Official Title: A Phase III, Randomized, Double-Blind, Placebo-Controlled, Multicenter Study to Evaluate Efficacy, Safety, Pharmacokinetics, and Pharmacodynamics of Satralizumab in Patients with Generalized Myasthenia Gravis

NCT Number: NCT04963270

Document Date: Statistical Analysis Plan Version 3: 16-Jan-2024

STATISTICAL ANALYSIS PLAN

STUDY TITLE:	**A PHASE III, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTICENTER STUDY TO EVALUATE EFFICACY, SAFETY, PHARMACOKINETICS, AND PHARMACODYNAMICS OF SATRALIZUMAB IN PATIENTS WITH GENERALIZED MYASTHENIA GRAVIS**
STUDY NUMBER:	WN42636
STUDY NAME:	Luminesce
VERSION NUMBER:	3
ROCHE COMPOUND(S):	Satralizumab (RO5333787)
EUDRACT NUMBER:	2020-004436-21
IND NUMBER:	151937
NCT NUMBER:	NCT04963270
PLAN PREPARED BY:	[redacted], Ph.D

STATISTICAL ANALYSIS PLAN APPROVAL

SPONSOR: **LEGAL REGISTERED ADDRESS:**	F. Hoffmann-La Roche Ltd Grenzacherstrasse 124 4070 Basel, Switzerland
DATE FINAL:	See electronic date stamp on the last page of this document

STATISTICAL ANALYSIS PLAN VERSION HISTORY

This Statistical Analysis Plan (SAP) was developed based on Roche SAP model document 28 February 2022.

SAP Version	Approval Date	Based on Protocol (Version, Approval Date)
2	See electronic date stamp below	Protocol version 5, dated 19 July 2023
1	30 May 2023	Protocol version 3, dated 10 Nov 2021

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE VERSION 3

Key changes to the SAP, along with the rationale for each change, are summarized below.

Section	Description of Change	Rationale for Change
4.2 Primary Endpoint Analysis	Pooling of North America and Europe for stratification by Region	Low number of participants from North America
4.2.3.1 Missingness of covariate data	Removed section	Does not apply to the collected data set
4.2.4 Sensitivity Analyses for Primary Endpoints	Removal of Sections: Impact of Imputations on Item Level, Impact of Imputation Modeling Approach, Impact of Outliers	Sections are not needed and have been replaced by FDA Requested Tipping Point Analysis
4.2.5 Supplementary Analyses for Primary Endpoint	Rephrased subgroup analyses and removed subgrouping by type of rescue therapy	Low number of rescue therapy administration
Table 6/Table 3	MuSK+ as a separate population to assess mean change in MG-ADL from baseline to Week 24	Low number of LRP4+ participants, MuSK+ has distinct pathophysiology and clinical picture (Gilhus et al. 2019)
4.3.4 Secondary Endpoints Under the count Estimand	Removed paragraph	No definition for repeated exacerbations is applicable
4.4.1 Additional Evidence of Efficacy in MuSK+ and LRP4+	Removed pooling analysis and replaced by MMRM combining DB+OLE, the section is now named sustained efficacy.	Aligning analysis to external analyses in gMG
4.4.2 Additional evidence of efficacy	Time-to-event analyses defined in the original protocol objectives are now detailed in this section	To align with protocol objectives

DB = double-blind; gMG = generalized myasthenia gravis; LRP4+ = low-density lipoprotein receptor-related protein 4 positive; MG-ADL = Myasthenia Gravis Activities of Daily Living; MMRM = mixed model repeated measures; MuSK+ = muscle-specific kinase positive; OLE = open-label extension.

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation or Term	Description
AChEI	acetylcholinesterase
AChR	acetylcholine receptor
AChR+	AChR-antibody seropositive (participants/population)
ADA	anti-drug antibody
AE	adverse event
ANCOVA	analysis of covariance
CCOD	clinical cut-off date
CRO	contract research organization
C-SSRS	Columbia Suicide Severity Rating Scale
C_{trough}	trough concentration
DB	double-blind
ECG	electrocardiogram
eCRF	electronic Case Report Form
gMG	generalized Myasthenia Gravis
ICE	intercurrent event
ICH	International Council on Harmonization
iDMC	independent Data Monitoring Committee
IL-6	interleukin-6
IST	immunosuppressant
ITT	intent to treat
IxRS	interactive voice/web-based response system
LRP4	low-density lipoprotein receptor-related protein 4
MG	Myasthenia gravis
MG-ADL	Myasthenia Gravis Activities of Daily Living
MGC	Myasthenia Gravis Composite
MG-QoL	Myasthenia Gravis Quality of Life
mITT	modified intent to treat population
MMRM	mixed model repeated measures
MuSK	muscle-specific kinase
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
Neuro-QoL	Quality of Life in Neurological Disorders
NSDCR	not study drug or condition related
OCS	oral corticosteroids
OLE	open-label extension

OP	overall population (ITT)
PD	pharmacodynamic
PK	pharmacokinetic
Q4W	every four weeks
QMG	Quantitative Myasthenia Gravis score
RO	receptor occupancy
$RO_{tr,ss}$	trough receptor occupancy at steady-state
SAE	serious adverse events
SAP	Statistical Analysis Plan
SD	standard deviation
SDCR	study drug or condition related
sIL-6R	soluble interleukin-6 receptor
SOC	standard of care

1. INTRODUCTION

This Statistical Analysis Plan (SAP) details the planned analyses and considerations of statistical methodology for Study WN42636. Study WN42636 is a Phase III, randomized, placebo-controlled, double-blind, multicenter clinical study to evaluate the efficacy, safety, pharmacokinetic (PK), and biomarker effects of satralizumab in participants with generalized Myasthenia Gravis (gMG). For background information about the study, please refer to the Study Protocol WN42636. The analyses described in this SAP will supersede any analysis described in the Protocol for regulatory filings.

1.1 OBJECTIVES AND ENDPOINTS

This study will evaluate the efficacy, safety, PK, and pharmacodynamics (PD) of satralizumab compared with placebo in participants with gMG on stable background therapy. In addition, the study will assess the long-term safety and efficacy of satralizumab during the open-label extension (OLE) period.

The study has the following objectives and corresponding endpoints (Table 1).

Table 1 Objectives and Corresponding Endpoints

Primary Efficacy Objective	Corresponding Endpoint
• To evaluate the efficacy of satralizumab versus placebo on function in daily life in the AChR+ population	• Mean change from baseline in total MG-ADL score at Week 24
Secondary Efficacy Objectives	**Corresponding Endpoints**
• To evaluate the efficacy of satralizumab versus placebo on function in daily life in the OP	• Mean change from baseline in total MG-ADL score at Week 24
• To evaluate the efficacy of satralizumab versus placebo in the AChR+ and OP on:	
– Function in daily life	• Percentage of participants with a ≥2-point reduction from baseline in total MG-ADL score at Week 24 [a]
– QMG, QoL, and Fatigue	• Mean change from baseline in QMG score, MG-QOL 15r total score and Neuro-QoL Fatigue Subscale total score at Week 24 • Percentage of participants with a ≥3-point reduction from baseline in QMG score at Week 24 [a]

Table 1 Objectives and Corresponding Endpoints

– Clinical status	• Mean change from baseline in total MGC score at Week 24 • Percentage of participants with a ≥3-point reduction from baseline in total MGC score at Week 24 [a]
– Disease severity	• Proportion of participants: • Who have achieved minimal symptom expression (total MG-ADL score of 0 or 1) at Week 24 [a] • With at least one gMG-related exacerbation between baseline and Week 24 • Receiving rescue therapy between baseline and Week 24
• To evaluate the durability of the efficacy of satralizumab versus placebo in the AChR+ population and the OP	• Duration (average number of consecutive months) of meaningful improvement, defined as ≥2-point reduction from baseline in total MG-ADL score [a]
Exploratory Efficacy Objectives	**Corresponding Endpoints**
• To evaluate the efficacy of satralizumab versus placebo in the AChR+ and OP on	
– Time-to-event	• Time to disease improvement as measured by: – ≥2-point reduction in total MG-ADL score or – ≥3-point reduction in QMG score or – ≥3-point reduction in total MGC score or – MG-ADL score of 0 or 1 (minimal symptom expression) • Time to disease worsening as measured by the time from baseline to: – Start of rescue therapy – MG-related exacerbation
• QoL	• Mean change from baseline to Week 24 in EuroQoL EQ-5D-5L Health Utility Index score [b]

Table 1 Objectives and Corresponding Endpoints

Safety Objective	Corresponding Endpoints
• To evaluate the safety of satralizumab versus placebo	• Incidence and severity of adverse events, with severity determined according to NCI CTCAE v5.0 grading • Change from baseline in targeted vital signs, ECG results, targeted clinical laboratory test results, and suicidality
Pharmacodynamic Objective	**Corresponding Endpoint**
• To confirm target engagement and pathway inhibition in response to satralizumab	• Absolute values and change from baseline in serum levels of biomarkers IL-6 and sIL-6R
Pharmacokinetic Objective	**Corresponding Endpoints**
• To investigate the PK of satralizumab by evaluating plasma exposure over 24 weeks	• Serum concentrations of satralizumab (mean and SD of C_{trough}) at specified timepoints [b] • Estimates of primary PK parameters (e.g., CL/F and V/F) and secondary PK parameters (e.g., AUC) derived using population-PK modeling [b]

