Parexel International

MedDay Pharmaceuticals

MD1003CT2016-01 MS-SPI2

Effect of MD1003 in progressive multiple sclerosis: a randomized double blind placebo-controlled study

Statistical Analysis Plan for Final Analysis of Study Part 1

Final Version 7.0

Parexel Project Number: 229829

MedDay Pharmaceuticals
MD1003CT2016-01 MS-SPI2 **Statistical Analysis Plan**

SPONSOR SIGNATURE PAGE

Approved by:

Frederick E. Munschauer, MSc., MD
Chief Medical Officer
MedDay Pharmaceuticals

10th February 2020
Date

Frederic Sedel, MD
Chief Scientific Officer
MedDay Pharmaceuticals

10th February 2020
Date

TP-GDO-WW-016-07 CONFIDENTIAL **Final Version 7.0**
Effective Date: 05 Sep 18 **Project Document Effective Date**: Date of last signature
Related to: SOP-GDO-WW-019 Page 2 of 61

Parexel SIGNATURE PAGE

Signatures below confirm that the review process was developed in accordance with SOP-GDO-WW-019 and that it approved for release.

This document has been approved and signed electronically on the final page by the following:

	Signatory
Author:	Peter Jakobs Project Role: Biostatistician

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AE	Adverse event
ATC	Anatomical Therapeutic Chemical
ATP	Adenosine triphosphate
BMI	Body mass index
CAC	Clinical Adjudication Committee
CAREQOL-MS	Caregiver Health-Related Quality of Life in Multiple Sclerosis
eCRF	electronic Case Report Form
CI	Confidence interval
CGI-I	Clinical Global Impression of Improvement
CGI	Clinician's Global Impression of Improvement
CSR	Clinical Study Report
C-SSRS	Columbia-Suicide Severity Rating Scale
DMT	Disease modifying therapy
DSMB	Data Safety Monitoring Board
ECG	Electrocardiography
EDSS	Expanded Disability Status Scale
EMA	European Medicines Agency
FAS	Full Analysis Set
FCS	Fully Conditional Specification
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
ITT	Intent-To-Treat
MedDRA	Medical Dictionary for Regulatory Activities
MITT	Modified Intent-To-Treat
MRI	Magnetic resonance imaging
MRS	Magnetic Resonance Spectroscopy
MS	Multiple sclerosis
MSQOL-54	Multiple Sclerosis Quality of Life-54
NAA/CR	N-acetylaspartate / creatine ratio
NA/AUS	North America / Australia
PPS	Per-Protocol Set
PMS	Progressive multiple sclerosis
PPMS	Primary progressive multiple sclerosis
PT	Preferred Term
RRMS	Relapsing-remitting multiple sclerosis
SAE	Serious adverse event
SAP	Statistical Analysis Plan

SD	Standard deviation
SDMT	Symbol Digit Modalities Test
SGI	Subject's Global Impression of Improvement
SOC	System Organ Class
SPMS	Secondary progressive multiple sclerosis
TEAE	Treatment emergent adverse event
TW25	Timed 25-Foot Walk

1 INTRODUCTION

Multiple sclerosis (MS) is a common neurological disease affecting more than 2 million people worldwide. It is an inflammatory autoimmune disease that damages the myelin of the central nervous system causing neurological impairment and, in many cases, severe disability.

Approximately 85% of all patients present with relapsing-remitting MS (RRMS), characterized by unpredictable acute episodes of neurological dysfunction (or 'relapses'), followed by variable recovery and periods of clinical stability. The remaining 15% of patients develop a sustained deterioration of their neurological function from symptom outset, termed primary progressive MS (PPMS). Over the chronic course of RRMS, approximately 50% of patients develop sustained deterioration with or without superimposed relapses; this form is called secondary progressive MS (SPMS). Progressive MS (PMS) includes both PPMS and SPMS.

Biotin, an essential co-enzyme for the production of adenosine triphosphate (ATP) by mitochondria, showed beneficial effects in patients with PMS when given in doses of 300 mg per day in open-label and double-blind trials. The current phase 3 study MD1003CT2016-01 MS-SPI2 has been designed to confirm these results (for MD1003 containing 100 mg Biotin to be taken three times a day) in PMS subjects without clinical evidence of a relapse in the previous 2 years.

This Statistical Analysis Plan (SAP) describes the planned final analyses for the randomized, double-blind, placebo-controlled study phase (comparing MD1003 to matching placebo) to be included in the Clinical Study Report (CSR).

This SAP covers study Part 1 and is based upon the following study documents:

- Study Protocol, version 4.0 (12 November 2018)
- electronic Case Report Form (eCRF) version 8.0 (18 Feb 2018)
- Data Safety Monitoring Board (DSMB) Charter, version 1.0 (04 May 2017).

2 STUDY OBJECTIVES

2.1 Primary Objective

The primary objective is to confirm the superiority of MD1003 at 300 mg/day over placebo to clinically improve patients with inactive PMS.

2.2 Secondary Objectives

Secondary objectives include:

- demonstration of increased time to 12-weeks confirmed Expanded Disability Status Scale (EDSS) progression for MD1003 compared to placebo
- comparison of MD1003 and placebo in clinical global impression of improvements assessed by investigators and subjects
- comparison of MD1003 and placebo in subjects' physical performance
- assessment of subjects' safety when treated with MD1003 in comparison to placebo.

2.3 Exploratory Objectives

Exploratory objectives include comparisons of MD1003 and placebo in brain Magnetic resonance imaging (MRI) measurements, neurological tests as well as subjects' and caregivers' health-related quality of life.

3 INVESTIGATIONAL PLAN

3.1 Overall Study Design and Plan

It was planned to randomize approximately 600 subjects (in approximately 90 centers located in North America, Europe and Australia); 642 subjects were actually randomized in 90 centers.

3.1.1 Visits Schedule and Evaluations

	M-1	**M0**	M3	M6	M9	M12	**M15**	M18	M21	M24	**M27**	Unscheduled Visits	Early Study Termination Visit
	Inclusion visit V1	**Randomization visit V2**	Follow-up visit V3	Follow-up visit V4	Follow-up visit V5	Follow-up visit V6	**Follow up visit V7**	Follow up visit V8	Follow up visit V9	Follow up visit V10	**Follow-up visit V11**		
STUDY WINDOWS***	+/- 15 days		+/- 10 days	+/- 10 days	+/- 10 days	+/- 10 days	+/- 15 days	+/- 15 days	+/- 15 days	+/- 15 days	+/- 15 days		
Informed consent	X												
Inclusion criteria	X	X											
Exclusion criteria	X	X											
Randomization		X											
Demographic Data	X												
Clinical examination*	X	X	X	X	X	X	X	X	X	X	X	X	X
EDSS	X	X	X	X	X	X	X	X	X	X	X	X	X
TW25	X	X	X	X	X	X	X	X	X	X	X	X	X
C-SSRS	X	X	X	X	X	X	X	X	X	X	X	X	X
Ambulatory activity monitoring		X	X	X	X	X	X	X	X	X	X	X	X
MS relapse questionnaire		X	X	X	X	X	X	X	X	X	X	X	X
SDMT		X					X				X		X
CGI-I (SGI & CGI)							X				X		X
MSQOL54 / CAREQOL-MS		X					X				X		X
Laboratory testing**	X			X			X				X	X	X
Blood-based pregnancy test[a]	X			X			X				X	X	X
Biotin blood dosage	X			X		X	X						
ECG		X[b]					X				X	X	X
Brain MRI		X		X			X				X	X[c]	
Concomitant medication / therapies	X	X	X	X	X	X	X	X	X	X	X	X	X
Study medication dispensation		X	X	X	X	X	X	X	X	X			
Compliance check			X	X	X	X	X	X	X	X	X	X	X
AEs reporting		X	X	X	X	X	X	X	X	X	X	X	X
Treatment assignment questionnaire							X						X

a: for women of childbearing potential; in addition, a urinary pregnancy test will be performed each month during the study and for 3 months after termination
b: triplicated ECG (Electrocardiography)
c: in case of relapse
*: physical examination and vital signs / **: CBC with differential, Metabolic panel (fasting), Hepatic function panel and Lipid panel (fasting) / *** One month is equal to 28 days

3.1.2 Randomization

At randomization visit M0, eligible subjects were centrally randomized to one of the two study treatments MD1003 or placebo. The underlying randomization list was generated with a 1:1 randomization ratio and stratified by center (study site) and MS disease history (SPMS / PPMS). The stratified randomization, however, was introduced by a study protocol amendment (protocol version 3.0). A small number of subjects (less than 5%) were randomized without any stratification under study protocol version 2.0.

3.1.3 Timing of Statistical Analysis

The statistical analysis of the randomized, double-blind, placebo-controlled study phase will be performed when all subjects have ended the double-blind period (planned to last from 15 to 27 months depending on the subject's date of randomization), i.e., all 6 conditions below are met:

1. All randomized subjects had their M15 visit or terminated the study
2. All subjects complete the next scheduled visit after condition 1 has been met; exceptions are subjects who have
   a. had their study termination visit or
   b. withdrawn their informed consent or
   c. been declared as lost to follow-up or
   d. been withdrawn from the study or
   e. deceased
3. Relevant data has been entered by site personnel and data has been transferred from external vendors (except unblinded randomization data)
4. Cleaning and reconciliation of relevant data has been performed
5. The agreed status of the database has been applied
6. Unblinded randomization data has been received and passed a quality control.

3.2 Efficacy Endpoints

3.2.1 Primary Efficacy Endpoint

The primary efficacy endpoint is a composite criterion which can be met in one of two ways: through confirmed improvement in EDSS or Timed 25-Foot Walk (TW25). The composite captures the improvement in ambulation which constitutes the main impaired function in subjects with progressive MS. Since the ambulation score in the EDSS scoring system can be somewhat insensitive to changes, the addition of a clinically meaningful decrease in TW25 may be more sensitive to demonstrate improvement of ambulation.

In more detail, the two criteria are:

- a decrease from baseline in EDSS at visit M12 confirmed at visit M15[1]: decrease of at least 1 point for baseline EDSS up[2] to 5.5 and of at least 0.5 point for baseline EDSS 6 to 6.5

or

- a decrease from baseline in TW25 of at least 20%[3] at visit M12 confirmed at visit M15[4]

with

- baseline EDSS is defined as the lowest (best) EDSS obtained at visits M-1 and M0.
- baseline TW25 is defined as the lowest (best) mean of TW25 attempts performed at the visits M-1 and M0, respectively, i.e.,
  $min(\text{mean of TW25 attempts at M - 1, mean of TW25 attempts at M0})$
- TW25 at visit M12 is defined as the mean of TW25 attempts at visit M12
- TW25 at visit M15 is defined as the mean of TW25 attempts at visit M15.

3.2.2 Secondary Efficacy Endpoints

3.2.2.1 Time to 12-Weeks Confirmed EDSS progression

12-weeks EDSS progression is defined as an increase of at least 1 point for baseline EDSS up[5] to 5.5 and of at least 0.5 point for baseline EDSS 6 to 6.5 with respective confirmation 12 weeks later (with time windows of ± 10 days up to 1 year after randomization, ± 15 days afterwards). The baseline EDSS value is defined as in Section 3.2.1.

Date of 12-weeks confirmed EDSS progression will be the first date of an EDSS progression (as defined above) that is confirmed 12 weeks later. Handling of missing scheduled EDSS assessments and censoring are described in Section 4.8.3.1.

Time to 12-weeks confirmed EDSS progression will be calculated as date of 12-weeks confirmed EDSS progression (or censoring) minus date of randomization plus 1; it will be expressed in weeks.

[1] i.e., defined EDSS decrease from baseline observed at visit M12 and at visit M15

[2] "EDSS 3.5 to 6.5 at inclusion" has been one of the inclusion criteria, but there may be few subjects with baseline EDSS (lowest EDSS obtained at visit M-1 and M0) slightly below 3.5; "up to 5.5 or below" is therefore used as baseline EDSS range for at least 1 point decrease

[3] No rounding of the TW25 decrease will be done before comparing it to 20%, i.e., a decrease in TW25 of 19.89% (or any other value < 20%) is not sufficient for fulfilling the TW25 criterion

[4] i.e., defined TW25 decrease from baseline observed at visit M12 and at visit M15

[5] Same rationale as for footnote 2

3.2.2.2 Clinical Global Impression of Improvement at Visit M15

The Clinical Global Impression of Improvement (CGI-I) is a 7-point scale that requires the clinician (CGI) and the subject (SGI) to assess how much the subject's disease has improved or worsened relative to the status at study baseline. It has the following categories:

1 very much improved
2 much improved
3 minimally improved
4 no change
5 minimally worse
6 much worse
7 very much worse.

CGI and SGI will be recorded at M15, M27 and early study termination visit.

3.2.2.3 Percentage Change from Baseline in TW25 at visit M15

Baseline TW25 is defined as the lowest mean of TW25 attempts performed at the visits M-1 and M0, respectively, i.e.,
$min(\text{mean of TW25 attempts at M - 1, mean of TW25 attempts at M0})$.

TW25 at visit M15 is defined as the mean of TW25 attempts at visit M15.

Percentage change from baseline in TW25 at visit M15 is defined as

$$\%\ change\ in\ TW25\ at\ M15 = \frac{TW25\ at\ M15 - baseline\ TW25}{baseline\ TW25} \times 100\%$$

i.e., negative values indication improvements.

3.2.3 Exploratory Efficacy Endpoints

3.2.3.1 Exploratory Brain MRI Efficacy Endpoints

Brain MRIs will be performed at visits M0, M6, M15, M27 and unscheduled visits in case of a relapse.

Exploratory brain MRI efficacy endpoints for visits M6 and M15 are

- whole brain volume
- change from baseline in thalamic volume
- percent change from baseline in brain volume
- change from baseline in cortical grey matter volume
- change from baseline in brain water content by Pseudo T2 relaxation time
- change from baseline in N-acetylaspartate / creatine ratio (NAA/Cr) in a subset of sites acquiring Magnetic Resonance Spectroscopy (MRS) (non-conventional sequences).

