Statistical Analysis Plan (Final)

Clinical Trial Protocol Identification No.	EMR700692_006
Title:	A multicenter, randomized, double-blind, placebo-controlled, parallel-group trial to investigate the efficacy and safety of different intra-articular (i.a.) dosages of sprifermin in subjects with primary osteoarthritis of the knee (FORWARD)
Trial Phase	II
Investigational Medicinal Product(s)	Sprifermin (AS902330)
Clinical Trial Protocol Version	Amendment No. 2 (Global), 05 Mar 2013
Statistical Analysis Plan Author	PPD
Statistical Analysis Plan Date and Version	Date and Version Number of this Statistical Analysis Plan: 11 August 2016/Final Version 1.0
Statistical Analysis Plan Reviewers	PPD, Merck Serono Sponsor Statistician PPD Sponsor Clinician representative

1 Signature Page

Statistical Analysis Plan: EMR700692_006

A multicenter, randomized, double-blind, placebo-controlled, parallel-group trial to investigate the efficacy and safety of different intra-articular (i.a.) dosages of sprifermin in subjects with primary osteoarthritis of the knee (FORWARD)

BioStata Representative:

PPD

Biostatistician:

Trial Biostatistician PPD
PPD
PPD

Backup Trial Biostatistician PPD
PPD

Merck Representative:

PPD
EMD Serono Inc. 45A Middlesex Turnpike, Billerica MA 01821, USA
PPD

Medical Responsible:

PPD

Merck Medical Representative: PPD,
EMD Serono Inc. One Technology Place, Rockland MA 02073,USA
PPD

Global Drug Safety Responsible:

Merck Representative: PPD
EMD Serono Inc. 45A Middlesex Turnpike, Billerica MA 01821, USA
PPD

Lead Statistical Programmer:

Merck Representative Statistical Programmer
PPD
EMD Serono Inc. 45A Middlesex Turnpike, Billerica MA 01821, USA
PPD

Biomarkers Lead:

Merck Biomarkers Representative
PPD
Merck KGaA, Frankfurter Str. 250, Postcode: F131/102, 64293 Darmstadt, Germany
PPD

2 Table of Contents

1 Signature Page 2
2 Table of Contents....4
3 List of Abbreviations and Definition of Terms 6
4 Modification History 10
5 Purpose of the Statistical Analysis Plan 10
6 Summary of Clinical Trial Features 11
7 Sample Size/Randomization....16
8 Overview of Planned Analyses....18
8.1 Analysis of data from DBPC treatment phase....20
8.2 Analysis of data from the follow-up phase....20
8.3 Analysis of safety data during conduct of the trial....20
8.4 Interim analysis....21
9 Changes to the Planned Analyses in the Clinical Trial Protocol....21
10 Analysis Sets....22
11 General Specifications for Statistical Analyses....27
12 Disposition of Subjects and Discontinuations....33
12.1 Protocol Deviations 34
13 Demographics and Other Baseline Characteristics....34
13.1 Demographics....34
13.2 Medical History 34
13.3 Other Baseline Characteristics....34
14 Previous or Concomitant Medications/Procedures....35
15 Treatment Compliance and Exposure....35
16 Endpoint Evaluation 36
16.1 Primary Efficacy Endpoint Analyses....36
16.2 Secondary Efficacy Endpoint Analyses....38
16.3 Pharmacokinetic Endpoint Analyses 40
16.4 Biomarker subgroups....42
16.5 Other Endpoint Analyses....44
17 Safety Evaluation....54
17.1 Adverse Events 54

17.1.1 All Adverse Events 54
17.1.2 Adverse Events Leading to Discontinuation 55
17.2 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events 56
17.2.1 Deaths 56
17.2.2 Serious Adverse Events 56
17.2.3 Acute Inflammatory Reactions (AIRs) 56
17.3 Clinical Laboratory Evaluation 57
17.4 Vital Signs 60
17.5 Physical Examination 61
17.6 ECG Evaluation 61
18 References 61
19 Appendices 62
19.1 List of endpoints 63
19.2 List of permitted pain medication 66
19.3 Handling of medication timing 67
19.4 Handling of TEAE timing 68
19.5 List of key statistical outputs 69
19.5.1 General Information Tables 69
19.5.2 Efficacy Tables 69
19.5.3 Safety Tables 70
19.5.4 Figures 71
19.5.5 Data Listings 71

3 List of Abbreviations and Definition of Terms

AAA	Anatomic Axis Angle (of knee malalignment, in degrees)
ADaM	Analysis Dataset Model
ADY	Analysis Day
AE	Adverse event
AESI	Adverse events of special interest
AIR	Acute inflammatory reaction
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANOVA	Analysis of variance
AST	Aspartate aminotransferase
ATC	Anatomic therapeutic class
AVISIT	Analysis Visit
BDS	Basic Data Structure
BLOKS	Boston-Leeds Osteoarthritis Knee Score
BMI	Body Mass Index (weight divided by height, squared [kg/m^2])
BUN	Blood urea nitrogen
BSAP	Biomarkers statistical analysis plan
CI	Confidence interval
CK	Creatine kinase
CRF	Case Report Form
CTR	Clinical Trial Report
CV	Coefficient of variation
DBPC	Double-blind placebo-controlled
DoFD	Day of First Dose
DoLD	Day of Last Dose
ECG	Electrocardiogram
ECM	Extracellular matrix
ECS	Enabling Clinical Sciences
eCRF	Electronic Case Report Form
ESR	Erythrocyte sedimentation rate
EudraCT	European Union Drug Regulating Authorities Clinical Trials

	(European clinical trials database)
FGF	Fibroblast growth factor
FGF-18	Fibroblast growth factor-18
FORWARD	FGF-18 Osteoarthritis Randomized trial With Administration of Repeated Doses
GCP	Good Clinical Practice
GED	Global Early Development
HLT	High-Level Term
hsCRP	High-sensitivity C-reactive protein
i.a.	Intra-articular(ly)
ICF	Informed Consent Form
ICH	International Conference on Harmonization
ICOAP	Measure of Intermittent and Constant Osteoarthritis Pain
IDMC	Independent Data Monitoring Committee
IL1RN	Interleukin 1 Receptor Antagonist
IMP	Investigational Medicinal Product
IND	Investigational New Drug
ITT	Intention-to-Treat
ITT2	Second Intention-to-Treat
IVRS	Interactive Voice Response System
JSW	Joint space width
KL	Kellgren-Lawrence
KOOS	Knee Injury and Osteoarthritis Outcome Score
KOOS QOL	Knee Injury and Osteoarthritis Outcome Score quality of life subscale
PPD	
LLN	Lower limit of normal (range)
LOCF	Last Observation Carried Forward
Max	Maximum
MCID	Minimally clinically important difference
mcg	Microgram(s)
MedDRA	Medical Dictionary for Regulatory Activities
Min	Minimum

mJSW	Minimum joint space width
mITT	Modified Intention-to-Treat
MOP	Manual of Operations
MOS SF-36	SF-36 Medical Outcomes Study Short Form-36 General Health Survey
MRI	Magnetic resonance imaging
m2ITT	Second Modified ITT
NRS	Numerical Rating Scale
NSAIDs	Nonsteroidal anti-inflammatory drugs
OA	Osteoarthritis
OAI	Osteoarthritis Initiative
OARSI	Osteoarthritis Research Society International
OMERACT-OARSI	Outcome Measures in Rheumatology – Osteoarthritis Research Society International (responder criteria)
PD	Pharmacodynamic(s)
PGA	Patient's Global Assessment
PGIC	Patient Global Impression of Change
PGX	Pharmacogenomics
PK	Pharmacokinetic(s)
POC	Proof of concept
PP	Per Protocol
PRO	Patient Reported Outcome
PT	Preferred Term
Q1	First quartile
Q3	Third quartile
QOL	Quality of life
QPD	Quantitative Pharmacology and Drug Disposition
QTc	Corrected electrocardiogram QT interval
rhFGF-18	Recombinant human fibroblast growth factor 18
SAE	Serious Adverse Event
SAP	Statistical analysis plan
SD	Standard deviation
SDTM	Study Data Tabulation Model

SNP	Single nucleotide polymorphism
SOC	System Organ Class
SOP	Standard Operating Procedures
TEAE	Treatment Emergent Adverse Event
TLFs	Tables, Listings and Figures
ULN	Upper limit of normal (range)
VAS	Visual analogue scale
W	Week
WBC	White blood cell
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index
WORMS	Whole Organ Magnetic Resonance Imaging Score

4 Modification History

Unique Identifier for SAP Version	Date of SAP Version	Author	Changes from the Previous Version
0.1	28Jun2013	PPD	First version
0.2	08Apr2014		Second version
0.3	21 Jul2014		Third version
0.4	18 Aug 2014		Fourth version
0.5	24 Sep 2014		Fifth draft
0.6	14 Oct 2014		Minor changes
0.7	13 July 2015		Changes include inclusion of biochemical markers and alignment with the imaging charter (inter- and intra-reader reliability testing)
0.8	3 August 2015		Implement comments from PPD and PPD
0.9	01Oct2015		Minor changes
0.10	15Oct2015		Minor changes
0.11	04Jan2016		Minor Changes, ready for finalization
0.12	21Apr2016		Major Changes, ready for finalization
0.13	14Jun2016		Minor Changes, ready for final review
0.14	10Aug2016		Minor changes, ready for final check before signing
1.0	11Aug2016		Final version

5 Purpose of the Statistical Analysis Plan

The purpose of this Statistical Analysis Plan (SAP) is to document technical and detailed specifications for the final analysis of data collected for protocol EMR700692_006. Results of the analyses described in this SAP will be included in the Clinical Trial Report (CTR). Additionally, the planned analyses specified in this SAP will be included in regulatory submissions or future

manuscripts. Any additional analyses performed to provide results for inclusion in the CTR, but not specified in this SAP, will be clearly identified in the CTR.

The statistical analysis will be performed by PPD in two steps as specified in Section 8. Analysis of pharmacokinetic data, antibody data, biomarkers as well as modeling and simulation will be performed by Merck-Serono. Details on the analysis of pharmacokinetic data (including antibody data), and exploratory biomarkers are specified in Sections 16.3 and 16.4, respectively. Analysis of antibodies to sprifermin is part of the safety evaluation and is specified in Section 16.3 and Section 17. The biomarkers statistical analysis plan (BSAP) and the Modeling and Simulation plan will be detailed by Merck-Serono in separate SAP documents. However, selected biomarker analyses (e.g., pharmacogenomics [PGX] and other biochemical markers), as presented in this SAP, will be performed by PPD. List of key Tables, Figures and Listings are included in Appendix 19.5.

6 Summary of Clinical Trial Features

Trial Objectives	**In subjects with primary osteoarthritis of the knee treated with different intra-articular (i.a.) dosages of Sprifermin:** **Primary** To evaluate structural changes in cartilage thickness in the total femorotibial joint of the target knee in terms of imaging by magnetic resonance imaging (MRI) **Secondary** • To evaluate different dose regimens of sprifermin • To evaluate changes in symptoms of osteoarthritis (OA) • To evaluate changes in structure in terms of imaging by X-ray • To evaluate other changes in cartilage morphology in terms of imaging by MRI (sub-regions) • To evaluate changes in physical functioning • To evaluate the safety of sprifermin • To evaluate the pharmacokinetics (PK) of sprifermin in serum and in synovial fluid following i.a. injection **Exploratory** • To evaluate responder criteria in symptoms • To explore treatment effect on OA pain and other patient-reported outcomes (PROs)

	• To evaluate changes in quality of life (QOL) • To explore treatment effect on the knee joint and bone structures by MRI • To evaluate long-term treatment effects • To explore the relationship between the change over time in synovial, serum, and urine biomarkers (e.g., cartilage and bone tissue turnover markers, inflammation markers) and the response to the drug • To evaluate potential associations of biomarkers (e.g., genetic variations or protein biomarkers measured at baseline) with drug response, and/or disease severity or disease progression • To explore the relationship between dosing (dose and regimen), cartilage structure, and clinical scores as a function of time, by means of modelling
Trial Endpoints	**Primary** Change from baseline in cartilage thickness in the total femorotibial joint as evaluated by MRI at 2 years **Secondary** • Changes from baseline in the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) total score and in the WOMAC pain, function, and stiffness index scores over 2 years • Change from baseline in the 20-meter walk test over 2 years • Change from baseline in the Patient's Global Assessment (PGA) over 2 years • Change from baseline in minimal Joint Space Width (JSW) in the medial and lateral compartments as evaluated by X-ray over 2 years • Change from baseline in cartilage thickness in the medial and lateral compartments over 2 years • Change from baseline in cartilage volume in the medial and lateral compartments over 2 years • Synovial fluid levels of sprifermin/FGF-18 • Serum levels of sprifermin/FGF-18 **Safety endpoints** • Nature, incidence and severity of local and systemic Adverse Events (AEs). • Incidence of acute inflammatory reactions (AIRs), defined as increase of pain by 30 mm on a 100 mm visual analogue scale (VAS) and a self-

	reported synovial fluid effusion (i.e., joint swelling) within 3 days following i.a. injection (see Section 7.4.1.1 of the protocol) • Changes in laboratory safety parameters, vital signs, 12-lead electrocardiogram (ECG) parameters, weight, and physical examinations • Occurrence of binding and neutralizing antibodies to sprifermin/FGF-18 **Exploratory** • Outcome Measures in Rheumatology – Osteoarthritis Research Society International (OMERACT-OARSI) responder rate at 2 years • Patient Global Impression of Change (PGIC) over 2 years • Changes from baseline in the scores of the Knee Injury and Osteoarthritis Outcome Score quality of life subscale (KOOS QOL) and the SF-36 Medical Outcomes Study Short Form-36 General Health Survey (MOS SF-36 questionnaire) over 2 years • Change from baseline in Knee Injury and Osteoarthritis Outcome Score (KOOS) Symptom Index over 2 years • Change from baseline in Numerical Rating Scale (NRS) pain score in the target and contralateral knee over 2 years using an 11-point NRS • Change from baseline in presence of pain in other joints over 2 years • Change from baseline in Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP) scores over 2 years • Change over time in structural features of the knee joint (e.g., synovium, menisci, bone, and other structures) as evaluated by MRI • Change from baseline in serum, urine, and synovial markers associated with administration of the compound • Baseline biochemical biomarkers and/or genetic markers associated with response to treatment or disease progression (response assessed by MRI, X-ray and/or questionnaire) • Changes in MRI outcomes, WOMAC scores, 20-meter walk test, PGA, PGIC, JSW, OMERACT-OARSI responder rate, KOOS QOL, KOOS Symptom Index, MOS SF-36, NRS pain score in the target and contralateral knee, presence of pain in other joints, ICOAP, safety laboratory values, and vital signs over the extended follow-up period and nature, incidence and severity of local and systemic AEs during the extended follow-up period (to be summarized descriptively) • Relationship between dosing (dose and regimen), cartilage structures, and clinical scores as a function of time