Table 2 Objectives and Corresponding Endpoints

Exploratory Pharmacokinetic Objective	Corresponding Endpoints
• To evaluate potential relationships between drug exposure and the efficacy and safety of satralizumab	• Relationship between selected covariates and exposure to satralizumab [b] • Relationship between serum concentration or PK parameters for satralizumab and efficacy endpoints, PD biomarkers, and safety endpoints [b]
Immunogenicity Objective	**Corresponding Endpoint**
• To evaluate the immune response to satralizumab	• Prevalence of ADAs at baseline and incidence of ADAs during the study
Exploratory Immunogenicity Objective	**Corresponding Endpoint**
• To evaluate potential effects of ADAs on efficacy, biomarker, safety, and PK endpoints	• Relationship between ADA status, biomarker, safety, or PK endpoints
Exploratory Biomarker Objective	**Corresponding Endpoint**
• To identify and/or evaluate biomarkers that are predictive of response to satralizumab, can provide evidence of satralizumab activity, or can increase the knowledge and understanding of disease biology	• Relationship between biomarkers in blood and efficacy endpoints [b]

AChR = acetylcholine receptor; AChR+ = AChR-antibody seropositive (participants/population); ADA = anti-drug antibody; AUC = area under the concentration–time curve; CL/F = apparent clearance; C_{trough} = trough concentration; gMG = generalized myasthenia gravis; IL-6 = interleukin-6; MG-ADL = Myasthenia Gravis Activities of Daily Living; MGC = Myasthenia Gravis Composite; MG-QOL 15r = Myasthenia Gravis Quality of Life 15 Scale (revised); NCI CTCAE v5.0 = National Cancer Institute Common Terminology Criteria for Adverse Events, Version 5.0; Neuro–QoL = Quality of Life in Neurological Disorders; OP = overall population; PD = pharmacodynamic; PK = pharmacokinetic; QMG = Quantitative Myasthenia Gravis; QoL = quality of life; sIL-6R = soluble interleukin-6 receptor; V/F = apparent volume of distribution.

[a] Participants who receive rescue therapy will be considered non-responders.

[b] Endpoints are mentioned in above table for alignment with the Study Protocol. They will not be analyzed as part of this SAP and will be presented separately.

Note: Efficacy, safety, pharmacokinetic, pharmacodynamics, and immunogenicity-related endpoints for adolescents enrolled in the study will be analyzed descriptively.

1.2 STUDY DESIGN

Study WN42636 has the following characteristics:

1. **Type/Design**: Phase 3, randomized, double-blind (DB), placebo-controlled, multicenter study. The randomization ratio is 1:1. Randomization will be stratified based on the following factors:

   – Baseline standard of care (SOC) treatment:
   - Acetylcholinesterase (AChEI) monotherapy and/or an oral corticosteroid (OCS)
   - A steroid-sparing immunosuppressant (IST) monotherapy or a combination of a steroid-sparing IST with other treatments (an AChEI and/or an OCS)

   – Auto-antibody type:
   - Acetylcholine receptor (AChR) antibody-positive
   - AChR-antibody-negative, which are either muscle-specific kinase (MuSK) antibody- or low-density lipoprotein receptor-related protein 4 (LRP4) antibody-positive.

   – Region:
   - North America
   - Europe
   - Rest of world

2. **Population:** Participants must fulfill the following selected inclusion criteria to be eligible for recruitment into the study:

   – Age ≥12

   – Confirmed diagnosis of gMG:
   - Documented history of myasthenic weakness
   - Myasthenia gravis (MG) severity of Myasthenia Gravis Foundation of America Class II, III, or IV at screening
   - The confirmation of the diagnosis has to be documented and supported by positive serologic test for one of the three antibody types: anti-AChR, anti-MuSK or anti-LRP4 at screening

   – A total Myasthenia Gravis Activities of Daily Living (MG-ADL) score of ≥5 at screening with more than 50% of this score attributed to non-ocular items

   – Ongoing gMG treatment at a stable dose (for precise listing refer to the Study Protocol)

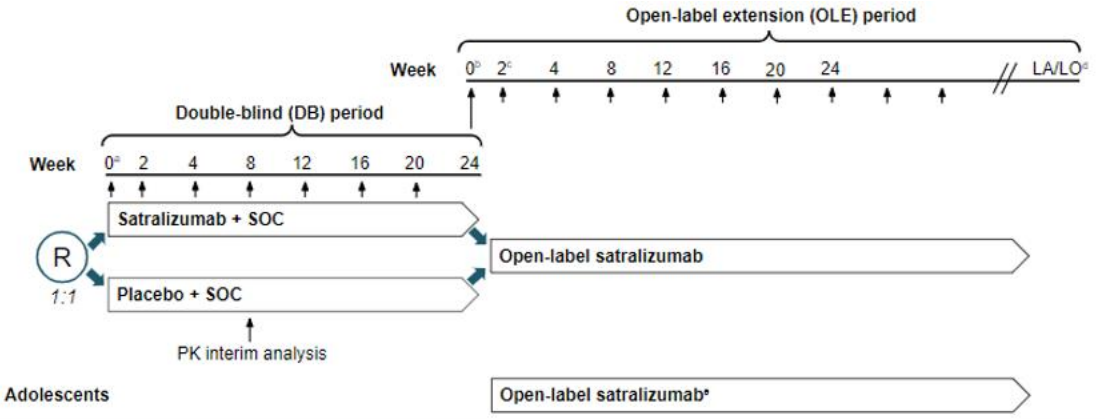
   – Female participants with childbearing potential agree to remain abstinent

3. **Duration:** The study will include a 28-day screening period, a 24-week DB treatment period, and approximately 2 years OLE period after the last participant initiates open-label treatment.

4. **Dosing of blinded study drug:** Blinded study drug will be administered subcutaneously to participants at Weeks 0, 2, 4, and every four weeks (Q4W) thereafter until the end of the DB period in addition to background treatments at a stable dose (see the Study Protocol, Section 4.3). The dosing regimen proposed for this study is 120 mg Q4W for participants ≤100 kg, and 180 mg Q4W for participants >100 kg.
5. **Reviews and Interim analyses:** An independent Data Monitoring Committee (iDMC) will conduct safety reviews approximately every 3 months. The dosing regimen was confirmed in a PK interim analysis performed after 40 participants (including approximately 20 receiving satralizumab) had been enrolled for at least 8 weeks in the trial. The purpose was to confirm whether the initial doses were sufficient, or whether higher (pre-defined) doses were necessary, to achieve target exposure and receptor occupancy (RO). This PK interim was conducted analytically by a PK contract research organization (CRO) and the dose recommendation was by the iDMC. All other interim analyses will be conducted by an independent Data Coordinating Center (iDCC) and the Sponsor will remain fully blinded. Interim analyses to stop for efficacy will not be performed. Futility analyses will be pre-specified in a separate Statistical Analysis Plan.
6. **Primary analysis:** The primary efficacy (final) analysis will be conducted once all participants which are part of the modified intent to treat (mITT) population (specified later in this document) have either reached Week 24 or withdrawn early from the study and data have been cleaned and verified, and the database has been locked.
7. As of the Global Protocol version 5, adolescents will be enrolled directly into the OLE period after completion of the screening period. This document specifies how these participants will be analyzed. They are not part of the primary analysis.

The study schema is shown in Figure 1.

Figure 1 Study Schema



LA = last assessment; LO = last observation; PK = pharmacokinetic; R = randomization; SOC = standard of care

[a] Week 0 baseline assessments will be collected pre-dose.

[b] Week 0 of OLE period coincides with Week 24 of DB period.

[c] Participants treated with active drug in DB period will be administered a placebo dose at Week 2 of the OLE period to maintain blinding of treatment assignment in the DB period.

[d] The length of OLE period is approximately 2 years after the last participant enters OLE or approximately 4 years after the first participant enters the OLE.

[e] Following global protocol amendment to version 5 adolescent participants will be enrolled directly into the OLE period (after completion of screening; see the Study Protocol, Section 3.1.3) and will receive open-label satralizumab SC loading doses at Weeks 0, 2, and 4 in the OLE, followed by maintenance doses Q4W thereafter during the OLE period.

1.2.1 Treatment Assignment

The study uses an interactive voice/web-based response system (IxRS) to randomize participants to two treatment arms and ensure blinding of both treating physician and the participant and the Sponsor. The IxRS implements a block randomization scheme with a block size specified in the Biostatistics Randomization Specification Document. Treatment assignments are stratified and hence balanced across the following baseline factors:

- **Region:** North America versus Europe versus Rest of World
- **Baseline SOC treatment:** AChEI monotherapy and/or an OCS versus a steroid-sparing IST monotherapy or a combination of steroid-sparing IST with other treatments (an AChEI and/or an OCS)
- **Auto-antibody type:** AChR-antibody positive versus AChR-antibody negative (includes MuSK or LRP4 antibody-positive)
- If a stratification factor level has a very low sample size compared to other levels, the Sponsor may group the levels together for the statistical analysis. Such a grouping would be determined before the data are unblinded.