3.2.3.2 Remote Monitoring of Ambulatory Activity

Remote monitoring of ambulatory activity will be assessed at visits M0 and each follow-up visits from M3 to M27, at the early study termination visit and at unscheduled visits.

Exploratory efficacy endpoints are

- average daily step count based on a 21 days period prior to the scheduled (M0 or post-baseline) visit recording steps continuously in the natural environment as part of routine daily activity; average is defined as the sum of daily steps from valid days within such a 21 days period divided by the number of valid days; a "valid" day is defined as a day with at least 130 steps. At least 3 valid days are required for a visit to be included in the statistical analysis.
- change from baseline in average daily step count.

3.2.3.3 MSQOL-54 and CAREQOL-MS

The Multiple Sclerosis Quality of Life 54 (MSQOL-54) and Caregiver Health-Related Quality of Life in Multiple Sclerosis (CAREQOL-MS) will be used at visits M0, M15 and M27 and at the early study termination visit.

MSQOL-54 summary scores (physical health, mental health) will be derived from individual items (Vickrey 1995, Vickrey 1997).

CAREQOL-MS sub-scores (physical stress/global health, social integration, emotion, need for assistance/emotional reactions) will be derived from individual items (Benito-León 2001).

Exploratory efficacy endpoints are

- change from baseline at visit M15 in MSQOL-54 summary scores
- change from baseline at visit M15 in CAREQOL-MS sub-scores.

3.2.3.4 Kurtzke Functional Systems Scores

Kurtzke Functional Systems Scores will be determined at M-1, M0 and each follow-up visits from M3 to M27, at the early study termination visit and at unscheduled visits.

Exploratory efficacy endpoints are

- change from baseline at visit M15 in Kurtzke Functional Systems Scores
- change from baseline at visit M15 in (Kurtzke total) EDSS.

3.2.3.5 Symbol Digit Modalities Test

The Symbol Digit Modalities Test (SDMT) will be performed at M0, M15 and M27 visits and at the early study termination visit.

Exploratory efficacy endpoints are

- change from baseline at visit M15 in SDMT score as a continuous change and as categorical change using the following categories:
  - decrease of at least 4 digits
  - -4 digits < change < 4 digits

- increase of at least 4 digits.

3.2.3.6 Neurofilament Blood Concentration

During the double-blind study part, blood samples for neurofilament concentrations will be taken at visits M-1, M6, M12, M15, M27 and analyzed in a central laboratory.

Exploratory endpoints related to neurofilament blood concentration are the absolute changes from baseline (M-1) in neurofilament blood concentration at visits M6, M12, M15 and M27.

3.3 Safety Endpoints

3.3.1 Extent of Exposure

Extent of exposure will be expressed in months and calculated as

$$\frac{\text{date of last intake of study treatment} - \text{date of first intake of study treatment} + 1}{365.25/12}.$$

3.3.2 Adverse Events

Adverse events (AEs) will be coded using the agreed version of the Medical Dictionary for Regulatory Activities (MedDRA).

Treatment-emergent adverse events (TEAEs) will be defined as those AEs that

- start on or after the date of first dose of study treatment and not later than the date of last dose of study treatment
- or worsen in severity on or after the date of first dose of study treatment and no later than the date of last dose of study treatment.

AEs with missing or partially missing onset dates will be assumed to be treatment-emergent, unless there is clear evidence (through comparison of partial dates) to suggest that the adverse event started prior to the first dose of study treatment.

3.3.3 Relapses

The investigators will report relapses by a relapse form to the Clinical Adjudication Committee (CAC) for validation and MedDRA coding purposes.

The CAC will review protocol-defined, non-protocol-defined and suspected relapses and will classify these as either relapse or no relapse.

3.3.4 Safety Brain MRI Endpoints

Brain MRIs will be performed at visits M0, M6, M15, M27 and at unscheduled visits in case of a relapse.

Safety brain MRI endpoints for visits M6 and M15 are:

- number of new or enlarging T2-weighted lesions
- presence of at least one new or enlarging T2-weighted lesion

- number of gadolinium-enhancing T1-weighted lesions
- presence of at least one gadolinium-enhancing T1-weighted lesion
- volume of T2-weighted lesions
- volume of non-enhancing T1-weighted lesions.

3.3.5 Clinical Laboratory Tests

Laboratory tests will be performed at visits M-1, M6, M15, M27, at the early study termination visit and at unscheduled visits.

Creatinine clearance will be calculated by using the formula (Cockcroft & Gault 1976):

$$CrClearance[\text{mL/min}] = \frac{(140 - age[\text{years}]) \times weight[\text{kg}]}{serum\ creatinine[\mu\text{mol/L}]} (\times 0.85 \text{ if female})$$

Laboratory test results will be classified as normal or abnormal according to the applicable laboratory's normal ranges (these may depend on the subject's age, sex, race and the date when the laboratory examines the sample).

3.3.6 ECG Endpoints

ECGs will be performed at visits M0, M15, M27, at the early study termination visit and at unscheduled visits. At visit M0, a triplicated ECG will be performed, and the mean of the 3 ECG results will be used. From the 3 clinical interpretations performed by the investigator and the Cardiac Safety Core Laboratory, the last one by the Cardiac Safety Core Laboratory will be used for data analysis. For post-baseline ECGs, the clinical assessment provided by the Cardiac Safety Core Laboratory will be used for data analysis.

Corrected QT interval will be calculated by using Fridericia's formula

$$QtcF = \frac{QT}{\sqrt[3]{RR}}$$

ECG endpoints for visit M15 and the early study termination visit are

- intervals PR, QRS, QT and QTcF and changes from baseline
- categories for QTcF: ≤450 ms, >450 ms, >480 ms and >500 ms
- categories for QTcF changes: ≤30 ms, >30 ms and >60 ms.

3.3.7 Columbia-Suicide Severity Rating Scale

Suicidal ideation and behavior will be assessed at visits M-1, M0 and each follow-up visits from M3 to M27, at the early study termination visit and at unscheduled visits using the Columbia-Suicide Severity Rating Scale (C-SSRS).

The total score of the C-SSRS can be 10 at maximum and will be derived as follows:

- each 'Yes' to the questions 1, 2, 3, 4, 5 of Suicidal Ideation, and each 'Yes' to the questions Actual attempt, Interrupted Attempt, and Aborted attempt. Preparatory acts or behaviors and Actual Suicide of Suicidal Behavior will count as 1 and the sum of all these 10 items will give the total score.

- sub-scales reflecting ideation and behavior will also be derived separately.
- at screening, the total score will be 9 at the maximum since the last question 'Actual suicide' will never be Yes.

3.3.8 Vital Signs

The following vital signs will be recorded at visits M-1, M0 and each follow-up visits from M3 to M27, at the early study termination visit and at unscheduled visits:

- body weight
- body temperature
- heart rate
- systolic and diastolic blood pressure.

3.4 Definition of Other Variables Used in the Statistical Analysis

3.4.1 Geographical Regions

There will be two geographical regions defined as follows:

- North America / Australia (NA/AUS): Canada, United States of America, Australia
- Europe: Belgium, Hungary, Poland, Turkey, Czech Republic, Germany, Italy, Spain, Switzerland, United Kingdom.

3.4.2 MS Disease History

For statistical analysis, the following data sources for MS disease history (PPMS or SPMS) will be used:

- eCRF data for subjects for whom disease history is available in the eCRF
- Interactive web response system data for subjects for whom disease history is not available in the eCRF.

3.4.3 Demographics and Baseline Characteristics

Age in years at time of consent is collected. If missing, it will be calculated as the number of complete years between a subject's birth date and the date of informed consent. Age will also be presented in classes: 18 to <=25, 26 to <=40, 41 to <=55 and 56 to <=65 and >65.

Body mass index (BMI) will be calculated as

$$BMI[\mathrm{kg/m^2}] = \frac{weight[\mathrm{kg}]}{(height[\mathrm{m}])^2}$$

In the summary of disease history, time since first appearance of the first symptom attributable to MS in years will be calculated as:

$$\frac{\text{date of informed consent} - \text{date of first symptom attributable to MS} + 1}{365.25}$$

Time since initial diagnosis of MS, time since last EDSS progression, time since last relapsing episode (if applicable), and time since discontinuation of fampridine (if applicable) will be calculated in the same way.

In case of partial date of first appearance of MS or initial diagnosis of MS (only year available or year and month), the earliest possible date will be used (e.g., 1st of January or first day of the month). In case of partial date of last EDSS progression, last relapsing episode or discontinuation of fampridine (only year available or year and month), the earliest possible date in the past will be used (e.g., 1st of January or first day of the month).

Medical history and current medical conditions will be coded using the latest available version of the MedDRA dictionary. In case of partial or completely missing start date and end date, if this is not possible to decide whether the medical condition is past or current (ongoing is also missing), then it will be considered current.

3.4.4 Previous and Concomitant Medications and Procedures

Previous and concomitant medications will be coded using the latest available version of the World Health Organization Drug Dictionary.

Medications and procedures will be classified either as previous or concomitant where

- previous includes medications that stopped prior to the date of randomization
- concomitant includes medications that are continuing at the date of randomization and medications that started at or after the date of first study treatment dose.

Medication start and stop dates will be compared to the date of randomization to allow medications to be classified as either previous or concomitant. If medication start or stop dates are missing or partial, the dates will be compared as far as possible with the date of randomization. Medications will be assumed to be concomitant, unless there is clear evidence to suggest that the medication stopped prior to the date of randomization.

Medications will be assessed by a sponsor physician as either disease modifying therapy (DMT), medication for MS symptoms or none of these. For DMTs, the following more detailed categories are defined:

- previous DMTs (DMTs that stopped prior to the date of randomization)
- DMT already present at the date of randomization
- DMT started after the date of randomization dose.

Following rules will be applied when counting previous / concomitant medications:

- if a subject receives the same previous / concomitant medication (i.e., same Anatomical Therapeutic Chemical (ATC) level 1, ATC level 4 and preferred name) more than once, they are only counted once under the count for that ATC level 1, ATC level 4 and preferred name
- if a subject receives more than one previous / concomitant medication in a particular ATC class, they will only be included once in the count for the ATC class, but will appear in the count for each applicable preferred term within the ATC class.

3.4.5 Duration of Study

Duration of study will be expressed in months and calculated as

$$\frac{\text{date of last visit or contact} - \text{date of informed consent} + 1}{365.25/12}$$

3.4.6 Study Treatment Compliance

Compliance with study treatment will be calculated

(1) for the double-blind study phase up to the M15 visit (including) and

(2) for the overall double-blind study phase.

For each of these two intervals, the following variables are derived on the subject level:

- study treatment duration in days based on data from the study medication compliance pages of the eCRF:
  - (1) $STDur_{(1)} =$ date of last study treatment intake up to the M15 visit (incl) − date of first study treatment intake in the double - blind study phase $+ 1$
  - (2) $STDur_{(2)} =$ date of last study treatment intake in the double - blind study phase − date of first study treatment intake in the double - blind study phase $+ 1$
- total number of capsules dispensed based on data from the study medication compliance pages of the eCRF:
  - (1) $TCD_{(1)} =$ sum of numbers of capsules dispensed prior to the M15 visit
  - (2) $TCD_{(2)} =$ sum of numbers of capsules dispensed prior to the last visit of the double - blind study phase
- total number of capsules returned based on data from the study medication compliance pages of the eCRF:
  - (1) $TCR_{(1)} =$ sum of numbers of capsules returned up to the M15 visit (incl)
  - (2) $TCR_{(2)} =$ sum of numbers of capsules returned in the double - blind study phase
- study treatment compliance expressed in percentages (note subjects are supposed to take 3 capsules per day):
  - (1) $STC_{(1)} = \frac{TCD_{(1)} - TCR_{(1)}}{3 \times STDur_{(1)}} \times 100$
  - (2) $STC_{(2)} = \frac{TCD_{(2)} - TCR_{(2)}}{3 \times STDur_{(2)}} \times 100.$

In addition, the following categorical study treatment compliance variables on the subject level will be derived:

(1) for the double-blind study phase up to the M15 visit (including) using the following categories:
- $STC_{(1)} < 70\%$
- $70\% \leq STC_{(1)} \leq 80\%$
- $80\% < STC_{(1)} < 120\%$
- $120\% \leq STC_{(1)} \leq 130\%$
- $STC_{(1)} > 130\%$.

(2) for the overall double-blind study phase, using the following categories:
- $STC_{(2)} < 70\%$
- $70\% \leq STC_{(2)} \leq 80\%$
- $80\% < STC_{(2)} < 120\%$
- $120\% \leq STC_{(2)} \leq 130\%$
- $STC_{(2)} > 130\%$.

3.4.7 Study Treatment Interruptions

Data for study treatment interruptions will be taken from the from the study treatment interruption part of the medication compliance pages of the eCRF.

Study treatment interruptions will be determined

(1) for the double-blind study phase up to the M15 visit (including) and

(2) for the overall double-blind study phase.

For each of these two intervals, the following variables are derived:

- number of study treatment interruptions (derived on the subject level)
- duration of study treatment interruptions (derived on the interruption level)
  (1) $STIDur_{(1)} =$ date of study treatment restarted $-$ date of last study treatment intake $+ 1$
  (for a study treatment interruption with date of last study treatment intake is $\leq$ date of the M15 visit and date of study treatment restarted is > date of the M15 visit date, then the latter will be replaced by the M15 visit date)
  (2) $STIDur_{(2)} =$ date of study treatment restarted $-$ date of last study treatment intake $+ 1$

In addition, the following categorical variables related to study treatment interruptions will be derived on the subject level:

(1) for the double-blind study phase up to the M15 visit (including) using the following categories:
- no study treatment interruption
- at least one study treatment interruption

(2) for the overall double-blind study phase, using the following categories:
- no study treatment interruption
- at least one study treatment interruption.

