based on the individual scores of the 7 Functional Systems including visual, brainstem, pyramidal, cerebellar, sensory, bowel and bladder, and cerebral, as well as on the walking range and the assistance needed to ambulate (Kurtzke JF et al. 1955; Kurtzke JF 1983). During the past years standardized training programs and certification of EDSS evaluating physician (EDSS assessor) have been developed and introduced to improve inter-rater and intra-rater reliability. For the Roche trials WA25046, WA21092 and WA21093 the EDSS step is assessed by trained neurologists who refer to the Neurostatus definitions in the booklet version of 04/10.2 (www.Neurostatus.net). Between Jan 1, 2011 and Jan 31, 2012 University Hospital Basel (USB) Experts reviewed 1082 EDSS assessments rated by 267 examining investigators at 160 study sites participating in the Roche trials WA25046, WA21092 and WA21093. They found in 23% of the cases inconsistencies in the last step of the assessment, namely the combination of the Functional Systems and the ambulation scores to the final EDSS step.

3. PROCESSES

3.1 FLOW CHART ON QUERY RESOLUTION (DETAIL IN 3.3)



3.2 ROLES & RESPONSIBILITIES

Responsible	Role
EDSS Evaluating Physician	Examining investigator / EDSS assessor who performs the neurological examination, documents the FS scores and assess EDSS steps. The examining Investigator will not be involved with any aspect of medical management of the patient and will not have access to patient data.
CRF Health (CRFHealth eCOA Solutions Company)	Provider of the TrialSlate and associated web portal (TrialManager). CRF Health is responsible for processing all data clarification forms (TrialDCFs) related to the EDSS assessment captured in the TrialManager.
Neurostatus Systems GmbH	Company responsible for the technical and administrative implementation of training and certification of physicians participating in projects using EDSS in multiple sclerosis. Neurostatus Systems is responsible to check for plausibility and inconsistencies of the EDSS assessments.
USB Expert	USB Experts are medical doctors working at the Department of Neurology of the USB specialized in the assessments of the EDSS. For contracts see attachment. Prof [redacted] has the oversight of the Expert Review process. He will appoint a named person as one of the USB Experts who will act as a single point of contact for Roche.
	USB expert are reviewing EDSS assessment of the scoring sheet. The USB expert is responsible for the content of the TrialDCFs. The USB expert is responsible for being available via email / telephone in order to respond to questions by the EDSS physician.
Monitor/CRA	Clinical Research Associate employed by a Clinical Research Organisation (CRO). Monitor / CRA log into TrialManager to check for TrialDCFs and to notify sites that a response is required from the examining investigator.

Roche Study Management Team (SMT)	The Study Management Team (SMT) reviews unresolved TrialDCFs and decides to close or continue to follow up the outstanding TrialDCFs

3.3 REPORTING AND REVIEW PROCESS IN WRITTEN FORM

Step 1: EDSS assessment by the examining investigator at the study sites

Functional system score, ambulation score and EDSS step (EDSS assessment) are assessed by examining investigators based on a standardized neurological examination using the Neurostatus Scoring booklet, version 04/10.2. These data are captured by using TrialSlate (an electronic data capturing device). Range checks are performed during data entry onto the device. The data are then transferred via LAN or Mobile network to TrialManager web portal (the database).

Step 2: Automated Consistency Check

Neurostatus Systems checks the data on TrialManager for plausibility and inconsistencies of the EDSS using automated consistency checks. The rules for these checks are given in the Neurostatus Scoring booklet, version 04/10.2. A scoring sheet consisting of the results of the EDSS assessments is generated by Neurostatus and uploaded to TrialManager. The scoring sheet will flag inconsistencies in the EDSS assessment.

If the scoring sheet doesn't identify inconsistencies, the assessment remains unchanged in TrialManager.

Step 3: Manual consistency check

The USB expert will review the scoring sheet with the EDSS assessments within 2 working days from upload to TrialManager.

The EDSS assessments with flagged inconsistencies in the scoring sheet are manually reviewed by the USB Expert.

- If after review by the USB expert, the flagged EDSS assessment is determined to be consistent then it will remain on TrialManager unchanged.
- If after review by the USB expert, the flagged EDSS assessments are confirmed to be inconsistent a manual query (TrialDCF) will be generated in TrialManager. The query in TrialManager will be reviewed and responded to by the examining investigator (See Step 4)

Step 4: TrialDCF

The TrialDCF are queries described in the study manual query process.