1.2.2 Data Monitoring

An iDMC will be used during the DB period and until the database lock for the primary analysis. The iDMC will perform periodic (approximately every 3 months scheduled in agreement with the iDMC) unblinded safety reviews and make recommendations on trial continuation or modification. The iDMC will be involved in the interim PK analysis as detailed in the iDMC charter. All summaries and analyses will be prepared by the iDCC and presented by treatment group for the iDMC's review. Members of the iDMC and iDCC will be external to the Sponsor and the study team and will follow the iDMC charter that outlines their roles and responsibilities. The Sponsor will remain blinded. Interactions between the iDMC and Sponsor will be carried out as specified in the iDMC Charter.

1.2.3 Global Protocol Version 5

In July 2023, the Global Protocol was amended to Version 5 resulting in a sample size reduction. Given the number of participants already recruited, screening of new participants was stopped for adult participants. Due to recruitment difficulties for adolescents related to placebo use in the double-blind period, adolescents will continue to be recruited and will enroll directly into the OLE period, once the protocol amendment is approved in their countries. Adolescents who are enrolled in the study after the last adult was randomized will not contribute to the primary analysis.

2. STATISTICAL HYPOTHESES AND SAMPLE SIZE DETERMINATION

2.1 STATISTICAL HYPOTHESES

The primary study objective is to demonstrate superiority of the experimental treatment (satralizumab) over the comparator treatment (placebo) on the primary endpoint defined as the change from baseline to Week 24 in the total MG-ADL score (Section 4.2.1) in the AChR-antibody seropositive (participants/population) (AChR +) population.

The primary efficacy analysis will compare the mean for the primary endpoint between the treatment arms using a two-sided Wald test at the α level of 0.05 following the estimand framework as described in Section 4.2.2.1.

The following null and alternative hypotheses will be tested:

$$H_0: \mu_{satralizumab} = \mu_{placebo} \text{ versus } H_1: \mu_{satralizumab} \neq \mu_{placebo}$$

For which $\mu_{satralizumab}$ and $\mu_{placebo}$ refer to the mean change from baseline to Week 24 in the total MG-ADL score in the satralizumab and placebo groups respectively.

The primary comparison of interest is the mean change difference for the primary endpoint (i.e., the estimated difference between the two treatment groups as detailed in Section 4.2.3.5).

2.2 SAMPLE SIZE DETERMINATION

Determination of sample size is based on all globally enrolled participants. In this study, approximately 185 participants will be enrolled and randomized in a 1:1 ratio to each treatment group (satralizumab or placebo). Randomization will be stratified by baseline SOC treatment, region, and auto-antibody type as described in the Study Protocol, Section 3.1.

The estimated sample size required to demonstrate efficacy with regard to the MG-ADL is based on the AChR+ population and the following assumptions:

- The primary hypotheses test in the AChR+ population is the difference between the placebo and satralizumab groups in the change from baseline to Week 24 in total MG-ADL score
- The assumed change from baseline to Week 24 in the placebo group is 2.3 points
- The assumed change from baseline to Week 24 in the satralizumab group is 4.3 points
- The assumed change from baseline accounts for approximately 10% of participants in the satralizumab group and approximately 20% of participants in the placebo group receiving rescue therapy
- The standard deviation (SD) of the change from baseline to Week 24 is 3.97 in both the placebo and satralizumab groups
- The assumed study treatment withdrawal rate is 10%

The assumptions are based on data reported in Howard et al. (2017). Due to changes in the treatment landscape the assumed mean change from baseline to Week 24 in the satralizumab arm is now assumed to be 4.3 (instead of 4).

Based on these assumptions and using a two-sided α level of 0.05, the sample size to achieve 85% power was estimated at 160 participants (80 per group) in the AChR+ population. Under the assumption that up to 25 participants meeting the study eligibility criteria will be AChR− (MuSK+ or LRP4+), the total study sample size will be approximately 185 participants. This sample size also provides approximately 80% power for an analysis of the difference between the placebo and satralizumab groups in the proportion of MG-ADL responders in the mITT and AChR+ population. This power calculation assumes a 50% response rate in the placebo group and a 75% response rate in the satralizumab group.

3. ANALYSIS SETS

The following populations are defined (Table 3):

Table 3 Analysis Sets

Population	Definition
Intent to treat (ITT)	All randomized participants. This excludes adolescents joining the study after the last adult participant is randomized.
modified intent to treat (mITT)	All participants that are part of the ITT and have a baseline and at least one post-baseline MG-ADL assessment during the double-blind period. This excludes adolescents joining the study after the last adult participant is randomized.
Overall Population (OP)	All randomized participants who are either AChR+ or AChR–.
AChR+ (AChR positive)	All participants in the study who are acetylcholine receptor antibody seropositive (AChR+).
AChR– (AChR negative)	All participants in the study who are not AChR+, i.e., participants who are either low-density lipoprotein receptor-related protein 4 (LRP4) or muscle-specific kinase (MuSK) antibody seropositive. At least one of the two has to be positive.
MuSK+ (MuSK positive)	All participants in the study who are MuSK+, but are not AChR+ and are not LRP4+.
Adolescent Population	All adolescent participants who entered the study.
Safety-evaluable	All participants randomly assigned to study treatment who received at least one dose of study treatment. This includes adolescents enrolled after the last adult participant is randomized.
Pharmacokinetic (PK) evaluable	All participants randomly assigned to study treatment who received at least one dose, and had sufficient sampling to permit PK evaluation. This includes adolescents enrolled after the last adult participant is randomized.
Immunogenicity	All participants randomly assigned to study treatment who received any study treatment with at least one post-dose anti-drug antibody assessment. This includes adolescents enrolled after the last adult participant is randomized.

4. STATISTICAL ANALYSES

The primary study objective is to demonstrate superiority of the experimental arm (satralizumab, 120 mg or 180 mg for participants with body weight ≤100 kg or >100 kg, respectively) over the control arm (placebo), administered for 24 weeks in participants with generalized myasthenia gravis on daily function.

The statistical analysis will follow the estimand framework aligned with the ICH E9 Addendum:

- The primary comparison of interest is the difference between the placebo and satralizumab groups (120 mg or 180 mg satralizumab for participants with body weight ≤100 kg or >100 kg, respectively) in the change from baseline to Week 24 in

the total MG-ADL score in the mITT population restricted to the AChR+ population, irrespective of treatment adherence or use of rescue medication (Section 4.2).

- The estimands for the secondary endpoints to be included in the hierarchical statistical testing process (Section 4.3.7) are defined in Section 4.3.

4.1 GENERAL CONSIDERATION

In addition to specific analyses and presentations that are detailed in the following sections, study results will be presented by treatment arm (with total when appropriate) and summarized according to the nature of the variables.

For continuous variables, using descriptive statistics, including the number of participants contributing to summary statistics, mean, standard deviation, median, and range as appropriate.

For categorical variables, using the frequency and proportion of participants falling into each category, grouped by treatment arm (and total). The percentages given in these tables will be rounded and therefore may not always sum to 100%. If a missing category is not presented in the data display, only participants with non-missing values for the parameter being assessed will be included in the percentage calculation.

All efficacy analyses will be performed on the mITT (or subpopulations of the mITT) unless otherwise specified. If additional subpopulations are specified (i.e. AChR+ or AChR–), then the mITT restricted to those subpopulations will be analyzed. The mITT ensures that each participant is able to contribute a minimal meaningful amount of information to evaluate the treatment effect of satralizumab. The ITT and mITT exclude adolescents joining the study after the last adult participant is randomized. Adolescent participants not part of the mITT will be analyzed separately.

The primary efficacy analysis will be performed on the AChR+ participants. Participants will be analyzed according to the treatment assigned at randomization by the IxRS.

All safety analyses will be performed in the safety-evaluable population, unless otherwise specified. Participants will be analyzed according to the first treatment they received after randomization.

The baseline value will be defined as the last available value recorded prior to the initiation of study treatment, unless otherwise specified.

4.2 PRIMARY ENDPOINT ANALYSIS

The primary endpoint analysis will be performed on the mITT population restricted to AChR+ participants. Participants will be analyzed according to the treatment assigned at randomization by the IxRS. Given the low number of participants from North America, the Region stratification factor will be pooled for participants from Europe and North America for the primary analyses.

The primary endpoint will be summarized by treatment group using tables, listings, and graphs, as appropriate.

This section first introduces the primary estimand, the sensitivity and subgroup analyses, the section that follows will describe the secondary endpoints and related estimands.

4.2.1 Definition of Primary Endpoint

The primary endpoint is the difference between the placebo and satralizumab groups in combination with stable background therapy in the change from baseline to Week 24 in the total MG-ADL score in AChR+ participants.

The MG-ADL was developed by Wolfe et al. (1999) to assess the degree of gMG symptoms (six items: diplopia, ptosis, difficulties with chewing, swallowing, talking, and respiratory problems) and functional limitations in carrying out activities of daily living (two items: ability to brush teeth or comb hair and impairment in the ability to arise from a chair) that have been shown to be present and clinically relevant in gMG patients. Each of the eight items is ranked on a 0-3 scale yielding a total score that ranges from 0 to 24, with higher scores indicating greater disease severity (see the Study Protocol, Appendix 6). The items of the MG-ADL were all derived from the original 13-item symptom list that comprises the clinician-rated Quantitative Myasthenia Gravis score (QMG) scale.