Trial Design	This is a multicenter, randomized, double-blind, placebo-controlled, parallel-group Phase II trial of sprifermin administered i.a. in subjects with primary OA of the knee and Kellgren-Lawrence (KL) grade 2 or 3. The trial consists of a screening period lasting up to 42 days, a 2-year (104-week) double-blind placebo-controlled (DBPC) treatment phase, which will start at randomization (Week 0/Visit 2) and a 3-year extended follow-up phase. Please see table 5 and 6 (Appendix E) in the clinical trial protocol. Please see below for specification of dose groups (Trial Product).
Number of Subjects	The trial will randomize approximately 545 subjects (109 per treatment group).
Trial Product	Sprifermin (also known as: recombinant human fibroblast growth factor 18 [rhFGF-18]; AS902330) 30 mcg or 100 mcg, administered intra-articularly (i.a.). Subjects will receive 4 cycles of treatment (each consisting of 3 once-weekly i.a. injections over 3 consecutive weeks) at intervals of 6 months. Subjects will be randomized in equal allocation to one of 5 treatment groups: • 4 cycles of sprifermin (100 mcg per injection, total dose 1200 mcg) • 2 cycles of sprifermin (100 mcg per injection) alternating with 2 cycles of placebo (total dose 600 mcg) • 4 cycles of sprifermin (30 mcg per injection, total dose 360 mcg) • 2 cycles of sprifermin (30 mcg per injection) alternating with 2 cycles of placebo (total dose 180 mcg) • 4 cycles of placebo (total dose 0 mcg)
Treatment and Trial Duration	Planned treatment duration is 2 years. Follow-up duration is 3 years.
Trial Periods (if applicable)	The screening period will comprise a period of 4 to 42 days during which a subject's eligibility will be determined, beginning at signature of the Informed Consent Form (ICF). During the screening visits, eligibility will be assessed and the target knee will be defined. At the baseline visit, eligibility will be confirmed, randomization will occur, and the first IMP injection will be given (Week 0/Visit 2). The last of the four cycles of 3 injections of IMP will begin on Week 78/Visit 17, and the last dose of IMP is scheduled to be administered at Week 80/Visit 19.

	Trial visits during the two-year DBPC treatment phase will take place at Weeks 0 (randomization and first IMP injection), 1, 2, 3, 12, 26, 27, 28, 29, 38, 52, 53, 54, 55, 64, 78, 79, 80, 81, 90, and 104. Major efficacy assessments will take place on the first day of each treatment cycle and then at the end of the DBPC treatment phase (Week 104/Visit 22). Minor efficacy assessments will take place 9 weeks after each treatment cycle. Adverse Events (AE) and concomitant medications will be reviewed and vital signs will be measured at all trial visits. Trial assessments during the 3-year extended follow-up phase will take place at 6-month intervals from Week 104/Visit 22 through Week 260/Visit 28.
Randomization and Blinding	Eligible subjects will be randomized to one of the 5 treatment parallel groups with an allocation ratio of 1:1:1:1:1. Randomization will be stratified by country. The central randomization service's IVRS will be used to randomize subjects to treatment. The allocated randomization number will only be used for dosing. The IVRS will also be used to assign a blinded treatment kit at each subsequent treatment visit.

7 Sample Size/Randomization

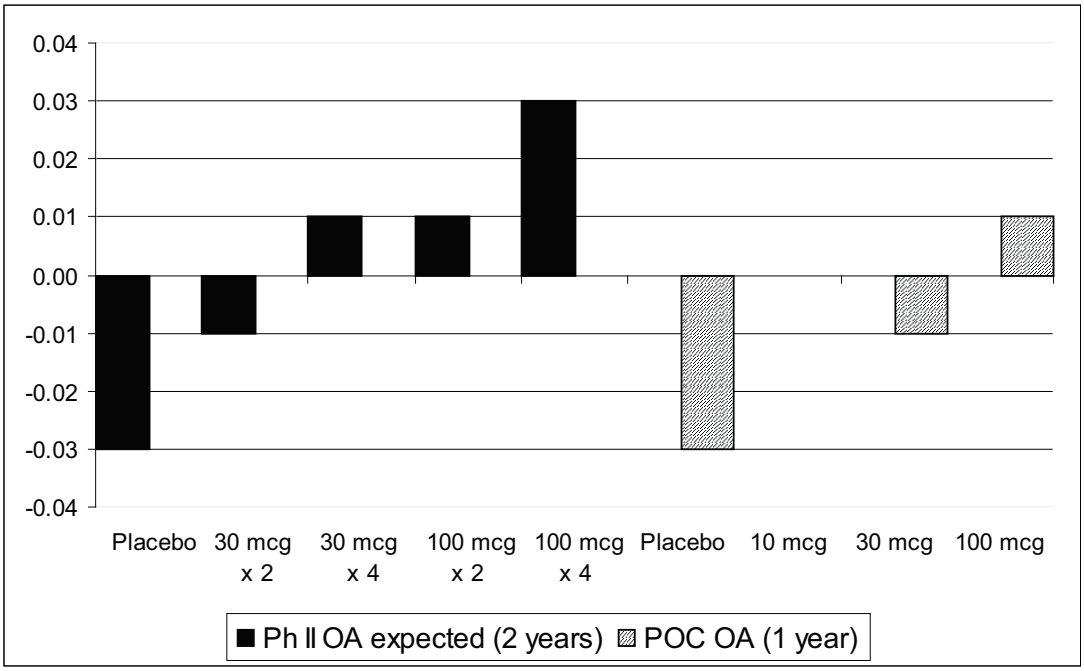
As described in protocol Section 8.1: The sample size was determined for the primary endpoint: Change from baseline in cartilage thickness in the total femorotibial joint as evaluated by quantitative MRI at 2 years.

This trial is set up to detect a potential dose relationship, based on the total dose of sprifermin received over the 2 years, in limiting the reduction of or increasing the cartilage thickness in the total femorotibial joint, assuming:

1. A mean change from baseline at 2 years:
    - -0.03 mm in the placebo group (i.e., 0 mcg over 2 years)
    - -0.01 mm in the sprifermin 30 mcg once a year group (i.e., 180 mcg over 2 years)
    - +0.01 mm in the sprifermin 30 mcg twice a year group (i.e., 360 mcg over 2 years)
    - +0.01 mm in the sprifermin 100 mcg once a year group (i.e., 600 mcg over 2 years)
    - +0.03 mm in the sprifermin 100 mcg twice a year group (i.e., 1200 mcg over 2 years)
2. A common SD (standard deviation) of 0.10 mm
3. A type-one error, α, of the primary efficacy analysis set at 5% two-sided
4. Normal distribution.

The above assumptions were based on available MRI data from the Osteoarthritis Initiative (OAI) (http://oai.epiucsf.org/datarelease/) and from the Phase I Proof of concept (POC) trial (28980).

Figure 1 **Expected Mean Changes in Cartilage Thickness (mm) in Total Knee at 2 Years, SD=0.10**



SD: standard deviation

A minor deviation from linearity has been introduced in these hypotheses based on the expectation of similar changes for the sprifermin 30 mcg twice a year, and sprifermin 100 mcg once a year groups. Indeed, according to the current knowledge of the biological mechanism of action of sprifermin, the treatment effect of the highest dose (100 mcg) given once a year could be similar to that of a lower dose (30 mcg) given twice a year. Moreover, the variability of the measure is large, and therefore a deviation from perfect linearity may be expected.

Calculations showed that:

- 55 evaluable subjects in each group (total 275 subjects) will ensure a power of 90% for detecting a linear dose relationship using a linear trend test,
- 76 evaluable subjects in each group (total 380 subjects) will ensure a power of 90% for detecting an overall treatment effect (meaning that at least one group differs from the other groups) as well.

Based on recent literature and other trial experience (see protocol references 45 and 46), the subject discontinuation rate is estimated to be close to 30% at two years. Assuming a 30% drop-out rate, a total of 545 subjects should be randomized (109 subjects per group) in order to detect an overall treatment effect.

In order to be able to demonstrate that structural improvement translates into symptomatic benefit, it is important that the study be adequately powered to detect significant differences in normalized WOMAC scores. With 76 evaluable subjects per treatment group, the power to show a statistically

significant difference between at least one active treatment group and the placebo group for the secondary WOMAC endpoints ranges from 69% to 87% for the total score, from 80% to 99% for the pain score and from 53% to 83% for the function score, depending on the expected mean difference from placebo (see Table 1).

Table 1 Power Calculations for Pairwise Differences (Sprifermin versus Placebo) on Normalized WOMAC Scores with 76 Evaluable Subjects per Treatment Group

	Expected mean difference from Placebo	Common SD	Power
Total Score - Normalized (range 0-100 points)	8 points	20	69%
	9 points		79%
	10 points		87%
Pain Score - Normalized (range 0-100 points)	**10 points**	22	80%
	15 points		99%
Function Score - Normalized (range 0-100 points)	7 points	21	53%
	9 points		75%
	10 points		83%

SD: standard deviation; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index

For reference, the minimally clinically important difference (MCID) in WOMAC scores is considered to be about 10% of the maximum score, which correspond to approximately 10 points for the total WOMAC score, 7 points for the WOMAC function index (activities of daily living) score, 2 points for WOMAC pain index score, and 1 point for the WOMAC stiffness index score) [1]. WOMAC scores that are considered MCID are highlighted in bold type in the table above.

Please note that the power calculations did not include adjustment for multiplicity. Randomization is described in Section 6 of the protocol.

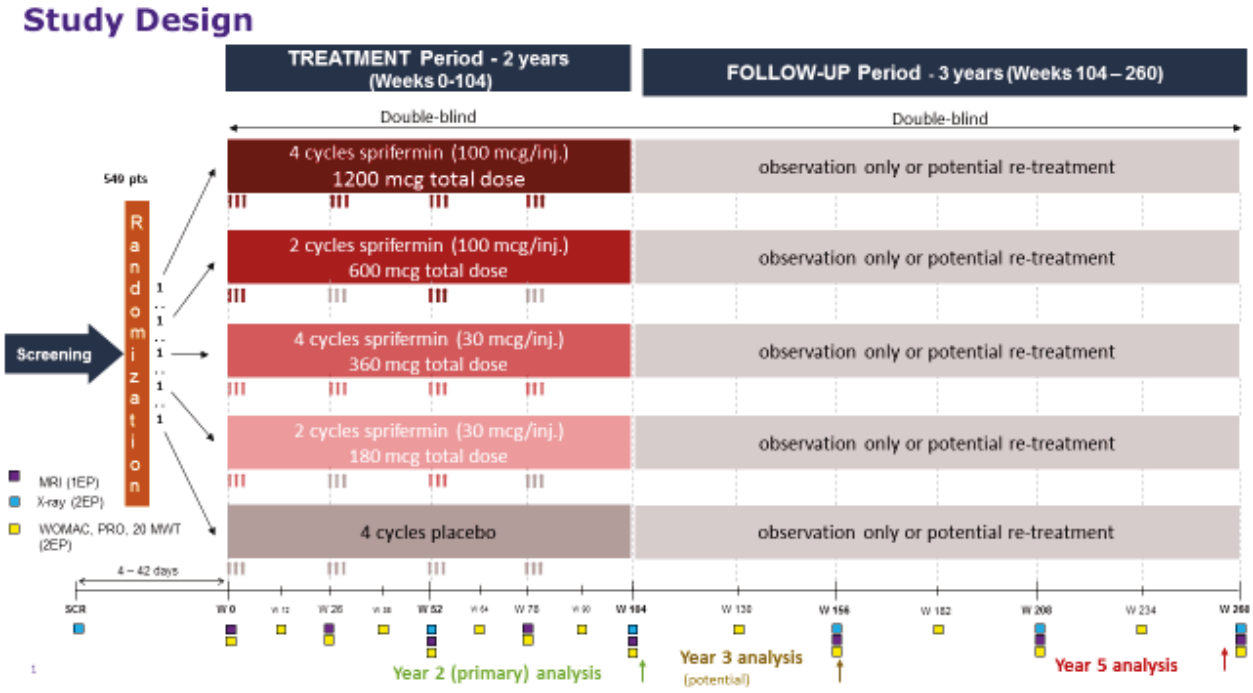
8 Overview of Planned Analyses

The protocol specifies a two-year treatment phase (Baseline to Year 2) followed by a three-year extension phase (Years 2 to 5). Subjects and sites will remain blind to treatment throughout the entire protocol (Baseline to Year 5). The statistical analysis will be performed in three steps (see "Study Design" figure below):

- Reporting of the DBPC treatment phase: After the last subject's last visit of the DBPC treatment phase, the data collected until the end of the DBPC treatment phase will be cleaned and released, a Blind Data Review Meeting will be held to clarify potential remaining issues before the database can be partially locked (this terminology is used here to differentiate data coming from the DBPC treatment phase vs. extended follow-up phase, but all processes for formal database lock will be followed); if all participants of the Blind Data Review Meeting are in agreement

that the database can be partially locked, then the database will be partially locked, the treatment code will be un-blinded, complete analyses of the DBPC phase will be performed. All data that is collected until week 104 (day 819:see table 3 for definition of target day and allowed windows) will be included in the statistical analysis of the partially locked database for the DBPC treatment phase. Each dataset will include variable to flag if a record will be included in analyses for the DBPC treatment phase. If records belong to days ≥ 820 they will be part of the long term follow-up analysis. Further description of database lock handling appears in the Data Management plan. The SAP has to be approved prior to database partial lock and breaking the blind. This will be referred to as 2-year analysis.