3.4.8 Protocol Deviations

Major protocol deviations are those deviations from the protocol likely to have an impact on the perceived efficacy and/or safety of study treatments.

Major protocol deviations for this study (and any action to be taken regarding the exclusion of subjects or affected data from specific statistical analyses, see Section 3.5) have been defined in the Protocol Deviation Specification document. Updates of the Protocol Deviation Specification document are allowed only before general unblinding for the final statistical analysis.

Of interest is the intake of off-study biotin as biotin is available as a food supplement in several countries. Biotin plasma concentrations will be measured from samples taken at visits M-1, M6, M12, M15 and M27 (or at early study termination) during the placebo-controlled study phase. Biotin plasma concentrations will be measured at a central laboratory not informed about the randomized study treatment and the results will be made available only after database lock and general unblinding for the final statistical analysis.

For subjects randomized to placebo, any biotin plasma concentration above 75 ng/ml is considered indicative for off-study intake of biotin and constitutes a major protocol deviation.

Similarly, if the visit M-1 biotin level exceeds 75 ng/ml in subjects randomized to MD1003, this is considered a major protocol deviation.

3.5 Analysis Sets

Decisions of inclusion of subjects in analysis sets described below will be made prior to unblinding of study treatments, documented by Parexel and approved by MedDay. For potential major protocol deviations for which final assessment requires knowledge of the randomized study treatment (see Section 3.4.6 for the example of off-study biotin intake for patients randomized to placebo), the respective analysis set will only be finally determined after general unblinding. The rules described in the SAP will be followed.

3.5.1 Screened Analysis Set

The Screened Analysis Set consists of all subjects who were assigned a subject number.

3.5.2 Intent-To-Treat Analysis Set

The Intent-To-Treat (ITT) Analysis Set comprises all subjects to whom study treatment was randomized. Statistical analyses will be based on study treatment groups as per randomization and MS disease history stratum used for the randomization (where applicable), irrespective of the study treatment actually received.

3.5.3 Modified Intent-To-Treat Analysis Set

The Modified Intent-To-Treat (MITT) Analysis Set is a subset of the ITT Analysis Set:

- subjects randomized to placebo with any biotin plasma concentration above 75 ng/ml at visits mentioned in Section 3.4.8 are excluded from the MITT Analysis Set
- subjects randomized to MD1003 with any biotin plasma concentration above 75 ng/ml at visit M-1 are excluded from the MITT Analysis Set.

3.5.4 Full Analysis Set

The Full Analysis Set (FAS) is a subset of the ITT Analysis Set, consisting of subjects who:

- received at least one dose of study treatment
- and had at least one EDSS or TW25 assessment at visits M-1 or M0
- and had at least one post-baseline EDSS or TW25 assessment prior to study treatment discontinuation in the randomized double-blind study phase.

3.5.5 Per-Protocol Analysis Set

The Per-Protocol Analysis Set (PPS) is a subset of the ITT Analysis Set, consisting of subjects who:

- did not discontinue study treatment prior to visit M15
- and had at least one EDSS or TW25 assessment at visits M-1 or M0
- and have EDSS or TW25 assessment at visits M12 and M15
- and have no major protocol deviations up to M15 visit (this includes off-study biotin use in subjects randomized to placebo).

3.5.6 Safety Analysis Set

The Safety Analysis Set comprises all subjects who received at least one dose of study treatment.

Statistical analyses will be based on study treatment subjects actually received: if a subject has received any dose of MD1003 or any biotin plasma concentration has been above 75 ng/dl, the subject will be assigned to the MD1003 treatment group within the Safety Analysis Set; otherwise the subject will be assigned to the placebo treatment group.

3.5.7 Roles of the Analysis Sets

The main analysis of the primary efficacy endpoint will be based on the ITT Analysis Set; sensitivity analyses based on the MITT Analysis Set, the FAS and the PPS will be performed to assess the robustness of conclusions to the choice of analysis set.

Analyses of secondary efficacy endpoints will be based on ITT Analysis Set (main analysis) and MITT Analysis Set as a sensitivity analysis. Analyses of exploratory efficacy endpoints will be based on the ITT Analysis Set.

Analyses of safety will be based on the Safety Analysis Set.

4 STATISTICAL METHODS

4.1 Data Quality Assurance

All tables, figures and listings to be included in the CSR will be independently checked for consistency, integrity and in accordance with PAREXEL's Standard Operation Procedures.

4.2 Data Summary Format

Continuous data will be summarized in terms of the number of subjects, mean, standard deviation (SD), median, minimum, maximum and number of observations, unless otherwise stated, e.g., additional lower and upper quartiles. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, lower quartile and upper quartile will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects providing data at the relevant time point (n), number of subjects with missing data, frequency counts and percentages. Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Percentages will be calculated using n as the denominator. Changes from baseline in categorical data will be summarized using shift tables where appropriate.

P-values greater than or equal to 0.0001, in general, will be presented to four decimal places. P-values less than 0.0001 will be presented as "<0.0001". Confidence intervals (CIs) will be presented to one more decimal place than the raw data.

For laboratory tests for which a detection limit exists, values below the detection limit (e.g.. < 0.35 kU/L) will be considered as equal to the limit (e.g., = 0.35) in the statistical analyses (except in listings where the reported value will be presented). When repeated analyses were performed for laboratory tests for a same visit, due to non-evaluable sample or in order to confirm results, the latest values will be taken into account in the statistical analyses (except in listings where all reported test results will be presented).

4.3 Software

Statistical outputs for inclusion into the CSR will predominantly be generated using SAS® version 9.3 or a later version in a validated environment (some planned statistical analyses require SAS® version 9.4).

4.4 Study Subjects

4.4.1 Disposition of Subjects

The number of subjects screened for entry into the study and the number and percentage of subjects with screening failures by major reasons will be summarized based on the set of all screened subjects.

The following summaries for disposition of subjects will be provided by study treatment group and overall for the ITT Analysis Set:

- number and percentage of subjects randomized on the study level per geographical region, country and site
- number and percentage of subjects randomized by disease history (PPMS/SPMS) and per DMT present at baseline (including its combinations)
- number and percentage of subjects randomized by disease history (PPMS/SPMS) and within geographical region and country
- number and percentage of randomized subjects not treated with any study treatment and primary reason for not getting treated
- number and percentage of subjects prematurely discontinuing study treatment and primary reason for premature discontinuation of study treatment
- number and percentage of subjects prematurely discontinuing the study and primary reason for premature discontinuation of study
- number and percentage of subjects by visit

number and percentage of subjects withdrawing informed consent.

Listings of eligibility details, randomization details, visit dates, study treatment discontinuation details (including reason for study treatment discontinuation) and withdrawal/study completion details (including reason for study discontinuation and duration of treatment prior to study discontinuation) will be provided.

4.4.2 Protocol Deviations

A summary of number and percentage of subjects with a major protocol deviation and type of protocol deviation will be provided by study treatment group for the ITT Analysis Set. A listing of all major protocol deviations by subject will be produced.

4.5 Analysis Sets

The following summaries on analysis sets will be provided:

- number and percentage of subjects in the Screened Analysis Set and ITT Analysis Set (Screened Analysis Set)
- number and percentage of subjects in ITT Analysis Set, MITT Analysis Set, FAS, PPS and Safety Analysis Set by study treatment group and overall (ITT Analysis Set).

A listing to include site and subject identifier, inclusion/exclusion flag for each analysis set, reason for exclusion from an analysis set (if applicable), randomized and actual study treatment will be provided.

4.6 Demographic and Other Baseline Characteristics

The following summaries of demographic and baseline characteristics will be based on the ITT analysis set and provided by study treatment group and overall:

- demographics age, sex, race, ethnicity, height, weight and BMI (for all subjects and separately for PPMS subjects and SPMS subjects)
- MS disease history (for all subjects and separately for PPMS subjects and SPMS subjects)
- childbearing status (for female subjects) and contraception method
- previous medical and surgical history by MedDRA System Organ Class (SOC) and Preferred Term (PT)
- continuing medical conditions by MedDRA SOC and PT
- pattern of progression by disease history (PPMS/SPMS) and by DMT present at the date of randomization.

4.7 Previous and Concomitant Medications and Procedures

The following summaries will be provided by study treatment group and overall based on the ITT Analysis Set:

- previous DMTs
- DMTs already present at the date of randomization
- DMTs already present at the date of randomization by disease history (SPMS/PPMS)
- DMTs started after the date of randomization
- DMTs started after the date of randomization, by disease history (SPMS/PPMS)
- previous medications for MS symptoms
- concomitant medications for MS symptoms
- other concomitant therapies
- concomitant procedures.

4.8 Efficacy Evaluation

4.8.1 Analysis and Data Conventions

4.8.1.1 Multi-Center Studies

Given the high number of centers with only few subjects per center, it has been decided to use geographical region (defined in Section 3.4.1) instead of center as stratification variable in statistical analysis procedures.

4.8.1.2 Handling of Missing Data

The description of handling missing data in the statistical analysis of primary and secondary efficacy endpoints is integrated in Sections 4.8.2 and 4.8.3.

Analyses of exploratory efficacy endpoints will be based on available data only.

Analyses of safety endpoints will be based on available data, except for safety data where partial or missing data will be imputed by a "worst case approach", for example:

- assume an AE is TEAE if AE onset / end dates are missing or inconclusively incomplete
- impute missing causality to study medication by "related"
- impute missing AE severity as "severe".

4.8.1.3 Multiple Comparisons/Multiplicity

Testing of null hypotheses for secondary efficacy endpoints may only be started if the null hypothesis for the primary efficacy endpoint on the ITT Analysis Set of all PMS subjects will have been rejected.

If the latter will be the case, then the following sequential conditional testing procedure will be applied to restrict the one-sided family-wise type-I-error probability for secondary efficacy endpoints by 0.025:

1. The null hypothesis for time to 12-weeks confirmed EDSS progression will be tested as described in Section 4.8.3.1:
   - if the null hypothesis cannot be rejected (upper limit of a two-sided 95% Wald-type CI limit for the hazard ratio ≥ 1), then the test procedure for secondary efficacy endpoints stops here
   - if the null hypothesis is rejected (upper limit of a two-sided 95% Wald-type CI limit for the hazard ratio < 1), then the test procedure for secondary efficacy endpoints continues with step 2 below.
2. The null hypothesis for CGI at visit M15 will be tested as described in Section 4.8.3.2:
   - if the null hypothesis for CGI at visit M15 cannot be rejected (one-sided p-value of the stratified van Elteren test ≥ 0.025), then the test procedure for secondary efficacy endpoints stops here
   - if the null hypothesis for CGI at visit M15 is rejected (one-sided p-value of the stratified van Elteren test < 0.025), then the test procedure continues with step 3 below.
3. The null hypothesis for SGI at visit M15 will be tested as described in Section 4.8.3.2:
   - if the null hypothesis for SGI at visit M15 cannot be rejected (one-sided p-value of the stratified van Elteren test ≥ 0.025), then the test procedure for secondary efficacy endpoints stops here
   - if the null hypothesis for SGI at visit M15 is rejected (one-sided p-value of the stratified van Elteren test < 0.025), then the test procedure for secondary efficacy endpoints continues with step 4 below
4. The null hypothesis for percentage change in TW25 from visit M0 to visit M15 will be tested as described in Section 4.8.3.3 (with one-sided significance level 0.025) and the test procedure for secondary efficacy endpoints ends.

In the analysis of the primary efficacy endpoint for subgroups based on MS disease history and subgroups based on geographical regions, a null hypothesis may only be rejected if the

null hypothesis for the primary efficacy endpoint on the ITT Analysis Set of all PMS subjects will have been rejected.

4.8.1.4 Interim Analyses

A Data Safety Monitoring Board (DSMB) will meet approximately every three months and review predefined interim analysis results. The DSMB maintains the ability to call for ad hoc sessions at any time.

Following each DSMB meeting, a notification will be provided to the Sponsor with a recommendation for study conduct.

4.8.1.5 Subgroup Analyses

To assess the treatment effect and the homogeneity of treatment effect of the primary and secondary efficacy endpoints, subgroup analyses defined by the following baseline characteristics will be performed (based on the imputation method used for the main analysis):

- MS disease history: SPMS or PPMS
- geographical region: North America / Australia or Europe
- EDSS at baseline: “up to 5.5” or “6 or above”
- age: ≤ overall median or > overall median
- sex: female or male
- concomitant physical therapy: no or yes
- BMI: ≤ overall median or > overall median
- use of rituximab or ocrelizumab at the date of randomization: no or yes
- use of other DMTs at the date of randomization: no or yes
- use of anti-spasticity drugs at the date of randomization: no or yes.

Details on the methods for subgroup analyses are provided in Sections 4.8.2.7 (for the primary efficacy endpoint) and 4.8.3 (for the secondary efficacy endpoints).

4.8.2 Analysis of the Primary Efficacy Endpoint

4.8.2.1 Main Statistical Analysis

The main analysis of the primary efficacy endpoint will be based on the ITT Analysis Set.

A logistic regression model with categorical fixed factors

- randomized study treatment (MD1003, placebo)
- disease history (SPMS, PPMS)
- geographical region (NA/Australia, Europe)

will be used to estimate and test the study treatment effect expressed as corresponding response probability odds ratio (a value >1 indicates a positive effect of MD1003 compared to placebo).

The null hypothesis within the logistic regression model will be

H_0: response probability odds ratio related to study treatment ≤ 1

and the alternative hypothesis will be

H_A: response probability odds ratio related to study treatment > 1.

The response probability odds ratio related to study treatment will be estimated together with a two-sided 95% profile likelihood CI and the null hypothesis will be rejected if the lower CI limit is larger than 1 (this constitutes a one-sided test with significance level 0.025).

