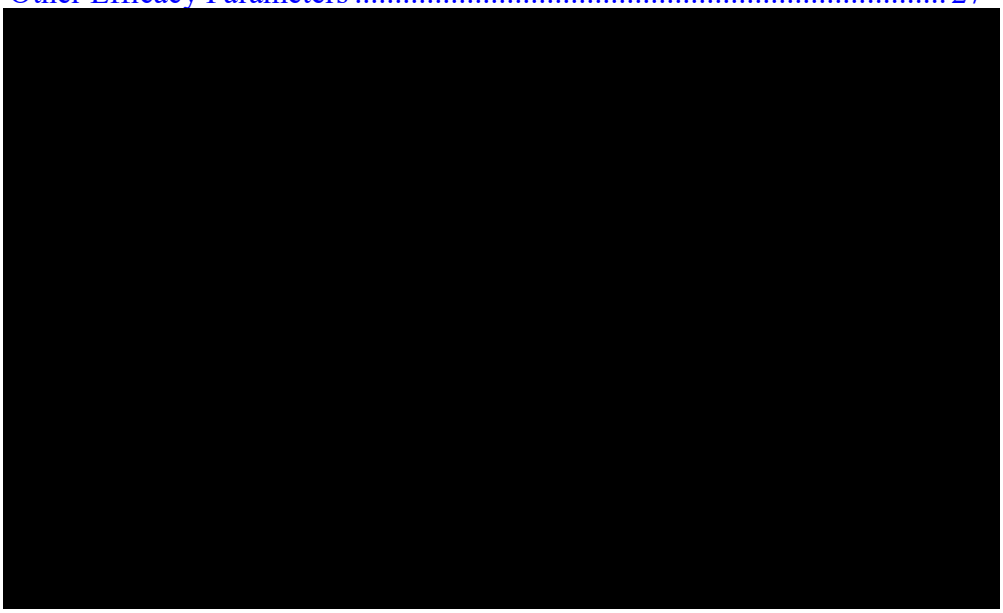
STATISTICAL ANALYSIS PLAN

Project Number:	GLPG1690
Study Number:	GLPG1690-CL-204
Study Title:	A Phase 2a randomized, double-blind, placebo-controlled, multi-center study to evaluate the efficacy, safety, and tolerability of orally administered GLPG1690 for 24 weeks in subjects with systemic sclerosis
Development Phase:	IIa
Status:	Final
Version:	1.00 **Date:** 5-Aug-2020
EudraCT:	2018-001817-33 **IND:** 140691
Sponsor:	Galapagos NV, Generaal De Wittelaan L11 A3, 2800 Mechelen, Belgium
[redacted] **Biostatistician:**	[redacted]
GLPG Biostatistician:	[redacted]
Medical Lead:	[redacted]

TABLE OF CONTENTS

Statistical Analysis Plan **1**
Table of contents **2**
Version History **6**
List of Abbreviations **7**
1. Introduction **9**
2. Study Design and Objectives **9**
2.1. Study Objectives 9
2.2. Study Endpoints 10
2.3. Study Design 11
2.4. CSP and CSP Amendments 12
2.5. Flow Chart 12
2.6. Sample Size Justification 12
2.7. Randomization and Blinding 13
3. General Methodology **13**
3.1. Analysis Sets 13
3.1.1. All Screened Analysis Set 13
3.1.2. All Randomized Analysis Set 13
3.1.3. Full Analysis Set 14
3.1.4. Per-Protocol Analysis Set 14
3.1.5. Safety Analysis Set 14
[redacted]
3.2. Randomized versus Actual Treatment Group 15
3.3. Analysis Periods and Analysis Time Points 15
3.3.1. Relative Day 15
3.3.2. Duration 15
3.3.3. Analysis Periods 15
3.3.4. Analysis Windows 16
3.4. Handling of Data 18
3.4.1. Handling of Missing Data 18
3.4.1.1. Handling of Missing Date Time Data 18
3.4.1.2. Handling of Missing Result Data 18
3.4.2. Handling of Values Below or Above a Threshold 18
3.4.3. Handling of Outliers 19
3.5. Presentation of Results 19
3.5.1. Presentation of Treatment groups 19
3.5.2. Calculation of Descriptive Statistics 19
3.5.3. Calculation of Percentages 19
3.5.4. Additional Considerations 20
4. Primary Analysis and Final Analysis **20**
5. Statistical Analyses **20**
5.1. Changes to the Planned Analyses, Not Covered by Protocol Amendments 20

5.2. Subject Information 21
5.2.1. Demographic and Baseline Disease Characteristics 21
5.2.2. Disposition Information 22
5.2.3. Protocol Deviations and Eligibility 22
5.2.4. Inclusion and Exclusion Criteria 23
5.2.5. Medical History and Concomitant Diseases 23
5.2.6. Prior and Concomitant Medications 23
5.2.6.1. Coding of Reported Terms 23
5.2.6.2. Classification of Medications 23
5.2.6.3. Calculation of Relative Days 24
5.2.6.4. Presentation of Results 24
5.2.7. Prior and Concomitant Non-Drug Therapies 24
5.2.7.1. Coding of Reported Terms 24
5.2.7.2. Classification of Non-Drug Therapies 24
5.2.7.3. Calculation of Relative Days 24
5.2.7.4. Presentation of Results 24
5.2.8. Exposure to IMP and Compliance 25
5.2.8.1. Derivation Rules 25
5.2.8.2. Presentation of Results 25
5.3. Efficacy Analyses 25
5.3.1. Level of Significance 26
5.3.2. Primary Efficacy Parameter 26
5.3.2.1. Definition 26
5.3.2.2. Derivation Rules 26
5.3.2.3. Analysis Methods 26
5.3.3. Sensitivity of Primary Endpoint 27
5.3.3.1. Derived mRSS 27
5.3.3.1.1. Definition 27
5.3.3.1.2. Derivation Rules 27
5.3.3.1.3. Analysis Methods 27
5.3.4. Other Efficacy Parameters 27



[redacted]

5.3.5. Subgroup Analyses 33
5.4. Safety Analyses 33
5.4.1. Adverse Events 34
5.4.1.1. Definition of Treatment-Emergent Adverse Events 34
5.4.1.2. Coding of Reported Terms 34
5.4.1.3. Allocation of Adverse Events to Analysis Periods 34
5.4.1.4. Treatment Relatedness 35
5.4.1.5. Presentation of Adverse Events and of Event Episodes 35
5.4.1.6. Worst-Case Selections 35
5.4.1.7. Calculation of Relative Days and Duration 36
5.4.1.8. Presentation of Results 36
5.4.1.9. EudraCT Adverse Events Reporting 36

[redacted]

5.4.3. Laboratory Safety 37
5.4.3.1. Available Data 37
5.4.3.2. Derivation Rules 38
5.4.3.3. Definition of Toxicity Grades 38
5.4.3.4. Definition of Non-Graded Abnormalities 40
5.4.3.5. Urinalysis Tests with Categorical Results 40
5.4.3.6. Treatment Emergent Principle 40
5.4.3.7. Worst-Case 41
5.4.3.8. Elevated Liver Function Test 42
5.4.3.9. Presentation of Results 42
5.4.4. Electrocardiogram 43
5.4.4.1. Available Data 43
5.4.4.2. Derivation Rules 43
5.4.4.3. Abnormalities 44
5.4.4.4. Presentation of Results 45
5.4.5. Vital Signs 45
5.4.5.1. Available Data 45
5.4.5.2. Abnormalities 45
5.4.5.3. Treatment Emergent Principle 46
5.4.5.4. Worst-Case Abnormality 46
5.4.5.5. Presentation of Results 46
5.4.6. Physical Examinations 47
5.4.7. Hospitalization and Meal intake 47

[redacted]

[redacted]

6. References 50
Appendix 1 52
Appendix 2 55
Appendix 3 56

VERSION HISTORY

SAP Amendment #	Date	Description of changes

LIST OF ABBREVIATIONS

AE	adverse event
ALT	alanine aminotransferase
AST	aspartate aminotransferase
[redacted]	[redacted]
BMI	body mass index
CI	confidence interval
CRF	case report form
[redacted]	[redacted]
CRO	contract research organization
CSP	clinical study protocol
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
DB	database
DBP	diastolic blood pressure
ECG	electrocardiogram
ED	early discontinuation
eGFR	estimated glomerular filtration rate
EoS	end of the study
FAS	Full-Analysis Set
[redacted]	[redacted]
FSH	follicle stimulating hormone
[redacted]	[redacted]
GGT	gamma glutamyl transferase
H	high, above the upper limit of the normal range
[redacted]	[redacted]
HDL	high density lipoprotein
HIV	human immunodeficiency virus
[redacted]	[redacted]
ICF	informed consent form
ICH	International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use
IEC	Independent Ethics Committee
[redacted]	[redacted]
IMP	investigational medicinal product
INF	infinity
IRB	Institutional Review Board
IWRS	interactive web response system
L	low, below the lower limit of the normal range
LDL	low density lipoprotein
LLN	lower limit of the normal range
LPA	lysophosphatidic acid
LPC	lysophosphatidylcholine
LS	least square
[redacted]	[redacted]
MCID	Minimal Clinically Important Difference

