Document title	**STATISTICAL ANALYSIS PLAN**
Study official title	**Efficacy and safety of 3 doses of S201086/GLPG1972 administered orally once daily in patients with knee osteoarthritis. A 52-week international, multi-regional, multi-center, randomized, double-blind, placebo-controlled, dose-ranging study.** **ROCCELLA Study**
Study brief title	**Efficacy of S201086/GLPG1972 in patients with knee osteoarthritis**
Test drug code	**S201086/GLPG1972**
Indication	**Osteoarthritis**
Development phase	**Phase 2**
Protocol code	**CL2-201086-002/GLPG1972-CL-201**
EudraCT Number	**2017-004581-10**
Universal Trial Number	**U1111-1205-0321**
Investigational New Drug Application Number	**133039**
Sponsor	**GALAPAGOS NV (US)** **Institut de Recherches Internationales Servier (I.R.I.S.) (ex-US)**
Date of the document	**29 September 2020**
Version of the document	**Final version 3.0**

CONFIDENTIAL

Signatories

Prepared by:	Signature	Date
[redacted] I.R.I.S. - FRANCE		
[redacted] - FRANCE		

Approved by:	Signature	Date
[redacted] I.R.I.S. - FRANCE		
[redacted] I.R.I.S. - FRANCE		
[redacted] GALAPAGOS - BELGIUM		
[redacted] GALAPAGOS - BELGIUM		
[redacted] I.R.I.S. - FRANCE		
[redacted] I.R.I.S. - FRANCE		
[redacted] I.R.I.S. - FRANCE		
[redacted] I.R.I.S. - FRANCE		
[redacted] GALAPAGOS - USA		
[redacted] GALAPAGOS - BELGIUM		

Follow up of versions

Version	Release date	Key modifications (*)	Impact
2.0		- Translation of the previous signed SAP into the new template of the SAP - Modification of the analysis set "Randomised Set" into "Modified Randomised Set" - More detailed information about sensitivity analyses	- All document - Section 2.1 - Section 3.4.2.2
3.0		- [redacted] - Addition of COVID-19 analyses (see Section 5.5)	- Sections 2.1 and 3.7 - Sections 3.2.1, 3.2.2, 3.4.2.2, 3.4.3 and 3.6.1

(*) Main changes as compared to the statistical analyses planned in the protocol for the first SAP signed version (1.0). Main changes from the previous signed version for the other SAP signed version(s).

Table of contents

List of tables

List of figures

List of abbreviations

%	:	percentage
µmol	:	micromole
AE	:	Adverse Event
ALT	:	ALanine (Amino)Transferase
ALP	:	ALkaline Phosphatase
ANCOVA	:	Analysis of COVAriance
AST	:	ASpartate (Amino)Transferase
AUC	:	Area Under the Curve
b.i.d.	:	bis in die (twice a day)
BMI	:	Body Mass Index
[redacted]	:	[redacted]
BP	:	Blood Pressure
bpm	:	beats per minute (heart rate unit)
CHMP	:	Committee for Medicinal Products for Human Use
CI	:	Confidence Interval
CL	:	Total CLearance
cm	:	centimetre
CPK	:	Creatine PhosphoKinase
CPMP	:	Committee for Proprietary Medicinal Products
CRO	:	Contract Research Organization
DBP	:	Diastolic Blood Pressure
DMC	:	Data Monitoring Committee
e.g.	:	exempla gratia (for example)
EAE	:	Emergent Adverse Event
ECG	:	ElectroCardioGram
e-CRF	:	electronic-Case Report Form
ELISA	:	Enzymes linked immunosorbent assay
EMA	:	European Medicines Agency
g	:	gram
G/L	:	Giga (10^9) per litre
GGT	:	Gamma-Glutamyl Transferase (Gamma-Glutamyl Transpeptidase)
h	:	hour
HbA1c	:	Glycated Haemoglobin
HBs	:	Surface antigen of Hepatitis B virus
HDL	:	High-Density Lipoprotein
HR	:	Heart Rate
i.e.	:	id est
I.R.I.S.	:	Institut de Recherches Internationales Servier
ICH	:	International Conference on Harmonization
im	:	intramuscular (route)
IME	:	Important Medical Event
IMP	:	Investigational Medicinal Product
IRS	:	Interactive Response System
IS	:	Included Set
IU	:	International Unit
IV	:	IntraVenous (route)

IVRS	:	Interactive Voice Response System
IWRS	:	Interactive Web Response System
kg	:	kilogram
L	:	Litre
LDH	:	Lactate DeHydrogenase
LDL	:	Low-Density Lipoprotein
LLN	:	Lower Limit of Normal laboratory reference range
LOCF	:	Last Observation Carried Forward
LLS	:	Lower Limit used to define potentially clinically Significant abnormal values
m	:	metre
Max	:	Maximum
MCH	:	Mean Corpuscular Haemoglobin
MCHC	:	Mean Corpuscular Haemoglobin Concentration
MCV	:	Mean Corpuscular Volume
MedDRA	:	Medical Dictionary for Regulatory Activities
mg	:	milligram
MI	:	Multiple Imputation
min	:	minute
Min	:	Minimum
mL	:	millilitre
mm	:	millimetre
mmHg	:	millimetre of mercury
mmol	:	millimole
MMRM	:	Mixed-effects Model for Repeated Measures
mRS	:	Modified Randomised Set
ms	:	millisecond
mV	:	milliVolt
NA	:	Not Applicable
NAE	:	Number of Adverse Events
NEAE	:	Number of Emergent Adverse Events
ng	:	nanogram
NIMP	:	Non Investigational Medicinal Product
NIS	:	Non Included Set
NPD	:	Number of Protocol Deviations
o.d.	:	omni die (every day)
PCSA	:	Potentially Clinically Significant Abnormal value
PMM	:	Pattern Mixture Model
po	:	per os (orally)
PPS	:	Per Protocol Set
PT	:	Preferred Term
PV	:	PharmacoVigilance
QTc	:	QT interval corrected for heart rate
RBC	:	Red Blood Cells
RS	:	Randomised Set
s	:	second
SAE	:	Serious Adverse Event
SAP	:	Statistical Analysis Plan
SBP	:	Systolic Blood Pressure
sc	:	subcutaneous (route)

SD	:	Standard Deviation
SE	:	Standard Error
SEAE	:	Serious Emergent Adverse Event
SOC	:	System Organ Class
SS	:	Safety Set
t.i.d.	:	ter in die (three times a day)
T/L	:	Tera (10^{12}) per litre
TLG	:	Tables, Listings and Graphs
TU	:	Therapeutic Unit
ULN	:	Upper Limit of Normal laboratory reference range
ULS	:	Upper Limit used to define potentially clinically Significant abnormal values
WBC	:	White Blood Cells
WHO	:	World Health Organization
WHO-DRL	:	World Health Organization, Drug Reference List

1. INTRODUCTION

This Statistical Analysis Plan details the planned analyses to be performed, in accordance with the main characteristics of the amended study protocol #2.
The templates for Tables, Listings and Graphs (TLG) are described in a separate document.

Of note, this SAP does not cover the pharmacokinetic and metabolism profiling data analyses described in the study protocol. These analyses are covered in separate analysis plan.

The content and format of this SAP conforms to Regulatory Authority Guidance including ICH E9 and is consistent with proposed consensus guidelines for the content of statistical analysis plans in clinical trials (Gamble *et al*, 2017).

1.1. Study objectives

Objectives:
The objectives of this study are to evaluate the efficacy and safety of 3 doses of S201086/GLPG1972 compared to placebo in patients with knee osteoarthritis (OA).

The primary objective of the study is to demonstrate the efficacy of at least one dose (among 3 doses) of S201086/GLPG1972 compared to placebo after 52 weeks of treatment in reducing cartilage loss measured by cartilage thickness using quantitative MRI (qMRI) of the central medial tibiofemoral compartment (cMTFC) of the target knee.

The secondary objectives are:
To assess the safety and tolerability of 3 doses of S201086/GLPG1972.

To assess the efficacy of 3 doses of S201086/GLPG1972 *versus* placebo after 52 weeks of treatment on:

- The proportion of structural progressors, defined as patients who had at least 8% cartilage loss in cMTFC between baseline and W052.
- Pain, function, and stiffness measured with WOMAC.
- Pain measured with a 100-mm visual analog scale (VAS).
- Patient global assessment (PGA) of disease activity measured with 100-mm VAS.
- Reduction of cartilage loss measured by cartilage thickness using qMRI of the total tibiofemoral compartment (tTFC) of the target knee.
- JSW measured by x-ray.

To assess the efficacy of 3 doses of S201086/GLPG1972 *versus* placebo after 28 and 52 weeks of treatment on bone area using qMRI of the medial femoral condyle surface of the target knee.

To assess the pharmacokinetics of S201086/GLPG1972 (and metabolite[s] if applicable).

To assess the efficacy of 3 doses of S201086/GLPG1972 *versus* placebo after 52 weeks of treatment on analgesic consumption.

The exploratory objectives are:
[redacted].

[redacted]

1.2. Study design

The study CL2-201086-002/GLPG1972-CL-201 is a phase II, international, multi-regional (Asia *versus* Rest of World), multicenter, randomised, double-blind, parallel-group (doses: 75, 150 and 300 mg/day of S201086/GLPG1972 and matching placebo), placebo-controlled, dose-ranging study of 52 weeks.

1.2.1. Study plan

The study is divided into the following periods:

- An up to 5-week **screening period** without study treatment from screening visit (ASSE) to inclusion visit (W000). The screening period (up to 5 weeks) allows enough time to obtain results from X-ray examination, electrocardiogram (ECG), laboratory examination, and to perform the baseline MRI (before IMP intake).
- A **double-blind treatment period** of 52 weeks (from W000 to W052 visit). Eligible patients are included and randomly assigned to receive 75 mg/day, 150 mg/day or 300 mg/day S201086/GLPG1972, or matching placebo on a 1:1:1:1 ratio.
- A 2-week **follow-up period (WEND)** (from W052 or prematurely withdrawn to WEND): each patient must have a study end visit 2 weeks after completed or discontinued the study (definitely stopping study treatment), unless the patient withdraws consent.