For descriptive purposes, the estimated effect size for MD1003 for the primary efficacy endpoint will be converted (by using overall response probability estimates for each study treatment group) from the response probability odds ratio point estimate above to point estimates as response probability ratio and response probability difference.

In addition to overall statistical significance for the primary outcome, both elements of the composite criterion should be consistent with the overall outcome, i.e., each component analysis must display a numerical advantage favoring the MD1003 treatment group but are not required to demonstrate statistical significance.

In case of less than 1% responders in at least one of the study treatment groups, a conditional exact logistic regression will be performed: an exact two-sided 95% CI for the response probability odds ratio will be obtained and the null hypothesis will be rejected if the lower CI limit is larger than 1 (this constitutes a one-sided test with significance level 0.025). In addition, an exact one-sided p-value will be reported.

The sample size has been determined to have a power of >90% to reject the null hypothesis with the test procedure described above if assuming response probabilities of 2% for placebo and 12% for MD1003. Note that in the previous MS-SPI study, the response probabilities were estimated as 0% for placebo and 12.6% for MD1003.

4.8.2.2 Handling of Missing Values in the Main Analysis

Detailed rules for handling of missing EDSS and/or TW25 values at visits Months 12 and 15 in the main analysis (Section 4.8.2.1) of the primary efficacy endpoint are included in Table 1 below:

Table 1: Handling of missing values for the main analysis of the primary efficacy endpoint

Baseline		Month 12		Month 15		Primary efficacy endpoint
EDSS	TW25	EDSS	TW25	EDSS	TW25	
missing	**missing**[6]	any	any	any	any	non-response
missing	*present*	any	any	any	any	depends on TW25
present	**missing**[7]	any	any	any	any	depends on EDSS
present	*present*	**missing**	**missing**	any	any	non-response

[6] there is no subject with both baseline EDSS and baseline TW25 missing
[7] there is no subject with missing baseline TW25

Baseline		Month 12		Month 15		Primary efficacy endpoint
EDSS	TW25	EDSS	TW25	EDSS	TW25	
present	*present*	any	any	**missing**	**missing**	non-response
present	*present*	**missing**	any	any	**missing**	non-response
present	*present*	any	**missing**	**missing**	any	non-response
present	*present*	**missing**	*present*	any	*present*	depends on TW25
present	*present*	any	*present*	**missing**	*present*	depends on TW25
present	*present*	*present*	**missing**	*present*	any	depends on EDSS
present	*present*	*present*	any	*present*	**missing**	depends on EDSS

The rationale for this single-imputation method of handling missing data (in essence: if missing data makes it impossible to determine the primary endpoint, then it will be imputed as "non-response") is that – considering the progressing population randomized for this study with almost no hope for spontaneous improvement over 15 months – non-response is the most likely outcome. The method may become anti-conservative if more subjects randomized to placebo will have an imputed non-response and there is a non-zero response probability for placebo.

4.8.2.3 First Sensitivity Analysis for Handling of Subjects with 12-Weeks Confirmed EDSS Progression

The main analysis will be repeated with the following modification of the primary efficacy endpoint definition (additional to the criteria specified in Section 3.2.1):

a subject will be handled as a non-responder in this sensitivity analysis, if

- the subject fulfilled the TW25 improvement criteria of the primary efficacy endpoint but not the EDSS improvement criteria of the primary efficacy endpoint (see Section 3.2.1) **and**
- the subject experienced a 12-weeks confirmed EDSS progression (as defined in Section 3.2.2.1) at any time during the double-blind study phase.

4.8.2.4 Sensitivity Analyses for Handling of Missing Values

Missing primary efficacy endpoint values are caused by combinations of missing baseline EDSS[8] and missing EDSS and TW25 values at M12 or M15 visits as described below in Table 2.

Table 2: Sources for missing primary efficacy endpoint values

Baseline		Month 12		Month 15		Primary efficacy endpoint
EDSS	TW25	EDSS	TW25	EDSS	TW25	
Missing	*present*	any	**missing**	any	any	*missing*
Missing	*present*	any	any	any	**missing**	*missing*
present	*present*	**missing**	**missing**	any	any	*missing*
present	*present*	any	any	**missing**	**missing**	*missing*
present	*present*	**missing**	any	any	**missing**	*missing*
present	*present*	any	**missing**	**missing**	any	*missing*

Sensitivity analyses will be performed for the primary efficacy endpoint to explore the influence of missing primary efficacy endpoint data (due to missing EDSS and/or missing TW25 at visits M12 and M15) and methods for handling these missing values on the treatment effect.

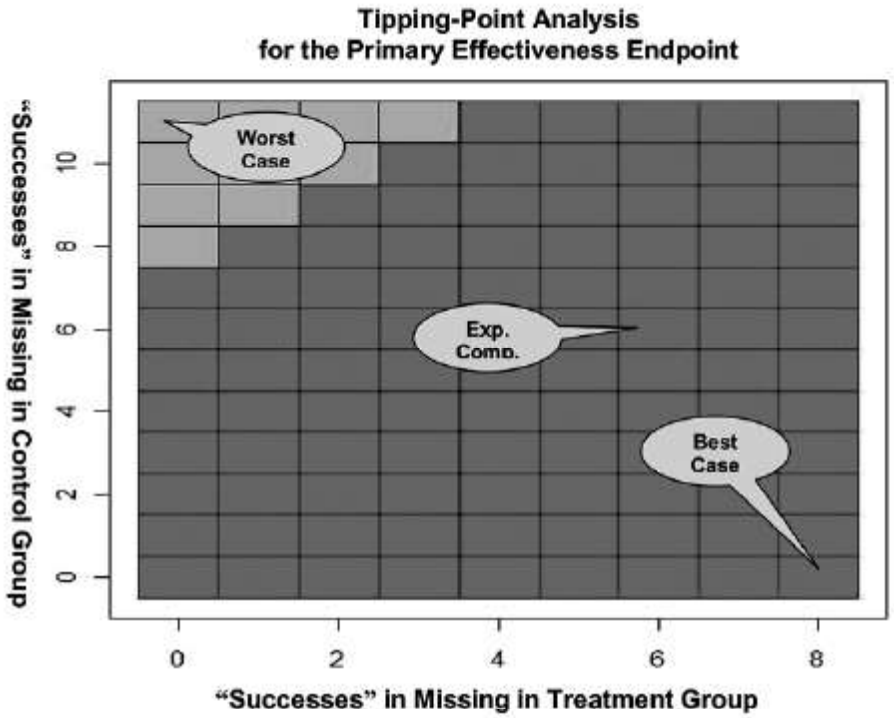
Tipping-point sensitivity analysis

Graphical displays, based on the "tipping-point" analysis introduced by Yan 2009, will be used to visualize the results of a set of sensitivity analyses (using different single "response" / "non-response" imputations for missing primary efficacy endpoint values) for comparison of the two study treatments. All possible combinations of single imputations of missing primary efficacy endpoint values as "response" or "non-response" in the MD1003 and placebo group will be evaluated by logistic regression with study treatment as the only factor.

Figure 1 below is an example graphic taken from Campbell 2011 with 8 missing endpoint values in the "Treatment Group" (so one could impute 0 to 8 successes) and 11 missing endpoint values in the "Control group" (so one could impute 0 to 11 successes), leading to a matrix of (9 times 12 =) 108 combinations to be tested for statistical significance; in the example, 10 combinations around the "worst case" (all missing values in the "Control Group" were imputed as successes and all missing value in the "Treatment Group" were imputed as failures.

[8] there is no subject with missing baseline TW25

Figure 1 Example for a tipping point analysis summary display



"Enhanced tipping-point displays" (Liublinska 2014) will be provided as compact summaries of conclusions drawn from different alternative assumptions.

Multiple imputation sensitivity analysis

This section is following Berglund & Heeringa 2014; this reference will also be used for implementation of the approach in SAS.

General approach

Multiple imputation methods replace each missing primary efficacy endpoint value with a set of m=10 (somewhat more than the default of 5 in SAS PROC MI) plausible values (based on a model predicting values for a missing data point based on available data). This set of values represents the uncertainty about the correct value to be imputed. The multiply imputed datasets (generated by SAS PROC MI) are then analyzed by the statistical procedure (only for asymptotic logistic regression and using the maximum likelihood parameter estimates and their standard errors) as complete datasets, followed by combining the results from these analyses (by using SAS PROC MIANALYZE).

Prediction model

The prediction model attempts to predict the primary efficacy endpoint ("response" or "non-response") based on available data (variables) that may have an influence on the primary efficacy endpoint.

The proposed prediction model for the primary efficacy endpoint is a logistic regression model with the following covariates:

- randomized study treatment: MD1003 or placebo

- MS disease history: SPMS or PPMS
- geographical region NA/AUS or Europe
- baseline EDSS: “up to 5.5” or “6 or above”
- baseline TW25 (continuous covariate)
- age at visit M-1 (continuous covariate)
- sex: female or male
- BMI at visit M-1 (continuous covariate)
- physical therapy after date of randomization and prior to visit M15 assessments: no or yes
- anti-spasticity drugs after date of randomization and prior to visit M15 assessments: no or yes
- DMT at date of randomization: no or yes
- DMT after date of randomization and prior to visit M15 assessments:
  - discontinuation of all DMT (in subjects with DMT present at date of randomization)
  - initiation of DMT (in subjects without DMT present at date of randomization)
  - or none of the above
- corticosteroid treatment after date of randomization and prior to visit M15 assessments: no or yes
- at least one MS relapse (confirmed by the adjudication committee) with onset after date of randomization and prior to visit M15 assessments: no or yes
- TW25 at visits M3, M6, M9 (continuous covariate)
- EDSS at visits M3, M6, M9: “up to 5.5” or “6 or above”.

Algorithm for the multiple imputation of missing values

Within the Bayesian framework, the task of imputing missing values is achieved by drawing random values from the posterior predictive distribution of the missing primary efficacy endpoint data (predicted by the logistic regression prediction model specified above). This posterior predictive distribution is a function of the observed data and regression parameters (or function of regression parameters).

As a monotone missing pattern cannot be expected (for example: there may be subjects with a missing TW25 value at visit M3 but a TW25 available at visit M12), the fully conditional specification (FCS) method will be used for dealing with arbitrary non-monotone missing data patterns. The FCS is based on an iterative algorithm; at each iteration and for each variable of the prediction model, there is a

- prediction step (P-step): the current (iteration) values of the observed and imputed values are used to derive the predictive distribution of the missing values
- and an imputation step (I-step): updated imputations are generated by draws from the predictive distribution defined by the updated logistic regression model.

When the last variable in the sequence (this is the primary efficacy endpoint) has been imputed, the algorithm cycles again through each variable, repeating the chain of

regression estimation and imputation draw steps. These cycles are repeated 10 times and finally there will *m*=10 draws from the predictive distribution for each missing primary efficacy endpoint value.

Analyzing multiply imputed datasets

Individual statistical analysis (as described in Section 4.8.2.1) will be performed for each of the *m*=10 imputed (complete) datasets and the results will be stored in a single well-specified output file. In particular, with multiple imputations per missing value, *m* different point and variance estimates for the log odds ratio related to study treatment will be computed.

Estimation and inference for multiply imputed datasets

As a final step, the *m*=10 estimates obtained in the individual statistical analyses of the multiply imputed (complete) datasets will be combined for making statistical inference.

Let $\hat{Q}_i$ and $\hat{W}_i$ denote the point and variance estimates for the log odds ratio related to study treatment from the *i*-th imputed complete data set, *i*=1, 2,…,*m*. Then the point estimate for the log odds ratio related to study treatment from multiple imputations, $\overline{Q}$, is the average of the *m* imputed (complete) datasets estimates:

$$\overline{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}_i .$$

Let $\overline{W}$ denote the average of the *m* ("within-imputation") variance estimates

$$\overline{W} = \frac{1}{m}\sum_{i=1}^{m}\hat{W}_i$$

and $\overline{B}$ the estimated ("between-imputation") variance of the point estimates

$$\overline{B} = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{Q}_i - \overline{Q})^2 .$$

Then the total variance *V* of the multiple imputation estimate $\overline{Q}$ for the log odds ratio related to study treatment is estimated as

$$\overline{V} = \overline{W} + (1 + m^{-1}) \cdot \overline{B} .$$

$\overline{Q}$ and $\overline{V}$ will be used for testing null hypothesis stated in Section 4.8.2.1 and constructing two-sided 95% CIs for the log odds ratio related to study treatment:

- test statistic

$$t = \frac{\overline{Q}}{\sqrt{\overline{V}}}$$

is approximately distributed as a *t*-distribution with degrees of freedom equal to

$$DF_{MI} = (m-1)\left(1 + \frac{\overline{W}}{(1 + m^{-1}) \cdot \overline{B}}\right)^2$$

- and the two-sided 95% CI for the log odds ratio related to study treatment is calculated as

$$\left(\overline{Q} - \sqrt{\overline{V}} \cdot t_{0.975,DF_{MI}}, \overline{Q} + \sqrt{\overline{V}} \cdot t_{0.975,DF_{MI}}\right).$$

Finally, the exponential function will be used for transforming point estimates and CI limits from the log odds ratio scale to the odds ratio scale.

Sensitivity analyses based on the MITT Analysis Set, FAS and PPS

Sensitivity analyses of the primary efficacy endpoint

- based on the MITT Analysis Set
- based on those subjects in the FAS for whom EDSS or TW25 is available at M12 and M15 and
- based on the PPS (the PPS, by definition, excludes subjects with missing primary efficacy endpoint data)

will be performed.

Note that the last two analyses are "completers analyses": hence, no handling of missing data is required – the results, however, are only valid under the strong assumption of "missing completely at random".