MedDRA	medical dictionary for regulatory activities
MMRM	mixed model repeated measures
mRSS	modified Rodnan skin score
N	normal, with the limits of the normal range
OC	observed case
OLE	open-label extension
[redacted]	[redacted]
PP	Per-Protocol
PT	preferred term
q.d.	once daily
QTcF	QT interval corrected for the heart rate using Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SBP	systolic blood pressure
[redacted]	[redacted]
SD	standard deviation
SE	standard error
[redacted]	[redacted]
SI	Standard International
SOC	System Organ Class
TC	Total cholesterol
TEAE	treatment-emergent adverse event
TLF	tables, listings and figures
ULN	upper limit of the normal range
VAS	visual analogue scale
WBC	white blood cell
WHO-DD	World Health Organization Drug Dictionary
WOCBP	women of childbearing potential

1. INTRODUCTION

This statistical analysis plan (SAP) defines the statistical analysis methods and data preparations to be used by [redacted] biostatistics department in the analysis and presentation of data for Galapagos protocol number GLPG1690-CL-204 entitled 'A Phase 2a randomized, double-blind, placebo-controlled, multi-center study to evaluate the efficacy, safety, and tolerability of orally administered GLPG1690 for 24 weeks in subjects with systemic sclerosis'.

The statistical analysis will process and present the results following the International Council for Harmonization (ICH) standards, particularly the ICH-E3, ICH-E6 and ICH-E9 guidelines, and [redacted] Global Biostatistics and Programming standard operating procedures (SOPs). This SAP is created based on the latest Clinical Study Protocol (CSP) as referred in section 2.4. Technical details on derivations and mock Tables, Listings and Figures (TLFs) will be presented in a separate document.

[redacted]

Only data collected for the GLPG1690-CL-204 (CL-204) will be stored in CL-204 SDTM and ADaM datasets, data collected after the date of first Investigational Medicinal Product (IMP) dose taken during the open-label extension phase of the study, which is described in protocol number GLPG1690-CL-206 (CL-206) will be excluded. The previous day before this date will be regarded as the end date of the CL-204 study for those subjects who start the CL-206 study, while the last contact date will be regarded as the end of CL-204 study for those subjects who do not roll over into CL-206 study. Data from the CL-204 and CL-206 protocols will be collected on the same database. At the end of the CL-204 study (i.e. when all subjects have met the CL-204 protocol definition of end of trial), a snapshot of the database will be used for the clinical study report (CSR) for CL-204. Any changes to the data and additional data that become available (only expected for concomitant medications and adverse events) captured after the data snapshot for CL-204 CSR will be reported in the CL-206 CSR, which will include cumulative data across both protocols.

2. STUDY DESIGN AND OBJECTIVES

2.1. STUDY OBJECTIVES

Primary Objective

- To evaluate the efficacy of GLPG1690 as evaluated by modified Rodnan skin score (mRSS) compared to placebo over 24 weeks for the treatment of subjects with systemic sclerosis

Secondary Objective

- To evaluate the safety and tolerability of GLPG1690 compared to placebo over 24 weeks in the treatment of subjects with systemic sclerosis

Other Objectives

[redacted]

2.2. STUDY ENDPOINTS

Primary Endpoint

- Change from baseline in mRSS over 24 weeks, i.e. over the entire treatment period as well as at each timepoint.

Secondary Endpoint

- Incidence of treatment-emergent adverse events (TEAEs), serious adverse events (SAEs), adverse events (AEs), and tolerability of GLPG1690 over 24 weeks

Other Endpoints

[redacted]

2.3. STUDY DESIGN

This is a randomized, double-blind, parallel-group, placebo-controlled, multi-center, Phase 2a study designed to evaluate the efficacy, safety, and tolerability of a 600 mg once daily (q.d.) dose of orally administered GLPG1690 over 24 weeks in approximately 30 adult subjects with a confirmed diagnosis of systemic sclerosis.

At Day 1/Visit 2 (baseline), subjects will be randomized in a 2:1 ratio to GLPG1690 600 mg q.d. taken as three film-coated tablets of 200 mg, or matching placebo q.d. administered for 24 weeks in addition to background treatment as defined in the protocol.

The subjects will visit the clinical study center at screening/Visit 1 (Day -28 to Day -1), Day 1/Visit 2 (baseline), Day 15 (Week 2/Visit 3), Day 29 (Week 4/Visit 4), Day 57 (Week 8/Visit 5), Day 113 (Week 16/Visit 7), Day 169 (Week 24/Visit 9), and, if applicable, the early discontinuation (ED) visit. At Weeks 12 (Visit 6) and 20 (Visit 8), a phone call will be made to assess safety. In addition, a follow-up visit will be planned 4 weeks after the last administration of IMP (Day 197 [Week 28]) (i.e. Follow-up Visit 1) and a second follow-up visit will be planned 12 weeks after the last administration of IMP (Day 253 [Week 36]) (i.e. Follow-up Visit 2). Additional unscheduled visits are allowed for any safety assessments if clinically indicated (see protocol Section 6.2). The end of the study (EoS) is reached when the last visit of the subject is performed. The assessments performed at each visit are detailed in the Schedule of Activities (see Appendix 1).

A schematic diagram of the clinical study design, procedures, and stages is provided in Figure 1 (section 2.5).

Each subject will be in the study for up to approximately 40 weeks (up to 4 weeks of screening, 24 weeks of treatment, and 12 weeks of follow-up).

Open-label Extension

Subjects who completed the Week 24 visit will be offered treatment with GLPG1690 in an optional open-label extension (OLE), provided regulatory and Independent Ethics Committee (IEC)/Institutional Review Board (IRB) approvals for such an extension is granted. In this case, the follow-up visits as scheduled in protocol GLPG1690-CL-204, will not be performed for these subjects.

The analysis methods of the OLE study is not presented in this SAP and will be described in a separate SAP.

2.4. CSP AND CSP AMENDMENTS

Protocol Versions	Date (ddMMMyyyy)
Version 1	31JUL2018
Version 2	20AUG2018
Amendment 1: Version 3	13NOV2018
Amendment 2: Version 4	31JUL2019
Amendment 3: Version 5	28APR2020

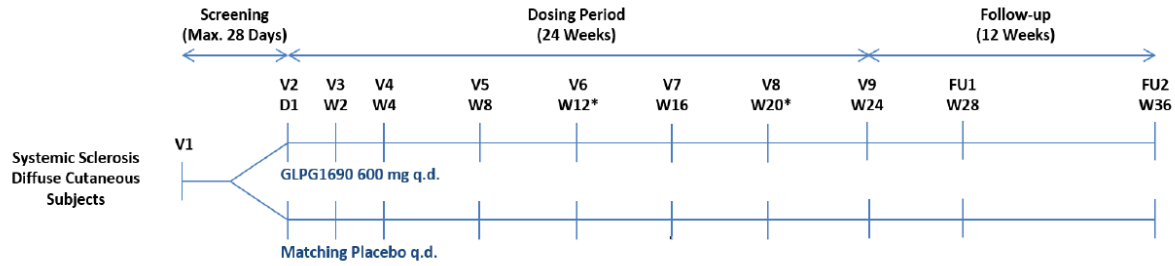
Protocol Amendments	Date (ddMMMyyyy)
[redacted] Country Specific Amendment	12MAR2019

This SAP was based on the latest version of the protocol that was available at the time of its finalization.

2.5. FLOW CHART

A schematic diagram of the clinical study design, procedures, and stages is provided in Figure 1.

Figure 1: Schematic Study Overview



D=Day, FU=Follow-up, V=Visit, W=Week.

* This will be a phone call to assess safety (no clinical study center visit).

2.6. SAMPLE SIZE JUSTIFICATION

Sample size was based on empirical considerations. The number of subjects (30) included in this study should give reasonable precision around the estimates derived for the efficacy evaluation.

The primary endpoint of the study is the change from baseline in mRSS over 24 weeks.

Considering 20 and 10 subjects in the GLPG1690 and placebo arm, respectively, and a common standard deviation (SD) of 5 on the change from baseline in mRSS, the Minimal Clinically Important Difference (MCID) of 4.7, and taking into account a 10% dropout, the probability to observe a treatment effect of more than 4 points is 63%.

2.7. RANDOMIZATION AND BLINDING

Randomization

This is a randomized, double-blind clinical study. At screening, subjects will be assigned a subject identification. When a subject is confirmed to be eligible for the clinical study, the subject will be randomized. Allocation of each subject to a given treatment will be done using a centralized electronic system (interactive web response system [IWRS]). Subjects will be randomized in a 2:1 ratio to receive GLPG1690 600 mg q.d. or matching placebo, respectively.