- Long term follow-up data will be analyzed and reported in addition to the DBPC treatment phase results. This will be referred to as 5-year analysis and will include data from baseline (baseline, day 1) until end of extended follow-up period (Year 5). To minimize bias during the extended follow-up phase, all site staff including Investigators, study monitors, X-ray and MRI readers, and subjects will remain blinded during the extended follow-up phase. Subjects will maintain their subject number from treatment phase to follow-up. X-ray and MRI readers will remain blinded during the study and after the study completion.



- In addition, a Year-3 analysis, repeating Year-2 analysis for primary and secondary efficacy endpoints as well as safety endpoints is planned.

No interim analysis is planned before the primary analysis defined at 2 years; however, an Independent Data Monitoring Committee (IDMC) will perform periodic reviews of semi-un-blinded data to evaluate the safety of the subjects on an ongoing basis. Detailed description of the safety monitoring is specified in a dedicated IDMC charter [2].

8.1 Analysis of data from DBPC treatment phase

An overview of endpoints and type of analysis for the DBPC treatment phase is provided in Section19.1. The analysis will report all data collected from the DBPC treatment phase up to 24 months (day 819), i.e. disposition, demographics, baseline characteristics, concomitant medication during and prior to DBPC treatment phase, compliance and exposure, all efficacy endpoints and safety parameters up to 24 months (day 819).

Moreover, PK-related analysis will be provided by Merck Serono Quantitative Pharmacology and Drug Disposition (QPD) and Enabling Clinical Sciences (ECS) and will be incorporated in the CTR (if at least one sample show sprifermin concentration). PK analyses and biomarker subgroups (genetic markers - Single nucleotide polymorphism [SNP] and subgroups based on biochemical marker cutoffs) are included in this SAP in Section 16.3 and Section 16.4, respectively. Additional analyses of other exploratory biomarkers will be described in the BSAP and will be reported separately in an Exploratory Biomarkers Report. Any modeling of the data will be further described in a separate Modeling and Simulation SAP and will be reported in a separate Modeling and Simulation Report.

Data used for reporting of the DBPC will be the data presented in Appendix 19.1.

Upon database release, protocol deviation and Analysis Set outputs will be produced and will be sent to Merck Serono for review. A Blind Data Review Meeting will be arranged to discuss the outputs and to decide which subjects and/or subject data will be excluded from certain analyses. Decisions made regarding the exclusion of subjects and/or subject data from analyses will be made prior to breaking the blind and will be documented and approved by Merck Serono.

The blind will be kept for investigators, X-ray and MRI readers, and subjects during the follow-up phase. Individual subject data listings revealing the treatment as well as tables from the CTR will not be shared with the site investigators or X-ray or MRI readers until completion of the follow-up phase.

8.2 Analysis of data from the follow-up phase

An overview of endpoints and type of analysis for the follow-up phase is provided in Appendix 19.1. In general, the same endpoints (for efficacy and safety) will be presented as for the treatment phase and the analysis of follow-up period (up to week 260) will be integrated combining all available data in the entire 5 years period (including data from the DBPC). No multiplicity adjustment for the integrated analysis will be done, as the efficacy evaluation on the integrated data is considered exploratory.

8.3 Analysis of safety data during conduct of the trial

The IDMC will perform periodic reviews of semi-unblinded safety data, e.g. adverse events, AIR, safety lab data and vital signs. Disposition and demographics will also be presented. The IDMC charter describes presentation of data.

8.4 Interim analysis

No interim analysis prior to Year-2 analysis is planned.

9 Changes to the Planned Analyses in the Clinical Trial Protocol

The following clarifications and changes (including deletion and additional analyses) compared to the protocol have been made prior to database lock and breaking the blind of the study:

Deletion:

Interleukin-6 (IL-6) is not measured (see Note to File "Selection of Biomarkers for Clinical Study EMR700692-006") and therefore was removed from the list of laboratory parameters (Section 7.4.3 of the protocol).

Clarifications:

1. The protocol is not specific on what deviations are considered major and resulting in subjects who are excluded from the PP Analysis (Section 8.4 of the protocol). Section 10 provides more details for definition of major protocol deviations and the process to identify them.

2. The protocol is not specific on how data will be included in the extended follow-up period analysis (Section 8.5.1 of the protocol). Section 8.2 provides more details on what data will be included in this analysis. It also specifies that all data from the 5 years period will be included in this integrated analysis.

3. Clarification on time points, scored regions, scoring system and analyses for Whole Organ Magnetic Resonance Imaging Score (WORMS) and (Boston-Leeds Osteoarthritis Knee Score) BLOKS are provided in Section 16.5 (Section 8.5.5 of the protocol).

Additions:

1. Analysis sets. The following additional analysis set is specified in Section 10. For the long term follow-up phase (following day 819), a separate analysis set is defined consisting of subjects from the ITT population who have at least one post-treatment quantitative MRI assessment or at least one WOMAC measurement in the follow-up period, i.e., the Second ITT (ITT2). The ITT2 Analysis Set will include all subjects from the ITT Analysis Set who have at least one MRI assessment available or one X-ray assessment available or one WOMAC measurement on or after day 820.

2. Year-3 Analysis: A Year-3 Analysis is planned and described in Section 8.

3. Sensitivity analysis using LOCF. The LOCF method as imputation for missing data on post baseline observations will be used as sensitivity evaluation with potential to be utilized in future studies. This is described in Section 11.

4. Adverse events of special interest: Injection site reactions, Neoplasms, Musculoskeletal disorders will be identified using pre-specified rules and will be summarized (as specified in Section 17.1).

5. Key efficacy endpoints will be summarized by baseline biomarker subgroups (as specified in Section 16.4) and by clinical subgroups. Some of the subgroups (as specified in Section 10, last part), were not pre-specified in the protocol. For countries with low accrual, pooling of subjects will be done by similarity. As of September 2014 (when study was fully enrolled) there were only 11 randomized subjects from the USA and one from Poland. Therefore, for the primary analysis of ANOVA, Poland will be pooled with the smallest Eastern European country, Romania. Czech Republic and Estonia, as well as Hong Kong, are counted separately, and the USA will be pooled with Denmark as USA+ Western Europe. For subgroup analyses there will be 3 regions: (1) Eastern European, (2) USA+ Western Europe, and (3) Hong Kong. This is specified in sections 10 and 16.1. Among these subgroups, 3 subgroups have been identified to have higher priority and described in Section 10).

6. Responder variables: subjects with at least 10-point improvement on WOMAC Total score, and subjects with change from baseline in total cartilage thickness ≥ 0.01mm and their combination (as specified in Section 16.5) are added.

7. Spearman's correlations between endpoints related to function and symptoms (e.g. WOMAC pain, WOMAC function, ICOAP, 20 meter walk test, KOOS symptoms, KOOS QoL, etc.) and structure related endpoints will be evaluated as specified in Section 16.5.

10 Analysis Sets

List of subjects to be included in the analysis sets will be presented in minutes from the Blind Data Review Meeting and will be part of the CTR (included as listing of major and minor protocol deviations by subject). Determination of final protocol deviation will be finalized before unblinding the study.

Screening Analysis Set

The Screening analysis set includes all subjects who provided informed consent, regardless of the subject's randomization and treatment status in the trial.

Intent-to-Treat Analysis Set (Full Analysis Set)

The Intention-to-Treat (ITT) Analysis Set will include all randomized subjects and will be reported in any analysis "as randomized" i.e. only the planned treatment regimen is used. (i.e., the planned treatment regimen rather than the actual treatment given in case of any difference). This analysis set will be used for all non-MRI efficacy evaluations (symptoms and X-ray based endpoints).

Modified Intention-to-Treat Analysis Set

The Modified ITT (mITT) Analysis Set will include all subjects from the ITT Analysis Set who have a baseline (prior to first injection) and at least one post-treatment quantitative MRI assessment available in the DBPC part (up to and including day 819). This analysis set will be used for all MRI efficacy evaluations.

Per Protocol Analysis Set

The Per-Protocol (PP) Analysis Set will include all subjects from the mITT Analysis Set who have been treated according to the clinical trial protocol and fulfill the following criteria:

- Compliance with all entry criteria (fulfilling all inclusion criteria and no exclusion criteria)
- Absence of major clinical trial protocol deviations with respect to factors likely to affect the efficacy of treatment, such as non-missing and valid baseline and post-baseline measurement of the primary efficacy endpoint (quantitative MRI), where the nature of such deviations will be defined before or at the Blind Data Review Meeting (before breaking the blind) for the reporting of the treatment phase, in collaboration between the medical, clinical and statistical responsible person (see also Section 8).

Major protocol deviations affecting efficacy include: Prohibited medication/procedures during the treatment period and inadequate treatment compliance i.e. deviations from the planned amount of active dose and schedule of treatment (out of allowed window as specified in the "Analysis Sets" document). In order to be compliant the inter-cycle window should be within 140 to 196 days, and the intra-cycle window should be 5 days or above. Subjects who have (actual total dose)/(planned total dose)*100 within 75-100 % will be considered treatment compliant. If subject was randomized to placebo group but instead received at least one dose of sprifermin, then this subject will be classified as having major protocol deviation. A comprehensive list of all deviations and the process to identify each type of deviation are provided in the "Analysis Sets" document that will be finalized prior to Year-2 database lock.

Additionally subjects identified in the Note To File: "NTF_QC of primary endpoint MRI_FORWARD_17MAY2016", will be excluded from the PP.

"Based on input from the IDMC (ad hoc meeting, 21 June 2016), subjects will be excluded from the Per Protocol analysis if they have significant pre-treatment MRI lesions that would interfere with analysis of the primary endpoint. Accordingly, the following 2 subjects will be excluded from Per Protocol analysis, although retained in ITT, mITT, and Safety analyses:

- PPD
- PPD

Second Intention-to-Treat Analysis Set

The Second ITT (ITT2) Analysis Set will include all subjects from the ITT Analysis Set who have at least one valid MRI assessment or one X-ray measurement available or one WOMAC measurement available after day 819 (i.e., during the 3-year Extended Follow-up Phase [Week 104, Year 2] and up to day 1849 [Week 260, Year 5]).

Safety Analysis Set

The Safety Analysis Set will include all subjects who receive at least one dose of trial treatment. Subjects will be analyzed according to the actual treatment they receive. In case of deviations from planned treatment, after un-blinding placebo patients receiving at least one active injection, will conservatively be allocated to the lowest of the possible planned Sprifermin dose groups. Please refer to the "Analysis Sets" document for more details.

Pharmacokinetic Analysis Set

The PK Analysis Set will include all subjects from the Safety Analysis Set who have at least one available serum or synovial fluid measurement.

PGX Analysis Set

The PGX Analysis Set will include subpopulation of subjects from the ITT Analysis Set with informed consent allowing for analyses of genetic data, and who have PGX measurement needed for the analysis specified in Section 16.4.

Tables and figures will be presented on separate analysis set as presented in Table 2 below:

Table 2 Overview of analyses by Analysis Set

	DBPC Treatment Phase				Follow-up Phase	Both phase
Analyses	**Intent-to-Treat Analysis Set (ITT)**	**Modified ITT (mITT)**	**Per Protocol Analysis Set (PP)**	**PGX Analysis Set (PGX)**	**Second ITT (ITT2)*** **	**Safety Analysis Set (Safety)**
Demography and Baseline Assessments (incl. other baseline characteristics section 13.3)	✓	✓	✓	✓	✓	✓
Concomitant Therapies	✓	✓	✓		✓	✓
Compliance and Exposure	✓	✓	✓	✓		✓
MRI Efficacy: Primary and secondary		✓	✓		✓	
Efficacy: Secondary (non MRI) with and without LOCF	✓		✓*		✓	
Exploratory endpoints – additional analyses end of section 16.5	✓**				✓**	
Subgroup analyses	✓				✓ *	
Biomarkers analyses section 16.4	✓(Biomarkers other than PGX)			✓(PGX only)		
Safety and Tolerability						✓

*Only for selected subgroups [subgroups 1-5 described in this section] and for the following endpoints: WOMAC total score, WOMAC sub scores, MRI endpoints (Cartilage thickness), x-ray endpoints (JSW), **including analysis of 10-point improvement on WOMAC scores and 0.01mm and 0.03mm increase on total cartilage thickness. ***Summary statistics only. Note: PK analysis will be based on PK analysis set.