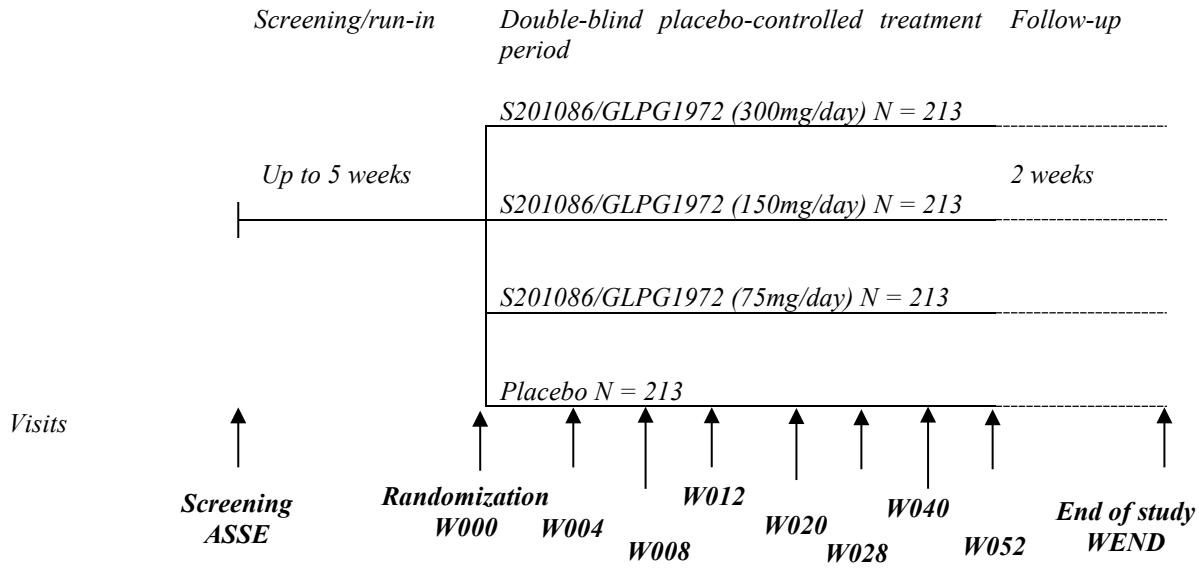
IMPs are supplied as 4 tablets orally once a day with a glass of water preferably in the morning (at the same time), corresponding to:

- S201086/GLPG1972 75 mg/day: 1 tablet of 75 mg + 3 matching tablets of placebo.
- S201086/GLPG1972 150 mg/day: 2 tablets of 75 mg + 2 matching tablets of placebo.
- S201086/GLPG1972 300 mg/day: 4 tablets of 75 mg.
- 4 matching tablets of placebo.

The appearance and taste of the tablets are the same for all study drugs throughout the study. The first IMP is taken at the site, during the W000 visit.

The study plan is shown in Figure (1.2.1) 1.

Figure (1.2.1) 1 - Study plan



1.2.2. Type of randomisation

The treatment randomisation and allocation are centralised using an Interactive Web Response System (IWRS) procedure. The treatment (S201086/GLPG1972 75 mg/day, S201086/GLPG1972 150 mg/day, S201086/GLPG1972 300 mg/day or placebo) is assigned at inclusion visit by a balanced (1:1:1:1), non-adaptive randomisation with stratification by geographic zone (Japan, South Korea & Taiwan and Rest of the World).

1.3. Determination of sample size

A total of 852 patients (213 per treatment group with a randomisation 1:1:1:1) will provide a minimal power of 70% to conclude that at least one dose of S201086/GLPG1972 is superior to placebo, assuming an ANCOVA model adjusted for multiple-testing by a Dunnett procedure, to detect a treatment effect on the primary efficacy endpoint of $\Delta = 0.0825$ mm (SD = 0.30 mm) at a 5% level (two-sided). Treatment effect parameters were derived from Lohmander *et al*, 2014.

2. ANALYSIS SETS / TREATMENT GROUPS

2.1. Analysis sets

- **Modified Randomised Set (mRS):**
  The modified Randomised Set (mRS) will be constituted of all included patients to whom a therapeutic unit was randomly assigned using IWRS. The mRS will be used for efficacy analyses. Patients will be analysed according to the randomised treatment.
- **Safety Set (SS):**
  The Safety Set (SS) will include all patients having taken at least one dose of IMP. The SS will be used for safety analyses. Patients will be analysed according to treatment actually received at inclusion.

[redacted]

The size of each analysis set, and reasons for exclusion will be described.
Listings of patients with their membership, or not, in each analysis set and of excluded patients with reasons for exclusion will be provided.

2.2. Treatment groups

The treatment groups will be:
- S201086/GLPG1972 75 mg.
- S201086/GLPG1972 150 mg.
- S201086/GLPG1972 300 mg.
- Placebo.

3. STATISTICAL METHODS

3.1. General considerations

3.1.1. Descriptive statistics

Descriptive statistics will be provided by treatment group, and all treatment groups pooled (for analyses on disposition of patients, baseline characteristics and patient follow-up, and on AEs).

For **discrete data**, number of observed values, number and percentage of patients per class will be presented. Unless otherwise specified in the TLG, no class "Missing" is considered.

For **continuous data**, number of observed values, mean, standard deviation, standard error of the mean, median, first and third quartiles, minimum and maximum will be presented.

3.1.2. General definitions

Unless specified otherwise in Sections 3.2 to 3.7, the following definitions will be considered:

- **Analysable value** will be defined as any non-missing value.
- **Baseline value** will be defined as the last analysable value prior to the first IMP intake (*i.e.* before or the same date as the first IMP intake date).

Note: In case of patient included and/or randomized but not treated (*i.e.* patients with treatment duration equal to 0): value at baseline is defined as the last analysable value prior or equal to date of inclusion visit.

- Post-baseline value will be defined as any value recorded at a given time point after baseline.
- Change from baseline will be defined as the arithmetic difference between a post-baseline value and the baseline value for a given variable at a given time point.
- Relative change from baseline will be defined as 100 * change from baseline / baseline value.

For safety endpoints (except adverse events), the following definition will be applied:

- A value is considered emergent under treatment if the assessment date is between the first IMP intake (excluded) and the last visit of the patient.

3.2. Disposition and baseline characteristics

Disposition of patients, main reasons for exclusion, patient follow-up and baseline characteristics will be described by treatment group, to assess their comparability, and overall.

Details concerning definitions on disposition of patients and baseline characteristics are provided in Appendix 5.2.

3.2.1. Disposition of patients

Disposition of patients, including reasons for withdrawal, will be summarized during the study, overall and by visit, in the mRS.

In order to assess the drop-out pattern between the treatment groups, the time to study discontinuation / time to premature IMP withdrawal will be described, in the mRS, using a Kaplan-Meier analysis.

Withdrawn and completed patients' characteristics will be described, in the mRS, and their comparability at inclusion visits assessed.

In addition, listings of adverse events leading to IMP withdrawal will be provided in the mRS for patients withdrawn for adverse events as well as for patients withdrawn but not for adverse events.

A listing of non-included randomised patients will be provided.

COVID-19 related analysis:
The study discontinuation / the premature IMP withdrawal related to COVID-19 will be summarized by visit.

3.2.2. Protocol deviations

Important protocol deviations before or at inclusion, as well as after inclusion, will be described in the mRS, by category of important deviations (based on ICH E3 guideline and ICH E3 Q&A).

COVID-19 related analyses:
A listing by subject and visit, giving an overview of the protocol deviations due to COVID-19 will be provided. This listing will be composed of:

- Subjects number.
- Visit (even if missing).
- Visit status (Remotely / Delayed).
- Missing assessment.
- Assessment done on paper (instead of ePRO).
- Assessments delayed.
- Withdrawal visit.

3.2.3. Demographic data and other baseline characteristics

Demographic data and other baseline characteristics such as life habits, history of the knee OA disease, baseline value of efficacy endpoints, vital signs, ECG parameters and serological status at baseline will be described in the mRS.

[redacted].

It is of note that QTcF interval at baseline will only be described in classes (≤ 450,]450 ; 480],]480 ; 500] and > 500ms, in accordance with ICH E14 guideline) and that the following continuous data will also be described in classes:

- Age ([40, 55[, [55, 65[, ≥ 65 years and < 65, ≥ 65 years (EudraCT analysis)).
- Body Mass Index (kg/m²) (< 18.5 / [18.5-25[/ [25-30[/ ≥ 30).
- Time since first diagnostic and Disease duration (years) (< 1 / [1-5[/ [5-10[/ [10-20[/ ≥ 20).

Moreover all previous treatments regardless of the disease specificity and previous treatments specific of knee OA disease will be described in the mRS, by anatomical therapeutic chemical (ATC) code, on-going non-chemical treatment at inclusion will also be described in the mRS (treatment without ATC code); as well as all medical history other than knee OA disease and surgical or medical procedures history, by MedDRA primary system organ class (SOC) and preferred term (PT).

3.3. Treatments of patients

Details concerning definitions on extent of exposure, treatment compliance and concomitant treatments are provided in Appendix 5.2.

3.3.1. Extent of exposure and treatment compliance

Extent of exposure (treatment duration (in weeks)) and treatment compliance (%) will be described in the mRS and the SS.

It is of note that treatment compliance (%) will also be described in classes (< 80, [80 ; 120] and > 120%).

3.3.2. Concomitant treatments

All OA specific (resp. non-specific) concomitant treatments taken at inclusion, during the treatment period will be described in the mRS, by ATC code.

3.4. Efficacy analysis

All efficacy analyses will be performed on the mRS. Patients will be analysed according to the randomised treatment.