4.8.2.5 Sensitivity Analyses in the Context of the Estimand Concept

The section refers to the draft International Council for Harmonisation (ICH) E9 (R1) Addendum on Estimands and Sensitivity Analysis (ICH 2017).

Estimand

For this study, the following description of estimand attributes is provided:

A. Population: subjects fulfilling inclusion/exclusion criteria for "non-relapsing but progressing multiple sclerosis (primary or secondary)"
B. Variable: improvement on EDSS or TW25 (details as defined in Section 3.2.1) at visit M12 and visit M15
C. Expected intercurrent events are:
    - events (occurring after the date of randomization and up to the M15 visit) related to the inflammatory component of MS:
        - any MS relapse confirmed by the Adjudication Committee
        - any new gadolinium enhancing T1-weighted lesion
        - any new or enlarging T2-weighted lesion
    - discontinuation of all DMT present at date of randomization and prior to visit M15 assessments (in subjects with DMT at date of randomization)
    - initiation of DMT prior to visit M15 assessments (in subjects without DMT at date of randomization)
    - corticosteroid treatment after date of randomization and prior to visit M15 assessments
    - end of study treatment prior to visit M15 assessments

D. Population-level summary: proportions of responders for each study treatment and logistic regression analysis results for comparing them between study treatments (details described in Sections 4.8.2.1 and 4.8.2.2).

Treatment policy strategy

Intercurrent events will be handled according to the treatment policy strategy, i.e., the intention to treat principle. Indeed, because of the distinct mechanism of action of MD1003 which targets neuro-degeneration (and not inflammation), it is considered that the occurrence of intercurrent events (as listed above under C) are assumed not to be related to the study treatments and, therefore, not relevant for the assessment of the effect of MD1003 compared to placebo. All measurements, regardless of intercurrent events other than premature end of study, are planned to be collected throughout the study. The value for the variable of interest will be used regardless of whether intercurrent events have occurred, and for subjects with premature end of study methods for handling missing data are specified in previous Sections 4.8.2.1 and 4.8.2.2.

The estimand described by the treatment-policy strategy is the effect "MD1003 + any intercurrent events" versus "placebo + any intercurrent events". Thus, dependent on the frequency of intercurrent events in both treatment arms, this estimand will capture a mixture of the effects of treatment and intercurrent events.

The estimand is therefore the effect of treatment conditions on the response probabilities in the targeted patient population, regardless of whether intercurrent events (as defined above in C) had occurred.

Sensitivity analyses related to the estimand concept

The robustness of the estimate derived for the estimand under the treatment policy strategy will be further assessed using pre-planned sensitivity analyses that will explore the potential influence of the following intercurrent events (as defined in more detail above) on the primary efficacy endpoint:

- subgroup analysis by occurrence of at least one MS relapse confirmed by the Adjudication Committee: no / yes
- subgroup analysis by at least one gadolinium enhancing T1-weighted lesion: no / yes
- subgroup analysis by at least one new or enlarging T2-weighted lesion: no / yes
- subgroup analysis by "active disease" defined by at least one MS relapse confirmed by the Adjudication Committee or at least one gadolinium enhancing T1-weighted lesion or at least one new or enlarging T2-weighted lesion: no / yes
- logistic regression analysis as described in Sections 4.8.2.1 and 4.8.2.2 with additional factor "active disease" and its interaction with study treatment
- subgroup analysis for "DMT subgroups" defined by
  - discontinuation of all DMT present (in subjects with DMT present at date of randomization)
  - initiation of DMT (in subjects without DMT present at date of randomization)
  - or none of the above

- logistic regression analysis as described in Sections 4.8.2.1 and 4.8.2.2 but with additional factor "DMT" (as defined above) and its interaction with study treatment
- subgroup analysis by treatment with corticosteroid: no / yes
- logistic regression analysis as described in Sections 4.8.2.1 and 4.8.2.2 but with additional factor "corticosteroid treatment" and its interaction with study treatment
- subgroup analysis by end of study treatment prior to visit M15: no / yes
- logistic regression analysis as described in Sections 4.8.2.1 and 4.8.2.2 but with additional factor "end of study treatment prior to visit M15" and its interaction with study treatment.

4.8.2.6 Additional Analyses Related to the Primary Efficacy Endpoint

The logistic model described in Section 4.8.2.1 will be used to analyze each component (EDSS, TW25) of the primary efficacy endpoint separately. In case of less than 1% responders in at least one of the study treatment groups, a conditional exact logistic regression will be performed. The handling of missing data for each component will follow the logic described in Section 4.8.2.2.

Number and percentage of subjects with

- EDSS or TW25 improvement
- EDSS improvement
- TW25 improvement

will be presented by visit and study treatment group, including two-sided 95% CIs for response probabilities to assess the development over time. For this "by visit" summary, only confirmation at the subsequent visit will be considered as a response, i.e., M3 confirmed M6, M6 confirmed at M9, M9 confirmed at M12, M12 confirmed at M15, M15 confirmed at M18, M18 confirmed M21, M21 confirmed at M24, M24 confirmed at M27.

4.8.2.7 Subgroup Analyses

A specific subgroup analysis will be performed for the subgroups defined by MS disease history (SPMS or PPMS) by using the logistic regression model of the primary efficacy endpoint separately for SPMS and PPMS subjects in the ITT Analysis Set. The two separate logistic regression analyses[9] will now have the factors study treatment group and geographical region and will provide the

- point estimate for the response probability odds ratio related to study treatment
- two-sided 95% profile likelihood CI for that odds ratio

separately for SPMS subjects and for PPMS subjects.

Respective null hypotheses will be rejected if the respective lower CI limit is larger than 1, but only if the null hypothesis for the primary efficacy endpoint based on the ITT Analysis Set of all PMS subjects will have been rejected.

[9] conditional exact logistic regression in case of less than 1% responders in at least one of the study treatment groups

A forest-plot displaying odds ratios comparing study treatment groups and related two-sided 95% confidence intervals will be provided for all subgroups defined in Section 4.8.1.5.

4.8.3 Analysis of Secondary Efficacy Endpoints

Multiplicity regarding testing of null hypotheses for secondary efficacy endpoints will be handled according to the sequential conditional testing procedure described in Section 4.8.1.3.

4.8.3.1 Time to 12-Weeks Confirmed EDSS Progression

Main analysis

The main analysis will be performed for the ITT analysis set.

Time to 12-weeks confirmed EDSS progression will be graphically presented using Kaplan Meier curves. If available, Q1 and median time to 12-weeks confirmed EDSS progression and 95% CIs for each study treatment group will be estimated according to the method by Brookmeyer & Crowley 1982.

A proportional hazards regression model stratified by MS disease history and geographical region will be used to estimate and test the effect of MD1003 relative to placebo on time to 12-weeks confirmed EDSS regression.

The model and its assumptions are described by:

$$h(t,x_{jk},\beta) = h_{0k} \cdot \exp(x_{jk} \cdot \beta)$$

$h(.)$ hazard function as a function of time t relative to date of randomization, study treatment of subject j in stratum k and the unknown regression parameter

$h_{0k}(t)$ unspecified baseline hazard function for stratum k at time t

x_{jk} study treatment for subject j in stratum k

β unknown model parameter (the log hazard ratio) for the study treatment effect (MD1003 relative to placebo) on the log-scale, to be estimated

The hazard ratio is a multiplicative constant $\lambda = \exp(\beta)$ comparing the hazard function for 12-weeks confirmed EDSS progression in the MD1003 treatment group to the hazard function for 12-weeks confirmed EDSS progression in the placebo treatment group. The latter is the baseline hazard function, which is allowed to vary across the four strata. However, the hazard ratio comparing treatments is assumed to be common across strata. A hazard ratio < 1 indicates decreased hazard for 12-weeks confirmed EDSS progression in the MD1003 study treatment group compared to the placebo treatment group.

The null hypothesis within the stratified proportional hazard regression model will be

H_0: hazard ratio $\lambda \geq 1$

and the alternative hypothesis will be

H_A: hazard ratio $\lambda < 1$.

The hazard ratio will be estimated together with a two-sided 95% Wald-type CI, and the null hypothesis will be rejected if the upper Wald-type CI limit is smaller than 1 (in addition, a one-sided Wald-test p-value will be reported). Ties will be handled by the "exact" method.

Summaries for 12-weeks confirmed time to EDSS progression by study treatment group will include:

- Kaplan-Meier plot
- number and percentage of subjects having a 12-weeks confirmed EDSS progression, number and percentage of subjects censored,
- estimated Q1 and median time to 12-weeks confirmed EDSS progression and two-sided 95% CI, if available
- progression-free probabilities at months 6, 9, 12, 15 and 18 estimated by the Kaplan-Meier method and its two-sided 95% CIs using the log-log transformation
- results of a stratified proportional hazards regression analysis (hazard ratio and its two-sided 95% CI)

Kaplan Meier plots by study treatment group will also be provided by disease history (SPMS/PPMS) and by geographical region.

Handling of missing data in the main analysis

For the analysis of time to 12-weeks confirmed EDSS progression, there are 3 types of "missing data":

A. administrative censoring caused by a subject completing the double-blind study phase without a 12-weeks confirmed EDSS progression and without any missing scheduled EDSS assessment: this constitutes administrative censoring and time to 12-weeks confirmed EDSS progression will be censored at the date of the last available scheduled EDSS assessment
   example1: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 5.5 at M18, 5.5 at M21
   example2: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 5.5 at M18, 6 at M21
   In both examples, M21 was the final visit of the double-blind study phase as this study phase was closed according to the rules specified in Section 3.1.3 → both examples are handled as censored at the M21 visit date

B. a subject prematurely discontinues the double-blind study phase without a 12-weeks confirmed EDSS progression and without any missing scheduled EDSS assessment prior to his/her end of double-blind study phase: time to 12-weeks confirmed EDSS progression will be censored at the date of the last available scheduled EDSS assessment
   example1: 5 at baseline, 5.5 at M3, 6 at M6, all subsequent EDSS assessments missing
   example2: 5 at baseline, 5.5 at M3, 5.5 at M6, all subsequent EDSS assessments missing
   → both examples are handled as censored at the M6 visit date

C. a subject has one or more missing scheduled EDSS assessments prior to the his/her end of the double-blind study phase

- C1 – a 12-weeks confirmed EDSS progression occurred prior to the first of such missing scheduled EDSS assessments: the missing EDSS assessments are irrelevant for the determination of the endpoint for such a subject, so no action required
  example1: 5 at baseline, 6 at M3, 6 at M6, missing at M9, missing at M12, 6 at M15
  → handled as 12-weeks confirmed EDSS progression at the M3 visit date
  example2: 5 at baseline, 6 at M3, 6 at M6, missing at M9, 5.5 at M12 → handled as 12-weeks confirmed EDSS progression at the M3 visit date
- C2 – the previous and subsequent scheduled EDSS assessments are available and **at most one of them** constitute a relevant EDSS increase (as defined in Section 3.2.2.1): the missing scheduled EDSS assessment(s) is/are or is/are viewed as irrelevant for the determination of the endpoint 12-weeks confirmed EDSS progression for such a subject (it can only occur at a visit later than those with missing EDSS assessments)
  example1: 5 at baseline, 5.5 at M3, missing at M6, 5.5 at M9
  → missing EDSS assessment at M6 is irrelevant
  example2: 5 at baseline, 5.5 at M3, missing at M6, missing at M9, 6 at M12
  → missing EDSS assessments at M6 and M9 are viewed as irrelevant; earliest time point for a 12-weeks confirmed EDSS progression can be M12 (if there would be a confirmation at M15)
  example3: 5 at baseline, 6 at M3, missing at M6, 5.5 at M9
  → missing EDSS assessment at M6 is viewed to be irrelevant as EDSS 5.5 at M9 does not constitute a relevant EDSS increase (as defined in Section 3.2.2.1); earliest time point for a 12-weeks confirmed EDSS progression can be M12 (if there would be a confirmation at M15)
- C3 – the previous and subsequent scheduled EDSS assessments are available and **both of them** constitute a relevant EDSS increase (as defined in Section 3.2.2.1): the missing scheduled EDSS assessment(s) is/are irrelevant, the subject is handled as having a 12-weeks confirmed EDSS progression at the earlier of the two visit dates
  example1: 5 at baseline, 6 at M3, missing at M6, 6 at M9
  → missing EDSS assessment at M6 is irrelevant, 12-weeks confirmed EDSS progression at the M3 visit date
  example2: 5 at baseline, 6 at M3, missing at M6, missing at M9, 6 at M12
  → missing EDSS assessments at M6 and M9 are irrelevant, 12-weeks confirmed EDSS progression at the M3 visit date.

First sensitivity analysis with additional censoring rule

The main analysis will be repeated with the following additional censoring rule:

- if a subject meets the EDSS-based response criteria at M12 and M15 for the primary efficacy endpoint (see Section 3.2.1), then the subject is not considered as having 12-week confirmed EDSS progression
- that subject's time to 12-weeks confirmed EDSS progression will be censored at the date of the last EDSS assessment at a scheduled visit in the double-blind study phase regardless of whether the criteria of 12-week confirmed EDSS progression are met or not.