For each subject at each clinical study center visit, the clinical study center will contact the IWRS for the appropriate treatment number to be assigned. The medication kit(s) will contain the relevant IMP for the period until the next clinical study center visit.

Subjects and the entire clinical study team, including the investigators, clinical study coordinators, [redacted] and sponsor personnel are blinded to treatment assignment.

Blinding and Unblinding

Blinded and packaged medication will be provided to the clinical center.

The blind can be broken by the investigator via IWRS only if the investigator deems it necessary for the safety of a subject. Further details are included in section 4.6.2 of the protocol.

Code-break information (via IWRS vendor/randomization list) will be provided to the bioanalytical laboratory responsible for plasma drug determination sample analysis, and to the sponsor pharmacovigilance lead for SAE reporting purposes.

Biostatistics team will be unblinded at the time of the primary analysis which is scheduled when all patients completed the Week 24 visit and data is locked for the primary analysis. Only study team members included in the analyses and reporting activities will become unblinded at the time of the primary analysis in case it is different from the CL-204 study full analysis with regards to timing.

3. GENERAL METHODOLOGY

3.1. ANALYSIS SETS

The analysis sets used in the statistical analyses are detailed in the following subsections. The analysis set will always be indicated in a subtitle in the TLF.

3.1.1. All Screened Analysis Set

All subjects who signed and dated an informed consent form (ICF) and underwent screening assessments to check whether or not they are eligible to participate in the clinical study.

3.1.2. All Randomized Analysis Set

All enrolled subjects who were randomized into the clinical study. Subjects will be analyzed according to the treatment arm to which they were randomized.

3.1.3. Full Analysis Set

The full analysis set (FAS) is defined as all randomized subjects who have received at least 1 intake of IMP. Subjects will be analyzed according to the treatment arm to which they were randomized.

3.1.4. Per-Protocol Analysis Set

The per-protocol set (PP) is defined as all FAS subjects who did not have a major protocol deviation that impacts the efficacy results (as defined in a protocol deviation plan). The determination of the PP population will be finalized and documented prior to unblinding. Any subjects having mRSS assessment at Week 24 outside of the original time window of [142, 183] study day will be excluded from the PP Set.

The per-protocol analysis is only planned if at least 10% of the subjects have a major protocol deviation. If the treatment code of a subject is broken (e.g., in case of a SAE), all efficacy data will be disqualified from the PP Set.

3.1.5. Safety Analysis Set

All randomized subjects who have received at least one intake of IMP. Subjects will be analyzed according to the treatment arm to which they were randomized.

[redacted]

3.2. RANDOMIZED VERSUS ACTUAL TREATMENT GROUP

For all subject data, the treatment group as assigned by the randomization will be used in the analysis (i.e., as-randomized analysis).

Differences between the randomized and actual treatment group will be explained in the CSR.

3.3. ANALYSIS PERIODS AND ANALYSIS TIME POINTS

3.3.1. Relative Day

The timing of an assessment or an event relative to a reference date will be calculated as follows:

$$\begin{aligned} &Relative\ day\ (DY) \\ &= concerned\ date - reference\ date\ + 1\ day, \text{ when concerned date } \geq \text{reference date} \\ &= concerned\ date - reference\ date, \text{when the concerned date<reference date} \end{aligned}$$

- The *reference date* is the date of the first IMP intake, unless specified otherwise.
- The *concerned date* could be the measurement date of the assessment, or the start or end date of the event.
- *Date* implies a complete date. If the date is (partially) missing, the relative day will be missing.

Exception: if the end date of an event is fully missing at database lock (in case of an interim analysis) or at the subject's study termination, the event will be considered as ongoing. The end date will be imputed by the database cutoff date or the date of last visit for cases where subject terminated study. In listing, the duration will be printed as "> X days".

3.3.2. Duration

The duration of an assessment or an event will be calculated as follows:

$$Duration\ (days) = end\ date - start\ date + 1\ day$$

- *Date* implies a complete date. If the date is (partially) missing, the duration will be missing.

Exception: if the end date of an event is fully missing at database lock (in case of an interim analysis) or at the subject's study termination, the event will be considered as ongoing. The end date will be imputed by the database cutoff date or the date of last visit for cases where subject terminated study. In listing, the duration will be printed as "> X days".

3.3.3. Analysis Periods

All event-type data (e.g., AEs) and assessments will be allocated to analysis periods according to Table 1.

Table 1: Analysis Periods

Analysis Period	Start Analysis Period	End Analysis Period
Screening	Date of signing the ICF.	Date of first IMP intake - 1 day
Treatment	Date of first IMP administration.	For subjects not entering CL-206 study: date of last intake of IMP + 30 days. For subjects who entered the CL-206 study: the date of last intake of IMP + 30 days or the date of first OLE IMP administration (excluded), whichever comes first.
Post-Treatment	End date of the treatment period + 1 day.	Date of last contact in CL-204 study. For subjects who entered the CL-206 study this will be the date of first OLE IMP administration.

ICF: informed consent form

The last analysis period will always end on the date of last contact, collected on the Study Completion/Discontinuation of Study eCRF page, whichever is earliest.

3.3.4. Analysis Windows

All assessments including data collected on unscheduled visits, will be allocated to analysis visits (time windows) based on the relative day of the assessment (see section 3.3.1) and according to the algorithm in Table 2. TFLs will present the analysis visits when analysis is derived.

Table 2: Analysis Windows

Analysis Visit label	Target Day	Interval Lower Bound	Interval Upper Bound
Vital signs, Clinical Laboratory tests			
Baseline*	1	-INF	1
Week 2	15	2	22
Week 4	29	23	43
Week 8	57	44	85
Week 16	113	86	141
Week 24	169	142	197
Week 28	198	198	+INF
12-Lead ECG			
Baseline*	1	-INF	1
Day 1	1	1	2
Week 2	15	3	64
Week 16	113	65	141
Week 24	169	142	197
Week 28	198	198	+INF
mRSS			
Baseline*	1	-INF	1
Week 4	29	2	43
Week 8	57	44	85
Week 16	113	86	141
Week 24	169	142	197
Week 28	198	198	+INF

Analysis Visit label	Target Day	Interval Lower Bound	Interval Upper Bound
[redacted]			

*: includes only assessments performed before first IMP administration.

Definition of Baseline

Generally baseline is defined as the last non-missing value before first IMP administration based on both date and time information (if available). If multiple values qualify as the last value, the mean of these values will be used as baseline value in the analysis.

[redacted]

For electrocardiogram (ECG), where triplicates (or duplicates) results are measured, the baseline is defined as the mean of the last recorded triplicates (or duplicates) results before first

IMP administration. If no triplicate (or duplicate) results are available before the first dose, the last value before first study drug administration will be used.

Selection of Visits

Per parameter, analysis visit and timepoint, the value closest to the target date will be used in analysis tables and figures. Other values will be listed only. If more than one value is located at the same distance from the target day, then the latest in time will be selected. If there are two values on the same day and no time indicating which one is last, the average of the two will be calculated.

3.4. HANDLING OF DATA

3.4.1. Handling of Missing Data

3.4.1.1. Handling of Missing Date Time Data

Generally no imputations will be done in case of missing date (time) fields, nor for the missing parts of partially known date (time) fields apart from partially missing initial diagnosis dates.

When the date of initial diagnosis is partially missing, the following rules will be used in order to determine the duration of disease:

- If the day of initial diagnosis is missing, then 01 will be assigned to the missing fields.
- If the month of initial diagnosis is missing, then January will be assigned to the missing fields.
- If date of initial diagnosis is completely missing, no imputation will be done.

Assessments with completely missing date (time) will be omitted from the analysis.

Event-type data (e.g. adverse events, concomitant medications) with missing date (time) will be allocated to analysis periods using a worst-case approach as explained in the respective sections.

3.4.1.2. Handling of Missing Result Data

No imputation will be done of missing results unless specified otherwise in corresponding sections (i.e. observed cases (OC) analysis).

3.4.2. Handling of Values Below or Above a Threshold

Values below (above) the quantification limit will be imputed by the value one unit smaller (larger) than the quantification limit itself. In listings, the original value will be presented.

Example: if the database contains the value “<0.04”, then for the descriptive statistics the value “0.03” will be used. The value “>1000” will be imputed by “1001”.

[redacted]

[redacted]

3.4.3. Handling of Outliers

There will be no outlier detection, all measured values will be included in the analyses [redacted]

[redacted]

3.5. PRESENTATION OF RESULTS

3.5.1. Presentation of Treatment groups

In the section Subject Information, a grand total "All Subjects" will be added to summarize all subjects over all treatment groups in tables.

3.5.2. Calculation of Descriptive Statistics

For continuous parameters, descriptive statistics will be presented when n≥2. When n=1, the summary statistics will be included in tables and figures. Only the summary statistics for n and mean will be displayed, other summary statistics will be left blank.