Analysis of Subgroups for

Key subgroup prioritization:

Based on current knowledge (see below justification), the following 3 subgroup analyses are of special interest and have higher priority than remaining subgroup analyses:

- KL grade: Grade 2, Grade 3 (prognostic factor based on literature, and OAI anlysis)
- Cartilage metabolism biomarker at baseline: urinary CTX-II (creatinine adjusted), categories defined in section 16.4 (≤ 350ng/mmol, > 350ng/mmol, identified in a post hoc analysis of the PoC study as a potential predictive biomarker)
- Joint tissue inflammatory biomarker at baseline: serum CRPM, categories defined in section 16.4 (≤ 10ng/mL, >10ng/mL, identified in a post hoc analysis of the PoC study as a potential predictive biomarker)

The following subgroups will be used as basis for comparisons with respect to descriptive summary tables for the primary and key secondary efficacy endpoints specified in Section 16.5. e.g., the subgroups age < 65 years and age ≥ 65 years will be evaluated by presentation of primary endpoint summary table in two versions, one for each age group.

Subgroup analysis will not be performed as factor analysis, but only by presenting summary tables of the primary and key secondary endpoints over time, per subgroup.

The following subgroups were specified in the protocol and will be used:

1. KL grade (Grade 2, Grade 3 separately),
2. sex (Male, Female),
3. age (< 65 years old, ≥ 65 years old),
4. body mass index (BMI; < 30 kg/m2, ≥ 30 kg/m2),
5. pooled country (Eastern Europe [including Poland], US + Western EU [Denmark], and Hong Kong),

The following subgroups were not specified in the protocol, but will be used.

Disease status at baseline (based on X-ray):

1. bilateral/unilateral (OA in both knees/OA in one knee only)
2. baseline medial mJSW in target knee (< 3mm, ≥ 3 mm)
3. baseline mal-alignment in target knee (< -3 or > 3 or [-3,3] degrees)

Pain medications [classified for 3 timing groups defined in Section 14]:

1. Pain medications (Yes, No) – any of 2, 3 or 4 defined below
2. Analgesics, Opiods in WHO ATC code group N02A (Yes, No)
3. Analgesics, Other analgesics and antipyretics in WHO ATC code group N02B (Yes, No)
4. Non-steroid anti-inflammatory and antirheumatic products in WHO ATC code group M01A (Yes, No)

Other subgroups:

Subgroups defined by SNPs and biochemical biomarkers cutoffs subgroups (see Section 16.4);

A listing of protocol-permitted concomitant pain medications and their respective washout times was provided in Appendix 19.2 . Prior to unblinding this list may be updated.

Subgroups will be defined by individual listing prior to database lock and prior to breaking the blind for the reporting of the treatment phase.

11 General Specifications for Statistical Analyses

It is planned to perform the statistical analysis in two steps. Approximately one month after the last subject reaches day 819 of the 2-year DBPC treatment phase, the database will be partially locked, the treatment code will be un-blinded. Thereafter, complete analyses will be performed, and a detailed CTR will be written based on the data collected up to the end of the DBPC treatment phase (day 819). Long term follow-up data (from Year 2 to Year 5) will be integrated and analyzed with the DBPC treatment phase data but reported separately from the DBPC treatment phase results. Investigators, other site staff, study monitors, and subjects will remain blinded to study treatment during the extended follow-up phase as described in Section 8. X-ray and MRI readers will remain blinded to study treatment until completion of the extended follow-up phase results.

A list of endpoints according to treatment phase is provided in Appendix 19.1.

Summary Statistics

Summary statistics will be presented per treatment group and where relevant for 'all subjects'. Summary statistics will be presented also per country. The summary statistics presented for quantitative variables will be the number of observations (n), the number of missing values (missing), mean, standard deviation (SD), median, first and third quartiles (Q1, Q3), and minimum (min) and maximum (max) values. The mean, SD, median, first and third quartiles values will be presented with one more decimal places as the raw data. The minimum and maximum values will be reported to the same number of decimal places as the raw data. In general, the maximum number of decimal places reported shall be four for any summary statistic.

The summary statistics presented for categorical data will be the number of observations (n), the number of missing values (missing), and the count and percentage of subjects in each category at the relevant time point(s). Percentage will be given with one decimal place.

The summary statistics presented for adverse events will be number of observations (n), the percentage of subjects (%), overall and per system organ class (SOC) and PT.

Apart from the trend test for AIRs, all confidence intervals (CI) and statistical tests will be two-sided. Calculation of differences between treatment groups of continuous endpoints and corresponding CIs, will be normal based using unequal variances. CIs will be presented with one more decimal place than the raw data. P-values greater than or equal to 0.001, in general, will be presented with three decimal places. However, if a p-value is only presented with four decimal

places (by SAS) it will not be rounded further, but will be presented with four decimal places. P-values less than 0.001 will be presented as "<0.001", while P-values greater than 0.999 will be presented as ">0.999".

Plots showing mean absolute change from baseline to each time point in key endpoints (structure and WOMAC) within each treatment group will be provided for the ITT analysis set for non- MRI endpoints, and for the mITT analysis set for MRI endpoints.

Age, sex and race will be reported on all by-subject listings unless otherwise specified. Sex will be abbreviated as: female (F) and male (M). Race will be abbreviated as: American Indian or Alaska Native (AI), Asian (AS), Black or African American (BL), Native Hawaiian or other Pacific Islander (NH), White (WH) and Other (O).

All report outputs will be produced using SAS® version 9.4 (or later) in a secure and validated environment. All report outputs will be provided to the Sponsor in PDF and true MS Word documents.

Baseline

Baseline value is defined as the last non-missing measurement taken before first treatment (including unscheduled visits). This will typically be measurements taken at Baseline visit (week 0) but if this measurement is missing the measurement taken at the Screening visit (day-42 to -4) will be used if available. See schedule of trial procedures in the protocol, Appendix E. For subjects randomized but not treated, date of randomization will be used as baseline.

If date of assessment is same as injection day and time is not collected, assessment is assumed to be collected prior to injection.

Time windows for Analysis Visit (AVISIT) to be used in Analysis Dataset Model (ADaM) datasets with Basic Data Structure (BDS)

The following tables describe the mapping to AVISIT, any reference to visits throughout the rest of the SAP, refers to this mapping.

Table 3 Target days for AVISIT

Study week window	Questionnaires/20M/Labs (for 2-5 years part) /NRS/SF36/Pain in other joints target day (min, max)~	qMRI and SF36/Pain in other joints and Labs (for the year-2 part) target day (min, max)~	X-ray, sqMRI, ECG target day (min, max)	AVISITN for ADQS	AVISITN for MRI	AVISITN for XRAY
Baseline *	1 (-42, 1)	1 (-42, 1)	1 (-42, 1)	1	1	1
Month 3/Week 12~	85 (2, 133)	183 (2, 224)	365 (2, 406) &	2	3	5
Month 6/Week 26 *	183 (134, 224)			3		
Month 9/Week 38~	267 (225, 315)	365 (225, 406)		4	5	
Month 12/Week 52 *	365 (316, 406)			5		
Month 15/Week 64~	449 (407, 497)	547 (407, 588)	729 (407, 819)	6	7	9
Month 18/Week 78 *	547 (498, 588)			7		
Month 21/Week 90~	631 (589, 679)	729 (589, 819)		8	9	
Month 24/Week 104	729 (680, 819)			9		
Month 30/Week 130	911 (820, 1001)	1093 (820, 1183)	1093 (820, 1183)	10	11	11
Month 36/Week 156	1093 (1002, 1183)			11		
Month 42/Week 182	1275 (1184, 1365)	1457 (1184, 1547)	1457 (1184, 1547)	12	13	13
Month 48/Week 208	1457 (1366, 1547)			13		
Month 54/Week 234	1639 (1548, 1729)	1821 (1548, 1849)	1821 (1548, 1849)	14	15	15
Month 60/Week 260	1821 (1730, 1849)			15		
LOCF1 (up to Month 24)	Last post-baseline measurement in DBPC (2, 819)			20 (for 2 years)		
LOCF2 (Up to Month 60)	Last post-baseline measurement in Follow-up period (2, 1849)			50 (for 5 years)		

*treatment cycle; Day 1 is first treatment day. If multiple assessments are available within each study week window, the one closest to target day will be used. At Baseline, week 26, 52 and 78 [injections visits] pre injection assessments will be considered for the corresponding study week window). & for X-ray and this timepoint with injections - only compare to first injection in cycle # 3. don't compare to first injection in cycle #2 to determine if used in analysis or not. ~For SF36, lab, height, weight (and BMI), and pain in other joint, in the 2-years period, assessments are scheduled as qMRI, so qMRI windowing will be used. For the 2-5 years the scheduling is like ADQS, therefore, ADQS windowing will be used for these assessments.

For vital signs, which are measured at all visits, two mappings will be used, one relative to injections and one relative to target days.

Table 3a Vital signs relative to injections

Study week window	To be used relative to injection visit independent of ADY	AVISITN for injection - related ADVS
Cycle 1/Week 0/pre-injection assessment/Baseline	Yes	1
Cycle 1/Week 1/pre-injection assessment	Yes	2
Cycle 1/Week 2/pre-injection assessment	Yes	3
Cycle 1/Week 3	Yes	4
Cycle 2/Week 0/pre-injection assessment	Yes	5
Cycle 2/Week 1/pre-injection assessment	Yes	6
Cycle 2/Week 2/pre-injection assessment	Yes	7
Cycle 2/Week 3	Yes	8
Cycle 3/Week 0/pre-injection assessment	Yes	9
Cycle 3/Week 1/pre-injection assessment	Yes	10
Cycle 3/Week 2/pre-injection assessment	Yes	11
Cycle 3/Week 3	Yes	12
Cycle 4/Week 0/pre-injection assessment	Yes	13
Cycle 4/Week 1/pre-injection assessment	Yes	14
Cycle 4/Week 2/pre-injection assessment	Yes	15
Cycle 4/Week 3	Yes	16

Table 3b Vital signs relative to target days

Study week window	ADVS target day (min, max)	AVISITN for relative visit ADVS
Baseline	1 (-42, 1)	1
Month 3/Week 12	85 (2, 133)	2
Month 6/Week 26	183 (134, 224)	3
Month 9/Week 38	267 (225, 315)	4
Month 12/Week 52	365 (316, 406)	5
Month 15/Week 64	449 (407, 497)	6
Month 18/Week 78	547 (498, 588)	7
Month 21/Week 90	631 (589, 679)	8
Month 24/Week 104	729 (638, 819)	9
Month 30/Week 130	911 (820, 1001)	10
Month 36/Week 156	1093 (1002, 1183)	11
Month 42/Week 182	1275 (1184, 1365)	12
Month 48/Week 208	1457 (1366, 1547)	13
Month 54/Week 234	1639 (1548, 1729)	14
Month 60/Week 260	1821 (1730, 1849)	15
LOCF1 (up to Month 24)	Last post-baseline measurement in DBPC (2, 819)	20
LOCF2 (Up to Month 60)	Last post-baseline measurement in Follow-up period (2, 1849)	50

Adjustment for Multiplicity

The main objective of this Phase II trial is to select a dose regimen for further development in Phase III; therefore, no adjustment for multiplicity (except for dunnett adjustment for multiple comparisons used in the ANOVA models) resulting from multiple endpoints will be made. Caution should be used in interpreting significant p-values ($p < 0.05$). Interpretation of the results will involve global consideration of all efficacy results obtained in combination with an appropriate judgment of clinical relevance and consistency across endpoints.

Missing Data

As a general rule, all data collected and available will be used in the analysis. In general no imputation of missing data will be applied independently of the statistical method used to analyze the data, see however below remarks for LOCF in summaries over time.

The number of subjects contributing data will decrease over time. In summary assessments by time-point within DBPC, the LOCF method on post baseline observations will be used as sensitivity evaluation, and will be defined as LOCF1 for the 2-year analysis, and LOCF2 for the

analysis of the extended follow-up. These will only be included in summary tables (not in proc mixed models) and are considered exploratory and will be utilized for phase III statistical consideration. In the remainder of the SAP, if not otherwise stated, LOCF refers to both LOCF1 and LOCF2. Details for derivation of LOCF1 and LOCF2 for corresponding endpoints are provided in Table 3, 3a, and 3b.

Exceptions are handling of missing data from questionnaires where the guide for analysis prescribes otherwise, e.g. calculation depending on whether only single items are missing or a complete questionnaire is missing. If only single items are missing, the rules defined by the authors of the questionnaires (if any) will be followed.

The number and percentage of subjects with missing data for the primary and the secondary endpoints medial and lateral cartilage thickness, will be summarized by treatment group and timepoint for the mITT Analysis Set.

Listing of data

In general, all data will be listed by treatment group, subject and time point. Age, sex and race will be displayed in all listings.