General definitions are provided in Section 3.1.2. *Details concerning efficacy endpoints definitions are provided in Appendix 5.2.*

3.4.1. Statistical hypotheses

For the comparison of each S201086/GLPG1972 dose regimen to placebo, the following null hypothesis, associated to the main analysis, will be tested associated:

$H_0: \mu_T = \mu_P \Leftrightarrow \mu_T - \mu_P = 0$ (no difference between S201086/GLPG1972 and placebo)

versus

$H_1: \mu_T \neq \mu_P \Leftrightarrow \mu_T - \mu_P \neq 0$ (difference between S201086/GLPG1972 and placebo)

Where μ_P and μ_T are the true adjusted means of the change from baseline in cartilage thickness in the central Medial TibioFemoral Compartment (cMTFC), assessed by qMRI on the target knee (central reading), to W052 (primary endpoint) under placebo and S201086/GLPG1972, respectively.

The type I error will be set at α = 5% (two-sided setting), which is consistent with the objective of demonstrating the efficacy of at least one dose (among 3 doses) of S201086/GLPG1972 compared to placebo.

3.4.2. Primary efficacy endpoint

Definition: The primary efficacy endpoint is defined as the change from baseline to W052 in cartilage thickness in the cMTFC assessed by qMRI on the target knee (central reading).

3.4.2.1. Primary analysis

Primary analysis: In order to meet the primary objective of the study, the efficacy of at least one dose of S201086/GLPG1972 as compared to placebo after 52 weeks of treatment in reducing cartilage loss in patients with knee OA will be assessed through the change from baseline to W052 in cartilage thickness as measured in the cMTFC on the target knee, in patients of the mRS. A restricted maximum likelihood (REML)-based, mixed-effects model for repeated measures approach (so called Mixed-effects Model for Repeated Measures – MMRM) using all longitudinal observations at each post-baseline visit (W028) (Mallinckrodt *et al*, 2013) will be used (main analysis).

The MMRM as a primary analysis will assume that patients would keep the benefit of the randomized treatment after study discontinuation (Missing At Random (MAR) hypothesis).

Analysis will include the fixed, categorical effects of treatment, region (Asia and Rest of the World), time and treatment-by-time interaction, as well as the continuous, fixed covariates of baseline and time-by-baseline interaction.

MMRM: Y = TREATMENT REGION BASELINE TIME TREATMENT x TIME BASELINE x TIME.

An unstructured (co)variance structure will be used to model the within-patient errors. If this analysis fails to converge, the following structures will be tested in the following order: Heterogeneous Toeplitz, Heterogeneous compound symmetry, Toeplitz then Compound symmetry. The first (co)variance structure yielding convergence will be used as the primary analysis.

The Kenward-Roger approximation will be used to estimate denominator degrees of freedom.

The assumptions underlying the model, as for instance, the normality and homoscedasticity of residuals and detection of outliers, will be checked. If necessary, sensitivity analyses other than those planned in the SAP could be carried out, in the framework of the validation of the assumptions underlying the model.

In the primary analysis, a value is defined as under treatment if the assessment date is before the Last IMP intake + 4 weeks (included)

The difference between treatment groups will be calculated as each S201086/GLPG1972 dose regimen minus placebo. Thus, considering the main analysis, a positive treatment difference will be in favour of S201086/GLPG1972 considered dose regimen.

Missing data handling: For patients for which there will be no post-baseline measurement of the primary endpoint (regardless the timing of discontinuation), as they cannot be considered through the MMRM, a Multiple Imputation (MI) procedure will be used to impute the missing evaluations, as a prior step, assuming that those patients would be in their randomized arm. A total of 100 imputed partially-complete data sets will be generated.

MI inference involves 3 consecutives phases:
1/ Imputation step:
The missing pattern is supposed to be monotone, so the regression method will be used to impute missing data. This will be performed under MAR hypothesis, by treatment group, using the region factor and taking into account the baseline score of the primary endpoint, based on patient which every post-baseline measurement available. It is of note that this imputation step might be preceded by one MI approach based on MCMC method, in case of arbitrary missing pattern. A total of 100 imputed partially-complete datasets will be generated.

2/ Analysis step:
The same analysis as described above will be applied to each of the 100 imputed datasets.
3/Combination step:
Statistical inferences will be generated by combining results from 100 analyses using Rubin's formulae (Rubin, 1987).
The multiple imputation estimator of the difference between each S201086/GLPG1972 dose and placebo is the average of the individual 100 estimators. The variance of the estimator is the combination of the between and within-imputation variability (Kenward & Carpenter, 2007).

It is of note that, for patients with only one missing post-baseline measurement and a monotone pattern, missing data will not be imputed. Those missing measurements will be handled through MMRM.

The multiple imputation approach relies on the MAR assumption. As we can never exclude the possibility for a not-missing-at-random (NMAR) missingness mechanism, sensitivity analyses to explore the impact of non-ignorable missingness on the primary efficacy analysis will be conducted (see sensitivity analyses defined in Section 3.4.2.2).

Multiplicity issues: In order to take into account the multiplicity of comparisons associated to the primary objective of the study (demonstration the efficacy of at least one S201086/GLPG1972 dose as compared to placebo on the primary efficacy endpoint), the Dunnett procedure will be applied to control the family-wise error rate.
The principle of this procedure consists in comparing each test statistic resulting from MMRM (corresponding to each comparison) to a critical value in the Dunnett distribution with $(m + 1)(n - 1)$ degrees of freedom (m corresponding to the number of comparisons *versus* placebo and n corresponding to the number of patients). The principle of the Dunnett procedure consists in examining the ordered test statistics starting with the most significant one. Here, in the case of 3 comparisons:

- If the most significant test statistic $t(3) \leq c3$ (Dunnett critical value for 3 comparisons) then both three null hypotheses will not be rejected and every p-values associated with 3 comparisons will be reported.
- Otherwise the hypothesis $H_0(3)$ corresponding to t(3) will be rejected and the next most significant test statistic t(2) will be compared to c2 (Dunnett critical value for 2 comparisons). If $t(2) \leq c2$, $H_0(2)$ and $H_0(1)$ will not be rejected and p-values associated to $H_0(2)$ and $H_0(1)$ with 2 comparisons will be reported.
- Otherwise the hypothesis $H_0(2)$ corresponding to t(2) will be rejected and the last test statistic t(1) will be compared to c1 (Dunnett critical value for 1 comparison). If $t(1) \leq c1$, $H_0(1)$ will not be rejected.
- Otherwise $H_0(1)$ will be rejected.

Statistical elements: For the comparison of each dose of S201086/GLPG1972 *versus* placebo, the following elements will be provided in a summary table:

- Estimate (standard error) of the difference between adjusted treatment group means.
- Two-sided 95% CI of the estimate.
- Two-sided adjusted p-value taking into account Dunnett procedure (to be compared to 0.05).

The consistency of the results between the Asian-region population and the non-Asian-region population (respectively between the Japanese-region population and the non-Japanese-region population) will be evaluated on primary endpoint, according to the Method 2 defined in Ministry of Health Labor and Welfare Notification (MHLW) Notification "Basic principles on Global Clinical Trials". Treatment effect estimates and confidence intervals will be provided, for each dose, in Asian-region population and non-Asian-region population (respectively in Japanese-region population and non-Japanese-region population). In case of a statistically significant overall treatment effect (in favor of S201086/GLPG1972) at a considered dose, the results will be considered consistent if the observed treatment effects in Asian-region and non-Asian-region (respectively in Japanese-region population and non-Japanese-region population) patients are in favour of S201086/GLPG1972.

3.4.2.2. Sensitivity analyses

To assess the robustness of the primary analysis results to the handling of missing data method, the following sensitivity analyses will be performed on the mRS:

- *Multiple Imputation:*

A MI approach based on the regression method will be used to impute values for each missing data at W028 and W052: A total of 100 imputed complete data sets will be generated. This will be performed under MAR hypothesis, by treatment group, using the region factor, the OARSI JSN score at baseline, the gender and the age in class (< 65, ≥ 65) and taking into account baseline score of the primary.
This imputation step might be preceded by one MI approach based on MCMC method by treatment groups, in case the initial dataset has not a strict monotone missing pattern.

Then, for each of the 100 complete data sets obtained, each S201086/GLPG1972 dose will be compared to placebo in the mRS, on the primary endpoint, and using one single two-way analysis of covariance (ANCOVA) models on the factors treatment and region with baseline as covariate and no interaction.

ANCOVA model Y = TREATMENT REGION BASELINE

Finally, the following elements, combining the results from the 100 complete data sets and associated to every comparison of each S201086/GLPG1972 dose *versus* placebo on the primary endpoint, using the same Dunnett procedure as for the main analysis strategy, will be provided in a summary table:

- Estimate (standard error) of the difference between adjusted treatment group means.
- Two-sided 95% CI of the estimate.
- Two-sided adjusted p-value taking into account Dunnett procedure (to be compared to 0.05).

- *Pattern Mixture model placebo-based imputation:*

This method assumes that after discontinuation, patients discontinued from the S201086/GLPG1972 group will exhibit an evolution of the disease similar to patients in the placebo group (MNAR hypothesis). The implementation will be performed according to (Ratitch & O'Kelly, 2011).
In order to implement the control-based pattern imputation, the imputation process will be broken into several sequences of MI, each sequence is intended to impute missing values at one time-point t only (W028 and W052). The input dataset should include all patients in placebo group, and only patients in S201086/GLPG1972 group that have missing values at time-point *t*. Thus, imputation model will be estimated using patients in placebo group only.

A MI approach based on the regression method will be used to impute one value for each missing data at W028 and W052: a total of 100 imputed complete data sets will be generated. This will be performed using the region factor, the OARSI JSN score at baseline, the gender and the age in class (< 65, ≥ 65) and taking into account baseline score of the primary endpoint.
This imputation step might be preceded by one MI approach based on MCMC method by treatment groups, in case the initial dataset has not a strict monotone missing pattern.