Descriptive statistics will include:

- the number of non-missing data points (n)
- the arithmetic mean
- the standard error (SE) and standard deviation (SD)
- the median, minimum and maximum
- 95% confidence interval (CI) of the mean (if indicated in the relevant section).

3.5.3. Calculation of Percentages

Frequencies and percentages will be generated for categorical parameters.

For event-type data (e.g. Adverse Events), the denominator will be all subjects in the analysis set and analysis visit/period. For other data (e.g. worst-case analysis of assessments), the denominator will be all subjects with post-baseline data for the parameter, in the analysis set and analysis visit/period.

3.5.4. Additional Considerations

All data collected except for assessment that were not done (and hence assessment date is missing) will be listed in data listings.

4. PRIMARY ANALYSIS AND FINAL ANALYSIS

There will be no interim analysis in the study. The primary analysis will be performed when all subjects have completed the Week 24 visit (or discontinued earlier), providing all tables, listings and figures based on the data collected and cleaned up to Week 24 visit. Every effort will be made to keep the subjects and investigators blinded to individual treatment assignments until completion of the final analysis (after the last subject has completed the last follow-up visit in CL-204 study). Upon the completion of the post-treatment period in the CL-204 study, the final analysis will be performed.

5. STATISTICAL ANALYSES

5.1. CHANGES TO THE PLANNED ANALYSES, NOT COVERED BY PROTOCOL AMENDMENTS

- All Enrolled Analysis Set will not be used for the analyses so not included in SAP.
- All Randomized Analysis Set will be used for the presentation and summary of the baseline assessments to better interpret the study data.
- Definition of Per-Protocol Analysis Set was slightly modified within this SAP text as unblinding will occur at the time of primary analysis, i.e. not at the same time of database lock.
- Definition of Per-Protocol Analysis Set was modified to exclude subjects who had mRSS assessment at Week 24 outside of the original time window of [142, 183] study day.
- [redacted]
- An additional sensitivity analysis for the primary endpoint was added where for subjects that did not have a Week 24 mRSS assessment at all within the extended window, but possibly in the Week 28 - Week 32 interval, their Week 28 result will be allocated to Week 24 analysis visit.

5.2. SUBJECT INFORMATION

Subject information will be tabulated using the FAS unless otherwise specified. No inferential testing will be performed, nor will p-values be provided.

Subject information will be tabulated with descriptive statistics per treatment group and overall.

5.2.1. Demographic and Baseline Disease Characteristics

The following parameters will be summarized:

- Gender: Male; Female
- Age at signing the ICF (years)
- Age, categorized (years): age ≤ 45; age > 45
- Ethnicity: Not Reported; Hispanic or Latino; Non Hispanic or Latino; Unknown
- Race: American Indian or Alaskan Native; Asian; Black or African American; Native Hawaiian or Other Pacific Islander; and White
- Is the Subject of Japanese Ancestry?: Yes; No
- Baseline Height (cm)
- Baseline Weight (kg)
- Body Mass Index BMI (kg/m²)

Age, height, weight, and BMI will be summarized using descriptive statistics. The number and percentage of subjects by age category, gender, ethnicity, and race will also be reported.

A summary of baseline information will be presented and summarize the following parameters:

- Duration of disease (years) $= \frac{\text{(date of first IMP intake)} - \text{(date of initial diagnosis)} + 1}{365.25}$.
- Duration of disease, categorized (years): duration < 2; duration ≥ 2
- Baseline total mRSS scores

Subjects with/without lung disease at baseline: presence of lung disease at baseline is defined based on Scleroderma Clinical Trials Consortium (SCTC) criteria identified based on medical history data. Subjects with lung disease at baseline will be defined as subjects with any medical history preferred term of 'Interstitial lung disease' or 'Systemic sclerosis pulmonary'

that were active at baseline (i.e. start date is prior to first dose of study medication and end date is after first dose of study medication or ongoing). All other baseline efficacy and safety parameters will be presented together with post-baseline results by timepoint tables.

For more details on the imputation of date of initial diagnosis, refer to section 3.4.1.1.

5.2.2. Disposition Information

The following tabulations will be provided:

- The number of subjects screened, randomized and not-randomized, with the reason for not being randomized. For all subjects screened only.
- Number (percent) of all randomized subjects per country and investigator by treatment and overall.
- The number of subjects in screened analysis set and number (percent) of subjects in each analysis set (randomized, FAS, PP, [redacted] and safety sets), on the "All randomized Analysis Set" as defined in section 3.1 by treatment and overall.
- The number (percent) of subjects per analysis window for mRSS as defined in section 3.3.4 by treatment and overall.
- The number (percent) of subjects who completed/discontinued the study treatment and the reasons for discontinuation by treatment and overall.
- The number (percent) of subjects who completed the study, subjects who completed the treatment period and entered the OLE study, subjects who completed the treatment period and entered Follow-up phase (i.e. did not entered the OLE study), subjects who completed Follow-up phase and subjects who discontinued the study and the reasons for discontinuation.

5.2.3. Protocol Deviations and Eligibility

Major protocol deviations are determined and recorded while the study is ongoing, and the list is finalized prior to unblinding for the Primary Analysis (although any possible protocol deviations collected between data cut for Primary Analysis and database lock for Final Analysis will continue to be reviewed). For more details, please refer to the Protocol Deviations Plan.

The number (percent) of subjects with major protocol deviations will be tabulated, overall and per class of deviation, by treatment group and overall.

All available information concerning major protocol deviations, deviations on eligibility criteria and subjects not treated will be listed. Protocol deviations leading to exclusion from any analysis set (see section 3.1.5) will be flagged.

5.2.4. Inclusion and Exclusion Criteria

All inclusion and exclusion criteria that subjects did not fulfil will be listed in detail on the All Screened Analysis Set.

5.2.5. Medical History and Concomitant Diseases

Medical history will be coded using the latest version of the medical dictionary for regulatory activities (MedDRA) as per the data validation manual (DVM). The number and percentage of subjects with any medical history will be summarized overall and for each predefined system organ class (SOC) and preferred term (PT). Frequency tabulations per SOC and PT will be provided for the medical history findings: (i.e., any event with a stop date prior to first IMP intake) as well as for the concomitant diseases (i.e., ongoing conditions or conditions with a stop date on or after first IMP intake).

5.2.6. Prior and Concomitant Medications

A prior medication is defined as any medication that is taken prior to the first dose of IMP. A concomitant medication is defined as any medication that has a stop date that is on or after the date of first dose of IMP or is ongoing. Medications that started prior to the first dose of IMP and have stop date on or after the first dose of IMP or is ongoing are defined as both prior and concomitant medications.

5.2.6.1. Coding of Reported Terms

All medications used within 12 weeks prior to and during the screening period will be collected on the electronic case report form (eCRF). All medications will be coded according to the latest World Health Organization Drug Dictionary (WHO-DD).

5.2.6.2. Classification of Medications

All prior and concomitant medication records will be categorized as follows, considering their date and flags indicating the relative timing versus study (IMP) start or end (before, after, ongoing…):

- Prior only: when the record ended before first IMP administration date.
- Concomitant only: when the record started on or after the first IMP administration date.
- Prior and concomitant: when the record started before the date of first IMP administration, and ended on or after this point, or continued, or if end date is missing.

When the start or end date of the prior and concomitant medication records are incomplete (and no flags indicating relative timing are available), the date of first IMP administration will be considered to the same level of information provided by these incomplete dates to categorize the timing of these records. This means that a record only having month and year will be categorized comparing only to the month and year of the date of first IMP administration. Worst case scenario will be used to categorize concomitant medications: if the same level of information provided by incomplete concomitant medication start date is identical first IMP administration, then the medication will be considered as both prior and concomitant, given end

date is not prior to first IMP administration. If the same level of information provided by incomplete concomitant medication end date is identical first IMP administration, then the medication will be considered as both prior and concomitant, given start date is not after to first IMP administration.

5.2.6.3. Calculation of Relative Days

For both the start and the end dates of the concomitant medication records, their day relative to the day of first IMP administration will be calculated as described in section 3.3.1.

5.2.6.4. Presentation of Results

A frequency tabulation of the anatomical therapeutic chemical (ATC) classes Level 4 by therapeutic subgroup (ATC Level 2) and anatomical main group (ATC Level 1) of the prior medications (defined as 'prior only' and 'prior and concomitant') will be provided as well as of the concomitant medications (defined as 'concomitant only' and 'prior and concomitant').

All medications will be listed in detail.

5.2.7. Prior and Concomitant Non-Drug Therapies

5.2.7.1. Coding of Reported Terms

All non-drug therapies used within 12 weeks prior to and during the screening period will be collected on the electronic case report form (eCRF). All non-drug therapies will be coded according to the latest MedDRA Dictionary.

A prior non-drug therapy is defined as any non-drug therapy that is taken prior to the first dose of IMP. A concomitant non-drug therapy is defined as any non-drug therapy that has a stop date that is on or after the date of first dose of IMP or is ongoing.