4.2.1.1 Definition of Primary Estimand

In accordance to the Addendum to International Council on Harmonization (ICH) E9 and to ensure alignment between trial objective, trial design, data collection, analysis, and interpretation, the estimand which describes the target of estimation and its five attributes including population, treatment condition of interest, variable (endpoint), intercurrent events (IECs; defined as events that can occur post-randomization and preclude or affect the interpretation of the variable) and summary of measure are described in Table 4.

Table 4 Attributes for the Estimand of the Primary Endpoint

Attribute	Description
Estimand	The primary endpoint will be evaluated irrespective of taking rescue therapy, treatment discontinuation due to SDCR, as though participants with treatment discontinuation due to NSDCR continued treatment and as though participants with missing data after SDCR, rescue therapy, treatment interruption due to infection continued with placebo.
Population	The population targeted are all participants recruited to the study (satisfying all inclusion criteria) and are part of the mITT restricted to AChR+.
Variable	MG-ADL total score
Treatments	Experimental Treatment: Satralizumab 120 mg (for participants with body weight ≤100 kg) or 180 mg (for participants with body weight >100 kg) and background therapy. Control Treatment: Placebo, administered for 24 weeks and background therapy.
Intercurrent Events	See next Section 4.2.1.2.
Summary of Measure	The difference in the change from baseline to Week 24 in MG-ADL total score. An estimate of the treatment effect will be computed using an analysis of covariance regression approach using the "Variable" as a response and adjusting for randomization stratification factors (background therapy, auto-antibody type, region) and the baseline "Variable" value. The computation will be performed on a completed data set which is obtained by a mixed model repeated measures imputation model.

AChR+ = AChR-antibody seropositive (participants/population); MG-ADL = Myasthenia Gravis Activities of Daily Living; mITT = modified intent to treat; NSDCR = not study drug or condition related; SDCR = study drug or condition related.

4.2.1.2 Intercurrent Events for the Primary Estimand

In this section, we will discuss specifically how ICEs are categorized and approaches for handling missing data due to ICEs.

In this study, three different ICEs are considered:

- Withdrawal from study treatment
- Receiving a rescue therapy
- Treatment interruption due to infection.

As described in Table 5, the ICEs are classified into two main categories:

- Study Drug or Condition Related ICE (SDCR ICE)
- Not Study Drug or Condition Related ICE (NSDCR ICE).

For the generic ICE “Withdrawal from study treatment” information on the cause and categorization into whether the interruption is SDCR or NSDCR is collected in the electronic Case Report Form (eCRF):

- NSDCR: Pregnancy, study terminated by Sponsor
- SDCR: Adverse Event (AE), death, lack of efficacy
- NSDCR or SDCR: Lost to follow-up, protocol deviation, withdrawal by subject, physician decision, other.

The last bullet collects cause that the investigator will be asked for categorization into NSDCR or SDCR in the eCRF.

Depending on to the classification of ICEs, we propose estimand strategies and statistical imputation methodology in handling SDCR- and NSDCR-ICEs in Table 5.

Table 5 Handling of ICEs for the Primary Endpoint

Intercurrent Event Classification	Study Drug or Condition Related (SDCR)	Not Study Drug or Condition Related (NSDCR)
Intercurrent event	Withdrawal from study treatment (AE, death, lack of efficacy).[1] Lost to follow-up, protocol deviation, withdrawal by subject, physician decision and other, that are specified for the individual participant as SDCR. Receiving a rescue therapy.[2] Treatment interruption due to infection.[3]	Withdrawal from study treatment (Pregnancy, study terminated by Sponsor). [1] Lost to follow-up, protocol deviation, withdrawal by subject, physician decision, other., that are specified as NSDCR for the individual participant.
Estimand Strategy	Treatment Policy In the treatment policy strategy, the occurrence of SDCR ICE is ignored and measurements of the variable of interest are used as such. In other words, the analysis does not make any statistical adjustment for SDCR ICE.	Hypothetical Approach The hypothetical approach envisages a scenario wherein it is presumed that NSDCR ICE would not have occurred. Thus, assumes that NSDCR ICE did not happen in the participants who consented to participate.
Data Collection	Regardless of ICE occurrence, the participant will be asked to remain in the study to be evaluated for the primary efficacy endpoint and data collection will continue as long as possible.	
Data Strategy	Data are used as available to estimate a treatment effect irrespective of the occurrence of SDCR ICE. This assumption will estimate the treatment effect by analyzing outcomes as they were observed, assuming rescue therapy is part of the standard treatment regimen and hence data are considered as recorded.	All data after an ICE is ignored.

Table 5 Handling of ICEs for the Primary Endpoint

Intercurrent Event Classification	Study Drug or Condition Related (SDCR)	Not Study Drug or Condition Related (NSDCR)
Imputation Strategy	If data are unavailable, or only partial data are available subsequent to a SDCR ICE, then data are used as available and the remainder is imputed using copy reference. Data are imputed using copy reference methodology (i.e., data are imputed from the placebo arm). The underlying assumption being that participants who discontinue the active arm did not get a benefit from the proposed treatment after the SDCR ICE. In other words, the participant profile following SDCR ICE tracks that of the reference arm but starting from the benefit already obtained.	Data are imputed using a missing at random assumption that is data will be imputed conditional on observed data. The imputation under a missing at random assumption estimates a treatment effect as if the ICE had not happened.

[1] For the generic ICE "Withdrawal from study treatment" information on the cause and categorization into whether the interruption is study drug or condition related (SDCR) or not SDCR (NSDCR) is collected in the electronic Case Report Form.

[2] Receiving rescue therapy might not result in a treatment interruption but is a relevant ICE in this study that always falls under the SDCR category.

[3] Treatment interruption due to infection will always be categorized as SDCR.

If a participant has recorded multiple ICEs, then the category of the first ICE will determine the handling strategy of all subsequent missing data.

The precise statistical imputation methodology is detailed in section on missingness and imputation.

For the efficacy analysis data entries will be mapped to the closest visit available. No visit will be mapped across their respective period for DB or OLE. If two data entries are mapped to the same visit, then the one closer to it and later will be used.

4.2.2 Main Analytical Approach for Primary Endpoint

4.2.2.1 Derivation of ICE's

To facilitate the analysis of the primary endpoint a data set containing all ICEs will be derived. For each participant it will contain a row for each ICE, the date and time of the ICE, the cause/description of the ICE and the categorization into SDCR/NSDCR.

4.2.3 Missing Data and Imputation

Obviously, the optimal strategy for dealing with missing data is to make every effort to obtain complete data during the conduct of the study. The Sponsor will work diligently and use a variety of methods to minimize the percentage of missing data in this trial. Nevertheless, there is likely to be a small percentage of missing data. In this context, approaches to handle missing data for the primary and secondary analysis are described in this section.

If a participant discontinues treatment the Sponsor will ensure all planned assessments are continued whenever possible to avoid unnecessary missingness.

The following missing data and imputation procedure applies to relevant baseline covariates (Stratification factors [auto-antibody type, background therapy and region]) and all longitudinal efficacy scores (MG-ADL, QMG, Myasthenia Gravis Composite [MGC], Myasthenia Gravis Quality of Life [MG-QoL] and Quality of Life in Neurological Disorders [Neuro-QoL]) including baseline.

In the following the resolution of different missingness types are described and will be addressed in the order listed.

4.2.3.1 Lateral Item Missingness

For specific lateral items such as right/Left-hand grip, right/left arm outstretched and right/left leg outstretched one side might be missing when the other is present. In such a case the mean lateral difference across all past visits for the participant will be computed, and the missing item will be imputed based on the contralateral side and the mean lateral difference.

4.2.3.2 Item Missingness

For the efficacy scores (MG-ADL, QMG, MG-QoL, MGC, NeuroQoL) items might be missing. This is the case if at any measurement visit date at least one item has been

recorded. In this case, missing items will be estimated using the proportional method of imputation based on participants' responses to the other items on the questionnaire.

The following rule is applied:

1.) If less than 50% of items (number of items available MG-ADL 4, QMG ≤7, MG-QoL ≤8, NeuroQoL: ≤4) are missing, the final score is computed as the sum of the non-missing items times the total number of items divided by the number of non-missing items. As an example, the sum of available items is 6, the number of available items is 9 and the total number of items is 13 then the result is 6*(13/9).

2.) In case more items than in the previous instruction are missing then data are considered fully missing and one of the next steps is applied.

The Sponsor will always trust the entered result of the Investigator but might query a site if a result seems incomplete or inconsistent. If a result contains a total score but some items remain incomplete the total score will be used in the statistical analysis. Imputation will only take place if items and total score are missing.

4.2.3.3 Longitudinal Missingness

In case data are missing for intermediate visits and missingness is not caused by an ICE then data will be imputed based on a missing at random assumption. This imputation will be performed at the same time imputation due to missingness of ICE is performed.

4.2.3.4 Imputation Due to Intercurrent Events

In the following we describe missingness and data imputation due to IECs. A detailed statistical justification for this imputation strategy can be found in Wolbers M et al. (2021).