Then, for each of the 100 complete data sets obtained, each S201086/GLPG1972 dose will be compared to placebo in the mRS, on the primary endpoint, and using one single two-way analysis of covariance (ANCOVA) models on the factors treatment and region with baseline as covariate and no interaction.

ANCOVA model Y = TREATMENT REGION BASELINE
Finally, the following elements, combining the results from the 100 complete data sets and associated to every comparison of each S201086/GLPG1972 dose *versus* placebo on the primary endpoint, using the same Dunnett procedure as for the main analysis strategy, will be provided in a summary table:

- Estimate (standard error) of the difference between adjusted treatment group means.
- Two-sided 95% CI of the estimate.
- Two-sided adjusted p-value taking into account Dunnett procedure (to be compared to 0.05).

- *Tipping point method:*

This method firstly relies on a regression approach under MAR assumption. This approach will be used to generate complete datasets. Several scenarios will be considered, each one consisting in a different penalized Multiple Imputation by treatment arm.
In order to implement the tipping point method, the imputation process will be broken into several sequences of MI, each sequence is intended to impute missing values at one time-point t only (W028 and W052).
A MI approach based on the regression method will be used to impute values for each missing data at W028 and W052: A total of 100 imputed complete data sets will be generated. This will be performed under MAR hypothesis, by treatment group, using the region factor, the OARSI JSN score at baseline, the gender and the age in class (< 65, ≥ 65) and taking into account baseline score of the primary. By subject, the first imputed values will be penalized under active treatment.

- *Subjects randomized under each S201086/GLPG1972 dose*: the first missing value of each subject will be worsened by δ_T (where δ_T = 0.00 to0.20, in step of 0.02).

The different scenarios will be constituted of each value of δ_T.
This imputation step might be preceded by one MI approach based on MCMC method by treatment groups, in case the initial dataset has not a strict monotone missing pattern.

Then, for each of the 100 complete data sets obtained by scenario, each S201086/GLPG1972 dose will be compared to placebo in the mRS, on the primary endpoint, and using one single two-way analysis of covariance (ANCOVA) models on the factors treatment and region with baseline as covariate and no interaction.

ANCOVA model Y = TREATMENT REGION BASELINE
Finally, the following elements, combining the results from the 100 complete data sets for each scenario and associated to every comparison of each S201086/GLPG1972 dose *versus* placebo on the primary endpoint, using the same Dunnett procedure as for the main analysis strategy, will be provided in a summary table:

- Estimate (standard error) of the difference between adjusted treatment group means.
- Two-sided 95% CI of the estimate.

Two-sided adjusted p-value taking into account Dunnett procedure (to be compared to 0.05).
All the scenarios will be separately analysed to explore under which condition the null hypothesis can no longer having be rejected.

- *Observed Cases (OC) analysis:*

Each dose of S201086/GLPG1972 will be compared to placebo in patients from mRS having a value of the primary endpoint at W052, using a single two-way ANCOVA model on the primary endpoint.

Analysis will include the fixed, categorical effects of treatment and region, as well as the continuous, fixed covariate of baseline and no interaction.

ANCOVA model Y = TREATMENT REGION BASELINE
Finally, the following elements, associated to every comparison of each S201086/GLPG1972 dose *versus* placebo on the primary endpoint, using the same Dunnett procedure as for the main analysis strategy, will be provided in a summary table:

- Estimate (standard error) of the difference between adjusted treatment group means.
- Two-sided 95% CI of the estimate.
- Two-sided adjusted p-value taking into account Dunnett procedure (to be compared to 0.05).

COVID-19 related sensitivity analyses:
The same analyses as the primary analysis will be performed to assess the robustness of the results on the definition of an under-treatment MRI by using several definitions of an under-treatment MRI:

- Covid-19 sensitivity 1: The value is defined as under treatment if the assessment date is before the Last IMP intake + 7 days (included).
- Covid-19 sensitivity 2: The value is defined as under treatment if the assessment date is before the Last IMP intake + 8 weeks (included).
- Covid-19 sensitivity 3: The value is defined as under treatment if the assessment date is at a post baseline visit.

At the time of the Statistical Blind Review, the number of BORDERLINE MRI will be evaluated. In case of more than X% of BORDERLINE MRI, a sensitivity analysis will be planned without those cases.

3.4.2.3. Supplementary analyses

For each treatment group, descriptive statistics will be provided for the primary endpoint (in terms of value at each visit and change from baseline to each post-baseline visit), overall and by regions, in patients of the mRS.

3.4.3. Secondary efficacy endpoints

Definitions: The secondary efficacy endpoints are:

- **The proportion of structural progressors at W052**, defined as patients who had at least 8% cartilage loss in cMTFC between baseline and W052.
- **The change from baseline to W052 in WOMAC** (Western Ontario and McMaster Universities Osteoarthritis Index) **total score and subscales scores for pain, function and stiffness**.
- **The change from baseline to W052 in pain in the target knee** measured with a 100-mm VAS.
- **The change from baseline to W052 in patient global assessment (PGA)** of disease activity measured with 100-mm VAS.
- **The proportion of OMERACT-OARSI responders at W052**, defined according to WOMAC and PGA as patients who had a high improvement in pain or in function ≥ 50% and absolute change ≥ 20 or, moderate improvement in at least 2 of the 3 following (Pham *et al*, 2004) :
  - Pain ≥ 20% and absolute change ≥ 10.
  - Function ≥ 20% and absolute change ≥ 10.
  - Patient's global assessment ≥ 20% and absolute change ≥ 10.
- **The change from baseline to W052 in cartilage thickness of the total tibiofemoral compartment of the target** knee using qMRI (centralized reading).
- **The change from baseline to W028 and to W052 in bone area of the medial femoral condyle surface** of the target knee using qMRI (centralized reading).
- **The change from baseline to W052 in JSW** of the target knee measured by x-ray (centralized reading).
- **Pain: Analgesic consumption:** at every visit up to W052.

Secondary analyses:

For the proportion of structural progressors at W052, the difference between each S201086/GLPG1972 dose and placebo will be studied in patients of the mRS using a logistic model, including the fixed, categorical effects of treatment, regions (Asia and Rest of the World), as well as the continuous, fixed covariates of cartilage thickness in the cMTFC at baseline. For handling all missing data, a MI on continuous data will be considered taking into account the same variables as the MI sensitivity analysis (Section 3.4.2.2). The definition of each discrete endpoint will then be applied to each generated dataset.

LOGISTIC model Y = TREATMENT REGION BASELINE

Finally, the following elements, combining the results from all the complete imputed data sets and associated to every comparison of each S201086/GLPG1972 dose *versus* placebo on the considered secondary endpoint, using the same Dunnett procedure as for the main analysis strategy, will be provided in a summary table:

- Estimate (standard error) of the adjusted odds ratio between treatment groups.
- Two-sided 95% CI of the estimate.
- Two-sided adjusted p-value taking into account Dunnett procedure (to be compared to 0.05).

For the proportion of OMERACT-OARSI at W040 and at W052, the difference between each S201086/GLPG1972 dose and placebo will be studied in patients of the mRS using a logistic model, including the fixed, categorical effects of treatment, regions (Asia and Rest of the World). For handling all missing data, a MI on continuous data will be considered taking into account the same variables as the MI sensitivity analysis (Section 3.4.2.2). The definition of each discrete endpoint will then be applied to each generated dataset.

LOGISTIC model Y = TREATMENT REGION

Finally, the following elements, combining the results from all the complete imputed data sets and associated to every comparison of each S201086/GLPG1972 dose *versus* placebo on the considered secondary endpoint, using the same Dunnett procedure as for the main analysis strategy, will be provided in a summary table:

- Estimate (standard error) of the adjusted odds ratio between treatment groups.
- Two-sided 95% CI of the estimate.
- Two-sided adjusted p-value taking into account Dunnett procedure (to be compared to 0.05).

For the change from baseline to W052 in JSW and the change from baseline to W028 in bone area, the difference between each S201086/GLPG1972 dose and placebo will be studied in patients of the mRS at W052 (respectively W028), using an ANCOVA. Analysis will include the fixed, categorical effects of treatment, regions (Asia and Rest of the World), as well as the continuous, fixed covariate of baseline. For handling all missing data at W052 (respectively at W028), a MI will be considered taking into account the same variables as the MI sensitivity analysis (Section 3.4.2.2).

ANCOVA model Y = TREATMENT REGION BASELINE

Finally, the following elements, combining the results from all the complete imputed data sets and associated to every comparison of each S201086/GLPG1972 dose *versus* placebo on the considered secondary endpoint, using the same Dunnett procedure as for the main analysis strategy, will be provided in a summary table:

- Estimate (standard error) of the difference between adjusted treatment group means.
- Two-sided 95% CI of the estimate.
- Two-sided adjusted p-value taking into account Dunnett procedure (to be compared to 0.05).

For other continuous secondary efficacy endpoints:

- The change from baseline to W052 in WOMAC total score and subscales scores of the target knee for pain, function and stiffness.
- The change from baseline to W052 in pain of the target knee.
- The change from baseline to W052 in PGA of disease activity of the target knee.
- The change from baseline to W052 in cartilage thickness of the total tibiofemoral compartment of the target knee.
- The change from baseline to W052 in bone area of the medial femoral condyle surface of the target knee.

The difference between each S201086/GLPG1972 dose and placebo (or between placebo and each S201086/GLPG1972 for WOMAC total score, subscales scores of pain, function and stiffness and VAS pain) will be studied in patients of the mRS, with the same strategy as the main analysis of the primary endpoint: MI for patients without any post-baseline value followed by a MMRM using all the longitudinal observations at each post-baseline visit (Section 3.4.2.1). Analysis will include the fixed, categorical effects of treatment, regions (Asia and Rest of the World), visit and treatment-by-visit interaction, as well as the continuous, fixed covariates of baseline, visit-by-baseline interaction.