5.2.7.2. Classification of Non-Drug Therapies

All prior and concomitant non-drug therapy records will be categorized as specified in section 5.2.6.2.

5.2.7.3. Calculation of Relative Days

For both the start and the end dates of the concomitant non-drug therapy records, their day relative to the day of first IMP administration will be calculated as described in section 3.3.1.

5.2.7.4. Presentation of Results

A frequency tabulation of the SOC and PT term of non-drug therapies of the prior therapies (defined as 'prior only' and 'prior and concomitant') will be provided. Similarly, a separate table will be provided for concomitant therapies (defined as 'concomitant only' and 'prior and concomitant').

All non-drug therapies will be listed in detail.

5.2.8. Exposure to IMP and Compliance

5.2.8.1. Derivation Rules

Derived Parameters: Extent of Exposure to IMP

- *Total treatment duration (days)* = last IMP administration date – first IMP administration date + 1 day.
- *Total treatment duration, excluding days off study IMP*: Number of days with any IMP administration.
- *Total treatment duration, fully compliant (days)*: Number of days with IMP administration, as planned per CSP.

Derived Parameters: Compliance

- *Overall compliance (%)* = $100 \text{ x } \left(\frac{\text{number of tablets actually used}}{\text{number of tablets that should have been used}}\right)$

Number of tablets that should have been used will be calculated as:

Total treatment duration * 3

- *Percentage days with any intake (%)* = $100 \text{ x } \left(\frac{\text{total treatment duration,excluding days off IMP}}{\text{total treatment duration}}\right)$
- *Percentage days fully compliant (%)* = $100 \text{ x } \left(\frac{\text{total treatment duration,fully compliant}}{\text{total treatment duration}}\right)$

Number of tablets actually used will be based on the exposure records collected in the database.

5.2.8.2. Presentation of Results

Summary statistics will be provided for each compliance and extent of exposure parameter on the Safety Set.

Frequency tables will be provided for the compliance parameters, using the following categories: <80%; 80% ≤ x < 100%; 100%; 100% < x ≤ 120%; >120%. Percentages will be calculated out of the number of subjects who were dosed.

5.3. EFFICACY ANALYSES

Efficacy analyses will be performed on the FAS. To investigate the impact of non-compliance and protocol violators, a Per-Protocol analysis will also be performed for the primary endpoint, if more than 10% of the subjects have at least one major protocol deviation leading to exclusion to the PP Analysis Set.

Tabulations will be shown per treatment group.

5.3.1. Level of Significance

All statistical tests will be 2-sided and performed at the 5% significance level, thereby providing 95% (2-sided) CIs. However, given the exploratory nature of this proof-of-concept study, the success of the study will be based on the evaluation of all analysis results, including safety profile and descriptive analysis of all efficacy endpoints.

5.3.2. Primary Efficacy Parameter

5.3.2.1. Definition

The primary endpoint of the study is the change from baseline in mRSS over 24 weeks. This is collected at the time points specified in the Schedule of Activities (see Appendix 1).

For individual subjects, clinically meaningful changes in skin score (MCID) are 4.7 skin score units and/or at least 20% change in overall mRSS. Only decreases from baseline will be deemed clinically meaningful changes in skin score.

5.3.2.2. Derivation Rules

Total mRSS score will be analyzed as reported by the investigator, considering that all individual scores are completed, no imputation will be done on the primary efficacy endpoint.

5.3.2.3. Analysis Methods

The following analyses will be done on the FAS as the primary analysis, and repeated on the PP Analysis Set, if more than 10% of the subjects have at least one major protocol deviation, to confirm the results of the primary analysis:

The mRSS and change from baseline in mRSS score (primary endpoint) will be summarized by descriptive statistics. Mean (±SE) plots over time (analysis visits based on time window) for both treatment groups, and similar mean (±SE) plots for the change from baseline will be presented. Additionally, spaghetti plot of change from baseline will be displayed with subjects presented by treatment.

Change from baseline in mRSS will be analyzed using a mixed-effects model for repeated measures (MMRM). The model will include treatment and visit as fixed effects, baseline mRSS and country as covariates, and treatment-visit as interaction terms. Subject will be included in the model as a random effect. The likelihood-based MMRM will be performed using a SAS PROC MIXED procedure (see Appendix 3), and the Kenward-Roger method for degrees of freedom will be applied. The restricted maximum likelihood (REML) method will be used for estimation. The Least Square (LS) means along with the associated SE and 95% CIs, as well as the difference in LS means and associated 95% CIs will be displayed at each time point. For exploratory purpose only, p-values will also be displayed for testing the difference in LS means at each timepoint. The p-value derived for the overall difference between treatment groups will also be displayed, as well as all p-values for fixed effects included in the model.

The variance-covariance matrix will be assumed to be unstructured. If the procedure does not converge then a compound symmetric variance-covariance matrix will be used instead.

The number and percentage of subjects with a clinically meaningful change in skin score (as defined in section 5.3.2.2) will be summarized by treatment. Incidence will be derived overall (at any post-baseline timepoint) and for each analysis visit.

5.3.3. Sensitivity of Primary Endpoint

5.3.3.1. Derived mRSS

5.3.3.1.1. Definition

If more than 10% of subjects have missing mRSS total score at Week 24, then the change from baseline in derived mRSS will be analyzed over 24 weeks to confirm the results of the primary efficacy analysis.

There will be an additional sensitivity analysis for subjects that did not have a Week 24 at all within the extended window, but possibly in the Week 28 - Week 32 interval. These subjects will be off IMP in that period. The primary endpoint analysis will be repeated including the closest mRSS assessment to the original Week 24 target (Day 169) up to and including Week 32 (Day 225).

5.3.3.1.2. Derivation Rules

Total mRSS score will be derived from individual mRSS scores at each visit if at least 14 of the 17 items are not missing. At each visit, the derived score will be:

(Sum of all non-missing items) * 17 / (number of non-missing items)

5.3.3.1.3. Analysis Methods

Derived mRSS will be analyzed, mean (±SE) plots over time and spaghetti plot will be produced, as specified in section 5.3.2.3.

5.3.4. Other Efficacy Parameters

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

5.3.5. Subgroup Analyses

Actual values and change from baseline in mRSS [redacted] will be summarized over the 24-weeks for the following subgroups:

- [redacted]
- mRSS baseline score: < 20 or ≥ 20 points

5.4. SAFETY ANALYSES

Safety analyses will be performed on the Safety Set.

No formal testing will be performed to compare the treatment groups.

Tabulations will be shown by treatment group.

5.4.1. Adverse Events

All adverse events (AEs) and changes in attributes (worsening and improvement) of AEs are reported in the database. An identification number serves to link the records considered by the investigator as describing the evolution of one and the same event.

5.4.1.1. Definition of Treatment-Emergent Adverse Events

The analysis of adverse events will be based on treatment-emergent events (TEAE). TEAEs are defined as

- For subjects not entering the OLE study: AEs having a start date equal or after the date of the first IMP administration and up to the last dose date + 30 days.

- For subjects who entered the OLE study: AEs having a start date equal or after the date of the first IMP administration and up to the date of last double-blinded IMP dose date + 30 days or the date of first OLE IMP administration (excluded), whichever comes first.

- And is either a newly reported event, or a worsening* of an existing event. Improvements are not considered treatment-emergent.

*Worsening is defined as worsening in at least one of the following attributes: seriousness, severity, relationship and/or action taken.

5.4.1.2. Coding of Reported Terms

All AE terms will be coded in the database using the MedDRA version 21.0 or higher coding dictionary.

All tables in this section will show the AE terms coded into PT grouped into body SOC. Subject listings will also show the reported terms. Any other coding levels will only be shown in a listing summarizing coding or if explicitly mentioned.

5.4.1.3. Allocation of Adverse Events to Analysis Periods

All AEs records will be placed into analysis periods considering their start date, aiming to report the incidence of these events only in the analysis period during which they started.

The general rule for allocation of AEs to analysis periods follows:

Analysis period start date ≤ AE start date ≤ analysis period end date

If the start date of an AE is missing or incomplete to a level preventing a clear allocation of the AE to one single analysis period and no flag indicating timing relative to study medication is available, a worst-case consideration (see below) will be done aiming to allocate the AE to one single analysis period, if possible. When a worst-case consideration is needed, the end date of the AE, if and as available, should also be considered; if such AEs clearly end on a given point, this will exclude the possibility to allocate the AE to an analysis period after that point.

- An AE which according to the available information of its start date could belong to the screening as well as to the treatment period will only be placed in the treatment period.

- An AE which according to the available information of its start date could belong to treatment period as well as to follow-up period will only be placed in the treatment period.

- An AE with a missing start date will be allocated to treatment period.

- If an AE can be allocated to both double-blind and open-label (i.e. CL-206 study) treatment period based on the above rules, then the event will be reported for CL-204 double-blind treatment period during CL-204 study analyses.