ANOVA models

ANOVA models are used for several endpoints. For each endpoint all available time-points except LOCF (specified in table 3) should be included and summarized. The general model specification for model that include 4 post-baseline measurements (i.e. qMRI) in SAS will be:

```
proc mixed data=p3 method=reml;
  class grpn vs usubjid CNTRGR1N ;
  model &var. = grpn vs grpn*vs base CNTRGR1N / solution;
  repeated vs / subject=usubjid type=&type. r;
  lsmeans grpn / adjust=dunnett diff=control('1') cl;
  lsmeans grpn*vs / adjust=dunnett diff=control('1' 'Week 26') cl;
  lsmeans grpn*vs / adjust=dunnett diff=control('1' 'Week 52') cl;
  lsmeans grpn*vs / adjust=dunnett diff=control('1' 'Week 78') cl;
  lsmeans grpn*vs / adjust=dunnett diff=control('1' 'Week 104') cl;
  lsmeans grpn*vs / slice=vs;
  contrast 'Linear trend (overall)' grpn -2 -1 0 1 2;
  contrast 'Linear trend at Week 26' grpn -2 -1 0 1 2 grpn*vs -2 0 0 0 -1 0 0
0 0 0 0 0 1 0 0 0 2 0 0 0;
  contrast 'Linear trend at Week 52' grpn -2 -1 0 1 2 grpn*vs 0 -2 0 0 0 -1 0
0 0 0 0 0 0 1 0 0 0 2 0 0;
  contrast 'Linear trend at Week 78' grpn -2 -1 0 1 2 grpn*vs 0 0 -2 0 0 0 -1
0 0 0 0 0 0 0 1 0 0 0 2 0;
  contrast 'Linear trend at Week 104' grpn -2 -1 0 1 2 grpn*vs 0 0 0 -2 0 0 0
-1 0 0 0 0 0 0 0 1 0 0 0 2;
```

```
run;
```

12 Disposition of Subjects and Discontinuations

This section presents specifications for reporting subject disposition and treatment/trial discontinuations. Additionally, procedures for reporting protocol deviations are provided.

Subject disposition will be presented for all subjects and by treatment group in terms of:

- Total number of subjects screened (i.e. subjects who signed informed consent)
- Number and percentages of subjects discontinuing the trial prior to randomization divided by main reason for discontinuation (did not meet all eligibility criteria, withdrew consent, other), by country and site, and overall.
- Number and percentages of randomized subjects.
- Number and percentages of exposed subjects
- Number and percentages of randomized subjects who completed the DBPC treatment phase. For the 2-year analysis a completer is a subject with no date of study discontinuation before day 819.
- Number and percentages of randomized subjects who discontinued the trial after randomization, during DPBC i.e. before day 819, and during extended follow up phase (grouped by treatment) and main reason (AE, death, lost to follow-up, etc).
- Number and percentages of randomized subjects who discontinued treatment after randomization, during DPBC and during extended follow up phase (grouped by treatment) and main reason (AE, death, lost to follow-up, etc).
- Number of subjects in each analysis set and percentages of total enrolled
- Number and percentage of subjects completing the extended follow up phase.

For subjects who discontinued the trial summary statistics will also be presented for time to study termination for the 2-year analysis this corresponds only to early discontinuation i.e. before day 819.

Subject disposition will also be presented per country.

Time to study termination in days will be calculated as follows:

=Date of study termination - Date of Randomization +1

A by-subject listing of Analysis Set details will be presented by treatment group including: center, subject identifier, inclusion/exclusion flag for each Analysis Set and reason for exclusion from each Analysis Set.

12.1 Protocol Deviations

A table will be produced summarizing, by treatment group, the number and percentage of subjects with a major protocol deviation, overall and by type of deviation (ITT and mITT Analysis Sets).

All protocol deviations will be listed and each will be marked whether it is major or minor. Major clinical deviations who affect efficacy will result in exclusion from the PP Analysis Set.

Major protocol deviations affecting efficacy include: Prohibited medication/procedures during the treatment period and inadequate treatment compliance i.e. deviations from the planned amount of active dose and schedule of treatment (out of allowed window as specified in the "Analysis Sets" document). Subjects who have (actual total dose)/(planned total dose)*100 within 75-100 % will be considered treatment compliant. If subject was randomized to placebo group but instead received at least one dose of sprifermin, then this subject will be classified as having major protocol deviation. Additionally subjects identified in the Note To File: "NTF_QC of primary endpoint MRI_FORWARD_17MAY2016", will be excluded from the PP.

13 Demographics and Other Baseline Characteristics

Demographics and other baseline characteristics will be presented descriptively for the analysis sets described in table 2 (Section 10). Moreover, demographics will be presented by country.

By-subject listings accompanying each demographic and baseline characteristics table will be produced.

13.1 Demographics

Demographic variables include age, sex, height, weight, BMI, race, ethnicity, and country.

13.2 Medical History

Medical history (other than knee OA) will be summarized descriptively by MedDRA SOC and PT.

13.3 Other Baseline Characteristics

Other Baseline Characteristics include the following variables:

- Alcohol and nicotine consumption
- Specification of target knee OA: (KL Grade 2 or 3, target knee location (Left or Right knee), medial and lateral mJSW, Anatomic axis angle (AAA) of knee alignment [in degrees calculated as 180 – ANGLE, where ANGLE is the value corresponding to xmtestcd = 'ANGLE'], response to Question 1 of the WOMAC pain index at screening, and time since diagnosis
- Contralateral knee: KL Grade (0, 1, 2, 3), medial and lateral mJSW

- Pain in other joints (yes/no, and if yes, how many joints and which joints specifically)
- All other attributes defined in subgroups analysis set in Section 10.

14 Previous or Concomitant Medications/Procedures

Medication start and stop dates will be compared to the date of first dose of study medication, see appendix 20.3, to allow medications to be classified as

(a) Previous,

(b) Concomitant, or

(c) Post-treatment medications.

All medications categories will be summarized and listed. Medications that start and stop prior to the date of first dose of study medication will be classified as previous medication. If a medication starts or stops on or after the date of first dose of study medication **and prior to last dose**, then the medication will be classified as Concomitant. Medications that start **after last injection** will not be considered concomitant and will be categorized as **Post-treatment** medications. If medication start and/or stop dates are missing or partial, the dates will be compared with the date of first dose of study medication. Medications will be assumed to be Concomitant, unless there is clear evidence (through comparison of partial dates) to suggest that the medication stopped prior to the first dose of study medication (in which case the medication will be assumed to be Previous medication) or started after last injection (in which case the medication will be assumed to be Post-treatment medication).

Pain and other medication such as NSAIDs taken in each of the three categories (a,b,c) described above will be summarized and listed. List of pain medication is attached in Appendix 19.2.

All medications will be coded using the latest version of the WhoDrug Anatomic therapeutic class (ATC) Medications using the indication or using the secondary level of the ATC code. Coding will be updated for the 5-year analysis.

Procedures will be listed only.

15 Treatment Compliance and Exposure

Exposure to sprifermin will be summarized descriptively in terms of:

- Duration of exposure = date of last injection – date of first injection +1
- Total dose (µg)= sum of (number of X µg dose injections * X µg)
- Listing of treatment exposure including time between injections within a cycle and time between last injection of previous cycle and first injection of next cycle will be provided.

Treatment compliance will be summarized descriptively in terms of:

- Mean number of injections by cycle and overall;

- Number of subjects (and percentage) receiving 1,2,3….12 injections; and
- Number of subjects (and percentage) with 75-100 % treatment compliance. Subjects who have (actual total dose)/(planned total dose)*100 within 75-100 % will be considered treatment compliant. Treatment compliance will not be summarized for the placebo group.
- Listing of subjects receiving wrong medication will be provided.

16 Endpoint Evaluation

This section presents specifications for analyzing the trial endpoints specified in the Clinical Trial Protocol to meet the trial objective. Endpoints not identified in the Clinical Trial Protocol are also presented here.

16.1 Primary Efficacy Endpoint Analyses

The primary endpoint for the DBPC treatment phase is the change from baseline in cartilage thickness in the total femorotibial joint as evaluated by quantitative MRI at 2 years in the mITT. Cartilage thickness of the total femorotibial joint will be calculated in two ways:

1. Average Cartilage Thickness (Total Volume divided by Total Surface Area): (MFTC.VC+LFTC.VC)/(cMF.tAB+cLF.tAB+MT.tAB+LT.tAB)

2. Total Cartilage Thickness (sum of cartilage thickness in medial and lateral compartment): MFTC_ThCtAB_aMe + LFTC_ThCtAB_aMe

These 2 variables are highly correlated, but since the first one (average) was used in the previous study as primary endpoint and power was based on this variable, this will be also the variable to be used as primary endpoint. Of these two (1) is considered the primary, and only this is summarized and analyzed, (2) is just derived in ADaM. Definitions of these variables are provided in the imaging transfer specifications document.

Summary statistics

Summary statistics for change from baseline in cartilage thickness in the total femorotibial joint will be presented both for the DBPC treatment phase and for the follow-up phase.

Summary statistics will be provided per timepoint (including LOCF) for:

- Cartilage thickness in the femorotibial joint as evaluated by quantitative MRI
- absolute change from baseline in cartilage thickness by quantitative MRI
- percentage change from baseline in cartilage thickness by quantitative MRI

Primary analyses of linear dose-relationship and treatment effect

The primary efficacy analyses will consist of testing the linear dose relationship at 2 years and the treatment effect and will be presented for the DBPC treatment phase.

The significance level will be set at 5% two-sided.

For the linear dose relationship testing, the null and alternative hypotheses will be conducted using the following linear contrast testing:

H0: -2 Mean (Placebo) - 1 Mean (180 mcg) + 0 Mean (360 mcg) + 1 Mean (600 mcg) + 2 Mean (1200 mcg) = 0
H1: -2 Mean (Placebo) - 1 Mean (180 mcg) + 0 Mean (360 mcg) + 1 Mean (600 mcg) + 2 Mean (1200 mcg) ≠ 0

where the doses shown are the total doses of sprifermin to be administered over the DBPC treatment phase.

The treatment effect on the primary endpoint will be assessed through dose-ranging using a repeated measurement analysis of variance (ANOVA) (using PROC MIXED in SAS) on absolute change from baseline, including the treatment group, the time point, and the (pooled) country as fixed factors and the baseline value as covariate and treatment by time point as interaction, see "ANOVA models" in section 11. Repeated measures over time will be accounted for using an 'unstructured' covariance pattern, and if the model fails to converge 'compound symmetry' will be used.

If the null hypothesis is rejected, the alternative hypothesis will be accepted and it will be concluded that there is a dose relationship versus placebo over time.

The analysis of the primary efficacy endpoint will also consist of testing the overall treatment effect at 2 years and will therefore only be conducted for the DBPC treatment phase.

The significance level will be set at 5% two-sided.

For the overall treatment effect testing, the null and alternative hypotheses will be:

H0: Mean (Placebo) = Mean (180 mcg) = Mean (360 mcg) = Mean (600 mcg) = Mean (1200 mcg)
H1: Mean (i) ≠ Mean (j) for some i ≠ j and i, j = Placebo or 180 mcg or 360 mcg or 600 mcg or 1200 mcg.

where the doses shown are the total doses of sprifermin to be administered over the DBPC treatment phase.

If the null hypothesis is rejected, the alternative hypothesis will be accepted and it will be concluded that there is at least one group that differs from the other groups.

Pairwise comparisons of absolute change from baseline in cartilage thickness (sprifermin treatment groups versus placebo) will be performed within the context of this modeling framework. For each pairwise comparison, the difference between treatments and the corresponding 95% CI and p-value will be presented. P-values (corresponding to Type 3 tests of fixed effects) will be reported for all covariates in the original 'Overall' model for all time points combined (i.e. baseline value, treatment, timepoint, treatment-by-timepoint interaction, country) and for all timepoints. Estimated coefficients, p-values and 95% CIs will be presented overall and at each timepoint for

(i) the dose relationship (linear trend); and, (ii) each pairwise comparison between dose level and placebo.

A plot of means and 95% CI including all time points (including LOCF) per treatment group for dose relationship will be presented. All treatment groups will be presented in one plot.

The primary efficacy analysis as well as secondary MRI endpoints will be performed on the mITT Analysis Set.

Sensitivity analysis using Per Protocol Analysis Set

In order to assess the robustness of the primary analysis, the test for linear dose-relationship and for the overall treatment effect will be repeated using the Per-Protocol Analysis Set.

Sensitivity analysis regarding fixed effects and covariates in the model

A non-parametric analysis will be conducted for the ordered data of cartilage thickness in the total femorotibial joint as an alternative method for the primary analysis. Data will be ordered by the magnitude of absolute change-from-baseline over 2 years during DBPC treatment phase using rank transformation. This will be done only for the mITT Analysis Set.

16.2 Secondary Efficacy Endpoint Analyses

Secondary endpoints related to structure:

Cartilage thickness

Cartilage thickness in the medial and lateral compartments will be summarised per timepoint (including LOCF) along with absolute and relative change from baseline. These will be calculated in two ways corresponding to the calculation of cartilage Average and Total thickness:

1. Medial
    a. MFTC.VC/(cMF.tAB+MT.tAB), primary
    b. MFTC_ThCtAB_aMe
2. Lateral:
    a. LFTC.VC/(cLF.tAB+LT.tAB), primary
    b. LFTC_ThCtAB_aMe

All these variables will be included in the ADAM. Only (1a) and (2a) will be used in summaries and analyses, the others (1b) and (2b) will only be derived in ADaM.

The absolute change from baseline in the medial and lateral compartments over 2 years will be analysed using a similar ANOVA model as the model used for the primary endpoint. Similarly to the analysis of the primary end-point, sensitivity analyses on ranked data will also be performed.

Cartilage Volume

Cartilage volume in the medial and lateral compartments as well as in the total femorotibial joint will be summarised per timepoint (including LOCF) along with absolute and relative change from baseline. These will be calculated as MFTC.VC, LFTC.VC, and MFTC.VC + LFTC.VC respectively.

The absolute change from baseline over 2 years will be analysed using a similar ANOVA model as the model used for the primary endpoint.

Joint Space Width

Minimal JSW (mJSW) in the medial and lateral compartments will be summarised per timepoint (including LOCF) as well as absolute and relative change from baseline for target knee as well as for contralateral knee.