MMRM model Y = TREATMENT REGION BASELINE TIME TREATMENT x TIME BASELINE x TIME

Finally, the following elements, combining the results from all the complete imputed data sets and associated to every comparison of each S201086/GLPG1972 dose *versus* placebo on the considered secondary endpoint, using the same Dunnett procedure as for the main analysis strategy, will be provided in a summary table:

- Estimate (standard error) of the difference between adjusted treatment group means.
- Two-sided 95% CI of the estimate.
- Two-sided adjusted p-value taking into account Dunnett procedure (to be compared to 0.05).

COVID-19 related sensitivity analyses:

For clinical endpoints (WOMAC total and subscores, PGA, VAS, OMERACT-OARSI), the same analyses as the secondary analyses will be performed to assess the robustness of the results on the collection of clinical endpoints:

- Covid-19 sensitivity: Inferential analyses will be provided as well at W040.

To be noticed, for clinical endpoints:

- For analyses at W040, all data up to W040 will be used in the analyses.
- For analyses at W052, all data up to W052, independently of the collection method (ePRO, paper) will be used in the analyses.

For protocol-permitted analgesic consumption, number and percentage of patients by treatment reported will be provided, overall and by treatment group.

For each treatment group, descriptive statistics will be provided for all secondary endpoints (in terms of value at each visit and change from baseline to each post-baseline visit for continuous endpoint), overall and by regions, in patients of the mRS.

3.4.4. Subgroups analysis

No formal subgroup analysis is planned for this study.

3.5. Exploratory analysis

[redacted]

3.6. Safety analysis

All safety analyses will be performed in the SS, by treatment group.

General definitions are provided in Section 3.1.2. Specific definitions are provided in Appendix 5.2 and provided in Specifications document.

3.6.1. Adverse events

Definition:

- Treatment Emergent Adverse Events (TEAE) are defined as all adverse events:
  - Which occur between the first IMP intake date (included) and the last visit of the patient,

    or
  - Which occur before the first IMP intake date and which worsen (in terms of intensity) or become serious according to the investigator opinion between the first IMP intake date (included) and the last visit of the patient.

Of note, in case of multiple information of the same event before the first IMP intake date, the information nearest to the first IMP intake date is taken into account.

- The seriousness and the relationship with the IMP of the adverse event are based on investigator opinion but sometimes also include sponsor decision to upgrade seriousness and/or relation to IMP (see TLG for details).

Analysis:
Number of events, number and percentage of patients reporting at least one event, presented by primary system organ class, and/or preferred term (depending on the analysis), will be provided for:

- Serious Adverse Events (SAEs), SAEs related to the study drug, SAEs leading to death and SAEs leading to death and related to the study drug over the study period according to the investigator or sponsor opinion.
- TEAE, TEAE leading to IMP withdrawal, TEAE requiring new treatment or increase of on-going treatment, TEAE requiring surgical or medical procedure, TEAE related to IMP, serious TEAE, severe TEAE, non-serious TEAE (EudraCT analysis) over the study period.

TEAE will be described according to the seriousness, the intensity, the relationship with the IMP, the action taken regarding the IMP, the requirement of added therapy, the outcome and the time to onset.

A listing of patients having experienced a non-serious TEAE not related to IMP and leading to IMP withdrawal will be provided.

In case of less than 20 events, listing of patients having experienced a non-fatal serious AE and listing of patients having experienced a non-serious AE (investigator opinion only) leading to IMP withdrawal will be provided.

COVID-19 related analysis:
A listing of patients reporting a coronavirus infection will be provided.

3.6.2. Clinical laboratory evaluation

Definition:

- A laboratory value is considered as analysable if non-missing and not flagged in the ClinTrial database as "not analysable".

Analysis:
For biochemical and haematological parameter, the following analyses will be performed, depending on the nature of considered endpoints (*i.e.*, discrete or continuous):

- Descriptive statistics on value at baseline, on value at each post-baseline visit under treatment, on last post-baseline value under treatment and on change from baseline to last post-baseline value under treatment.
- Number and percentage of patients with at least one high/low emergent abnormal value under treatment, according to the laboratory reference ranges and to the cut-offs for PCSA values.
- Number and percentage of patients with at least one emergent CTCAE value under treatment.
- Laboratory parameters classified (number and percentage of patients in each class) according to these reference ranges and cut-offs and using shift tables from baseline to the worst (high and/or low) values under treatment.
- Abnormalities and toxicities grades (according to the CTCAE grading) of the actual values using shift tables of the worst-case abnormality/toxicity grade versus the baseline abnormality/toxicity grade.

Moreover, listings of patients with out-of-range or PCSA analysable values emergent under treatment and of non-analysable values excluded from analyses will be provided.

To assess the potential of the drug to cause severe liver damage, possible Hy's Law cases will be identified. These subjects are defined as having any elevated AT (AST or ALT) of > 3×ULN, ALP < 2×ULN, and associated with an increase in bilirubin ≥ 2×ULN at one visit. Description of each component of the Hy's Law cases will also be provided.
This table will allow the assessment of drug-induced liver injury (DILI) based on the above parameters.

3.6.3. Vital signs, clinical examination and other observations related to safety

3.6.3.1. Vital signs and clinical examination

Definition:
The following vital signs and clinical examination will be analysed:
- Weight (kg).
- BMI (kg/m²).
- SBP (mmHg).
- DBP (mmHg).
- Heart Rate (bpm).

Analysis:
They will be described, in terms of value at baseline, value at each post-baseline visit under treatment as well as in terms of change from baseline to each post baseline visit under treatment and to last post-baseline value under treatment.

Number of emergent abnormal values, number and percentage of patients with at least one emergent abnormal value, based on SBP, DBP and Heart Rate will be provided.

Vital signs classified (number and percentage of patients in each class) according to the cut-offs for abnormal values and using shift tables from baseline to the worst (high and/or low) values under treatment will be provided.

3.6.3.2. Electrocardiogram

Definition:
- A value is considered as analysable if non-missing value obtained from an interpretable ECG.
- An ECG is considered as interpretable if the quality of the trace is "Correct", with "Minor problems" or "Missing data".

Analysis:
ECG parameters will be described, in terms of value at baseline, value at each post-baseline visit under treatment and last post-baseline value under treatment; as well as, for continuous endpoints, in terms of change from baseline to each post baseline visit under treatment and to last post-baseline value under treatment. Moreover values and changes form baseline of corrected QT interval will be described in classes, considering thresholds defined in IC E14 (*i.e.*, ≤ 450,]450 ; 480],]480 ; 500] and > 500 ms for values, and ≤ 30,]30 ; 60] and > 60 ms for changes). QT interval will be described in the overall SS, in the Men of the SS and in the Women of the SS. Of note, for the analysis on the Women of the SS, the thresholds will be adapted (*i.e.*, ≤ 470,]470 ; 480],]480 ; 500] and > 500 ms for values).

Two listings of participants having at least one post baseline abnormal clinically significant event that has been considered as well as a significant change with baseline will be provided. One will concern ECG values and changes from baseline at each visit (including unscheduled), and the other will concern ECG morphological abnormalities.

Number and percentage of patients having at least one emergent clinically significant ECG during the 52-week double-blind treatment period will be described. Number and percentage of patients having at least one clinically significant ECG abnormality, overall and by class and subclass of Minnesota codelist will also described, using value at baseline and at each post-baseline visit under treatment (except for WEND).

3.6.3.3. Special situations

The total number of special situations, their distribution per type and per association or not with an AE will be described in the SS. Moreover, special situations without AEs associated will be described in the SS.

3.7. Biomarkers analysis

[redacted]

4. INTERIM ANALYSIS

No interim analysis will be performed.

5. APPENDICES

5.1. General analytic definitions

Definitions below correspond to calculation rules for first and last IMP intake dates.

5.1.1. First and last IMP intake dates

For patients having taken at least one dose of IMP, the dates of first and last IMP intake on the analysis period will be defined as follows:

- The date of the first IMP intake at the first visit performed within the analysis period.
- The date of the last IMP intake at the last visit performed within the analysis period.

Note: Visits with both missing first and last IMP intake dates and with number of returned tablets equal to number of tablets delivered at the previous visit (or with estimated number of tablets taken equal to 0) will not be taken into account.

After selection of the dates of first and last IMP intake as defined above, if these dates are missing or incomplete, substitution rules will be applied to identify baseline value, values under treatment and emergent adverse events.

5.2. Specific analytic definitions and data handling conventions

5.2.1. Disposition and baseline characteristics

- **Disposition**

The **Time to study discontinuation** is defined, in weeks, as follows:
Time to study discontinuation (weeks) = (Last date - Randomisation date)/7.02
with the Last date defined as:

- The last available date, for patients lost to follow-up.
- The death date, for patients withdrawn for death.
- The withdrawal date, for patients withdrawn for other reasons.
- The date of the last visit, otherwise.

The following classes will be defined for time to study discontinuation: [0-4[weeks, [4- 8[weeks, [8-12[weeks, [12-20[weeks, [20-28[weeks, [28-40[weeks, [40-52[weeks and ≥ 52 weeks.

- **History of knee OA**

The **time since the first diagnostic** (years) is defined as:
(Date of screening visit (ASSE) – Date of initial diagnosis of knee OA + 1) / 365.25
The result is rounded to the nearest integer. If the day and/or month of date of diagnosis is missing, the calculation is done using the 1st day of the month or the 1st day of the year

The **disease duration** (years) is defined as:
(Date of screening visit (ASSE) – Date of diagnosis based on the clinical and radiological criteria of the American College of Rheumatology + 1) / 365.25
The result is rounded to the nearest integer. If the day and/or month of date of diagnosis is missing, the calculation is done using the 1st day of the month or the 1st day of the year

- **Medical history other than studied disease and surgical or medical procedures history**

The existence of a history (Yes/No) is defined from the presence, or not, of a Primary system organ class and/or Preferred term.

- **Previous treatments**

The previous treatments are the treatments for osteoarthritis disease in the past, or any previous treatments which could interfere with the IMP or the study assessments in the last 6 months before the screening.