5.4.1.4. Treatment Relatedness

Following ICH-E3 (ICH), the originally reported relatedness to IMP of an AE will be dichotomized as follows:

- *Not IMP related*: all non-missing weaker levels of relatedness than 'possibly IMP related'.

- *IMP related*: 'possibly IMP related' and all stronger levels of relatedness (this class also includes any missing IMP relatedness, as a worst-case consideration).

Only this dichotomized relatedness will be used in tables. Relatedness as originally reported will only be listed.

5.4.1.5. Presentation of Adverse Events and of Event Episodes

All AEs tables will only show TEAEs and will show counts of the number of subjects with TEAEs. The number of event episodes will be shown only in tables where explicitly mentioned.

AEs starting before first IMP administration or later than 30 days after the last dose of IMP but before first OLE IMP administration (not TEAEs) will only be listed. AEs reported on or after the date of first OLE IMP administration will be included in the CSR produced after CL-206 study has been completed.

AEs leading to death will be presented in a separate listing.

5.4.1.6. Worst-Case Selections

When cross-tabulating AE preferred terms versus an AE attribute (like severity), only the worst-case within each same preferred term, same subject and same analysis period will be considered, i.e., when a same subject has more than once the same AE preferred term reported in the same treatment group, the subject will be counted only once and will be shown under the worst outcome (like the worst severity for that AE in the concerned treatment period).

5.4.1.7. Calculation of Relative Days and Duration

For each AE record, its start day in study (the day of the AE start date relative to the date of first IMP administration), its start day in the analysis period, and its duration (in days) will be calculated and shown only in listings.

See sections 3.3.1 and 3.3.2 for the calculation of relative days and duration respectively.

5.4.1.8. Presentation of Results

A summary table will be provided, showing then number (percentage) of subjects with:

- at least one TEAE,
- at least one IMP-related TEAE,
- any serious TEAE,
- any TEAE leading to death,
- any TEAEs by worst severity,
- TEAE leading to IMP discontinuation.

Frequency tabulations, by SOC and PT, of the number (percent) of subjects with a TEAE will be presented by treatment group. Similar tables will be provided for the following groups:

Severity ('Grade 1 - Mild', 'Grade 2 – Moderate', 'Grade 3 – Severe', 'Grade 4 – Life Threatening', and 'Grade 5 – Death')

- Related TEAEs
- Related TEAEs by severity
- Serious TEAEs.

Serious adverse events, AEs leading to death, AEs leading to IMP and trial discontinuation will be listed.

5.4.1.9. EudraCT Adverse Events Reporting

For the purpose of EudraCT reporting, the following tabulations will be added:

Frequency tabulations, by SOC and PT, of the number (percent) of subjects with non-serious TEAE will be presented. A similar table will be provided for the all serious TEAEs (counting events), as well as a table for non-serious TEAEs reported in at least 5% of the subjects in any treatment group.

[redacted]

5.4.3. Laboratory Safety

5.4.3.1. Available Data

Laboratory tests scheduled are described in section 6.6.2 of the protocol.

The statistical analyses will only present results in Standard International (SI) units. Other units will not be presented.

Hematology, coagulation, and serum/plasma chemistry data provided by the central lab will be used in tables and figures. For hematology parameters, when both absolute and percentage of white blood cells (WBC) are available, only the absolute values will be presented in tables and figures.

Urinalysis tests will be presented as part of the descriptive statistics tables.

Results of serology testing, pregnancy test, and FSH will only be listed.

5.4.3.2. Derivation Rules

Fasted and Non-Fasted Results

Laboratory parameters that are sensitive to the fasting status: glucose, triglycerides.

These parameters will be tabulated with descriptive statistics by fasting/non-fasting and toxicity grade. Laboratory results for which the fasting status is missing, will be considered as taken non-fasted.

5.4.3.3. Definition of Toxicity Grades

Toxicity grades will only be derived for laboratory tests for which toxicity gradings are available.

For the analysis values of the following continuous laboratory parameters, toxicity grades will be determined as implemented in the next table (Common Terminology Criteria for Adverse Events (CTCAE), v5.0: November 27, 2017). Analysis results scoring below the lowest grade limits are defined to correspond to grade 0.

Table 4: Laboratory Toxicity Grades

Parameter	Direction of abnormal	Grade 1	Grade 2	Grade 3	Grade 4
Alanine amino transferase	high	>ULN-3 x ULN	>3 x ULN -5 x ULN	>5 x ULN - 8 x ULN	>8 x ULN
Albumin (g/L)	low	<LLN-30	<30-20	<20	-
Alkaline phosphatase	high	>ULN-2.5 x ULN	>2.5 x ULN -5 x ULN	>5 x ULN -20 x ULN	>20 x ULN
Aspartate amino transferase	high	>ULN-3 x ULN	>3 x ULN -5 x ULN	>5 x ULN -8 x ULN	>8 x ULN
Bilirubin [total]	high	>ULN-1.5 x ULN	>1.5 x ULN -3 x ULN	>3 x ULN -10 x ULN	>10 x ULN
Calcium, below (mmol/L)	low	<LLN-2.0	<2.0-1.75	<1.75-1.5	<1.5
Calcium, above (mmol/L)	high	>ULN-2.9	>2.9-3.1	>3.1-3.4	>3.4
Cholesterol (mmol/L)	high	>ULN-7.75	>7.75-10.34	>10.34-12.92	>12.92
Creatine kinase	high	>ULN-2.5 x ULN	>2.5 x ULN -5 x ULN	>5 x ULN -10 x ULN	>10 x ULN
Creatinine	high	>ULN-1.5 x ULN	>1.5 x ULN -3 x ULN or >1.5 x baseline -3 x baseline	>3 x ULN -6 x ULN or >3 x baseline	>6 x ULN
eGFR/ Creatinine clearance	low	<LLN-60	<60-30	<30-15	<15

Parameter	Direction of abnormal	Grade 1	Grade 2	Grade 3	Grade 4
(mL/min/1.73 m2)					
Eosinophils absolute count	high	>ULN	-	-	-
Eosinophils relative count (%)	high	>7	-	-	-
Gamma-glutamyl transferase	high	>ULN-2.5 x ULN	>2.5 x ULN -5 x ULN	>5 x ULN -20 x ULN	>20 x ULN
Glucose [fasting], below (mmol/L)	low	<LLN-3.0	<3.0-2.2	<2.2-1.7	<1.7
Hemoglobin, below (g/L)	low	<LLN-100	<100-80	<80	-
Hemoglobin, above (g/L)	high	Increase >0-20	Increase >20-40	Increase >40	-
Lymphocytes, below (10^9/L)	low	<LLN-0.8	<0.8-0.5	<0.5-0.2	<0.2
Lymphocytes, above (10^9/L)	high	-	>4.0-20.0	>20.0	-
Neutrophils, below (10^9/L)	low	<LLN-1.5	<1.5-1.0	<1.0-0.5	<0.5
Potassium, below (mmol/L)	low	<LLN-3.0	-	<3.0-2.5	<2.5
Potassium, above (mmol/L)	high	>ULN-5.5	>5.5-6.0	>6.0-7.0	>7.0
Sodium, below (mmol/L)	low	<LLN-130	<130-125	<125-120	<120
Sodium, above (mmol/L)	high	>ULN-150	>150-155	>155-160	>160
Triglycerides fasting and non-fasting (mmol/L)	high	>1.71-3.42	>3.42-5.7	>5.7-11.4	>11.4
Partial thromboplastin time [activated or not specified]	high	>ULN-1.5 x ULN	>1.5 x ULN -2.5 x ULN	>2.5 x ULN	-
International normalized ratio	high	>1.2 x ULN -1.5 x ULN	>1.5 x ULN -2.5 x ULN	>2.5 x ULN	-
Platelets (10^9/L)	low	<LLN-75	<75-50	<50-25	<25
White blood cells, below (10^9/L)	low	<LLN-3	<3-2	<2-1	<1
White blood cells, above (10^9/L)	high	-	-	>100	-

Legend:
[Specifications].
(Specific unit into which the defined grades apply).
Baseline value is defined as abnormal if the value is not within the normal range. For directional toxicity of high an abnormal baseline will be a baseline value which is higher than the upper limit of the reference range. For directional toxicity of low, an abnormal baseline will be a baseline value which is lower than the lower limit of the reference range.
x LLN / x ULN = times the lower / times the upper limit of the normal range.
Below / above = differentiate grade definitions in the low / high sense for a same parameter.

The following parameters will be analyzed and categorized based on pre-defined criterion without using the above defined grading system:

- Prothrombin time: >ULN-1.5 x ULN; >1.5 x ULN - 2.5 x ULN; >2.5 x ULN

- Glucose [fasting] (mmol/L): >7.2

- High density lipoprotein (mmol/L): <1.554

- Low density lipoprotein (mmol/L): >4.144

The analysis of these parameters will be performed in the same way as the analysis of the parameters with standard toxicity grade.