1. For each participant the time of the first ICE is determined using the ICE data set. For each efficacy score a data set is created where all subsequent data to the first ICE is removed. If a participant has multiple ICEs between which normal dosing of study drug is attained all the data are nevertheless removed in a first step, i.e., data that would be dealt under treatment policy is still removed in this step. This ensures that the imputation model is fitted on data that is completely unaffected by ICEs.
2. A mixed model repeated measures (MMRM) model is fitted as an imputation model to address all the missingness due to ICEs. The MMRM is modeling change from baseline of the efficacy score (separate MMRMs will be fitted for each score) and has the following covariates: treatment group, visit, treatment group x visit, baseline stratification factors (background therapy, auto-antibody type [where this applies], region), baseline efficacy score and baseline efficacy score x visit. (“x” refers to complete interactions). The MMRM will use an unstructured covariance overall, in case of convergence issues the covariance structure will be simplified to compound symmetry. The resulting model will serve as a model for all imputations of longitudinal data in the next step.

3. The MMRM model implies a multivariate normal imputation distribution conditional on baseline covariates and conditional on observed longitudinal data that is not affected by an ICE. The implied conditional mean of this normal distribution is used to impute missing data. Participants with an ICE but subsequent data available which falls under treatment policy will use the observed data and will not be imputed. Participants on the satralizumab arm with missing data due to an ICE in category SDCR will be imputed using the imputation model and their treatment group covariate will be artificially set to placebo (copy reference imputation), this way their data are imputed as if they had not had a benefit from the treatment after the ICE. This imputation procedure will result in a complete data set for each efficacy score. All the efficacy analyses will be conducted on the completed data sets.

4.2.3.5 Analysis and Inference of the Primary Endpoint

The treatment effect on the primary endpoint is estimated using an analysis of covariance (ANCOVA) with response variable change in MG-ADL from baseline to Week 24 adjusting for treatment group and the same covariates as for the imputation model.

The primary treatment effect estimate is the regression coefficient of the treatment group.

The estimate of variance coming from the ANCOVA will underestimate the true variance as it ignores the variance due to the imputation. To correct the underestimation of the variance the jackknife is applied to all previous steps (2-3) of the imputation modeling, that is an imputation model is fit on each data set with one participant removed. The p-value of the treatment effect is derived through the Wald test.

4.2.4 Sensitivity Analyses for Primary Endpoints

The robustness of the primary method of estimation will be explored using a series of sensitivity and supplementary estimators based on varying different aspects of the primary endpoint, including sensitivity analyses exploring robustness of departures from statistical assumptions and supplementary analyses investigating the trial data as listed in this section. The p-values from sensitivity and supplementary analyses of the primary endpoint are for descriptive purpose only, and there will be no multiplicity adjustment for these analyses.

We propose a series of sensitivity analyses to further assess the potential dependence of the results of the primary analysis on missing values. However, additional sensitivity analyses may be considered if appropriate.

4.2.4.1 Impact of Rescue Therapy

To assess the impact of rescue therapy in AChR+ participants all values after the administration of rescue therapy are replaced by the worst possible MG-ADL assessment that a particular participant has had prior to receiving rescue therapy. This

imputation overwrites the estimand specified imputation strategy. The impact will be estimated using an MMRM.

4.2.4.2 Impact of Major Protocol Deviations

To assess the impact of major protocol deviations, the primary analysis will be run excluding participants with a major protocol deviation of which an impact on the efficacy outcome is expected. A separate analysis for each type of major protocol deviation is conducted:

1. Violation of the inclusion criteria of the MG-ADL score at screening
2. MG-ADL not performed in the order specified per protocol at DB Week 0 and/or Week 24
3. MG-ADL assessment performed by non-qualified site personnel at DB Week 0 and/or Week 24
4. Failure to maintain dual assessor approach.

4.2.4.3 Tipping Point Analysis

A δ tipping point analysis will be conducted on all participants who discontinue satralizumab treatment. The MG-ADL value after stopping treatment will be replaced by the last available assessment prior to stopping plus an additive constant delta. The treatment effect will be re estimated for varying values of delta $(0,\infty)$ until the first delta at which the primary analysis treatment effect p-value will exceed 5%. The minimal value of delta at which this happens is referred to as the tipping point. This imputation will overwrite the estimand specified imputation strategy.

4.2.5 Supplementary Analyses for Primary Endpoint(s)

In addition to the primary analysis for each efficacy score at Week 24, the same analysis will be performed for each other clinical visit date: Week 2, Week 4, Week 8, Week 12, Week 16 and Week 20.

Subgroup analyses for the primary endpoint will be performed to explore whether the treatment effect on the primary endpoint is consistent across subgroups. The influence of baseline and demographic characteristics on the treatment effect among participants will be explored via exploratory subset analyses for the following factors:

- Sex
- Age at diagnosis
- Race
- Region
- Background therapy
- Baseline disease severity and characteristics
- Dose level.

Treatment effects within each subgroup will be examined separately using the imputed data of the primary estimand and the model specified previously for the primary endpoint.

Forest plots will be generated displaying the estimated difference (the effect size of the primary treatment effect), 95% confidence intervals, and the corresponding p-value for each subgroup, with the overall result also included at the bottom. These subsidiary analyses are intended to provide reassurance that the observed treatment effect is consistent across all participant subgroups.

The study is not powered to detect differences between subgroups and any observed patterns should be interpreted extremely cautiously, owing to the smaller numbers and increased chance of type I error.

All listed subgroup analyses will be presented in a forest plot showing the effect size of the primary treatment effect in that subgroup.

4.3 SECONDARY ENDPOINTS ANALYSES

Secondary endpoints are based on the following scores (see the Study Protocol for details):

- QMG: The QMG is a 13-item direct physician assessment scoring system that quantifies disease severity based on impairments of body functions and structures. Each item is quantitatively assessed and scored from 0 to 3 (where 3 represents the most severe), providing a total QMG score ranging from 0 to 39 (Clinical Review Report)
- MG-QoL: The MG-QOL-15 is a disease-specific health-related quality of life measure that consists of 15 items: mobility (9 items), symptoms (3 items), and contentment and emotional well-being (3 items). Items are scored on a Likert scale from 0 to 4 with the total score ranging from 0 to 60 (Burns et al. [2008])
- MGC: The MGC is a composite measure consisting of items drawn from the MG-ADL (chewing, swallowing, speech, and breathing), QMG (diplopia and ptosis), and Manual Muscle Test (hip, neck, facial, and deltoid strength) in an effort to include both clinician- and participant-reported elements in a single measure (Burns et al. [2008]). Each of the ten items contribute to a total score ranging from 0 to 50, with higher values indicating increasing symptom severity (Burns et al. [2010])
- Neuro-QoL: The Neuro–QoL Fatigue scale is a short form that is part of a collection of instruments and item banks, developed through a National Institute of Neurological Disorders and Stroke–sponsored initiative to evaluate the health-related quality of life of adults and children diagnosed with neurological disorders. It consists of eight items, each using a 5-level Likert scale ranging between 1=never to 5=always, with a 7-day recall period (Cella et al. [2012]).

4.3.1 Estimands for Secondary Endpoint Analyses

In addition to the primary estimand, the analysis of the secondary endpoints will rely on the responder estimand and the Duration Estimand. Except for the difference in estimands with regards to composite strategies, the analysis of secondary endpoints is the same as for the primary endpoint and will make use of the same imputation scheme.

4.3.1.1 ICE Strategies for Estimands for Secondary Endpoints

Endpoints depending on different estimands than the primary will still be computed on the imputed data set of the primary estimand and their ICE strategies will be implemented on top of the imputed data set:

- Responder Estimand: Same imputation as for primary estimand, if a participant requires rescue therapy the participant is categorized as non-responder (composite strategy). For example, for the secondary endpoint percentage of participants with a ≥2-point reduction from baseline in total MG-ADL score at Week 24, the dataset with MG-ADL score imputed as in Section 4.2.3.4. Completed MG-ADL will be used to identify the participants with a ≥2 point reduction from baseline. On top of that, if participants require rescue therapy the participant is categorized as non-responder
- Duration Estimand: Same imputation as for primary estimand, if a participant requires rescue therapy the participant will be imputed as not having an endured treatment effect after administration of rescue therapy (composite strategy).

The following table lists all secondary endpoints, the corresponding estimand, and relevance as secondary endpoint (Table 6). All populations are to be understood as the mITT restricted to the mentioned population. LRP4 & MuSK refers to the AChR- population.