The absolute change from baseline over 2 years will be analysed using an ANOVA model similar to the primary endpoint for target knee only.

Secondary endpoints related to function and symptoms:

WOMAC Total Score

WOMAC total scores (and WOMAC sub scores, as described below) will be displayed descriptively after being transformed to a scale of 0 to 100 using the formula described below.

All items have 11 possible answers ranging from 0 (no problems) to 10 (extreme problems). Given items I1 to Ix as the answers to each questionnaire item on a given subscale and the raw score as the *arithmetic mean* of the respective non-missing answers of items I1 to Ix, the subscale summary score will be calculated as:

$$\left[\frac{(mean\ score\ of\ subscale)}{10}\right] \cdot 100$$

The WOMAC total score, and absolute changes from baseline in the WOMAC total score for the target knee will be presented with summary statistics for all timepoints (including baseline, post-treatment, and LOCF).

Absolute change from baseline in WOMAC total score over 2 years will be considered a continuous variable and will be analysed using a model similar to the ANOVA model used for the primary endpoint.

If an entire questionnaire is missing for a time point, no imputations will be performed for that time point. If some items are missing, imputation for Total score (and for sub scores) will be performed using same principal used for KOOS.

WOMAC Sub Scores

WOMAC sub scores (weight-bearing pain score (sum of items: walking, going up and down stairs, and standing), pain, function and stiffness) will be tabulated descriptively as described above for Total score.

Absolute change from baseline in pain, function, stiffness, and weight-bearing pain, over 2 years will be analysed using a model similar to the ANOVA model described under primary endpoint.

WALK test

Time needed to complete the 20-meter walk tests as well as absolute change from baseline will be presented with summary statistics per timepoint (including LOCF). No imputation for missing data will be done. Absolute change from baseline in time needed to complete the 20-meter walk will be analysed using a model similar to the ANOVA model described under primary endpoint.

Patient Global Assessment (PGA)

The Patient Global Assessment is based on subject's answer to the question "Considering all the ways your osteoarthritis of the knee has affected you during the last 48 Hours, select the number that best describes the impact of your knee osteoarthritis on your daily life", and can take on values between 0-10 (0=None, 10=Extreme), for summaries the values are rescaled to 0-100 by multiplication with 10.

The Patient Global Assessment will be presented with summary statistics per timepoint (including LOCF) as well as absolute change from baseline. Mean score will be presented as described for continuous variables, and also number and percentages per score will be presented. No imputation for missing data will be done (except LOCF).

16.3 Pharmacokinetic Endpoint Analyses

Pharmacokinetics (PK)

Analyses relating to this endpoint will be conducted by Merck Serono QPD and ECS using the PK Analysis Set.

The data will be reviewed for implausible concentration data. Exclusion of implausible values will be decided by the responsible PK/Pharmacodynamic(s) (PD) representative and/or the Clinical Pharmacology Representative and will be documented in the CTR. If necessary, the subject will be excluded from analysis.

Quality control procedures for the pharmacokinetic analysis will be performed in accordance with the QPD Standard Operating Procedures (SOPs) at Merck Serono.

In case no measurable sprifermin could be detected either in serum or in synovial fluid this will be described in the CTR but no Tables, Listings and Figures (TLFs) will be provided.

Pharmacokinetics based on concentrations measured in the serum

Blood samples for the quantification of sprifermin are collected pre-dose at the baseline visit Week 0/Visit 2, 2 hours after the last dose of each cycle and at the end of the DBPC treatment phase (Visit 22).

Serum levels of sprifermin (pg/mL) will be determined under the responsibility of QPD and will be stored in the PPD . QPD will produce a separate bio-analytical report. The lower limit of quantification will be 100 pg/mL.

Before DBPC database lock reconciliation of blinded data will be performed. After DBPC database lock final data will be mapped to SDTM and returned to Merck Serono Data Management.

QPD will be responsible for the descriptive statistics on the serum concentrations of sprifermin and TLFs will be provided for the CTR, as applicable.

The following descriptive statistics will be presented for the PK Analysis Set by treatment group and time-point: Number (N) of observations, mean, SD, median, min, max, and coefficient of variation (%CV), if applicable. Individual and mean concentrations will be plotted by treatment group and time-point, if applicable.

Pharmacokinetics based on concentrations measured in the synovial fluid

Synovial fluid for the quantification of sprifermin is collected at visits where IMP is injected. Synovial fluid will be collected as part of the i.a. injection procedure, thus pre-dose synovial fluid samples are collected prior each IMP injection.
Synovial fluid levels of sprifermin will be determined under the responsibility of QPD and will be stored in the PPD . QPD will produce a separate bioanalytical report. The lower limit of quantification will be 150 pg/mL.

Before DBPC database lock reconciliation of blinded data will be performed. After DBPC database lock final data will be mapped to SDTM and returned to Merck Serono Data Management. QPD will be responsible for the descriptive statistics on the synovial concentrations of sprifermin and TLFs will be provided for the CTR, as applicable. The following descriptive statistics will be presented for the PK Analysis Set by treatment group and time point: Number (N) of observations, mean, SD, standard error of the mean (SEM), median, min, max, and %CV, if applicable. Individual and mean concentrations will be plotted by treatment group and time-point, if applicable.

Occurrence of antibodies against FGF18 (FGF18ab)

Analyses relating to this endpoint will be conducted by Merck Serono QPD and ECS using the Safety Analysis Set and will be reported in the Safety Section of the CTR.

Blood samples for the determination of the presence of anti-FGF-18 are collected at the baseline visit Week 0/Visit 2, at the first dose of each cycle, 1 week after the last dose of each treatment cycle (at Weeks 3, 29, 55 and 81/Visit 5, 10, 15 and 20) and at the end of the DBPC treatment phase (Week 104/Visit 22).

The presence of anti-FGF-18 antibodies will be determined under the responsibility of QPD and will be stored in the PPD . QPD will produce a separate bioanalytical report.

Before DBPC database lock reconciliation of blinded data will be performed. After DBPC database lock final data will be mapped to SDTM and returned to Merck Serono Data Management. QPD will be responsible for the data analyses and TLFs for the CTR. The data will be summarized by overall frequency and by treatment and time point, if applicable.

16.4 Biomarker subgroups

PGX markers

SNPs located in the Interleukin 1 Receptor Antagonist (IL1RN) gene are known to be linked to disease severity and predictive of progression [3, 4]. Subgroups based on distinct genotypes resulting from the combination of the SNPs rs9005 and rs315952 in the IL1RN cluster will be considered as exploratory objective based on a retrospective analysis of subjects with a defined haplotype in the sprifermin Phase 1b PoC study (28980), suggesting the potential identification of subgroups of subjects who can be classified as progressors and/or responders to treatment. Considering these results, the following biomarker subgroups (as well as the subgroups in the single SNPs) will be evaluated in this study:

Subgroup A: rs9005 GG/rs315952 TT

Subgroup B: rs9005 A carriers (AG or AA)/rs315952 TT

Subgroup C: rs9005 GG/rs315952 C carriers (CT or CC)

Subgroup D: rs9005 A carriers/rs315952 C carriers

These subgroups will be evaluated by summarizing structure (MRI and X-ray) and symptom (WOMAC) endpoints by treatment groups in the PGX population (n, number of missing values, mean, SD, median, Q1, Q3, min, and max). Supporting figures will also be provided. To facilitate informal comparisons between active treatment groups and placebo group, the mean difference from placebo in absolute change including 95% CIs will be calculated and presented in tables and figures by timepoint. If treatment effect seems consistent across treatment groups, analyses pooling the different treatment groups will also be performed. The occurrences of AIRs (frequency and 95% exact confidence interval) will be summarized by genetic groups described above.

Demographic and baseline characteristics as specified in Section 13 will be summarized for the PGX population.

Biochemical markers of joint tissue inflammation

The following blood-based biochemical markers are known to be linked to joint inflammation and ECM turnover through the upregulation of MMP observed in OA: C1M, C3M, and CRPM [5]. hsCRP is a measure of acute inflammation and a general marker of systemic inflammation [5].

Two subgroups resulting from categorizing each of the inflammation markers at baseline (and all possible 2-way pairwise combinations of two of them [i.e. C1M, C3M, CRPM, C1M*C3M, C1M*CRPM, and C3M*CRPM], to support the hypothesis that subjects with high localized inflammation show low structural and symptomatic response to Sprifermin) as marker falling into highest tertile (=33% of all subjects in FORWARD with highest marker at baseline), and marker falling into the two lower tertiles will be evaluated separately for structure (MRI and X-ray) and symptom (WOMAC) endpoints by treatment groups in the ITT population (n, number of missing values, mean, SD, median, Q1, Q3, min, and max). The general systemic inflammation marker hsCRP will be used as a reference for the overall inflammation status and will be evaluated for its association with the other markers.

In addition, Baseline CRPM will be analysed with cutoff of 10 (Baseline CRPM ≤ 10ng/mL, and >10 ng/mL).

Supporting figures will also be provided. To facilitate informal comparisons between active treatment groups and placebo group, the mean difference from placebo in absolute change including 95% CIs will be calculated and presented in tables and figures by timepoint. If treatment effect seems consistent across treatment groups, analyses pooling the different treatment groups will also be performed.

The occurrences of AIRs (frequency and 95% exact confidence interval) will be summarized by subgroups for C1M, C3M, and CRPM.

Biochemical markers of cartilage metabolism

The following biochemical markers of cartilage metabolism have been used to describe the metabolic status of cartilage by analysis of data out of previous clinical studies with sprifermin as well as in non-clinical investigations. These are urinary CTX-II (creatinine adjusted), a catabolic marker of collagen degradation [6, 7], and proC2 alias PIINP, an anabolic marker of collagen synthesis [8].

Subgroups resulting from categorizing of each of these 2 biomarkers at Baseline (low CTX-II vs. high CTX-II, and high proC2 vs. low proC2, to support the hypothesis that subjects with high anabolic or low catabolic activity show low structural and symptomatic response to Sprifermin) as marker CTX-II falling into lowest tertile (=33% of all subjects in FORWARD with lowest marker at baseline), and marker CTX-II falling into the two higher tertiles. For marker proC2: proC2 falling into highest tertile (=33% of all subjects in FORWARD with highest marker at baseline), and marker falling into the two lower tertiles) will be evaluated separately for structure (MRI and

X-ray) and symptom (WOMAC) endpoints by treatment groups in the ITT population (n, number of missing values, mean, SD, median, Q1, Q3, min, and max).

In addition, Baseline CTX-II will be anlaysed with cutoff of 350 (Baseline CTX-II ≤ 350ng/mmol, and >350ng/mmol).

Supporting figures will also be provided. To facilitate informal comparisons between active treatment groups and placebo group, the mean difference from placebo in absolute change including 95% CIs will be calculated and presented in tables and figures by timepoint. If treatment effect seems consistent across treatment groups, analyses pooling the different treatment groups will also be performed.

16.5 Other Endpoint Analyses

Outcome Measures in Rheumatology

Based on the assessments of WOMAC pain subscore, WOMAC function subscore, and PGA, ranges are defined for absolute and percent changes from baseline that correspond to "high improvement" and "moderate improvement".

OMERACT-OARSI response is defined [9] as either

i) high improvement (relative change ≤ -50% and absolute change ≤ -20) in WOMAC pain or WOMAC function or
ii) moderate improvement (relative change ≤ -20% and absolute change ≤ - 10) in at least 2 of the 3 domains (WOMAC pain score, WOMAC function score, and PGA score).

Responder rate per domain will be summarized per treatment and analysis time point (including LOCF) along with absolute and percentage change-from-baseline in WOMAC pain, WOMAC function, and PGA score.

No imputation for missing data will be done.

Responder analyses for 2 years will be analysed using Logistic regression (SAS Proc Genmod), for outcomes response, high improvement, and moderate improvement. Treatment and pooled country will be included in the model:

```
proc genmod data=qs2 desc ;

  class grpn CNTRGR1N;

  model aval = grpn CNTRGR1N / dist = bin link = logit type3;

  estimate '30µg x2 vs Placebo'  grpn -1 1 0 0 0 / exp ;

  estimate '30µg x4 vs Placebo'  grpn -1 0 1 0 0 / exp ;

  estimate '100µg x2 vs Placebo' grpn -1 0 0 1 0 / exp ;
```

```
  estimate '100µg x4 vs Placebo' grpn -1 0 0 0 1 / exp ;

run ;
```

This analysis will be performed only if more than 10% of all subjects in the ITT analysis set satisfy condition i) or ii) above.

Patient Global Impression of Change (PGIC)

The PGIC score is based on subject's answer to the question: "Since getting the first injection of study treatment in your knee how would you describe the change in overall symptoms for this knee" and can take on values between 'Very much improved' 'Very much Worse' with 7 possible choices. The PGIC will be summarised by treatment and analysis timepoint (including LOCF). No imputation for missing data will be done.

Knee Outcome in Osteoarthritis Score (KOOS)

In this study the following two knee-related subscales of the KOOS are collected:

- Symptoms (items S1 to S5);
- Quality of life (items Q1 to Q4).

All items have five possible answers ranging from 0 (no problems) to 4 (extreme problems). Given items I1 to Ix as the answers to each questionnaire item on a given subscale and the raw score as the *arithmetic mean* of the respective non-missing answers of items I1 to Ix, the subscale summary score will be calculated as:

$$\left[\frac{(mean\ score)}{4}\right] \cdot 100$$

As long as at least 50% of the subscale items are answered for each subscale for a specific visit, a mean score can be imputed using the mean of the non-missing items for that visit. If more than 50% of the subscale items are omitted, the response is considered invalid and no subscale score should be calculated. For KOOS Symptoms, this means that 3 items must be answered; for KOOS QOL, 2 items in order to calculate a subscale score.