Previous treatment is defined as any treatment with associated stop date strictly inferior to the first IMP intake date.

Note: In case of patient included and/or randomised but not treated (*i.e.* patients with treatment duration equal to 0), previous treatment is defined as any treatment with associated stop date strictly inferior to date of inclusion visit.

Only treatment with an Anatomical therapeutic chemical classification and/or a Preferred name is considered.

The lists of the considered ATC codes for specific previous treatments are the following:
- 3692.0 Corticosteroids for systemic use.
- 4174.0 Systemic Analgesics without NSAIDs, Glucosamine and Hyaluronic acid.
- 4175.0 Non-steroidal anti-inflammatory drugs (NSAIDs) without Glucosamine Hyaluronic acid and Chodroitin.
- 4176.0 Glucosamine.
- 4177. 0 Hyaluronic acid.
- 4178.0 Chondroitin.

Other ATC codes in the ClinTrial database are for non-specific previous treatments.

5.2.2. Treatments of patients

- **Extent of exposure and treatment compliance**

Global duration (weeks) is defined as:
(Date of the last visit - Date of screening + 1) / 7.02
The result is rounded to one decimal place.

The **treatment duration** (weeks) is defined as:
(Date of the last IMP intake (for the considered period) - Date of the first IMP intake (for the considered period) + 1) / 7.02
The result is rounded to one decimal place.

Notes:
- For patients with no dose of IMP (for the considered period according to the general note in Section 2.1), the duration is null.
- For patients with missing or incomplete date of first or/and last IMP intake (for the considered period) before substitution, the duration is not calculated.

The treatment exposure (weeks) is defined as:
(Treatment duration (for the considered period) - Overall duration of interruption (for the considered period)) / 7.02, with the overall duration of interruption defined as:
Sum (for the considered period) of (Date of IMP restarted - Date of last IMP intake before interruption - 1 (*)).

(*) or the number of days of interruption in case of missing or incomplete date of IMP restarted or date of last IMP intake before interruption.
The result is rounded to one decimal place.

The treatment compliance (%) is defined as:
(Sum of number of tablets taken / Sum of number of tablets to be taken) x 100 (for the considered period)
with:

- Number of tablets taken = Estimated number of tablets taken, or if not completed, Number of tablets dispensed - Number of tablets returned (for the considered period).
- Number of tablets to be taken = Number of tablets prescribed per day x (Last visit date - First visit date) (for the considered period).

Note: Compliance is not calculated in case of missing information.

- **Concomitant treatments**

The existence of a concomitant treatment (Yes/No) is defined from the presence, or not, of an Anatomical therapeutic chemical classification and/or Preferred name.

The **periods** considered **for the analysis** are:

- At inclusion for which treatments with start date ≤ inclusion date and stop date ≥ inclusion date or missing are taken into account.
- During the treatment period for which treatments:
  - With start date ≥ first IMP intake date and < last IMP intake date, or
  - With start date ≤ first IMP intake date and stop date ≥ first IMP intake date or missing

  are taken into account.

Concomitant treatments could be considered in one or several of the possible analysis periods.

The lists of the considered ATC codes for specific concomitant treatments are the following:

- 3692.0 Corticosteroids for systemic use.
- 4174.0 Systemic Analgesics without NSAIDs, Glucosamine and Hyaluronic acid.
- 4175.0 Non-steroidal anti-inflammatory drugs (NSAIDs) without Glucosamine Hyaluronic acid and Chodroitin.
- 4176.0 Glucosamine.
- 4177.0 Hyaluronic acid.
- 4178.0 Chondroitin.

5.2.3. Efficacy analysis

- **WOMAC score**

The WOMAC total score is available in the SDTM dataset, and corresponds to the sum of the items 1 to 24.
The WOMAC subscore of pain is calculated with the sum of the items 1 to 5 (subscale of pain). If one item is missing, the subscore is not calculated.
The WOMAC subscore of stiffness is calculated with the sum of the items 6 and 7 (subscale of stiffness). If one item is missing, the subscore is not calculated.
The WOMAC subscore of function is calculated with the sum of the items 8 to 24 (subscale of function). If one item is missing, the subscore is not calculated.
For each score, no decimal will be kept.

5.2.4. Exploratory analysis

No more information.

5.2.5. Safety analysis

5.2.5.1. Adverse events

Each **medical concept of adverse event coded according to the internal "multiple medical concept" process** is taken into account as a single adverse event in the statistical analysis. The modalities of the adverse event (onset and end dates, intensity, seriousness, action taken, additional therapy, relationship, outcome...) replicated by default to each medical concept are also taken into account in the statistical analyses.

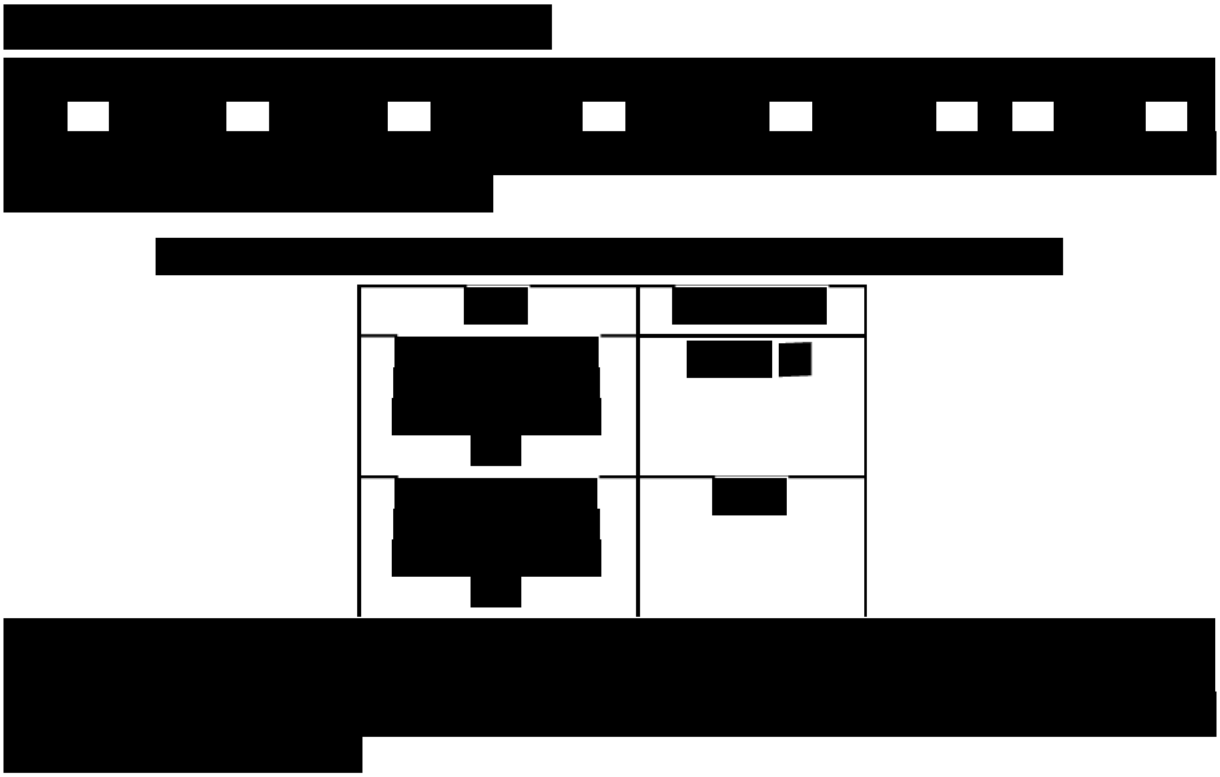
5.2.5.2. Clinical laboratory evaluation

See Specifications document.

5.2.5.3. Vital signs, clinical examination and other observations related to safety

See Specifications document.

5.2.6. [redacted]



5.2.6.2. Management of multiple samples

In case of multiple samples, the value at the visit corresponds to the last reliable value at the visit, based on sampling date and time.

5.3. Statistical methods details

No more information than in Section 3.4.2.

5.4. Software and programming codes

Statistical analyses will be performed using SAS®/PC Software version 9.4.

5.4.1. Multiple imputation using the regression method

- *Imputation step:*

```
PROC SORT data=<name_of_the_table>;
   by Treatment;
RUN;
```

- Step of producing monotone missing patern with MCMC

```
PROC MI data=<name_of_the_table> seed=2010861972 nimpute=100 out=out_mcmc noprint;
        mcmc impute=monotone;
        var Baseline W028 W052;
        by Treatment;
RUN;
```

- Step of imputing the monotone missing data

```
PROC SORT data=outmcmc;
        by _imputation_Treatment Region Patient;
RUN;

PROC MI data=<entry_table> seed=2010861972  nimpute=1 out=table_out noprint;
        class Region;
        monotone reg(Baseline W028 W052/details);
        var Region Baseline W028 W052;
        by _imputation_Treatment;
RUN;
```

- *Inference step:*

```
PROC SORT data=table_out;
   by _imputation_ Treatment Region Patient Timepoint;
RUN;
```

- Step for the estimate (standard error) of the difference and two-sided 95% CI of the estimate

```
ods output diffs=work.diff;
PROC MIXED data=table_out order=internal;
        class Treatment Region;
        model <name_of_the_variable Y> = Treatment Region Baseline / ddfm=kr;
        repeated Visit/type=UN  subject = Patient ;
        lsmeans Treatment*Visit / diff=control("4" "W052") alpha=0.05 cl;
        by _imputation_;
RUN;
```

- *Combination step:*

- Step for the estimate (standard error) of the difference and two-sided 95% CI of the estimate

```
PROC SORT data= diff;
        by Treatment;
RUN;

PROC MIANALYZE data= diff;
        modeleffects ESTIMATE;
        stderr STDERR;
        by Treatment;
RUN;
```