Sensitivity analyses for handling of missing values

As censoring rules for handling missing values of types A and B may introduce informative censoring and too long censored observation times, the following alternative rules for types A and B will be used to assess such potential bias by a sensitivity analysis (censoring for EDSS improvements at M12 and M15 will not be applied):

A. administrative censoring caused by a subject completing the double-blind study phase without a 12-weeks confirmed EDSS progression and without any missing scheduled EDSS assessment: this constitutes administrative censoring and time to 12-weeks confirmed EDSS progression will be censored at the date of the last available scheduled EDSS assessment not constituting a relevant EDSS increase (as defined in Section 3.2.2.1)
   example1: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 5.5 at M18, 5.5 at M21
   example2: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 5.5 at M18, 6 at M21
   In both examples, M21 was the final visit of the double-blind study phase as this study phase was closed according to the rules specified in Section 3.1.3 → example 1 handled as censored at the M21 visit date, example 2 handled as censored at the M18 visit date

B. a subject prematurely discontinues the double-blind study phase without a 12-weeks confirmed EDSS progression and without any missing scheduled EDSS assessment prior to his/her end of double-blind study phase:
   - B1 – the last available scheduled EDSS assessment constitutes a relevant EDSS increase (as defined in Section 3.2.2.1): the subject will be handled as having a 12-weeks confirmed EDSS progression and the date of the 12-weeks confirmed EDSS progression will be the date of the last available scheduled EDSS assessment
     example: 5 at baseline, 5.5 at M3, 6 at M6, all subsequent EDSS assessments missing → handled as 12-weeks confirmed EDSS progression at the M6 visit date
   - B2 – the last available scheduled EDSS assessment does not constitute a relevant EDSS increase (as defined in Section 3.2.2.1): this will be handled as censored subject, and the censoring date will be the date of the last available scheduled EDSS assessment
     example: 5 at baseline, 5.5 at M3, 5.5 at M6, all subsequent EDSS assessments missing → handled as censored at the M6 visit date.

Multiple imputation sensitivity analysis

Censoring for EDSS improvements at M12 and M15 will not be applied for this sensitivity analysis. The same general approach for multiple imputation as outlined in Section 4.8.2.4 will be followed. The proposed prediction model for missing EDSS assessments, dichotomized as either "relevant EDSS increase from baseline" (as defined in Section 3.2.2.1) or "no relevant EDSS increase from baseline" is a logistic regression model with the following covariates:

- randomized study treatment: MD1003 or placebo
- MS disease history: SPMS or PPMS
- geographical region NA/AUS or Europe
- baseline EDSS: "5.5 or below" or "6 or above"
- baseline TW25 (continuous covariate)
- age at visit M-1 (continuous covariate)
- sex: female or male
- BMI at visit M-1 (continuous covariate)
- physical therapy after date of randomization and prior to the date of the first missing EDSS assessment: no or yes
- anti-spasticity drugs after date of randomization and prior to the date of the first missing EDSS assessment: no or yes
- DMT at date of randomization: no or yes
- DMT after date of randomization and prior to the date of the first missing EDSS assessment:
  - discontinuation of all DMT (in subjects with DMT present at date of randomization)
  - initiation of DMT (in subjects without DMT present at date of randomization)
  - or none of the above
- corticosteroid treatment after date of randomization and prior to the date of the first missing EDSS assessment: no or yes
- at least one MS relapse (confirmed by the adjudication committee) with onset after date of randomization and prior to the date of the first missing EDSS assessment: no or yes
- TW25 (continuous covariate) at the scheduled visit with the first missing EDSS assessment; if also missing, then the latest available TW25 at a previous scheduled visit
- relevant EDSS increase (as per SAP Section 3.2.2.1) at M3 Visit: no or yes
- relevant EDSS increase (as per SAP Section 3.2.2.1) at M6 Visit: no or yes
- relevant EDSS increase (as per SAP Section 3.2.2.1) at M9 Visit: no or yes
- relevant EDSS increase (as per SAP Section 3.2.2.1) at M12 Visit: no or yes
- relevant EDSS increase (as per SAP Section 3.2.2.1) at M15 Visit: no or yes
- relevant EDSS increase (as per SAP Section 3.2.2.1) at M18 Visit: no or yes
- relevant EDSS increase (as per SAP Section 3.2.2.1) at M21 Visit: no or yes

- relevant EDSS increase (as per SAP Section 3.2.2.1) at M24 Visit: no or yes
- relevant EDSS increase (as per SAP Section 3.2.2.1) at M27 Visit: no or yes.

The algorithm for the multiple imputation of missing EDSS assessments will be FCS as described in Section 4.8.2.4 and the multiply imputed datasets will be analyzed by the stratified proportional hazards model. The multiple imputation point estimate and two-sided 95% CI for the log hazard ratio will be obtained (similar as described in Section 4.8.2.4 for the log odds ratio) and finally transformed to point estimate and CI for the multiple imputation hazard ratio.

Sensitivity analysis based on the MITT Analysis Set

Sensitivity analysis based on the MITT Analysis Set (otherwise following the main analysis) will be conducted.

Subgroup analyses

A specific subgroup analyses will be performed for the subgroups defined by MS disease history (SPMS or PPMS): there will be analyses (Kaplan-Meier plots and proportional hazards regression model) of time to 12-weeks confirmed EDSS progression (as per the main analysis) separately for SPMS and PPMS subjects in the ITT Analysis Set. The two separate proportional hazard regression analyses will now be stratified by geographical region and will provide the

- point estimate for the hazard ratio related to study treatment
- two-sided 95% CI for that hazard ratio

separately for SPMS subjects and for PPMS subjects.

Respective null hypotheses will be rejected if the respective lower CI limit is larger than 1, but only if the null hypothesis for time to 12-weeks confirmed EDSS progression based on the ITT Analysis Set of all PMS subjects will have been rejected.

A forest-plot displaying hazard ratios comparing study treatment groups and related two-sided 95% CIs for the hazard ratios will be provided for all subgroups defined in Section 4.8.1.5.

4.8.3.2 Analysis of Clinical Global Impression of Improvement

The following summaries will be provided by study treatment group:

- number and percentage of subjects for each CGI category
- number and percentage of subjects for each SGI category.

Let $\pi_{CGI,i,MD1003}$, $\pi_{SGI,i,MD1003}$, $\pi_{CGI,i,placebo}$ and $\pi_{SGI,i,placebo}$ denote the probabilities for category i, i=1,..,7 of CGI and SGI for MD1003 and placebo, respectively, at visit M15.

Null hypothesis for CGI at visit M15 is stated as

$H_{0,\,CGI}$: the distributions in both study treatment groups are the same, i.e.,

$$\pi_{CGI,i,MD1003} = \pi_{CGI,i,placebo} \text{ for all } i = 1,..,6$$

with alternative hypothesis for CGI at visit M15

$H_{A,\,CGI}$:
the distribution in the MD1003 treatment group is stochastically strictly smaller (less) than the distribution in the placebo treatment group, i.e.,

$$\sum_{i=1}^{j}\pi_{CGI,i,MD1003} \geq \sum_{i=1}^{j}\pi_{CGI,i,placebo} \text{ for all } j = 1,..,6 \text{ and}$$

$$\sum_{i=1}^{k}\pi_{CGI,i,MD1003} > \sum_{i=1}^{k}\pi_{CGI,i,placebo} \text{ for at least one } k \text{ in } \{1,..,6\}.$$

Null hypothesis for SGI at visit M15 is stated as

$H_{0,\,SGI}$: the distributions in both study treatment groups are the same, i.e.,

$$\pi_{SGI,i,MD1003} = \pi_{SGI,i,placebo} \text{ for all } i = 1,..,6$$

with alternative hypothesis for SGI at visit M15

$H_{A,\,SGI}$: the distribution in the MD1003 treatment group is stochastically strictly smaller (less)
than the distribution in the placebo treatment group, i.e.,

$$\sum_{i=1}^{j}\pi_{SGI,i,MD1003} \geq \sum_{i=1}^{j}\pi_{SGI,i,placebo} \text{ for all } j = 1,..,6 \text{ and}$$

$$\sum_{i=1}^{k}\pi_{SGI,i,MD1003} > \sum_{i=1}^{k}\pi_{SGI,i,placebo} \text{ for at least one } k \text{ in } \{1,..,6\}.$$

CGI and SGI at visit M15 will be compared between study treatment groups using the non-parametric one-sided van Elteren test stratified by MS disease history and geographical region (one-sided significance level 0.025). The one-sided van Elteren test is a stratified version of the one-sided Wilcoxon (Mann-Whitney) test: ranks and Wilcoxon scores will be generated separately within each stratum and the four by-stratum statistics will be weighted by the stratum sizes.

In addition, MD1003 and placebo will be compared by the terms $P(CGI_{MD1003} < CGI_{placebo})$ and $P(SGI_{MD1003} < SGI_{placebo})$. For example, $P(CGI_{MD1003} < CGI_{placebo})$ can be interpreted as the probability of achieving a better CGI outcome (lower CGI score) for a randomly chosen subject of the MD1003 group compared to a randomly chosen subject of the placebo group. These quantities are estimated by comparing CGI (SGI) at visit M15 for all n_{MD1003} subjects to CGI (SGI) at visit M15 of all $m_{placebo}$ subjects: counting the pairs where CGI (SGI) in the MD1003 group is better (smaller) and dividing this count by the number of all pairs: $n_{MD1003} \times m_{Placebo}$. Two-sided 95% CIs for $P(CGI_{MD1003} < CGI_{placebo})$ and $P(SGI_{MD1003} < SGI_{placebo})$ will be provided as well.

Handling of missing data in the main analysis

The general multiple imputation approach as outlined in Section 4.8.2.4 will be applied. The proposed prediction model for missing CGI (SGI) values at visit M15, is a logistic regression model with the following covariates:

- randomized study treatment: MD1003 or placebo
- MS disease history: SPMS or PPMS
- geographical region NA/AUS or Europe
- baseline EDSS: "up to 5.5" or "6 or above"
- baseline TW25 (continuous covariate)
- age at visit M-1 (continuous covariate)
- sex: female or male
- BMI at visit M-1 (continuous covariate)
- physical therapy after date of randomization and prior to visit M15 assessment: no or yes
- anti-spasticity drugs after date of randomization and prior to visit M15 assessment: no or yes
- DMT at date of randomization: no or yes
- DMT after date of randomization and prior to the date of visit M15 assessment:
  - discontinuation of all DMT (in subjects with DMT present at date of randomization)
  - initiation of DMT (in subjects without DMT present at date of randomization)
  - or none of the above
- corticosteroid treatment after date of randomization and prior to visit M15 assessment: no or yes
- at least one MS relapse (confirmed by the adjudication committee) with onset after date of randomization up to visit M15 assessment: no or yes
- TW25 at visits M3, M6, M9, M12 (continuous covariate)
- EDSS at visits M3, M6, M9, M12, M15: "up to 5.5" or "6 or above".

A refinement of the prediction model may be proposed at the Blinded Data Review Meeting prior to unblinding of the randomization.

Derivation of multiple imputation p-value based on the stratified van Elteren test

The algorithm for the multiple imputation of missing CGI (SGI) values at visit M15 will be FCS as described in Section 4.8.2.4 and the multiply imputed datasets will be analyzed by van Elteren tests stratified by MS history and geographical region: the difference between observed stratified sum of Wilcoxon scores and the expected stratified sum of Wilcoxon scores together with its standard deviation will be obtained from each multiply imputed dataset. The combination of these results and the final derivation of multiple imputation p-values for CGI (SGI) will be performed in analogy to the description in Section 4.8.2.4.

Derivation of multiple imputation probability estimates

For each multiply imputed dataset, the estimate for $P(CGI_{MD1003} < CGI_{placebo})$ its standard deviation will be obtained. The combination of these results and the final derivation of multiple imputation point estimate and two-sided 95% CI for $P(CGI_{MD1003} < CGI_{placebo})$ will be performed in analogy to the description in Section 4.8.2.4. The same applies for $P(SGI_{MD1003} < SGI_{placebo})$.

Sensitivity analyses based on the MITT Analysis Set

A sensitivity analysis based on the MITT Analysis Set will be performed.

Subgroup Analysis

A forest-plot displaying $P(SGI_{MD1003} < SGI_{placebo})$ estimates comparing study treatment groups, the related two-sided 95% CIs and the one-sided p-value of the stratified van Elteren test will be provided for the subgroups defined in Section 4.8.1.5. The same applies for CGI.

4.8.3.3 Analysis of TW25

A summary of TW25 values and percentage changes from baseline will be provided by visit and study treatment group. Note: If a subject performed only one of the two TW25 attempts at a visit, then the available TW25 value will be used as the subject's mean value for that visit.

Null hypothesis for percentage change in TW25 from baseline (defined in 3.2.2.3) to visit M15 is stated as

H_0: the distributions in both study treatment groups are the same

with alternative hypothesis:

H_A: the distribution in the MD1003 treatment group is stochastically strictly smaller (less) than the distribution in the placebo treatment group, i.e.,

$$P_{MD1003}(\%\text{change in TW} \leq x) \geq P_{placebo}(\%\text{change in TW} \leq x) \text{ for all } x \text{ and}$$

$$P_{MD1003}(\%\text{change in TW} \leq y) > P_{placebo}(\%\text{change in TW} \leq y) \text{ for some } y.$$

The percentage change in TW25 from baseline to visit M15 will be compared between study treatment groups using the non-parametric one-sided van Elteren test stratified for MS disease history and geographical region (same approach as described in Section 4.8.3.2, one-sided significance level 0.025).

In addition, the Hodges-Lehman point estimate for shift of distributions and its two-sided 95% CI will be provided.

Handling of missing data in the main analysis

If both TW25 values at a visit are missing due to inability to perform the trials because of MS worsening, then the following single imputation rule will be applied:

- the maximum TW25 value observed in the ITT Analysis Set (at any scheduled visit and in any study treatment group) will be identified

- impute that value + 1 second for the missing TW25 value at that visit.