This imputation is equivalent to taking the arithmetic mean of the non-missing items. Summary scores range from 0 to 100, with zero representing no knee problems and 100 representing extreme knee problems. An overall summary score (both subscales combined) will not be calculated.

Summary statistics for summary scores, and its absolute change from baseline, will be presented for each subscale, by treatment group and timepoint.

An accompanying by-subject listing will be provided including: treatment; country; age; sex; BMI; response for each KOOS item at each visit; each subscale summary score at each analysis time point; change from baseline in each subscale summary score at each visit (ITT Analysis Set).

The Medical Outcomes Survey Short Form 36 (MOS SF-36) Score

The MOS SF-36 consists of the following 8 sub-scores (with corresponding items within each sub-score are scored on 2-, 3-, 5-, or 6-point scale) and will be transformed to a 0-100 scale as described in [10] and in the table below:

- general health (GH, based on 5 items [Q1,Q11a, Q11b, Q11c, Q11d]),
- physical function (PF, based on 10 items [Q3a-Q3j]),
- physical role limitations (RP, based on 4 items [Q4a-Q4d]),
- bodily pain (BP, based on 2 items [Q7, Q8]),
- vitality (VT, based on 4 items [Q9a, Q9e, Q9g, Q9i]),
- social function (SF, based on 2 items [Q6, Q10]),
- emotional role limitations (RE, based on 3 items [Q5a-Q5c])
- mental health (MH, based on 5 items [Q9b, Q9c, Q9d, Q9f, Q9h]).

and a total score (based on 36 items [all above+Q2]).

Each scale is directly transformed into a 0-100 scale on the assumption that each question carries equal weight. All of the scales are scored so that the least health has a value of 0 and the greatest health has a value of 100.

Table 4 Scoring transformation for MOS SF-36

Sub score	# of items and range of scale	Total possible score	Items need to be reversed (max-item+1)	Normalizing Algorithm to 0-100= [(item[or average of items]min)/range]*100
GH	5 items with 1-5	5-25	Q1, Q11b, Q11d	100*(average-1)/4
PF	10 items with 1-3	10-30		100* (average-1)/2
RP	4 items with 1-5	4-20		100* (average-1)/4
BP	2 items with 1-6 and 1-5	2-11	Q7, Q8	100*{[(Q7-1)/5]+[(Q8-1)/4]}/2
VT	4 items with 1-5	4-20	Q9a, Q9e	100* (average-1)/4
SF	2 items with 1-5	2-10	Q6	100* (average-1)/4

RE	3 items with 1-5	3-15		100* (average-1)/4
MH	5 items with 1-5	5-25	Q9d, Q9h	100* (average-1)/4
Total	35 items	35-156	10 items	NA
Total score	**36 items***	**36-161**	**11 items ****	100*{[sum of 10 items (item-1)/2]+[sum of 25 items (item-1)/4]+[sum of 1 item (item-1)/5]}/36

*Q2 is only included in Total score (range of Q2: 1-5); **Q2 need to be reversed.

As long as at least 50% of the sub score items are answered for each sub score for a specific visit, a mean score can be imputed using the mean of the non-missing items for that visit. If more than 50% of the sub score items are omitted, the response is considered invalid and no sub score should be calculated. Similarly, same principle will be applied to Total score.

Sub scores and total score will be summarised by treatment and timepoint (including LOCF) along with the absolute change from baseline. Individual items as well as derived sub scores will be presented in listings.

The Numeric Rating Scale (NRS) for pain score

The NRS pain score in the target and contralateral knee is based on subject's answer to the question "Please indicate the worst pain in the past 24 hours in the target (contralateral) knee" and can take on values between 0-10 (0=No Pain, 10=Worst Pain imaginable).

The NRS will be presented with summary statistics by treatment and timepoint (including LOCF) as well as absolute change from baseline for target knee and contralateral separately. No imputation for missing data will be done. Individual items will be listed.

Pain in other joints

Pain in other joints will be summarised per joint and for number of subjects having pain in at least one joint. Data will be summarised by treatment and timepoint (including LOCF). No imputation for missing data will be done.

The Intermittent and Constant OA Pain (ICOAP) score

ICOAP is based on subject's response to 11 questions designed to assess OA knee pain that is "constant" and pain that is "intermittent" [11].

To calculate the "constant" pain subscale, the sum of items 1 to 5 (each on the scale of 0 [not at all/I don't have pain] to 4 [Extremely]) will be derived as total pain score with a range of 0-20. This score can be transformed to a score out of 100 using the following formula: (total/20) x 100. Similarly, to calculate the "intermittent" pain subscale the sum of items 6 to 11 (each on the scale

of 0 [not at all/I don't have pain] to 4 [Extremely/very often]) will be derived as total pain score with a range of 0-24. This score can be transformed to a score out of 100 using the following formula: (total/24) x 100. Total pain score is the sum of "constant" pain subscale and "intermittent" pain subscale (range 0-44). Total pain score can be transformed to a score out of 100 using the following formula: (total/44) x 100.

If there are 3 or more items missing for each subscale for a visit, the response is considered invalid. If there are less than 3 items missing, the missing item can be replaced with the mean of the response to other items within the same subscale.

Total pain score and "constant" and "intermittent" pain subscores will be summarised by treatment and timepoint (including LOCF) as will absolute change from baseline.

Individual items will be listed.

Change over time in structural features of the knee joint (e.g., synovium, menisci, bone, and other structures) as evaluated by semiquantitative MRI

The protocol specifies that both WORMS and BLOKS will be used for semi-quantitative assessment of the target knee MRI, but does not specifically state at which of the 8 MRI-acquiring timepoints WORMS and BLOKS will be used. Moreover, as many WORMS and BLOKS parameters are redundant, and not all BLOKS measures convey good longitudinal data [12], it was decided to evaluate MRI using WORMS and partial BLOKS (cartilage + meniscal parameters) only at the following time points: Baseline, Year 1, 2 and 3.

The WORMS compartments and subregions and scoring details are provided in Table 5 and Table 6, respectively below.

BLOKS (cartilage + meniscal parameters) will be evaluated using regions specified in Table 7.

Table 5 WORMS compartments and subregions

Compartment	**Subregion**
Medial femoro-tibial joint (MFTJ)	MFc: Medial Femur central MFp: Medial Femur posterior MTa: Medial Tibia anterior MTc: Medial Tibia central MTp: Medial Tibia posterior
Lateral femoro-tibial joint (LFTJ)	LFc: Lateral Femur central LFp: Lateral Femur posterior LTa: Lateral Tibia anterior LTc: Lateral Tibia central LTp: Lateral Tibia posterior
Patello-Femoral Joint (PFJ)	MP: Medial Patella LP: Lateral Patella MFa: Medial Femur anterior LFa: Lateral Femur anterior

Table 6 WORMS scores

WORMS score	Variable	Compartment	Derivation
Cartilage	Cartilage signal, Cartilage morphology (only morphology in TLFs)	MFTJ LFTJ PFJ Entire knee	**MFTJ, LFTJ, PFJ** Sum of the regions of each compartment **Entire knee** Sum of the 3 compartments
Bone Marrow Edema	Bone Marrow Edema	MFTJ LFTJ PFJ Entire knee	**MFTJ, LFTJ, PFJ** Sum of the regions of each compartment **Entire knee** Sum of the 3 compartments
Subarticular cysts	Subarticular cysts	MFTJ LFTJ PFJ Entire knee	**MFTJ, LFTJ, PFJ** Sum of the regions of each compartment **Entire knee** Sum of the 3 compartments
Bone attrition	Bone attrition	MFTJ LFTJ PFJ Entire knee	**MFTJ, LFTJ, PFJ** Sum of the regions of each compartment **Entire knee** Sum of the 3 compartments
Osteophytes	Osteophytes	MFTJ LFTJ PFJ Entire knee	Sum of the regions of each compartment as described below: **MFTJ (**central femur, central tibia, posterior femur, posterior tibia) **LFTJ** (central femur, central tibia, posterior femur, posterior tibia) **PFJ** (superior patella, inferior patella, lateral patella, medial patella, anterior femur medial, anterior femur lateral, anterior tibia medial, anterior tibia lateral) **Entire knee** Sum of the 3 compartments
Ligaments	Anterior Cruciate Ligament (ACL) Posterior Cruciate		

	Ligament (PCL) Medial Collateral Ligament (MCL) Lateral Collateral Ligament (LCL)		
Menisci-Damage	Menisci Medial Anterior Menisci Medial Body Menisci Medial Posterior Menisci Lateral Anterior Menisci Lateral Body Menisci Lateral Posterior		
Menisci – Subluxation	Menisci Medial Menisci Lateral	Medial extrusion Lateral extrusion	**Medial, Lateral**
Hoffa-Synovitis	Hoffa fat pad intensity	Entire knee	
Effusion-Synovitis	Joint effusion	Entire knee	N/A
Synovial Cysts and Bursae	Anserine Bursitis Infra- or Prepatellar Bursitis Lateral Meniscal Cyst Medial Meniscal Cyst Tibiofibular Cyst Baker's Cyst		N/A

Each of the WORMS scores for the entire knee and for each compartment will be tabulated per time point, as well as the absolute change from baseline. Analyses will be performed on a whole knee level and separately for the compartments given in table 6 for all subgroups. Synovial cysts and bursae are summarized using shift tables. For all but Baker´s cysts, the values are "Present"/"Absent". For Baker´s cysts the possible values are "0", "POP1", "POP2", and "POP3" (1=small size Baker's, 2=medium size Baker's, and, 3=large size Baker's).

In addition, for the purpose of comparison of sprifermin doses vs. placebo, absolute differences from placebo in mean change from baseline in WORMS scores for the entire knee, as well for each compartment, for each sprifermin doses will also be tabulated at Week 52, Week 104 and Week 156. For the WORMS scoring of the target knee joint, the following information will be listed, in addition to core variables: time point, date of the MRI, for each score: variables used for the derivation, scores per compartment, absolute change from baseline.

For the calculation of the WORMS scores:

- If a region is graded "-9" (Unscorable) or "0.1" (Progression unscorable, coded as "NA" in the data), then the region's grade is set to missing. No imputation for missing data for WORMS will be done.

- If a region is graded "0.5" (Worsening (within grade), coded as "+") or "-0.5" (Improvement (within grade), coded as "-"), then the region's grade is set to the last available existing grade within the previous visits, for the purpose of calculating change from baseline on numeric scale.

The partial BLOKS scoring system will be assessed using subregions for cartilage and menisci defined in Table 7. The scoring approach of knee features to be analyzed is described in Table 7.

Table 7 BLOKS scores

Cartilage	10 different scores possible (0.0, 1.0, 1.1, 2.0, 2.1, 2.2, 3.0, 3.1, 3.2, 3.3), with the first number representing the score for the size of cartilage loss (% of surface area, 0: none; 1: <10%; 2: 10-75%; 3: >75%), and the second number representing the score for the full thickness cartilage loss (% full thickness, 0: none; 1: <10%; 2: 10-75%; 3: >75%). Each score is therefore a categorical variable, and is derived for 14 different regions, matching the WORMS regions of table 5.
Meniscal extrusion	Score of 0-3 applied for amount of extrusion in 2 locations: Medial meniscus (medial extrusion), Lateral meniscus (lateral extrusion)
Meniscus tear	Signal change, type of tear and cyst are scored as present (score 1) or absent (score 0) in 6 regions and maceration is scored as absent/partial/progressive (follow-up only)/complete (38 variables in total): - Medial meniscus anterior (intrameniscal signal, vertical tear, horizontal tear, complex tear, maceration, meniscal cyst), - Medial meniscus body (intrameniscal signal, vertical tear, horizontal tear, complex tear, maceration, meniscal cyst), - Medial meniscus posterior horn (intrameniscal signal, vertical tear, horizontal tear, complex tear, root tear, maceration, meniscal cyst), - Lateral meniscus anterior (intrameniscal signal, vertical tear, horizontal tear, complex tear, maceration, meniscal cyst), - Lateral meniscus body (intrameniscal signal, vertical tear, horizontal tear, complex tear, maceration, meniscal cyst),

	- Lateral meniscus posterior horn (intrameniscal signal, vertical tear, horizontal tear, complex tear, root tear, maceration, meniscal cyst)

Cartilage:

Shift tables will be used to summarize the change in cartilage score from baseline to all post baseline time points, for each region. The number and percentage of subjects in each cartilage category (0.0, 1.0, 1.1, 2.0, 2.1, 2.2, 3.0, 3.1, 3.2, 3.3) and the number with missing values will be presented at each time point according to their baseline classification, by treatment group. Additionally the scores +/-.0, +/-.1, +/-.2, and +/-.3, are possible and as for WORMS the result from the previous visit will be used instead. If both parts of the score are missing these will be considered missing, otherwise (only one part is missing), since the missing part is imputed from previous visit it will not be counted as missing. NA is counted as missing.

Meniscal Extrusion:

Both meniscal extrusion parameters have scores ranging from 0 to 3.

Summary statistics for the individual parameters, and their absolute change from baseline, will be presented, by treatment group and timepoint. Similarly, summary score will be determined for lateral and medial extrusion locations only.

Meniscal Tear:

All 38 meniscal tear parameters apart from maceration score as either absent (0) or present (1). Maceration scores as absent/partial/complete/progressive(FU only).

Shift tables for the individual parameters, will be presented, by treatment group and timepoint.