5.4.2. Multiplicity adjustment: Dunnett procedure

```
data adjust_pval;
set table_pval;
        pval3_max=1-probmc("dunnett2", abs(tvalue3),.,(3+1)*(233-1),3);
        pval3_int=1-probmc("dunnett2", abs(tvalue3),.,(2+1)*(233-1),2);
        pval3_min=1-probmc("dunnett2", abs(tvalue3),.,(1+1)*(233-1),1);
        pval2_max=1-probmc("dunnett2", abs(tvalue2),.,(3+1)*(233-1),3);
        pval2_int=1-probmc("dunnett2", abs(tvalue2),.,(2+1)*(233-1),2);
        pval2_min=1-probmc("dunnett2", abs(tvalue2),.,(1+1)*(233-1),1);
        pval1_max=1-probmc("dunnett2", abs(tvalue1),.,(3+1)*(233-1),3);
        pval1_int=1-probmc("dunnett2", abs(tvalue1),.,(2+1)*(233-1),2);
        pval1_min=1-probmc("dunnett2", abs(tvalue1),.,(1+1)*(233-1),1);
    if ((tvalue1 < tvalue2) and (tvalue2 < tvalue3)) then do;
        pval3_end = pval3_max;
        if pval3_max <0.05 then do;
                pval2_end = pval2_int;
                if pval2_int < 0.05 then do;
                        pval1_end = pval1_min;
                end;
                if pval2_int >= 0.05 then do;
                        pval1_end = pval1_int;
                end;
        end;
        if pval3_max >= 0.05 then do;
                pval2_end = pval2_max;
                pval1_end = pval1_max;
        end;
    end;
    if ((tvalue2 < tvalue1) and (tvalue1 < tvalue3)) then do;
        pval3_end = pval3_max;
        if pval3_max < 0.05 then do;
                pval1_end = pval1_int;
                if pval1_int < 0.05 then do;
                        pval2_end = pval2_min;
                end;
                if pval1_int >= 0.05 then do;
                        pval2_end = pval2_int;
```

```
                end;
        end;
        if pval3_max >= 0.05 then do;
                pval1_end = pval1_max;
                pval2_end = pval2_max;
        end;
    end;
    if ((tvalue1 < tvalue3) and (tvalue3 < tvalue2)) then do;
        pval2_end = pval2_max;
        if pval2_max < 0.05 then do;
                pval3_end = pval3_int;
                if pval3_int < 0.05 then do;
                        pval1_end = pval1_min;
                end;
                if pval3_int >= 0.05 then do;
                        pval1_end = pval1_int;
                end;
        end;
        if pval2_max >= 0.05 then do;
                pval3_end = pval3_max;
                pval1_end = pval1_max;
        end;
    end;
    if ((tvalue3 < tvalue1) and (tvalue1 < tvalue2)) then do;
        pval2_end = pval2_max;
        if pval2_max <0.05 then do;
                pval1_end = pval1_int;
                if pval1_int < 0.05 then do;
                        pval3_end = pval3_min;
                end;
                if pval1_int >= 0.05 then do;
                        pval3_end = pval3_int;
                end;
        end;
        if pval2_max >= 0.05 then do;
                pval1_end = pval1_max;
                pval3_end = pval3_max;
        end;
    end;
    if ((tvalue2 < tvalue3) and (tvalue3 < tvalue1)) then do;
        pval1_end = pval1_max;
        if pval1_max <0.05 then do;
                pval3_end = pval3_int;
                if pval3_int < 0.05 then do;
                        pval2_end = pval2_min;
                end;
                if pval3_int >= 0.05 then do;
                        pval2_end = pval2_int;
                end;
        end;
```

```
        if pval1_max >= 0.05 then do;
                pval3_end = pval3_max;
                pval2_end = pval2_max;
        end;
    end;
    if ((tvalue3 < tvalue2) and (tvalue2 < tvalue1)) then do;
        pval1_end = pval1_max;
        if pval1_max <0.05 then do;
                pval2_end = pval2_int;
                if pval2_int < 0.05 then do;
                        pval3_end = pval3_min;
                end;
                if pval2_int >= 0.05 then do;
                        pval3_end = pval3_int;
                end;
        end;
        if pval1_max >= 0.05 then do;
                pval2_end = pval2_max;
                pval3_end = pval3_max;
        end;
    end;
run;
```

5.5. COVID-19 Risk Assessment

Document title	**COVID-19 RISK ASSESSMENT**
Study official title	**Efficacy and safety of 3 doses of S201086/GLPG1972 administered orally once daily in patients with knee osteoarthritis. A 52-week international, multi-regional, multi-center, randomized, double-blind, placebo-controlled, dose-ranging study.** **ROCCELLA Study**
Study brief title	**Efficacy of S201086/GLPG1972 in patients with knee osteoarthritis**
Test drug code	**S201086/GLPG1972**
Indication	**Osteoarthritis**
Development phase	**Phase 2**
Protocol code	**CL2-201086-002/GLPG1972-CL-201**
EudraCT Number	**2017-004581-10**
Universal Trial Number	**U1111-1205-0321**
Investigational New Drug Application Number	**133039**
Sponsor	**GALAPAGOS NV (US)** **Institut de Recherches Internationales Servier (I.R.I.S.) (ex-US)**
Date of the document	**29 September 2020**
Version of the document	**Final version 2.0**

CONFIDENTIAL

Table of contents

1. INTRODUCTION

COVID-19 outbreak has rapidly evolved into a global pandemic. The impact of COVID-19 on patient journeys represents a new risk for trial results interpretation.
Some of its impacts are direct, *e.g.* infections and deaths. Others are indirect but still deeply concerning, *e.g.* increased demands on the health service, travel restrictions and measures of social distancing, leading to clinical site closures, treatment interruptions/discontinuations and delayed/missed trial visits.
Patients within a clinical trial are unequally concerned depending on situations:

- Patients having completed the study before COVID-19 pandemic.
- Patients still ongoing during COVID-19 pandemic.

According to the EMA "Guidance on the management of clinical trials during the COVID-19 (coronavirus) pandemic", Sponsors are advised to perform a risk assessment based on accumulating trial data to evaluate the implications on recruitment, loss of patients during the trial, ability to record data and ability to interpret the treatment effect.

ROCCELLA study is the first study on patients suffering from knee OsteoArthritis (kOA) evaluating the one-year structural effect of the S201086/GLPG1972 drug on 932 patients randomized and included.

In this study, Risk assessment was performed by the Sponsor on aggregated and blinded data and has been focused on quality and reliability of the data from a trial conduct perspective and has been considered in terms of impact on study conduct (e.g. treatment discontinuations), delay in assessments and missing data arising from the COVID-19 pandemic on the analysis and interpretation of the data. The ICH E9 (R1) addendum on estimands and sensitivity analyses provide a comprehensive approach to articulate this impact analysis.

For ROCCELLA study, the pandemic spread occurred when the study was fully recruited so it impacted mainly the end of patient follow-up.
In the study, as of 18 September 2020, at least 185 patients (around 20% of included and randomized patients) have at least one visit impacted by the COVID-19 pandemic with:

- around 1% of W040 visits impacted.
- around 18% of W052 visits impacted.
- around 20% of WEND (follow-up) visits impacted.

The data collection process during the COVID-19 pandemic, that allows this Risk assessment, is documented in the Data Management Plan.

All the indicators presented in this Risk Assessment are based on an almost final database (ADaM version of 18th September 2020), This allows to take adequate decision for this study.

2. COVID-19 IMPACT ASSESSMENT

2.1. Direct impact affecting trial participants

2.1.1. COVID-19 infection

COVID-19 infections and associated symptoms can have a direct impact on patient's safety, and participation in the considered study.

In ROCCELLA study 10 patients had reported a coronavirus infection during their follow-up (High Level Term based on MedDRA 23.0). Of those, 2 patients had reported a COVID-19 pneumonia, 6 had reported a COVID-19 infection and 2 had reported a Suspected COVID-19.

One patient having reported a COVID-19 infection died. This patient took [redacted] last IMP on the [redacted], developed COVID-19 symptoms on [redacted], was hospitalized on [redacted] and died on [redacted].

See Section 3.2.3 for the impacts in the SAP.

2.2. Indirect impact affecting clinical trial conduct

2.2.1. Recruitment impact

The pandemic had no impact on the recruitment as the study was fully recruited.

2.2.2. Patient retention

In ROCCELLA study less than 1% of included and randomized patients withdrew the study (at W052) with a withdrawal reason linked to the COVID-19 pandemic.

Around 6% of patients did not attend the WEND visit due to COVID-19 pandemic: the visit was either not done or, for the majority of these patients (around 67 %), the visit was done remotely.

2.2.3. Study treatment accessibility

In ROCCELLA study, IMP were dispensed at visits W000, W004, W008, W012, W020, W028 and W040. As previously mentioned, few patients had their W040 visit impacted by the COVID-19. For patients not able to come on site, IMP was sent directly to the patient.

No IMP interruption was reported/observed for visits impacted by COVID-19 pandemic except for 2 patients (1 AE "COVID-19 pneumonia" that led to temporarily IMP interruption and 1 patient who finished all the IMP dispensed before performing the W052 with some delay).

2.2.4. Standard of care modifications

The knee OsteoArthritis (kOA) standard of care is a combination of non-pharmacological and symptomatic pharmacological treatments which should be personalized to the need of the individual patient. Despite several investigations of potential therapies, there are currently no approved disease modifying drugs. Effective and safe long-term treatment for most OA patients is thus not available, creating a clear unmet medical need for structural protection (Karsdal *et al*, 2016).

GLPG1972/S201086 is a disease-modifying osteoarthritis drug candidate administered as a once daily oral dose to reduce cartilage loss in knee OA patients. ROCCELLA primary's objective is to demonstrate the efficacy of at least one dose as compared to placebo in reducing cartilage loss of the target knee (structural protection) in patients suffering from knee OA.