- The USB expert is responsible for the content of the TrialDCF
- CRF Health is responsible for processing all TrialDCFs related to the EDSS assessment captured in TrialManager.

- The Monitor / CRA will log into TrialManager at least once per week for any new TrialDCFs. If new TrialDCFs have been issued, sites will be notified by the monitor/ CRA that a response to the TrialDCF is required by the examining investigator within two weeks via TrialManager.
- If the examining investigator does not respond to the initial TrialDCF within the two weeks as stated above, the monitor/CRA will contact the examining investigator to request to resolve the outstanding TrialDCF as per the agreed monitoring plan.
- The Examining investigator and/or monitor/CRA can ask for support by the USB expert via email or telephone to ask any question to enable resolution of the TrialDCF.
- Call service times are weekly for European countries on every Tuesday 3-4 p.m.
    - CET, biweekly on Tuesdays 9-10 a.m. CET for Australia and Asia and 5-6 p.m. CET for the Americas. The content of these discussions during the calls will be documented and archived by the USB Expert.
- Detailed information on call in numbers and email information will be provided by the USB expert.
- All outstanding TrialDCFs will be followed up by the monitor/CRA as per the agreed monitoring plan. The Study Management Team (SMT) will review unresolved TrialDCFs (as defined in study Integrated Data Review Plan (iDRP)) and will decide to close or continue to follow up the outstanding TrialDCFs.
- Any TrialDCFs that are not answered/resolved will have their status changed as defined by CRFHealth (following authorisation from the Roche SMT during the course of the study (see Step 6).

Step 5: Manual re-check

All answered TrialDCFs will be re-checked by the same USB Expert who issued the initial TrialDCF. TrialManger will generate a report to notify the USB expert of the response from examining investigator at the site and he/she will review the response.

Step 6: Resolution of TrialDCF in TrialManager

If the TrialDCFs is resolved, the EDSS assessment will be considered final and CRFHealth will implement and verify the change in TrialManger

The final status of a TrialDCF will be either:

1. “Resolved with change and closed” – TrialDCF is answered and revised data are entered by the examining investigator into TrialManager
2. “Resolved without change and closed” – TrialDCF is answered by the Examining Investigator who confirms his original assessment in the TrialDCF.
3. “Unresolved and closed” – TrialDCF is not answered and the SMT authorises to close due to unresponsiveness

In all cases the examining investigator remains the final decision maker on the EDSS assessment.

3.4 HANDLING OF EDSS ASSESSMENTS CHECKED PRIOR TO THE IMPLEMENTATION OF THE PROCESS OUTLINED IN THIS SOP

All previous EDSS assessments, that had gone through the previous automated TrialDCF process, whether they had been changed or not as a result of this process will go through this new revised Neurostatus Systems edit check process again and the process steps outlined in this current SOP 2-6 will be applied. Automated and new TrialDCF data will be stored in TrialManager.

4. GLOSSARY & DEFINITIONS

ICH GCP guideline	International Council for Harmonisation good clinical practice guidelines
CRF Health	CRFHealth eCOA Solutions Company
CRA	Clinical research associate
CRO	Contract research organisation
EDSS	Expanded Disability Status Scale
FSS	Functional system score
iDRP	Integrated Data Review Plan
MS	Multiple sclerosis
Neurostatus Systems GmbH	Company responsible for the technical and administrative implementation of training and certification of physicians participating in projects using EDSS in MS.
Principal Investigator	Lead site investigator
SMT	Study Management Team
TrialSlate	An electronic data capture device provided by CRFHealth
Trial DCF	Trial data clarification form (this resides in TrialManager)
TrialManager	Webportal Database which collects all questionnaires completed on the trialslate including EDSS assessments of the study patients
USB	University Hospital Basel

USB expert	Medical doctor who works at the University Hospital Basel (USB), specialized in the assessment of the EDSS and responsible for the quality control of inconsistencies, queries, teaching of the EDSS and certification process.

5. REFERENCES

Kurtzke JF. A new scale for evaluating disability in multiple sclerosis. Neurology 1955;5:580-583.

Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983;33(11):1444-1452.

Neurostatus, Version 04/10.2. www.Neurostatus.net

Signature Page for Statistical Analysis Plan -
System identifier: RIM-CLIN-473022

Approval Task	[redacted] Company Signatory 20-Mar-2023 15:24:01 GMT+0000