For remaining missing TW25 values, the general multiple imputation approach as outlined in Section 4.8.2.4 will be applied. The proposed prediction model for missing TW25 values at visit M15, is a linear regression model with the following covariates:

- randomized study treatment: MD1003 or placebo
- MS disease history: SPMS or PPMS
- geographical region NA/AUS or Europe
- baseline EDSS: “up to 5.5” or “6 or above”
- baseline TW25 (continuous covariate)
- age at visit M-1 (continuous covariate)
- sex: female or male
- BMI at visit M-1 (continuous covariate)
- physical therapy after date of randomization and prior to visit M15 assessment: no or yes
- anti-spasticity drugs after date of randomization and prior to visit M15 assessment: no or yes
- DMT at date of randomization: no or yes
- DMT after date of randomization and prior to the date of visit M15 assessment:
  - discontinuation of all DMT (in subjects with DMT present at date of randomization)
  - initiation of DMT (in subjects without DMT present at date of randomization)
  - or none of the above
- corticosteroid treatment after date of randomization and prior to visit M15 assessment: no or yes
- at least one MS relapse (confirmed by the adjudication committee) with onset after date of randomization and prior to visit M15 assessment: no or yes
- TW25 at visits M3, M6, M9, M12 (continuous covariate)
- EDSS at visits M3, M6, M9, M12, M15: “up to 5.5” or “6 or above”.

A refinement of the prediction model may be proposed at the Blinded Data Review Meeting prior to unblinding of the randomization.

The algorithm for the multiple imputation of missing TW25 values at visit M15 will be FCS as described in Section 4.8.2.4 (an additional condition is that negative TW25 imputations must be avoided: in case a negative TW25 value is predicted by the regression model, it will be replaced by the minimum observed value in the respective study treatment group). The percentage change in TW25 from baseline to visit M15 will be calculated for all subjects with multiply imputed TW25 values at visit M15.

Derivation of multiple imputation p-value based on the stratified van Elteren test

The multiply imputed datasets will be analyzed by van Elteren tests stratified by MS history and geographical region: the difference between observed stratified sum of Wilcoxon

scores and the expected stratified sum of Wilcoxon scores together with its standard deviation will be obtained from each multiply imputed dataset. The combination of these results and the final derivation of multiple imputation p-value for percentage change in TW25 from baseline to visit M15 will be performed in analogy to the description in Section 4.8.2.4.

Derivation of multiple imputation location shift estimates

For each multiply imputed dataset, the Hodges-Lehmann point estimate for location shift and its standard deviation will be obtained. The combination of these results and the final derivation of multiple imputation point estimate and two-sided 95% CI for the location shift in percentage change in TW25 from baseline to visit M15 will be performed in analogy to the description in Section 4.8.2.4.

Sensitivity analyses based on the MITT Analysis Set

A sensitivity analysis based on the MITT Analysis Set will be performed.

Subgroup Analysis

A forest-plot displaying Hodges-Lehmann location shift estimates comparing study treatment groups, related two-sided 95% CIs and the one-sided p-value of the stratified van Elteren test will be provided for the subgroups defined in Section 4.8.1.5.

4.8.4 Analysis of Exploratory Efficacy Endpoints

Analysis of exploratory efficacy endpoints will be based on the ITT Analysis Set.

4.8.4.1 Analysis of Exploratory Brain MRI Endpoints

Exploratory brain MRI data will be summarized by visit and study treatment group. The respective endpoints, defined in Section 3.2.3.1, will be analyzed by the non-parametric van Elteren test stratified by MS disease history and geographical region.

4.8.4.2 Analysis of Remote Monitoring of Ambulatory Activity

A summary of average daily step counts and changes from baseline in average daily step counts will be provided by visit and study treatment group.

The change from baseline in average daily step counts will be compared between study treatment groups using a mixed model repeated measures approach with fixed effects

- MS disease history
- geographical region
- study treatment group
- month of study visit
- visit-by-study treatment group interaction

and random subject effect.

The covariance matrix for the random subject effect will be block diagonal (with each block corresponding to a subject) with unstructured non-zero block diagonal elements as

defaults. An alternative structure for the block diagonal part of the covariance matrix for the random subject effect is the spatial power law structure in order to reduce the number of parameters to be estimated. Akaike's Information Criteria and Schwarz' Bayesian Criterion are used to assist in the selection process for the covariance structure. A converging model with the better criterion value could be used as justification for using that covariance structure.

Hypotheses related to study treatment are expressed as linear combinations of fixed effects given above. Least squared means point estimates and two-sided 95% CIs will be provided.

4.8.4.3 Analysis of MSQOL-54 and CAREQOL-MS

MSQOL-54 summary scores and CAREQOL-MS sub-scores as well as changes from baseline will be summarized by visit and study treatment group. Changes from baseline to visit M15 will be analyzed by the non-parametric van Elteren test stratified by MS disease history and geographical region.

4.8.4.4 Analysis of Kurtzke Functional System Scores

Descriptive summaries of Kurtzke Functional System Scores and change from baseline will be presented by visit and study treatment group.

Change from baseline to visit M15 will be analyzed for each Kurtzke Functional System Score by the non-parametric van Elteren test stratified by MS disease history and geographical region.

4.8.4.5 Analysis of Symbol Digit Modalities Test

SDMT as well as changes from baseline (including the categorized version defined in Section 3.2.3.5) will be summarized by visit and study treatment group. SDMT at visit M15 and changes from baseline to visit M15 will be analyzed by the non-parametric van Elteren test stratified by MS disease history and geographical region.

4.8.4.6 Analysis of Neurofilament Blood Concentration

Neurofilament blood concentration as well as absolute change from baseline will be summarized by visit and study treatment group.

Absolute changes from baseline in neurofilament blood concentration will be compared between study treatment groups using a mixed model repeated measures approach with fixed effects

- MS disease history
- geographical region
- DMT at date of randomization (no or yes)
- study treatment group
- month of study visit
- visit-by-study treatment group interaction

and random subject effect.

The covariance matrix for the random subject effect will be block diagonal (with each block corresponding to a subject) with unstructured non-zero block diagonal elements as

defaults. An alternative structure for the block diagonal part of the covariance matrix for the random subject effect is the spatial power law structure in order to reduce the number of parameters to be estimated. Akaike's Information Criteria and Schwarz' Bayesian Criterion are used to assist in the selection process for the covariance structure. A converging model with the better criterion value could be used as justification for using that covariance structure.

Hypotheses related to study treatment are expressed as linear combinations of fixed effects given above. Least squared means point estimates and two-sided 95% CIs will be provided.

4.9 Safety Evaluation

All safety summaries and analyses will be based on the Safety Analysis Set. Listings will include raw and derived safety data.

4.9.1 Analysis of Extent of Exposure

A summary of the duration of treatment will be provided by study treatment group.

4.9.2 Analysis of Duration of Study

A summary of the duration of study (length of follow-up) will be provided by study treatment group.

4.9.3 Analysis of Adverse Events

The following summaries will be provided:

- overall summary of the number and percentage of subjects reporting any TEAE, any TEAE causally related to study medication, any TEAE causally related to study procedure, any serious TEAE, any serious TEAE causally related to study medication, any TEAE leading to treatment discontinuation, any TEAE leading to death and number of deaths (the number of respective AEs will also be reported)
- number and percentage of subjects reporting a TEAE by treatment group, SOC, and PT
- number and percentage of subjects reporting a TEAE by treatment group, severity, SOC and PT
- number and percentage of subjects reporting a treatment-emergent adverse event by treatment group, causality to study treatment, SOC and PT
- number and percentage of subjects reporting a treatment-emergent adverse event by treatment group, causality to study procedure, SOC and PT.

Adverse event summaries will be ordered in terms of decreasing incidence for SOC, and PT within SOC, in the MD1003 treatment group, and then similarly by decreasing frequency in the Placebo group, and then alphabetically for SOC, and PT within SOC.

For each subject and each adverse event, the worst severity recorded will be attributed and used in the by-severity summaries. Similarly, the worst causality (most related to treatment) will be attributed and used in the by-causality summaries. If severity or causality is missing, the worst case will be assumed.

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

The following summaries will be provided:

- number and percentage of subjects reporting treatment emergent SAEs by treatment group, SOC, and PT
- number and percentage of subjects reporting treatment emergent SAEs assessed as causally related to study medication by treatment group, SOC, and PT
- number and percentage of subjects reporting TEAEs leading to death by treatment group, SOC, and PT
- number and percentage of subjects reporting TEAEs leading to study treatment discontinuation by treatment group, SOC, and PT
- number and percentage of subjects reporting a TEAEs of suicidal ideation or behavior by treatment group, SOC, and PT.

In particular, the following listings will be provided:

- listing of all deaths that occurred during the study
- listing of all SAEs
- listing of all AEs leading to discontinuation of study treatment
- listing of all AEs of suicidal ideation or behavior.

4.9.5 Analysis of Relapses

Annualized rates of relapses reported by investigators will be calculated by study treatment group as shown below:

$$annual\ relapse\ rate = \frac{sum\ of\ relapses\ reported\ by\ investigators}{sum\ of\ days\ at\ risk\ for\ a\ relapse} * 365.25$$

For each subject, "days at risk for a relapse" is defined as date of subject's last visit in the double-blind study part minus the date of randomization + 1". The "sum of days at risk for a relapse" is the sum across all subjects of the respective study treatment group.

The annualized relapse rates based on only relapses confirmed by the Adjudication Committee will also be calculated accordingly.

The following summaries of relapses reported by investigators will be provided by study treatment group:

- number and percentage of subjects with at least one relapse
- Kaplan-Meier curve for time from randomization date to onset date of subject's first relapse
- number of relapses per subject
- total number of relapses and for those relapses:
  - type of relapse (protocol defined, non-protocol defined and suspected relapse)
  - number and percentage of relapses treated with steroid
- annualized relapse rate (all, protocol defined, non-protocol defined and suspected relapse).

Further summaries related to relapses will be provided by study treatment group:

- Kaplan-Meier curve for time from randomization to onset date of subject's first relapse confirmed by the Adjudication Committee
- annualized rate of relapses confirmed by the Adjudication Committee
- relapse interpretation compared between investigator and Adjudication Committee.

4.9.6 Analysis of Safety Brain MRI Endpoints

Safety Brain MRI endpoints will be summarized descriptively by visit and study treatment group.

4.9.7 Analysis of Clinical Laboratory Tests

The following summaries will be provided by laboratory test and study treatment group:

- laboratory test value and change from baseline by visit
- number and percentage of subjects experiencing the worst of normal, abnormal not clinically meaningful or abnormal clinically meaningful laboratory test result as determined by the Site Investigator
- number and percentage of subjects shifting from baseline category to worst post-baseline value (categories of normal, abnormal not clinically significant, or abnormal clinically significant laboratory test result).

4.9.8 Analysis of ECG results

The following summaries of ECG reporting will be provided:

- ECG parameter and change from baseline by treatment group and time point
- number and percentage of subjects with QT or QTc intervals exceeding predefined upper limit (>450ms, >480ms and >500ms), the number and percentage of subjects with change from baseline in QT or QTc intervals exceeding predefined upper limit (>30ms, >60ms) and the combination of these 2 conditions
- number and percentage of subjects with ECG abnormalities by treatment group
- shift table for ECG abnormalities from baseline by treatment group.

A listing of ECG data with abnormal values flagged will be provided.

4.9.9 Analysis of the Columbia-Suicide Severity Rating Scale

The following summaries will be provided by visit and study treatment group:

- C-SSRS total score and change from baseline
- suicidal ideation and suicidal behavior component scores.

4.9.10 Analysis of Vital Signs

The following summaries of vital signs (including BMI) will be provided by visit and study treatment group:

- vital sign and change from baseline

- number and percentage of subjects experiencing normal, abnormal not clinically significant, or abnormal clinically significant vital sign.

4.10 Other Analyses

4.10.1 Treatment Assignment Questionnaire

Treatment assignment questionnaire is performed at the end of the randomization period or at study termination visit to assess whether the subject and the investigator guessed which treatment the subject was receiving.

The following tables will be presented by study treatment group

- a comparison table of the subject versus the investigator interpretation of study treatment
- a table of the subject and investigator interpretation of study treatment by whether or not the subject had an adverse event.

A listing of the treatment assignment questionnaire will be provided.

4.10.2 Study Treatment Compliance

The following tables will be presented by study treatment group:

- summary of study treatment compliance in the double-blind study phase up to the M15 visit (including)
- summary of study treatment compliance in the overall double-blind study phase.

A listing of compliance results by subject will be provided.

4.10.3 Study Treatment Interruptions

The following tables will be presented by study treatment group.

For the double-blind study phase up to the M15 visit (including):

- summary of number of study treatment interruptions per subject
- summary of subjects with no / at least one study treatment interruption
- summary of durations of study treatment interruptions.

For the overall double-blind study phase:

- summary of number of study treatment interruptions per subject
- summary of subjects with no / at least one study treatment interruption
- summary of durations of study treatment interruptions.

A listing of study treatment interruptions by subject will be provided, including the reason for study treatment interruption.

4.11 Changes in the Conduct of the Study or Planned Analysis

Study Protocol 4 (dated 12-Nov-2018) has been the last study protocol version prior to this final version of the SAP. There were no changes in the conduct of the study since then, but there was a mistake in Study Protocol 4: the randomization was in fact stratified by disease history and center (study site), not by disease history and geographical region as stated.

Relevant changes in the planned statistical analyses of primary and secondary efficacy endpoints are summarized in Table 3 below.