Table 6 Secondary Endpoints and Corresponding Estimand and Relevance

Secondary Endpoint	Population	Relevance	Estimand Type (Generic)
Mean change from baseline in MG-ADL score at Week 24	OP, LRP4 & MuSK	confirmatory	**Primary (continuous) estimand:** Difference in mean change of "Score" from baseline in "Population", treated with satralizumab vs placebo, in combination with stable background therapy, irrespective of taking rescue therapy, treatment discontinuation due to SDCR, as though participants with treatment discontinuation due to NSDCR continued treatment and as though participants with missing data after SDCR, rescue therapy, treatment interruption due to infection continued with placebo.
Mean change from baseline in MG-ADL score at Week 24	MuSK	exploratory	
Mean change from baseline in QMG score at Week 24	AChR+, OP	confirmatory	
	LRP4 & MuSK	exploratory	
Mean change from baseline in MG-QOL 15r total score at Week 24	AChR+	confirmatory	
	OP	supportive	
	LRP4 & MuSK	exploratory	
Mean change from baseline in MGC total score at Week 24	AChR+	confirmatory	
	OP	supportive	
	LRP4 & MuSK	exploratory	
Mean change from baseline in Neuro-QoL Fatigue Subscale total score at Week 24	AChR+	confirmatory	
	OP	supportive	
	LRP4 & MuSK	exploratory	
	AChR+, OP	supportive	

Table 6 Secondary Endpoints and Corresponding Estimand and Relevance

Secondary Endpoint	Population	Relevance	Estimand Type (Generic)
Percentage of participants with a ≥2-point (and more) reduction from baseline in total MG-ADL score at Week 24	LRP4 & MuSK	exploratory	**Responder Estimand:** The difference in proportion in population with a ≥x-point reduction from baseline in Score at Week 24 between participants treated with satralizumab vs placebo, in combination with stable background therapy, as though participants with treatment discontinuation due to NSDCR continued treatment and as though participants with missing data after SDCR, treatment interruption due to infection continued with placebo as though participants receiving rescue therapy were not showing a reduction from baseline.
Percentage of participants with a ≥3-point (and more) reduction from baseline in QMG score at Week 24	AChR+, OP	supportive	
	LRP4 & MuSK	exploratory	
Percentage of participants with a ≥3-point (and more) reduction from baseline in total MGC score at Week 24	AChR+, OP	supportive	
	LRP4 & MuSK	exploratory	
Proportion of participants who have achieved minimal symptom expression (total MG-ADL score of 0 or 1) at Week 24	AChR+, OP	supportive	
	LRP4 & MuSK	exploratory	
Proportion of participants with at least one gMG related exacerbation between baseline and Week 24	AChR+, OP	supportive	
	LRP4 & MuSK	exploratory	

Duration (# consecutive visits) that participants show at least 2 point reduction from baseline MG-ADL	AChR+, OP	supportive	**Duration Estimand:** The difference in the mean duration of meaningful improvement at Week 24 in population between participants treated with satralizumab vs placebo, in combination with stable background therapy, as though participants with treatment discontinuation due to NSDCR continued treatment and as though participants with missing data after SDCR, treatment interruption due to infection continued with placebo as though participants receiving rescue therapy were not showing a further reduction from baseline.
	LRP4 & MuSK	exploratory	

AChR = acetylcholine receptor; AChR+ = AChR-antibody seropositive (participants/population); gMG = generalized Myasthenia Gravis; LRP4 = low-density lipoprotein receptor-related protein 4; MG-ADL = Myasthenia Gravis Activities of Daily Living; ; MGC = Myasthenia Gravis Composite; MG-QoL = Myasthenia Gravis Quality of Life; MuSK = muscle-specific kinase; NSDCR = not study drug or condition related; OP = overall population; QMG = Quantitative Myasthenia Gravis score; QoL = quality of life; SDCR = study drug or condition related; Exacerbation = increase of non-ocular MG-ADL item by at least two points from baseline or achieving non-ocular item value of 3 when baseline was below 3.

4.3.2 Secondary Endpoints Under the Primary (Continuous) Estimand Framework

Secondary endpoints falling under the primary estimand framework (Table 6) will be analyzed the same way as the primary endpoint. The ANCOVA model will be analyzed by adjusting for the baseline value of the corresponding score, i.e., QMG change from baseline to 24 weeks will be adjusted for baseline QMG (instead of baseline MG-ADL).

The estimands for the secondary endpoints will be derived on the imputed data set of the primary estimand if the corresponding endpoint is solely based on the raw score.

4.3.3 Secondary Endpoints Under the Responder Estimand

To analyze secondary endpoints under the responder estimand framework (Table 6) each participant will have to be labelled as a responder or non-responder. This is done using the complete data set from the Imputation procedure and by applying the definition of the secondary endpoints. For a rationale on those responder definitions refer to the Study Protocol sections on MG-ADL, QMG and MGC responder definitions, which employ minimum clinically important difference derived in the relevant clinical literature. According to ICE specification of the responder estimand, participants that received rescue therapy will be imputed as non-responders (composite strategy) irrespective of the actual score in the completed data set.

A Cochran-Mantel-Haenszel test over the treatment groups stratifying for baseline stratification factors is the treatment effect estimator for the responder estimand. The estimate is the difference in proportion at Week 24.

4.3.4 Secondary Endpoints Under the Duration Estimand

The duration estimand is only relevant for the analysis of the number of consecutive visits a participant shows an MG-ADL reduction of at least 2 points from baseline. Based on the completed data set, the first time a participant shows a 2-point or more reduction from baseline, the duration is measured until Week 24 or until the visit when MG-ADL = is no longer reduced by 2 points from baseline MG-ADL. The duration is the difference in weeks between the two visits defining the start and end (or Week 24) of reduction from baseline. If the participant shows a reduction for only one visit, the count is set to 1 week. If the participant required rescue therapy, the count is stopped prior to the visit at which the participant received rescue therapy (composite).

The count variable (the number in weeks by consecutive visits) will be modeled using an ANCOVA adjusting for the same covariates as the imputation model and baseline MG-ADL. The treatment effect estimator is the regression coefficient of the treatment group. The estimate is the difference in the mean duration of meaningful improvement at Week 24.

4.3.5 Secondary Endpoint of Rescue Therapy

The endpoint of proportions of participants receiving rescue therapy during the double−blind period will analyze the variable that encodes whether a participant received rescue therapy during the double-blind period or not. Treatment policy is used, that is if a participant stops study drug but receives rescue therapy during the safety follow-up and this occurs within 24 weeks of baseline then this is counted as having received rescue therapy. Stop of study drug and no further information available will be imputed as not having received rescue therapy.

A Cochran-Mantel-Haenszel test over the treatment groups stratifying for baseline stratification factors is the treatment effect estimator for the proportions of participants receiving rescue therapy. The estimate is the difference in proportion at Week 24.

4.3.6 Population of Imputation Models

All analyses on the AChR+ population will use an imputation model fitted on AChR+ only. Analyses on OP, MuSK, or LRP4 will use the OP to fit the imputation model.

4.3.7 Hierarchy of Secondary Endpoints

The primary and secondary endpoints will be tested using a hierarchical gatekeeping procedure. Where the hierarchical gatekeeping procedure is used, if any test result is not statistically significant, formal testing of subsequent endpoints will not occur. If the primary endpoint is statistically significant, the confirmatory secondary endpoints are tested in the following order:

1. Mean change from baseline in QMG score in AChR+ participants at Week 24
2. Mean change from baseline in MG-QOL 15r total score in AChR+ at Week 24
3. Mean change from baseline in MGC total score in AChR+ participants at Week 24
4. Mean change from baseline in MG-ADL score in the OP at Week 24
5. Mean change from baseline in QMG score in the OP at Week 24
6. Proportion of AChR+ participants receiving rescue therapy between baseline and Week 24
7. Mean change from baseline in MG-ADL score in MuSK+ and LRP4+ at Week 24
8. Mean change from baseline in Neuro-QoL Fatigue Subscale total score in AChR+ participants at Week 24.

4.3.8 Software for Computation of Primary and Secondary Endpoints Analysis

The implementation of the previous described analyses will be conducted using the programing language R. The estimand framework will be implemented using the rbmi package. The statistical analysis will be conducted using the programing languages SAS and R.

4.4 EXPLORATORY ENDPOINTS ANALYSES

Exploratory endpoints analyses are listed in Table 6.

4.4.1 Sustained Efficacy

To demonstrate evidence of sustained efficacy of satralizumab, the mean change from baseline to OLE Week 48 will be fitted using an MMRM for both the MG-ADL and the QMG in OP, AChR+ and AChR− participants. This analysis will include participants who at the time of the clinical cutoff (CCOD) have been enrolled for at least 40 weeks, i.e. with at least an OLE Week 16 visit for those still part of the study.

4.4.2 Additional Evidence of Efficacy

4.4.2.1 Time to Disease Improvement and Disease Worsening

Time-to-event analyses of the following endpoints will be conducted on mITT participants during the double-blind period.

Time to disease improvement (composite endpoint):

- ≥2 point reduction in total MG-ADL score
- ≥3 point reduction in QMG score
- ≥3 point reduction in MGC score
- MG-ADL score of 0 or 1 (minimal symptom expression status)
- Combination (whichever happens first) of reduction by ≥2 points in total MG-ADL or ≥3 points in total QMG score

Time to disease worsening:

- Time to first use of rescue therapy
- Time to MG-related exacerbation.

Participants with no event during the double-blind period will be censored on the day of their last double-blind period visit. The time of each event will be derived from the actual time point the event first occurs compared to the timepoint of baseline. The first four endpoints will use a composite strategy for rescue therapy, i.e., participants will be coded as not having had an event if they receive rescue therapy during the double-blind period and they will be censored at the time they receive rescue therapy. The treatment effect will be estimated using a log-rank test. Kaplan-Meier curves will be used to plot the data and listings will contain survival estimates at regular time points.

4.4.2.2 Quality of Life Analyses

The mean change in EuroQoL EQ-5D-5L Health Utility Index score from baseline to Week 24 will be computed by arm and compared.

4.4.3 Dose Reduction of Oral Corticosteroids and Immunosuppressants

A descriptive overview of background therapy dose reduction from OLE Week 12 onward will summarize the proportion of participants with an ongoing tapering or substantial dose reduction and/or discontinuation.