5.4.3.4. Definition of Non-Graded Abnormalities

Non-graded abnormalities will be determined only for the parameters having no definitions of toxicity grades (see section 5.3.3.3).

For laboratory tests provided by the laboratory, the position of the actual analysis values versus their normal ranges will be determined directly by using the position indicator provided in the database as reported, expressing the classes for these analysis values as low (L), normal (N) or high (H). L, N and H are further referred to as non-graded abnormalities.

5.4.3.5. Urinalysis Tests with Categorical Results

Results of urinalysis with qualitative results will be tabulated by time point. No toxicity grading or non-graded abnormalities will be derived.

5.4.3.6. Treatment Emergent Principle

Toxicity Grades

A post-baseline toxicity grade 1, 2, 3 or 4 is defined as treatment-emergent when higher than the toxicity grade of the baseline result. If the baseline result is missing, a post-baseline toxicity grade 1, 2, 3 or 4 will always be considered as treatment-emergent.

For prothrombin time parameter a post-baseline value meeting the defined abnormality criteria is regarded as treatment-emergent if the post-baseline criteria meet a higher degree of abnormality than the baseline value (e.g. post-baseline value meeting >1.5-2.5 x ULN category while baseline meeting >ULN-1.5 x ULN category) or the baseline value is missing or does not meet any abnormality criteria. For glucose [fasting], high density lipoprotein and low density lipoprotein a post-baseline abnormal category will be regarded as treatment emergent, if the baseline value is missing or does not meet the defined abnormality criteria.

Non-graded Abnormalities

A post-baseline non-graded abnormality class L or H is defined as treatment-emergent when it differs from the abnormality class of the baseline result. If the baseline result is missing, a post-baseline abnormality L or H will always be considered as treatment-emergent.

5.4.3.7. Worst-Case

Toxicity Grading

The worst-case post-baseline toxicity grade 0, 1, 2, 3 or 4 will be determined per subject, per parameter (and sense, if below and above) during the on-treatment analysis period, using all non-missing post-baseline records within that period (including unscheduled and follow-up visits).

The worst-case toxicity grade is the highest toxicity grade scored for the parameter (in each sense, if below and above).

Non-graded Abnormalities

The following worst-case post-baseline abnormalities L, N or H will be determined per subject, per parameter and for the on-treatment analysis period, using all non-missing post-baseline records within that period (including unscheduled and follow-up visits):

- L = low: at least one post-baseline result is classified as L.
- N = normal: all post-baseline results are classified as N.
- H = high: at least one post-baseline result is classified as H.

If, for a subject, both L and H are reported, the subject will be counted twice in the table: once with a worst-case L and once with a worst-case H.

5.4.3.8. Elevated Liver Function Test

To assess the potential of the IMP to cause severe liver damage, possible Hy's Law cases will be identified. Count and percentage of subjects who have at least one assessment meeting the criteria defined in Table 5 will be summarized in the on-treatment analysis period.

Table 5: Hepatotoxicity

Parameter	Thresholds
AST / ALT combination	AST > 1.5- <3 x ULN AST >= 3- <5 x ULN AST >= 5- <8 x ULN AST >= 8 x ULN ALT > 1.5- <3x ULN ALT >= 3- <5 x ULN ALT >= 5- <8 x ULN ALT >= 8 x ULN ALT and/or AST > 1.5 -<3 x ULN AST and /or ALT >= 3- <5 x ULN AST and /or ALT >= 5- <8 x ULN AST and /or ALT >= 8 x ULN
AST / ALT / bilirubin combination	Bilirubin >= 1.5 x ULN AND AST or ALT >= 3 x ULN Bilirubin >= 2 x ULN AND AST or ALT >= 3 x ULN

Hepatotoxicity values will also be listed including the timing of occurrence.

5.4.3.9. Presentation of Results

No formal inferential statistics (p-values) will be provided.

Continuous parameters will be summarized by means of descriptive statistics (including 95% CI of the mean change from t-test) by parameter, treatment group and analysis visit. Actual values and changes from baseline will be tabulated separately.

Figures of the mean (± SE) actual values over time and of the mean (± SE) change from baseline will be prepared for all parameters.

All laboratory abnormalities after the first dose of study medication up to 30 days after the last dose, if the subject did not enroll in the OLE study, or the earliest between the last IMP intake + 30 days during the double-blind study and the first dose of IMP in the OLE study (excluded), if the subject continues in OLE, will be included.

Abnormalities of the actual values will be presented as shift tables of the worst-case abnormality versus the baseline abnormality/toxicity grade. The table will be created per parameter and treatment group for the on-treatment analysis period. The results of non-graded abnormalities and toxicities grades will be shown separately.

A frequency table of the number (percentage) of subjects with treatment-emergent worst-case abnormalities per parameter and treatment group for the on-treatment analysis period will be presented. The results of non-graded abnormalities and toxicities grades will be shown separately.

Urinalysis tests for which no normal range is available will be tabulated separately as categorical data.

Clinical laboratory values will also be listed including normal ranges and indicating if value is out of range. Separate listings will also be created to include all laboratory values of a subject who has had at least one value meeting CTCAE grading 3 or higher category.

5.4.4. Electrocardiogram

5.4.4.1. Available Data

All subjects will have a standard 12-lead electrocardiogram (ECG) performed at the time points specified in the Schedule of Activities (see Appendix 1). The following ECG parameters will be analyzed: heart rate (bpm), PR interval, QRS interval, uncorrected QT interval (ms), morphology, QT interval corrected for the heart rate using Fridericia's formula (QTcF) (ms).

5.4.4.2. Derivation Rules

Derived Parameters

The ECG parameters will be provided in the clinical database, no imputation will be done.

Handling of ECGs Measured in Triplicate

If ECG is collected in triplicates (or duplicates), the following approach will be taken.

The mean of the triplicate (or duplicate) ECG values will be calculated for each individual ECG parameter, without rounding the result. These calculated means will constitute the analysis values; any derivation (e.g. change from baseline, assignment of abnormalities) and statistic will be based on the mean value of the triplicates (or duplicates).

When a single ECG is performed, the actual results of the single ECG will be summarized.

The values of the individual triplicate (or duplicate) assessments will be listed, i.e. will not be summarized or graphically presented.

5.4.4.3. Abnormalities

The actual analysis values and changes from baseline of the QT and QTcF parameters will be categorized into the abnormality classes as defined in ICH E14:

Table 6: Abnormalities on ECG Parameters

Parameter	Abnormality	Limits
Abnormalities on actual values		
QT and QTcF (ms)	QT* ≤ 450 450 < QT* ≤ 480 480 < QT* ≤ 500 QT* > 500	≤ 450 450 < value ≤ 480 480 < value ≤ 500 > 500
Abnormalities on change from Baseline		
QT and QTcF (ms)	QT* change ≤ 30 30 < QT* change ≤ 60 QT* change > 60	≤ 30 30 < value ≤ 60 > 60

* Indicate which QT(c) parameter is analyzed in the label

Worst-Case Abnormality

The worst-case post-baseline categorized actual analysis value and the worst-case categorized change from baseline for QT and QTcF will be determined per subject, per parameter, using all non-missing post-baseline records (including unscheduled and follow-up visits) for the on-treatment analysis period.

The worst-case categorized actual analysis value is the category corresponding to the highest post-baseline actual value.

The worst-case change from baseline is the category corresponding to the largest increase (positive change) from baseline.

For the overall interpretation the worst-case post-baseline will be determined per subject, per parameter, and for on-treatment analysis period, using all non-missing post-baseline records (including unscheduled and follow-up visits) as follow:

- Normal: if all the post-baseline results are normal
- Abnormal: if at least one post-baseline result is abnormal.

Treatment Emergent Abnormalities

Actual value: An abnormal post-baseline abnormality is defined as treatment-emergent when the abnormality is worse compared to the abnormality at baseline. If the baseline result is missing, a post-baseline result outside of the normal range will always be considered as treatment-emergent.

An abnormal category for change from baseline is always treatment-emergent.

5.4.4.4. Presentation of Results

No formal inferential statistics (p-values) will be presented.

Continuous parameters will be summarized by means of descriptive statistics (including 95% CI of the mean change from t-test) by parameter, treatment group and analysis visit. Actual values and changes from baseline will be tabulated separately.

All ECG abnormalities after the first dose of study medication up to 30 days after the last dose, if the subject did not enroll in the OLE study, or the earliest between the last IMP intake + 30 days during the double-blind study and the first dose of IMP in the OLE study (excluded), if the subject continues in OLE, will be included.

Abnormalities of the actual values will be presented as shift tables of the worst-case abnormality versus the baseline abnormality. The table will be created per parameter, treatment group for the on-treatment analysis period.

A frequency table of the number (percentage) of subjects with treatment-emergent worst-case abnormalities per parameter, treatment group for the on-treatment analysis period will be presented. Overall interpretation as assessed by the investigator will also be summarized.