Since there is no standard of care for reducing structural progression, no modification could have occurred during Covid-19 pandemic. Thus, there is no impact on the evaluation of structural progression.

Some secondary objectives of ROCCELLA study are to assess the efficacy of the drug on pain and function (symptomatic effect). Given the SOC for KOA is based on symptomatic treatment, a potential impact of the pandemic on this treatment (e.g. interruption of physical therapy) cannot be ruled out. Also, symptom assessment may also have been impacted. However, this impact is difficult to quantify.

See Section 3.2.2.3 for the impacts in the SAP.

2.2.5. Life habits modifications

Patients who still were in the study during the pandemic may have been impacted by government or state department travel restrictions and sheltering, limiting their physical activity. This limitation may have had an impact on their symptom assessment hence on some of the secondary outcomes measured during the study. See Section 3.2.2.3 for the impacts in the SAP.
Moreover, as the cartilage loss is a slow process, it is considered that the daily movement decrease has no impact on the structural evaluation.

2.2.6. Study visits and endpoints measurements

Due to the COVID-19 pandemic, some visits/assessments of ongoing patients during the COVID-19 pandemic phase could have been impacted in several ways.
In ROCCELLA study these impacts are:

- An estimated number of 10% of W052 visits and around 4% of WEND visits were done remotely. During those visits, the ECG, laboratory samples, vital signs are generally not collected (for only 5 W052 visits and 2 WEND visits site staff visited the patient at home and some assessments were done whereas the visit was done remotely). The clinical endpoints (collected via an ePRO on site according to the clinical study protocol) are collected on paper at home and flagged in the clinical trial database.
- Around 25% of W052 visits done with a delay leading to protocol deviations as regards to the time-window allowed in the clinical study protocol. During those visits, the assessment planned according to the clinical study protocol, but with a delay leading to a protocol deviation.

Of note, in some cases, visits have been done as scheduled but, due to inaccessibility to imaging centers, a delay in some assessments (*e.g.* MRI, X-Ray…) could be observed. Moreover, in some cases, the visits could have been done remotely with some assessments done with a delay (*e.g.* MRI, X-Ray).

In a very few cases, some assessments have been performed at home by the site employees (*e.g.* vital signs). The remote and/or delayed visits could have an impact on the clinical and structural evaluation.

On the primary efficacy endpoint (MRI), the observed delays are as follows:

- Around 25% of W052 MRI have been performed more than 7 days after the last IMP intake.
- Around 20% of W052 MRI have been performed more than 4 weeks after the last IMP intake.
- Around 15% W052 MRI have been performed more than 8 weeks after the last IMP intake.

See Section 3.2.2.3 for the impacts in the SAP.

3. IMPLICATIONS AND MITIGATIONS

3.1. Sample size

The recruitment was completed in June 2019 around 8 months before the pandemic. 932 patients were randomized and included, 10% more than the planned sample size. From the operational perspective, it was not possible to restart the recruitment. No sample size reassessment was performed.

3.2. Statistical analysis

This section sets out the additions that will be made to the statistical analysis in the light of the impact presented above. These modifications will be implemented in the final version of the SAP.

3.2.1. Baseline characteristics and patient dispositions

As some patients withdrew for a reason related to the COVID-19 pandemic, a table summarizing the study premature withdrawal related to COVID-19 will be provided.

As some visits are impacted by the COVID-19 pandemic (*i.e.* done remotely or performed with some delay), a listing by subject and visit, giving an overview of the protocol deviations due to COVID-19 will be provided.

3.2.2. Efficacy analyses

3.2.2.1. Primary analysis update

The primary efficacy endpoint is defined as the change from baseline to W052 in cartilage thickness in the cMTFC assessed by qMRI on the target knee (central reading).

As planned in the protocol, the primary analysis is a MMRM that will include the fixed, categorical effects of treatment, region (Asia and Rest of the World), time and treatment-by-time interaction, as well as the continuous, fixed covariates of baseline and time-by-baseline interaction.

According to the clinical study protocol, no data was supposed to be collected after the end of study visit (WEND) which could occur up to 4 weeks after the W052 planned date. Indeed, the W052 could be done up to 7 days after the W052 planned date, and the WEND visit has to be done 2 weeks (+/- 7 days) after the W052 of premature withdrawal visit.

The only visit with MRI assessment planned that is impacted by the COVID-19 pandemic is the W052 visit. Due to this situation, some MRI assessments have been performed with delay according to the W052 planned visit date, from few days to more than 3 months. Thus, MRI assessment done too far away from the last IMP intake could not reflect the targeted assessment planned per protocol.

In order to deal with those delayed assessments due to the COVID-19 pandemic, it has been decided:

- To consider in the primary analysis MRI exams done not later than 4 weeks (included) after the last IMP intake
- To consider as missing in the primary analysis MRI exams done beyond 4 weeks after the last IMP intake. For those MRI, the same strategy of handling of missing data as in the primary analysis (planned in the protocol) will be applied.

3.2.2.2. Sensitivity analyses update

In order to assess the robustness of the primary analysis results to this time-window choice for MRI assessment, some sensitivity analyses will be performed:

- Covid-19 sensitivity analysis 1:
  - To consider in the primary analysis MRI done before 7 days (included) after the last IMP intake.
  - To consider as missing in the primary analysis MRI done after 7 days after the last IMP intake. For those MRI, the same strategy of handling of missing data as in the primary analysis will be applied.
- Covid-19 sensitivity analysis 2: The value is defined as under treatment if the assessment date is before the Last IMP intake + 8 weeks (included)
  - To consider in the primary analysis MRI done before 8 weeks after the last IMP intake.
  - To consider as missing in the primary analysis MRI done after 8 weeks after the last IMP intake. For those MRI, the same strategy of handling of missing data as in the primary analysis will be applied.
- Covid-19 sensitivity analysis 3:
  - To consider in the primary analysis all MRI done after the last IMP intake.

3.2.2.3. Secondary analyses update

For structural objectives of ROCCELLA study (endpoints derived from MRI assessment or from X-Ray assessment), the same strategy as the primary analysis will be applied for consistency.

For clinical objectives of ROCCELLA study, as previously mentioned data collected during COVID-19 pandemic (especially W052 data) could be impacted in several ways:

- Standard of care modifications.
- Paper questionnaires instead of ePRO questionnaires.
- Questionnaires filling at home instead of on-site.
- Possible delay for some questionnaires assessments.

All the planned secondary analyses linked to symptomatic evaluation will be kept as specified in the protocol (consideration of all data collected at W052 visit).

In order to evaluate the impact of considering such data in those analyses, analysis of those endpoint will be also performed at W040 (previous visit where symptomatic evaluations are collected and where only 1% of the visits have been impacted by COVID-19 pandemic).

3.2.3. Safety analyses

Due to the low number of patients infected (Section 2.1.1), a listing of all patients having a COVID-19 infection will be provided, with at least the CTCAE grading, the action taken regarding this infection and the study status linked to this infection (*e.g.* leading to withdrawal).

3.3. Interim analysis

Not applicable.

© I.R.I.S. / GALAPAGOS - 29 September 2020 - Confidential

4. REFERENCES

EMA - Guidance on the management of clinical trials during the covid-19 (coronavirus) pandemic - Version 3 28/04/2020

FDA - Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Public Health Emergency - Guidance for Industry, Investigators, and Institutional Review Boards - March 2020

ICH E9 - Statistical Principles for Clinical Trials - Adopted by CPMP, March 1998, issued as CPMP/ICH/363/96/step 5.

ICH E9 - Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1) Final version - Adopted on 20 November 2019

Karsdal MA, Michaelis M, Ladel C, *et al*. Disease-modifying treatments for osteoarthritis (DMOADs) of the knee and hip: lessons learned from failures and opportunities for the future. Osteoarthritis Cartilage. 2016;24(12):2013-2021. doi:10.1016/j.joca.2016.07.017

Points to consider on implications of Coronavirus disease (COVID-19) on methodological aspects of ongoing clinical trials - Adopted by CHMP 26 June 2020, issued as EMA/158330/2020 Rev. 1

6. REFERENCES

Gamble C, Krishan A, Stocken D, *et al*. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343. doi:10.1001/jama.2017.18556.

Kenward M & Carpenter J. (2007). Multiple imputation: current perspectives (Vol. 16(3)). Statistical Methods in Medical Research.

ICH E3 - Structure and Content of Clinical Study Reports – Adopted by CPMP, December 1995, issued as CPMP/ICH/137/95/step 5.

ICH E3 Q&A - Structure and Content of Clinical Study Reports - Questions & Answers (R1) - July 2012, issued as EMA/CHMP/ICH/435606/2012.

ICH E9 - Statistical Principles for Clinical Trials - Adopted by CPMP, March 1998, issued as CPMP/ICH/363/96/step 5.

ICH E14 – The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic drugs – Adopted by CHMP, May 2005, issued as CHMP/ICH/2/04.

Lohmander L. Stefan, Scarlett Hellot, Don Dreher, Eduard F. W. Krantz, Dawie S. Kruger Ali Guermazi Felix Eckstein (2014). Intraarticular Sprifermin (Recombinant Human Fibroblast Growth Factor 18) in Knee Osteoarthritis: A Randomized, Double-Blind, Placebo-Controlled Trial, Arthritis & Rheumatology, https://doi.org/10.1002/art.38614.

Mallinckrodt C, Lin Q, & Molenberghs M. (2013). A structured framework for assessing sensitivity to missing data assumptions in longitudinal clinical trials. Pharmaceutical Statistics, 12, 1-6.

Pham T, van der Heijde D, Atlman R, Anderson J, Bellamy N, Hochberg M, *et al*. (2004). OMERACt-OARSI initiative: Osteoarthritis Reseach Society International set of responder criteria for ostoarthritis clinical trials.

Ratitch B, & O'Kelly M. (2011). Implementation of pattern-mixture models using standard SAS/STAT Procedures. PharmaSUG.

Rubin. (1987). Multiple Imputation for Nonresponse in Surveys. Wiley.