We consider the following ratios:

- Ongoing corticosteroid tapering: Number of participants who reduce their dose after OLE Week 12 and do not increase their dose subsequently, divided by number of participants who reduce their dose after OLE Week 12

- Successful corticosteroid tapering and discontinuation: Number of participants who either reduce their dose after OLE Week 12 to physiologic dose level of ≤7.5 mg prednisone daily dose equivalent or discontinue corticosteroid, divided by number of participants who reduce their dose after OLE Week 12

- Ongoing immunosuppressant (IST) tapering: Number of participants who reduce their dose after OLE Week 12 and do not increase their dose subsequently, divided by number of participants who reduce their dose after OLE Week 12

- Successful IST tapering and discontinuation: Number of participants who substantially reduce their dose after OLE Week 12 (at least 25% dose reduction) or withdraw from IST, divided by number of participants who reduce their dose after OLE Week 12

Of note: This analysis will be conducted on CCOD data set used for the primary analysis and will include only participants who have at least 24 weeks of OLE period data collected. The analysis will be repeated at the end of the study to provide a more complete picture.

4.5 SAFETY ANALYSES

The safety analysis population will consist of all randomized participants who received at least one dose of study drug, with participants grouped according to treatment and first dose received after randomization.

The main analyses will be done with a while on treatment strategy for two analyses sets: double-blind and overall satralizumab treatment period, which refers to the time in which participants are either randomized to satralizumab or switch to satralizumab at the beginning of the OLE period. The two analysis sets are defined as follows:

Analysis Set for the Double-Blind: All data up to the first dose in OLE, or up to 4 weeks after the last dose of the study drug (for participants who permanently discontinued the double-blind treatment), or last contact with the participant, or CCOD, whatever occurs first.

Analysis Set for the 'Overall Satralizumab Treatment': All data from the first dose of satralizumab (in the DB for the participants receiving satralizumab and in the **OLE for participants receiving placebo during the DB period**) until up to 4 weeks after the last dose of the satralizumab (for participants who permanently discontinued satralizumab), or last contact with the participant, or CCOD, whatever occurs first.

In addition, the data collected in the remaining period of the safety follow-up will be described separately and will use the analyses set for safety follow-up: all data later than 4 weeks after the last dose of study drug, for participants who permanently discontinued from the study drug.

Safety will be assessed through summaries of exposure to study treatment, AEs, changes in targeted laboratory test results, and changes in vital signs and electrocardiograms (ECGs).

Study treatment exposure (such as treatment duration, total dose received, and number of doses and dose modifications) will be summarized with descriptive statistics.

AE analyses will be performed on key subgroups including, but not limited to, age, race and region.

4.5.1 Extent of Exposure

Exposure to study drug information will be descriptively summarized by treatment as follows:

- Treatment duration (in days)
- Total number of administrations
- Total cumulative dose (mg)
- Frequencies of participants in each dose level
- Percentage of patients with a dose change due to weight changes.

Exposure to concomitant medications will be shown in a list overall and by treatment group showing the total number of uses of concomitant medications.

4.5.2 Adverse Events

All verbatim AE terms will be coded using the Medical Dictionary for Regulatory Activities version that is current at the time of the analysis (AE intensity will be graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events, version 5.0 (NCI CTCAE, v5.0) grading scale or according to Table 11 in Section 5.3.3 of the

Study Protocol (mild/moderate/severe/life-threatening/death), if the event is not specifically listed in the NCI CTCAE. For each treatment group, the frequency of each AE preferred term will be defined as the number of participants experiencing at least one occurrence of the AE. Each table will present the overall number of events and percentage of participants experiencing at least one AE.

Additional analyses for the 'overall satralizumab treatment' analysis set, summary of AE per 100 participant years will be provided. In this analysis multiple events will be counted each time they occur.

Percentages will be based on the number of participants in the safety-evaluable analysis population. In summary tables, AEs will be sorted by body system (in decreasing order of overall incidence), then by preferred term (in decreasing order of overall incidence).

The following safety information will be summarized by treatment group for the double-blind treatment period:

- AEs, AEs by intensity, AEs related to study drug
- Deaths
- Serious adverse events (SAEs), SAEs related to study drug
- AEs leading to discontinuation of study treatment
- AEs leading to dose modifications (dose interruption)
- Protocol-specified adverse events of special interest
- Infections
- Injection reactions.

The following data handling rules will be applied for all AE summary tables:

- Events that are missing both at onset and at end dates will be considered to have started after the first dose of study drug and the duration will be set to missing.
- If the onset date is missing, and the end date is on or after the first dosing date or unresolved or missing, then the event will be considered to have started after the first dose of study drug.

4.5.3 Laboratory Data

Laboratory data will be summarized by treatment group using descriptive statistics of absolute values and change from baseline values. In addition, the frequency of participants with abnormal laboratory values will be summarized by treatment group, visit, and baseline status.

4.5.4 Vital Signs

Vital signs assessments include systolic blood pressure, diastolic blood pressure, and pulse rate measured throughout the study. Vital sign measurements will be summarized by treatment group using descriptive statistics of absolute values and change from

baseline values. In addition, the frequency of participants with abnormal results will be summarized by treatment group, visit, and baseline status.

4.5.5 ECGs

ECG data will be summarized by treatment group for each assessment visit using descriptive statistics of absolute values and change from baseline values for the following parameters:

- Heart rate
- QRS duration
- RR interval
- PR interval
- QT interval

In addition, ECG overall interpretations will be summarized by treatment group and visit.

4.5.6 Columbia Suicide Severity Rating Scale (C-SSRS)

The Columbia Suicide Severity Scale (C-SSRS) is an assessment tool used to assess the lifetime suicidality of a participant (C-SSRS at baseline) as well as any new instances of suicidality (C-SSRS since last visit). The structured interview prompts recollection of suicidal ideation, including the intensity of the ideation, behavior, and attempts with actual or potential lethality.

- Suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent will be summarized by treatment group at each assessment visit. In addition, change from baseline to worst post-baseline assessment in suicidality.
- Categories will be summarized by treatment group.

4.6 OTHER ANALYSES

4.6.1 Summaries of Screened Population

A summary of screened participants and reasons for screen failures will be summarized by the IxRS report.

4.6.2 Summaries of Study Conduct

The numbers of participants who enroll in the study, discontinue from the study, and complete the study will be summarized overall and by treatment arm for each period (DB and OLE extension). Reasons for premature study withdrawal will be listed and summarized. Enrollment and major protocol deviations will be listed and evaluated for their potential effects on the interpretation of study results.

Number of participants with IECs, their category, and timing will be summarized.

Missingness for each efficacy measure irrespective of ICE will be summarized.

The impact of COVID-19 infections on missed visits, missed laboratories, AE underreporting, and study drug discontinuation due to COVID-19 infections will be summarized.

4.6.3 Summaries of Treatment Group Comparability/Demographics and Baseline Characteristics

Demographic and baseline characteristics will be summarized by treatment group using means, standard deviations, medians and ranges for continuous variables and proportions for categorical variables, as appropriate. Summaries will be presented overall and by treatment.

4.6.4 Overview of Intercurrent Events

Number of participants with ICEs and timing will be analyzed and their cause.

The following summaries will be provided:

- Number of ICEs split by type and NSDCR/SDCR and treatment
- Listing of ICEs

4.6.4.1 Analysis of the Adolescents Population

The adolescents population will be analyzed separately from the mITT. Due to the low number anticipated in this group it will be analyzed descriptively.

4.6.5 Pharmacokinetic Analyses

The PK analysis population consists of all participants in the safety analysis set with at least one valid post-dose concentration result with a dosing record and sampling time. The trial will evaluate the PK characteristics of satralizumab treatment over 24 weeks by summary statistics and non-linear mixed effects analysis (population PK) based on PK samples from both the DB and OLE periods.

The serum concentration at each sampling timepoint will be described by dose with means, medians, range and standard deviation and coefficient of variation of trough concentration (C_{trough}) irrespective of whether participants receive rescue therapy, change in baseline therapy for MG, miss a dose, or if study drug administration is delayed, or if they withdraw from treatment before data collection at Week 24. Mean serum-concentration-versus-time curves will be plotted.

Non-linear mixed effects analysis will be performed to analyze the satralizumab concentration–time data collected in the trial. The model to be used was previously developed on the basis of PK data from adult healthy volunteer and adult and adolescent participants with neuromyelitis optica spectrum disorder. Further model development may be undertaken if needed to achieve a satisfactory description of the data, and the data from this study may be pooled with data from other studies with satralizumab. Population and individual PK and exposure parameters will be generated based on the model. Covariate analysis, including demographic factors and anti-drug

antibody (ADA) status, will also be performed. The results of the pop PK analysis will be reported separately from the Clinical Study Report.

Additional exploratory PK analyses may be conducted as appropriate. The relationship between PK and efficacy, safety or pharmacodynamic (PD) endpoints may also be explored. These analyses will be reported separately.