5.4.5. Vital Signs

5.4.5.1. Available Data

The following vital signs parameters will be analyzed: Systolic Blood Pressure (SBP) (mmHg), Diastolic Blood Pressure (DBP) (mmHg), heart rate (bpm), respiratory rate (breaths/minute), and body temperature (°C). Vital signs will be analyzed regardless of assessment position. If there are more values available for a subject taken by different positions for a visit, then selection rules for the analyses will be applied as described in Section 3.3.4. No adjustment to the temperature will be made to account for the possible different method of collection.

5.4.5.2. Abnormalities

The relevant vital sign values at each visit will be classified based on the reference ranges found in Table 7 below.

Table 7: Normal Ranges for Vital Signs

Systolic blood pressure (mmHg)	Diastolic blood pressure (mmHg)	Heart rate (bpm)	Tympanic temperature (°C)	Respiratory rate (breaths per min)
90≤ SBP ≤140	45≤ DBP ≤90	50≤ HR ≤100	35.5≤ t° ≤37.5	12≤ RR ≤20

5.4.5.3. Treatment Emergent Principle

A post-baseline abnormality class L or H is defined as treatment-emergent when it differs from the abnormality class at baseline. If the baseline result is missing, a post-baseline abnormality L or H will always be considered as treatment-emergent.

5.4.5.4. Worst-Case Abnormality

The following worst-case post-baseline abnormalities low, below the lower limit of the normal range (L), normal, with the limits of the normal range (N) or high, above the upper limit of the normal range (H) will be determined per subject, per parameter and for the on-treatment analysis period, using all non-missing post-baseline records (including unscheduled and follow-up visits):

- L = low: at least one post-baseline result is classified as L.
- N = normal: all post-baseline results are classified as N.
- H = high: at least one post-baseline result is classified as H.

If, for a subject, both L and H are reported, the subject will be counted twice in the table: once with a worst-case L and once with a worst-case H.

5.4.5.5. Presentation of Results

No formal inferential statistics (p-values) will be derived.

Continuous parameters will be summarized by means of descriptive statistics (including 95% CI of the mean change from a t-test) for all vital sign parameters and analysis visits, by treatment group and overall for subjects in the applicable Safety Set. Actual values and changes from baseline will be tabulated separately.

All abnormalities after the first dose of study medication up to 30 days after the last dose, if the subject did not enroll in the OLE study, or the earliest between the last IMP intake + 30 days during the double-blind study and the first dose of IMP in the OLE study (excluded), if the subject continues in OLE, will be included.

Abnormalities of the actual values will be presented as shift tables of the worst-case abnormality versus the baseline abnormality. The table will be created per parameter and treatment group for the on-treatment analysis period.

A frequency table of the number (percentage) of subjects with treatment-emergent worst-case abnormalities per parameter, treatment group for the on-treatment analysis period will be presented.

Mean (± SE) plots over time for both treatment groups, with heart rate, respiratory rate, and blood pressure by position on a new page, will be displayed.

5.4.6. Physical Examinations

Abnormal post baseline physical examination results (including gynaecological exam results for female subjects) will be presented in a listing.

5.4.7. Hospitalization and Meal intake

All information recorded in the eCRF pages "Meal Information" and "Hospitalization Details" will be listed.

[redacted]

[redacted]

[redacted]

[redacted]

6. REFERENCES

Bazett, H. (1920 (7)). An analysis of the time-relations of electrocardiograms. *Heart*, 353–370.

Fridericia, L. (1920). Die systolendauer im elektrokardiogramm bei normalen menschen und bei herzkranken. 53:469-486.

ICH-E3. (December 1995). *Structure and content of clinical study reports. Step 4 Guideline.*

ICH-E6. (17 July 1996). *Guideline for good clinical practice. Step 5 Guideline.*

ICH-E9. (5 February 1998). *Statistical principles for clincal trials. Step 4 guideline.*

ICH-E14. (12 May 2005). *The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs.*

[redacted]

APPENDIX 1

EVENT	SCREENING	TREATMENT PERIOD									FOLLOW-UP[j]	
Study Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**ED[a]**	**FU1**	**FU2**
Study Day (D) or Week (W)[b] ± days (d)	**D-28 to D-1**	**D1**	**W2 ±2d**	**W4 ±2d**	**W8 ±4d**	**W12[c] ±4d**	**W16 ±4d**	**W20[c] ±4d**	**W24[i] ±4d**		**W28 ±7d**	**W36 ±7d**
Informed consent	✓											
Inclusion/exclusion criteria	✓	✓										
Demographics	✓											
Medical history/concurrent illnesses	✓											
FSH test	✓											
Randomization		✓										
Phone call						✓		✓				
Pregnancy test[d]	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	
Physical examination	✓	✓	✓	✓	✓		✓		✓	✓	✓	
Vital signs	✓	✓	✓	✓	✓		✓		✓	✓	✓	
12-Lead ECG[e]	✓	✓	✓				✓		✓	✓	✓	
Serology	✓											
Clinical laboratory tests	✓	✓	✓	✓	✓		✓		✓	✓	✓	
mRSS	✓	✓		✓	✓		✓		✓	✓	✓	

EVENT	SCREENING	TREATMENT PERIOD									FOLLOW-UP[j]	
Study Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**ED[a]**	**FU1**	**FU2**
Study Day (D) or Week (W)[b] ± days (d)	**D-28 to D-1**	**D1**	**W2 ±2d**	**W4 ±2d**	**W8 ±4d**	**W12[c] ±4d**	**W16 ±4d**	**W20[c] ±4d**	**W24[i] ±4d**		**W28 ±7d**	**W36 ±7d**
[redacted]		✓					✓		✓	✓	✓	
[redacted]		✓					✓		✓	✓	✓	
[redacted]		✓					✓		✓	✓	✓	
[redacted]		✓					✓		✓	✓		
[redacted]		✓							✓	✓		
[redacted]		✓	✓	✓	✓		✓		✓	✓		
[redacted]		✓	✓	✓			✓		✓	✓	✓(✓)	(✓)
[redacted]		✓							✓	✓	(✓)	(✓)
Dispense subject diary		✓	✓	✓	✓		✓		✓	✓	✓	
Collect subject diary			✓	✓	✓		✓		✓	✓	✓	✓
Dispense IMP		✓	✓	✓	✓		✓					
Review IMP compliance			✓	✓	✓		✓		✓	✓		

EVENT	SCREENING	TREATMENT PERIOD									FOLLOW-UP[j]	
Study Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**ED[a]**	**FU1**	**FU2**
Study Day (D) or Week (W)[b] ± days (d)	**D-28 to D-1**	**D1**	**W2 ±2d**	**W4 ±2d**	**W8 ±4d**	**W12[c] ±4d**	**W16 ±4d**	**W20[c] ±4d**	**W24[i] ±4d**		**W28 ±7d**	**W36 ±7d**
Dose IMP		q.d. throughout the treatment period										
AE assessment	Throughout the study											
Concomitant medication assessment and documentation	Throughout the study											

FU=follow-up, QoL=quality of life.
[a] ED visit if applicable.
[b] Week is defined as 7 days.
[c] This will be a phone call to assess safety (no clinical study center visit).
d Serum pregnancy test at screening, urine pregnancy test at all other visits.
e Triplicate ECGs will be performed irrespective of IMP intake at Visit 1 (screening), within 30 minutes before IMP intake and between 2 to 3 hours after IMP intake at Visit 2 (baseline) and Visit 3. A single ECG will be performed within 30 minutes before IMP intake at Visits 7 and 9, and irrespective of IMP intake at ED (if applicable) and Follow-up Visit 1.

[redacted]

[i] If Visit 9 (Week 24) cannot be conducted within the time window of ±4 days, the visit window can be extended to ±28 days (i.e. can be extended or shortened up to 28 days) to enable conduct of the visit on site. If Visit 9 is conducted after Week 24, additional regular phone calls at Week 24 and every 14 days thereafter should be implemented, until an on-site Visit 9 can take place, or until Week 32, to evaluate the subject's safety. In addition, if any of the follow-up visits cannot be performed on-site, a phone call to evaluate the subject's safety should be performed on the date of the visit. The information to be collected during these phone calls is detailed in Protocol Section 6.6.
[j] The FU visits need to be scheduled in relation to the last dose of IMP. FU1 should be 28 days ±7 days after the last IMP administration, and FU2 should be 84 days ±7 days after the last IMP administration.

[redacted]

APPENDIX 3

The below SAS code represents the analysis model used for the primary efficacy variable analysis.

The variance-covariance matrix will be assumed to be unstructured. If the procedure does not converge then a compound symmetric variance-covariance matrix will be used instead [redacted].

Signature Page for glpg1690-cl-204-sap 15149

Approval	[redacted] Biostatistician Biometrics 06-Aug-2020 07:26:08 GMT+0000

Approval	[redacted] 07-Aug-2020 11:19:13 GMT+0000

Signature Page for glpg1690-cl-204-sap 15149