Table 3: Relevant changes in planned statistical analyses from last study protocol version to final SAP

Study Protocol version 4	Final SAP 6.0	Justification for change
Intent-to-treat population	Intent-to-treat analysis set	Changed per ICH terminology
Not included	Modified intent-to-treat analysis set	Included to allow sensitivity analyses excluding subjects with high biotin plasma concentrations at baseline or post-baseline (the latter if randomized to placebo)
Full analyzable set	Full analysis set	Changed per ICH terminology
Per Protocol population	Per-protocol analysis set	Changed per ICH terminology and criteria more specific for the primary efficacy endpoint
Two-sided significance levels of 0.05	One-sided significance levels of 0.025	Study objective is demonstration of MD1003's superiority over placebo, i.e., a clearly one-sided hypothesis setting; one-sided approach simplifies the sequential conditional testing procedure for the multiple secondary efficacy endpoints
Not included	First sensitivity analysis of the primary efficacy endpoint for handling of subjects with 12-weeks confirmed EDSS progression	Analysis requested by the FDA and implemented as first sensitivity analysis of the primary efficacy endpoint
Only single imputation of missing primary efficacy endpoint data (as non-responders)	Additional sensitivity analyses (tipping point and multiple imputation) for handling of missing primary efficacy endpoint data	Allowing more insight in the effect of single imputation in the main analysis of the primary efficacy endpoint
Multiple imputation for secondary efficacy endpoints described only vaguely	Much more detailed description for handling of missing secondary efficacy endpoint data (including	Allowing a consistent approach across primary and secondary efficacy endpoints, in particular regarding multiple imputation

Study Protocol version 4	Final SAP 6.0	Justification for change
	multiple imputation models); for TW25 change from baseline, a rule for visits with both TW25 trials missing due to inability caused by MS worsening has been introduced	models); considering available information on the reason for missing TW25 values
Logistic regression with factors MS history and center	Logistic regression with factors MS history and geographical region	As the number of centers is too large for the number of subjects, replacing factor center by geographical region stabilizes the statistical analysis.
Not included	Use of a conditional exact logistic regression analysis in case of less than 1% responders in at least one of the study treatment groups	Results of the asymptotic logistic regression analysis may not be valid with such a sparse response probability
Handling of intercurrent events (relapses and steroid treatment during the study) only briefly mentioned	More extended and detailed sensitivity analyses within the "estimand" framework as described in the ICH E9 Addendum	Update according to new regulatory guidelines
Not included	First sensitivity analysis of time to 12-weeks confirmed EDSS progression for handling of subjects with EDSS improvement at M12 and M15	Analysis requested by the FDA and implemented as first sensitivity analysis of this secondary efficacy endpoint
Secondary efficacy endpoints CGI, SGI and TW25 change from baseline to M15 visit analyzed by van Elteren test stratified by MS history	Secondary efficacy endpoints CGI, SGI and TW25 change from baseline to M15 visits analyzed by van Elteren test stratified by MS history and geographical region	Consistency with randomization and analysis of primary and other secondary efficacy endpoints
Not included	Additional analyses related to the primary efficacy endpoint and analysis of time to 24-weeks confirmed EDSS progression	Requested by the EMA

In addition, for the exploratory efficacy endpoints related to remote monitoring of ambulatory activity, the relevant period prior to each post-baseline visit has been extended from 7 to 21 days, but only including "valid" days with at least 130 steps and requiring at

least 3 valid days for a visit to be included in the statistical analysis. This approach has been recommended by the investigator coordinator to stabilize the endpoint.

5 REFERENCES

Benito-León J, Rivera-Navarro J, Guerrero AL, de Las Heras V, Balseiro J, Rodríguez E, Belló M, Martínez-Martín P; caregiver quality of life in multiple sclerosis (CAREQOL-MS) Study Group. The CAREQOL-MS was a useful instrument to measure caregiver quality of life in multiple sclerosis. J Clin Epidemiol 2011; 64; 675-686

Berglund P, Heeringa S. Multiple Imputation of Missing Data Using SAS. SAS Institute Inc. 2014

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics 1982; 38; 29-41

Campbell G, Pennello G, Yue L. Missing data in the regulation of medical devices. Journal of Biopharmaceutical Statistics 2011; 21; 180-195

Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron 1976; 16; 31-41

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use: Draft ICH E9 (R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. 2017

Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. Statistics in Medicine 2014; 33; 4170-4185

Vickrey BG, Hays RD, Genovese BJ, Myers LW, Ellison GW. Comparison of a generic to disease-targeted health-related quality-of-life measures for multiple sclerosis. J Clin Epidemiol 1997; 50; 557-569

Vickrey BG, Hays RD, Harooni R, Myers LW, Ellison GW. A health-related quality of life measure for multiple sclerosis. Qual Life Res 1995; 4; 187-206

Yan X, Lee S, Li N. Missing data handling methods in medical device clinical trials. Journal of Biopharmaceutical Statistics 2009; 19; 1085-1098

6 ADDITIONAL ANALYSES FOR EUROPEAN MEDICINES AGENCY FOLLOWING SCIENTIFIC ADVICE

In addition to analyses described in the previous sections of this SAP, analyses described in this section will be provided to European Medicines Agency (EMA) as advised in a Scientific Advice provided in February 2019.

Additional analyses related to the primary efficacy endpoint

Analyses (descriptively and by applying statistical methods described in Sections 4.8.2.1 and 4.8.2.2) assessing the durability of MD1003's effect and the potential impact of discordant EDSS versus TW25 outcomes at visits M12 and M15 for:

- TW25 or EDSS improvement (as defined in Section 3.2.1) at M12 confirmed at M18[10]
- TW25 or EDSS improvement (as defined in Section 3.2.1) at M9 confirmed at M15[11]
- modified primary efficacy endpoint considering subjects with discordant responses as failures; a discordant response is defined as:
  - response on EDSS at M12 and M15 but worsening (increase in TW25 by more than 20% compared to baseline) of TW25 at M12 and M15 visits
  - or response on TW25 at M12 and M15 but worsening (increase in EDSS from baseline as defined for the EDSS related secondary endpoint) of EDSS at M12 and M15 visits.

Additional sensitivity analysis related to secondary efficacy endpoints

Time to 24-weeks confirmed EDSS progression will be analyzed (in addition to time to 12-weeks confirmed EDSS progression).

24-weeks EDSS progression is defined as an increase of at least 1 point for baseline EDSS up to 5.5 and of at least 0.5 point for baseline EDSS 6 to 6.5 with respective confirmation 12 and 24 weeks later (with time windows of ± 10 days up to 1 year after randomization, ± 15 days afterwards). The baseline EDSS value is defined as in Section 3.2.1.

Date of 24-weeks confirmed EDSS progression will be the first date of an EDSS progression (as defined above) that is confirmed 12 and 24 weeks later. Handling of missing data EDSS assessments and censoring are described further below. Time to 24-weeks confirmed EDSS progression will be calculated as date of 24-weeks confirmed EDSS progression (or censoring) minus date of randomization plus 1; it will be expressed in weeks.

[10] i.e., defined EDSS or TW25 decrease from baseline observed at visit M12 and at visit M18

[11] i.e., defined EDSS or TW25 decrease from baseline observed at visit M9 and at visit M15

For the main analysis of time to 24-weeks confirmed EDSS progression, there are 3 types of "missing data":

A. administrative censoring caused by a subject completing the double-blind study phase without a 24-weeks confirmed EDSS progression and without any missing scheduled EDSS assessment: this constitutes administrative censoring and time to 24-weeks confirmed EDSS progression will be censored at the date of the last available scheduled EDSS assessment
example1: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 5.5 at M18, 5.5 at M21
example2: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 5.5 at M18, 6 at M21
example3: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 6 at M18, 6 at M21
In all 3 examples, M21 was the final visit of the double-blind study phase when this study phase was closed according to the rules specified in Section 3.1.3 → all 3 examples are handled as censored at the M21 visit date

B. a subject prematurely discontinues the double-blind study phase without a 24-weeks confirmed EDSS progression and without any missing scheduled EDSS assessment prior to his/her end of double-blind study phase: time to 24-weeks confirmed EDSS progression will be censored at the date of the last available scheduled EDSS assessment
example1: 5 at baseline, 5.5 at M3, 6 at M6, 6 at M9, all subsequent EDSS assessments missing
example2: 5 at baseline, 5.5 at M3, 5.5 at M6, 6 at M9, all subsequent EDSS assessments missing
example3: 5 at baseline, 5.5 at M3, 5.5 at M6, 5.5 at M9, all subsequent EDSS assessments missing
→ all 3 examples are handled as censored at the M9 visit date

C. a subject has one or more missing scheduled EDSS assessments prior to the his/her end of the double-blind study phase

- C1 – a 24-weeks confirmed EDSS progression occurred prior to the first of such missing scheduled EDSS assessments: the missing EDSS assessments are irrelevant for the determination of the endpoint for such a subject, so no action required
example: 5 at baseline, 6 at M3, 6 at M6, 6 at M9, missing at M12, missing at M12, 5 at M15 → handled as 24-weeks confirmed EDSS progression at M3 visit date
- C2 – the previous and subsequent scheduled EDSS assessments are available and **at most one of them** constitute a relevant EDSS increase (as defined in Section 3.2.2.1): the missing scheduled EDSS assessment is or is viewed as irrelevant for the determination of the endpoint 24-weeks confirmed EDSS progression for such a subject (it can only occur at a visit later than those with missing EDSS assessments)
example1: 5 at baseline, 5.5 at M3, missing at M6, 5.5 at M9
→ missing EDSS assessment at M6 is irrelevant

example2: 5 at baseline, 5.5 at M3, missing at M6, missing at M9, 5.5 at M12
→ missing EDSS assessments at M6 and M9 are irrelevant
example3: 5 at baseline, 5.5 at M3, missing at M6, missing at M9, 6 at M12
→ missing EDSS assessments at M6 and M9 are viewed as irrelevant; earliest time point for a 24-weeks confirmed EDSS progression can be M12 (if there would be confirmations at M15 and M18)
example4: 5 at baseline, 6 at M3, missing at M6, missing at M9, 5.5 at M12
→ missing EDSS assessments at M6 and M9 are viewed to be irrelevant as EDSS 5.5 at M12 does not constitute a relevant EDSS increase (as defined in Section 3.2.2.1); earliest time point for a 24-weeks confirmed EDSS progression can be M15 (if there would be confirmations at M18 and M21)

- C3 – the previous and subsequent scheduled EDSS assessments are available and **both of them** constitute a relevant EDSS increase (as defined in Section 3.2.2.1): the missing scheduled EDSS assessment(s) is/are viewed as irrelevant, the subject is handled as having a 24-weeks confirmed EDSS progression at the earlier of the two visit dates
  example1: 5 at baseline, 6 at M3, missing at M6, 6 at M9
  → missing EDSS assessment at M6 is viewed as irrelevant, 24-weeks confirmed EDSS progression at M3 visit date
  example2: 5 at baseline, 6 at M3, missing at M6, missing at M9, 6 at M12
  → missing EDSS assessments at M6 and M9 are viewed as irrelevant, 24-weeks confirmed EDSS progression at M3 visit date.

The analysis of time to 24-weeks confirmed EDSS progression will be performed for the ITT analysis set:

- graphically presented using Kaplan Meier curves by study treatment group
- if available, Q1 and median time to 24-weeks confirmed EDSS progression and 95% CIs for each study treatment group will be estimated according to the method by Brookmeyer & Crowley 1982
- a proportional hazards regression model stratified by MS disease history and geographical region will be used to estimate and test the effect of MD1003 relative to placebo on time to 24-weeks confirmed EDSS regression
- Kaplan Meier plots by study treatment group will also be provided by disease history (SPMS/PPMS) and by geographical region.

Sensitivity analysis for handling of missing values

As censoring rules above for missing values of types A and B may introduce informative censoring and too long censored observation times, the following alternative rules for types A and B will be used to assess such potential bias by a sensitivity analysis:

A. administrative censoring caused by a subject completing the double-blind study phase without a 24-weeks confirmed EDSS progression and without any missing scheduled EDSS assessment: this constitutes administrative censoring and time to 24-weeks confirmed EDSS progression will be censored at the date of the last available scheduled

EDSS assessment not constituting a relevant EDSS increase (as defined in Section 3.2.2.1)
example1: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 5.5 at M18, 5.5 at M21.
example2: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 5.5 at M18, 6 at M21.
example3: 5 at baseline, 5 at M3, 5 at M6, 5.5 at M9, 5 at M12, 5.5 at M15, 6 at M18, 6 at M21
In all 3 examples, M21 was the final visit of the double-blind study phase when this study phase was closed according to the rules specified in Section 3.1.3 → example 1 handled as censored at the M21 visit date, example 2 handled as censored at the M18 visit date, example 3 handled as censored at the M15 visit date

B. a subject prematurely discontinues the double-blind study phase without a 12-weeks confirmed EDSS progression and without any missing scheduled EDSS assessment prior to his/her end of double-blind study phase:

- B1 – the last two available scheduled EDSS assessments both constitute a relevant EDSS increase (as defined in Section 3.2.2.1): the subject will be handled as having a 24-weeks confirmed EDSS progression and the date of the 24-weeks confirmed EDSS progression will be the date of the earlier of the two last available scheduled EDSS assessments
  example: 5 at baseline, 5.5 at M3, 6 at M6, 6 at M9, all subsequent EDSS assessments missing → handled as 24-weeks confirmed EDSS progression at the M6 visit date
- B2 – the last available scheduled EDSS assessment constitutes a relevant EDSS increase (as defined in Section 3.2.2.1) and the previous scheduled EDSS assessment does not constitute a relevant EDSS increase: the subject will be handled as having a 24-weeks confirmed EDSS progression and the date of the 24-weeks confirmed EDSS progression will be the date of the last available scheduled EDSS assessment
  example: 5 at baseline, 5.5 at M3, 5.5 at M6, 6 at M9, all subsequent EDSS assessments missing → handled as 24-weeks confirmed EDSS progression at the M9 visit date
- B3 – the last available scheduled EDSS assessment does not constitute a relevant EDSS increase (as defined in Section 3.2.2.1): this will be handled as censored subject, and the censoring date will be the date of the last available scheduled EDSS assessment
  example: 5 at baseline, 5.5 at M3, 5.5 at M6, 5.5 at M9, all subsequent EDSS assessments missing → handled as censored at M9 visit date

PAREXEL International Electronic Signature Page

This page is the manifestation of the electronic signature(s) used in compliance with PAREXEL International's electronic signature policies and procedures and in compliance with applicable regulations.

UserName: Jakobs, Peter (jakobsp)
Title: Senior Director, Biostatistics, GDO
Date: Monday, 10 February 2020, 03:56 PM GMT Standard Time
Meaning: Document contents approved.
==========