An early PK data release will be conducted to ensure expedient sample analysis and model building by PK experts. This early release has a CCOD approximately 7 weeks prior to the CCOD for the primary analysis. The data will be released to the PK analysis group only after the primary analysis CCOD and the PK team receiving the data will be strictly separate from the team working on the analysis of efficacy data and will not communicate with them until data has been unblinded. In this early release the PK team will only receive PK-, anti-drug antibody (ADA) data and post-baseline IL6 and sIL-6R data, data of which association to clinical response is unknown. The details of this early release are described in a clinical pharmacology access charter.

4.6.6 Immunogenicity Analyses

The immunogenicity analysis population will consist of all participants with at least one post-dose ADA assessment. Participants will be grouped according to treatment received or, if no treatment is received prior to study discontinuation, according to treatment assigned.

The numbers and proportions of ADA-positive participants and ADA-negative participants at baseline (baseline prevalence) and after drug administration (post-baseline incidence) will be summarized by treatment group (and dose, if the iDMC concludes that dose levels should be increased). When determining post-baseline incidence, participants are considered to be ADA-positive (also called treatment-emergent ADA) if they show treatment-induced ADA response or treatment-enhanced ADA response. Participants who are ADA-negative or have missing data at baseline but develop an ADA response following study drug exposure have a treatment-induced ADA response. Participants who are ADA-positive at baseline and the titer of one or more post-baseline samples is at least 4-fold (0.60 titer unit) greater than the titer of the baseline sample have a treatment-enhanced ADA response. Participants are considered to be ADA-negative if they are ADA-negative at baseline or have missing data at baseline and all post-baseline samples are negative, OR if they are ADA-positive at baseline but (a) do not have any post-baseline samples with a titer that is at least 4-fold (0.60 titer unit) greater than the titer of the baseline sample, or (b) where all post-baseline samples are negative or missing (i.e., treatment unaffected).

The percentage of participants who have positive or negative ADA results for satralizumab will be tabulated. PD and safety will be summarized by anti-satralizumab antibody (i.e., satralizumab ADA) status. The impact of ADA on PK will be reported in the Pop PK report.

In addition, immunogenicity analyses may also be performed in the subgroups detailed in the Study Protocol, Section 6.4.8.

4.6.6.1 Further Analyses of ADA Classification

ADA responses are further classified as transient ADA response or persistent ADA response as follows and summarized using descriptive statistics:

- Transient ADA response: ADA-positive result detected
  a) At only one post-baseline sampling timepoint (excluding last timepoint), OR
  b) At 2 or more time points during treatment where the first and last ADA-positive samples are separated by a period $<$16 weeks, irrespective of any negative samples in between
- Persistent ADA response: ADA-positive result detected
  a) At the last post-baseline sampling timepoint, OR
  b) At 2 or more time points during treatment where the first and last ADA-positive samples are separated by a period $\geq$16 weeks, irrespective of any negative samples in between.

4.6.6.1.1 Pharmacodynamic Analyses

Serum interleukin-6 (IL-6) and soluble interleukin-6 receptor (sIL-6R) levels will be summarized by treatment group (and by dose, if the iDMC concludes that dose levels should be increased) and timepoint graphically and descriptively, as appropriate.

4.6.7 Analyses of China Subpopulation

The China subpopulation will include all participants enrolled at National Medical Products Administration-recognized sites. Results from these analyses will be summarized in a separate Clinical Study Report.

4.7 INTERIM ANALYSES

4.7.1 Safety Review

Periodic reviews of the safety data will be conducted by an iDCC with the results reviewed by the iDMC on an approximately 3-monthly basis until the last participant has completed the DB period.

4.7.2 Optional Interim for Futility

To adapt to information that may emerge during the course of this study, the Sponsor may choose to conduct one interim analysis for futility. Below are the specifications in place to ensure the study continues to meet the highest standards of integrity when an optional interim analysis is executed.

The study will not be stopped for positive efficacy as a result of the interim analysis.

An interim for futility using a threshold at a confidential cutoff value specified in a separate interim SAP was conducted on the 24 July 2023 and the iDMC confirmed no modification to the study.

4.7.3 Pharmacokinetic Interim Analysis

An interim analysis of PK data will be performed when approximately 30 participants (to include approximately 15 participants from the satralizumab group), have completed a minimum of 8 weeks of DB treatment. The purpose of the interim analysis is to confirm that the achieved exposure to satralizumab (and predicted trough receptor occupancy at steady-state [$RO_{tr,ss}$]) is within the predicted range. Participants with gMG with body weights representative of the overall gMG population, both in terms of range and approximate proportion, are expected to be included in the PK interim analysis dataset. The Sponsor may decide to postpone the conduct of the PK interim analysis to allow for inclusion of adolescent participants in the PK interim analysis.

A dose decision framework (including a pre-specified alternative higher dose, in case exposures are lower than predicted) will be defined in an analysis plan prior to study start (PK interim analysis plan). This plan will set out the criteria for predicted exposure and RO under which the decision for an adapted dose will be made. Predictions for exposure and RO in gMG using both the initial and pre-defined higher doses will be made using the existing popPK model, and the regimen that more closely achieves the target exposure and RO will be recommended by the iDMC.

This interim PK analysis will be performed by an external CRO while the Sponsor, participants, and investigators will remain blinded. The review by the external CRO will be restricted to PK and ADA data only, not safety or efficacy data, and no information that would reveal individual treatment assignments will be shared with the Sponsor. The external CRO will perform blinded simulations for exposure and RO, which will be made available to the iDMC, and will collaborate with the iDMC, such that the iDMC can make a dose recommendation on the basis of these simulations. Following this review, the iDMC will communicate its recommendation to the Sponsor.

Study recruitment will continue during the review period. Should a dose change be warranted, any participants treated at the original dose level will continue in the study on randomized treatment at the revised dose level, but will not contribute to the primary analysis. Additional participants will be recruited to ensure that the number of participants eligible for inclusion in the primary analysis reaches the target sample size. No type I error multiplicity correction will be performed as only PK data, unrelated to efficacy outcome, will be used for interim decision-making.

The PK interim took place on the 2 September 2022 and the iDMC recommended to continue the study without dose modification.

5. SUPPORTING DOCUMENTATION

This section is not applicable since there are no additional supporting documents.

APPENDIX 1: CHANGES TO PROTOCOL-PLANNED ANALYSES

This SAP clarifies the main analysis mentioned in the protocol with regards to the imputation strategy. The protocol applies an MMRM directly for the final analysis. In this SAP the final analysis is split into two transparent steps: Imputation and estimation of the treatment effect. The imputation uses an MMRM and the treatment effect is estimated through an ANCOVA.

6. REFERENCES

Burns T, Conaway M, Sanders, D. The MG Composite: A valid and reliable outcome measure for myasthenia gravis. Neurology 2010;74:1434-40.

Burns TM, Conaway M, Cutter G, et al. Less is more, or almost as much: A 15-item quality-of-life instrument for myasthenia gravis. Muscle and Nerve 2008;38:957-63.

Cella D, Lai JS, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. Neurology 2012;78:1860-7.

Clinical Review Report: Eculizumab (Soliris): Alexion Pharma Canada Corporation: Indication: Adult patients with generalized myasthenia gravis [Internet]. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; 2020 Dec.

Columbia Suicide Severity Scale. Available at: https://cssrs.columbia.edu/ (Assessed on 8 April 2022).

Data Analysis Plan for the PK Interim Analysis of Study WN42636, Version 1.0, Report no. 1111404 (by Hanna Silber Baumann).

Gilhus NE, Skeie GO, Romi F, Lazaridis K, Zisimopoulou P, Tzartos S. Myasthenia gravis - autoantibody characteristics and their implications for therapy. Nat Rev Neurol. 2016;12(5):259-268.

Harmonization, I. C. (n.d.). ICH E9 (R1) addendum on estimands and Sensitivity Analysis in Clinical Trials to the guideline on statistical principles for clinical trials. EMA/CHMP/ICH/436221/2017.

Howard J, Utsugisawa K, Benatar M, et al. Safety and efficacy of eculizumab in antiacetylcholine receptor antibody-positive refractory generalised myasthenia gravis (REGAIN): a phase 3, randomised, double-blind, placebo-controlled, multicentre study. Lancet Neurol 2017;16:976–86.

ICH E9, I. C. (n.d.). ICH E9 (R1) addendum on estimands and Sensitivity Analysis in Clinical Trials to the guideline on statistical principles for clinical trials. EMA/CHMP/ICH/436221/2017.

Interim Statistical Analysis Plan for the Futility Analysis: A Phase III, Randomized, Double-Blind, Placebo-Controlled, Multicenter Study to Evaluate Efficacy, Safety, Pharmacokinetics, and Pharmacodynamics of Satralizumab in Patients with Generalized Myasthenia Gravis. May 2023.

Rbmi package. Available at: https://github.com/insightsengineering/rbmi (Assessed on 6 April 2022).

Wolbers M, Noci A, Delmar P, et al. Reference-based imputation methods based on conditional mean imputation. arXiv preprint arXiv. 2021, 2109.11162.

Wolfe GI, Herbelin L, Nations SP, et al. Myasthenia gravis activities of daily living profile. Neurology 1999;52:1487-9.

Signature Page for Statistical Analysis Plan WN42636 (LUMINESCE) Satralizumab v3
System identifier: RIM-CLIN-518123

Approval Task	[redacted] Company Signatory 16-Jan-2024 17:09:26 GMT+0000