An accompanying by-subject listing for all BLOKS scores will be provided.

Baseline protein markers and/or genetic markers associated with response to treatment or disease progression (response assessed by MRI and/or questionnaire)

This analysis will be done by Merck Serono and is described in a separate analysis plan. Exception is subgroup analyses by biomarker subgroups as specified in Section 16.4.

Relationship between dosing (dose and regimen), cartilage structure, and clinical scores as a function of time

This analysis will be done by Merck Serono Exploratory Medicine and is described in a separate analysis plan.

Subgroup analysis

The endpoints mentioned below will be presented for the subgroups defined in Section 10 and will be conducted only for the DBPC treatment phase for the ITT analysis set.

Endpoints:

- Cartilage thickness in the total femorotibial joint
- Cartilage thickness in the medial and lateral compartments
- WOMAC total score
- WOMAC pain and function sub scores
- 20 meter walk test
- Lateral and medial JSW (mJSW) in target knee

Data will be presented (descriptive only) as described for the primary efficacy endpoint.

Additional analyses for primary and secondary endpoints

In addition to analysing continuous variables, it is important to identify subjects who achieve meaningful clinical response (for structure and symptoms). Therefore the following variables will be derived and analysed:

Subjects who achieve meaningful clinical response for structure are considered responders if they have increase from baseline of at least 0.01 mm in MRI total femorotibial cartilage thickness. Similarly, subjects who achieve meaningful clinical response for symptoms are considered responders if they have a decrease from baseline of at least 10 points in Total WOMAC. In addition, a combination of the above 2 mentioned responder variables will be explored as potential composite responder variable (a decrease from baseline of 10 points for Total WOMAC and an increase from baseline of 0.01 mm for MRI Total femorotibial cartilage thickness).

Number and percent of subjects with total femorotibial joint thickness increase of at least 0.01mm from baseline will be summarized for all timepoints (including LOCF).

Number and percent of subjects with 10-point improvement (on a scale of 0-100) from baseline will be summarized for WOMAC Total and WOMAC sub scores at all timepoints (including LOCF).

Number and percent of subjects with 10-point improvement of Total WOMAC and increase of at least 0.01mm in total femorotibial joint thickness will be summarized at all timepoints (including LOCF).

Spearman's correlations and corresponding p-values between endpoints related to function and symptoms (WOMAC, 20 meter walk test, KOOS) and structure related endpoints (MRI total, medial and lateral thickness) will be calculated for all timepoints including for change from

baseline in the ITT analysis set. The number and proportion of non-missing and missing values will also be reported for each correlation.

17 Safety Evaluation

The subsections in this section include specifications for summarizing safety endpoints that are common across clinical trials such as adverse events, laboratory tests and vital signs. Antibodies to sprifermin are part of the safety evaluation and analyses will be done by QPD and ECS as specified in Section 16.3. Safety analyses will be done on the safety analysis set and according to the as-treated principle.

17.1 Adverse Events

Adverse Events will be coded using latest version of MedDRA at the end of each of 2 reporting periods (2 year and 5 year analyses).

For 2 years analysis: Treatment-emergent adverse events (TEAEs) are defined as those adverse events that either start or worsen in severity on or after the date of first dose of study treatment (first injection) until 30 days from last dose of study treatment (last injection). Where dates are missing or partially missing, adverse events will be assumed to be treatment-emergent, unless there is clear evidence (through comparison of partial dates) to suggest that the adverse event started prior to the first dose of study treatment or after the subject's last treatment (+30 days). See appendix 20.4 for details.

For 5 years analysis: All AEs from the complete duration of the study will be summarized (but not TEAEs, as these are not relevant per definition above). There will be no summaries of TEAEs for the 5 years analysis only AE

Pre-treatment emergent adverse events, i.e. adverse events starting between date of informed consent and until date of first dose of study treatment will be tabulated separately.

Adverse events of special interest (AESI) will include the following:

Injection site reactions will be summarized by treatment group and PT. These will be identified using higher level term (HLT) "Injection site reactions".

Similarly, musculoskeletal complaints (eg. Back pain, neck pain), will be identified using pre-specified SOC "Musculoskeletal and connective tissue disorders" and will be summarized by treatment group and by PT.

In addition, Neoplasms will be identified using pre-specified SOC "Neoplasms benign, malignant and unspecified (including cysts and polyps)" and will be summarized by treatment group and by PT.

17.1.1 All Adverse Events

AEs will be summarized as follows:

- Overview of all AEs
- TEAE by SOC and PT
- TEAE by severity
- TEAE by relationship
- Local TEAE by SOC and PT
- Local TEAE by severity
- Local TEAE by relationship
- Systemic TEAE by SOC and PT
- Systemic TEAE by severity
- Systemic TEAE by relationship
- Any serious adverse events (SAEs) further specified in Section 17.2.2
- Any non-serious AEs (at least 5% in any treatment group) by SOC and PT
- Most frequent TEAEs (at least 5% in any treatment group) by PT only
- Treatment emergent Injection Site Reactions by PT only (as described in Section 17.1)
- Treatment emergent Musculoskeletal disorders by PT only
- Treatment emergent Neoplasms by PT only

All AE and TEAE summaries will report the number and percentage of subjects, overall and by SOC and PT. AESI will be summarized by subjects (and %).

Each summary will be ordered alphabetically for SOC, and decreasing frequency for PT within SOC overall.

Subjects can contribute to more than one relationship or severity count per SOC/PT. In this case, the strongest relationship or worst severity will be used. If severity or relationship is missing, the worst case will be assumed.

Clinical trial.gov and EudraCT requirements: Summary table for non-serious AEs by SOC and PT (excluding SAE) with at least 5% in any treatment group will be provided.

A by-subject listing of all adverse events (including non-treatment-emergent events) will be provided. This listing will be presented by treatment group and will include: country, subject identifier, age, sex, race, target knee (left/right), adverse event (SOC, PT, and verbatim term), date of onset, date of resolution, duration, severity, seriousness, action taken, outcome and causal relationship to treatment.

17.1.2 Adverse Events Leading to Discontinuation

The following summaries will be presented

- TEAE leading to discontinuation
- Local TEAE leading to discontinuation
- Systemic TEAE leading to discontinuation

A by-subject listing of TEAEs leading to discontinuation will be provided.

The time to discontinuation due to an AE in days will be calculated as follows:

$$Date\ of\ discontinuation\ due\ to\ an\ AE - Date\ of\ first\ injection + 1$$

There are two types of discontinuations caused by an AE, the criteria are: action = drug withdrawn OR other action = led to study termination. These are summarized collectively as discontinuations due to AE. The date of discontinuation is the first date where any of the two types of discontinuations occurs.

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

17.2.1 Deaths

All AEs leading to death will be listed.

17.2.2 Serious Adverse Events

Serious adverse events will be summarized in terms of

- All SAEs
- All treatment emergent SAEs
- All treatment emergent SAEs leading to Death
- Local treatment emergent SAEs
- Systemic treatment emergent SAEs
- Serious TEAEs by relationship

By-subject listings will be provided for the all SAEs (treatment-emergent or not.)

17.2.3 Acute Inflammatory Reactions (AIRs)

Incidence of AIRs will be summarized.

AIR is defined as: increase of pain by 30 mm on a 100 mm VAS and a self-reported synovial fluid effusion (i.e., joint swelling) within 3 days following i.a. injection.

Synovial fluid effusions during the treatment period are reported in the subject diary by the subject answering 'Yes' to the question 'Swelling x Day1 before/after injection'. If this occurs within 3 days after injection, then the subject is classified as having a self-reported synovial fluid effusion within 3 days following i.a. injection.

Subjects are classified as experiencing an increase of pain by 30 mm on a 100 mm VAS from pre-dose to post-dose when the following occurs within 3 days following i.a. injection:

$$Pain\ VAS\ (highest\ value\ recorded\ within\ 3\ days\ following\ i.a.injection) - Pain\ VAS\ (last\ available\ value\ before\ i.a.injection) \geq 30mm$$

Information on pain and swelling as recorded in diary will be summarized by treatment group per injection, cycle and overall as follows:

- Number and percentage of subjects with self-reported synovial fluid effusion within 3 days following i.a. injection, for each cycle at Week 1, 2 and 3 (Yes, No);
- Number and percentage of subjects with increase of pain by 30 mm on a 100 mm VAS from pre-dose to post-dose within 3 days following i.a. injection, for each cycle at Week 1, 2 and 3 (Yes, No);
- Number and percentage of subjects experiencing an AIR, for Week 1, 2, 3 and cycle for each cycle (Yes, No).
- Number and percentage of subjects experiencing an AIR, overall (at any of the 12 injections) (Yes, No).
- Number and percentage of subjects experiencing single, multiple AIRs, or Not experiencing any AIR, overall (at any of the 12 injections).
- Number of total AIRs observed at each injection, cycle and overall (number of events, not subjects)

Fisher's exact tests will be used to test for a difference in treatment groups in the proportion of subjects experiencing AIRs. In addition, Cochran-Armitage exact trend test will be used also to evaluate whether increasing occurrences of AIRs are associated with increasing dose/injections. The number of subjects with non-missing and missing data for each item will also be presented. A by-subject listing will be provided for AIRs.

17.3 Clinical Laboratory Evaluation

The following safety laboratory parameters are measured

Hematology

- White blood cell (WBC) count
- Lymphocytes (absolute count, %)
- Monocytes (absolute count, %)

- Neutrophils (absolute count, %)
- Eosinophils (absolute count, %)
- Basophils (absolute count, %)
- Red blood cells
- Hematocrit
- Hemoglobin
- Platelets

Clinical biochemistry

- Aspartate aminotransferase (AST)
- Alanine aminotransferase (ALT)
- Alkaline phosphatase (ALP)
- Total protein
- Total bilirubin
- Creatine kinase (CK)
- Creatinine
- Calcium
- Phospate
- Sodium
- Potassium
- Magnesium
- Urea (Blood urea nitrogen (BUN))
- Glucose (not fasting)
- Chloride
- Albumin
- High-sensitivity C-reactive protein (hsCRP)
- Erythrocyte sedimentation rate (ESR)

Urinalysis

- Dipstick parameters
- pH
- Specific gravity

- Protein
- Glucose
- Ketones
- Bilirubin
- Blood
- Urobilinogen
- Leukocytes

Antibodies to sprifermin /FGF-18 (if data available)

- Binding antibodies
- Neutralizing antibodies

Coagulation parameters

- Partial thromboplastin time (PTT)
- International Normalized Ratio (INR)

Pregnancy test: Serum during screening, urine at other visits

Hematology, clinical biochemistry and urinalysis parameters and their corresponding change from baseline variables will be summarized by parameter, treatment and timepoint.

Analysis of antibodies to sprifermin will be provided by Merck Serono QPD and presented by number of observations (n), number of missing values (missing), and the count and percentage of subjects with positive findings of antibodies by treatment and time point, and will be included in the Safety Section of the CTR accordingly as specified in Section 16.3. In case of too low numbers of anti-FGF18 antibodies only a listing will be generated.

Reference ranges from the central laboratory will be used to define Lower limit of normal (LLN) and Upper limit of normal (ULN). For the purposes of laboratory parameter summaries, 'low', 'normal' and 'high' are classified as follows: low < LLN; LLN ≤ normal ≤ ULN; high > ULN.

A shift table will be used to summarize for each laboratory parameter the number and percentage of subjects experiencing low, normal, high or missing values at each timepoint according to their baseline classification, by treatment group.

Mean values of continuous parameters will be plotted per timepoint and corresponding change from baseline. One symbol per treatment group will be used. Plots will also present reference range limits. Plots will be split by gender whenever reference limits are gender-dependent.

Scatter plots of the following continuous parameters will be provided for the 2 years analysis (up to LOCF1):

- baseline vs largest post-baseline value

- baseline vs lowest post-baseline value

Plots will include 95% CI around the mean and present reference range limits.

Coagulation and pregnancy parameters collected during screening only will be listed.

Laboratory values that are outside the normal range will also be flagged in the data listings, along with corresponding normal ranges. Baseline is the last measurement prior to the first dose of any trial treatment. Subjects without post baseline laboratory samples will be excluded from analyses with respect to values after the baseline.

In addition, the last available laboratory measurement (LOCF1 for 2 years analysis and LOCF2 for 5 years analysis) will be presented.

17.4 Vital Signs

The following vital signs will be summarized by treatment and timepoint together with absolute change from baseline in the same manner specified for laboratory parameters. All analysis are for timepoints defined in table 3a and 3b.

- sitting systolic and diastolic blood pressure
- heart rate
- body temperature

All parameters will be listed by subject and visit, including all relevant criteria (as specified below) and flags.

The maximum changes of blood pressure measurements from baseline to maximum shifts (with pre-specified categories for increase and decrease separately) after start of 1st injection will be summarized (number of subjects and percentages) and grouped by baseline category as follows (missing values will be defined as a separate category):

Table 8 Blood pressure categories

Baseline category	Post-baseline category
Systolic baseline: < 140 mmHg, ≥ 140 mmHg	Increase of: ≤20 mmHg, > 20-40 mmHg, > 40 mmHg
Systolic baseline: < 140 mmHg, ≥ 140 mmHg	Decrease of: ≤20 mmHg, > 20-40 mmHg, > 40 mmHg
Diastolic baseline: ≤ 90 mmHg, >90 mmHg	Increase of: ≤10 mmHg, > 10-20 mmHg, > 20 mmHg

Diastolic baseline: ≤ 90 mmHg, > 90 mmHg	Decrease of: ≤10 mmHg, > 10-20 mmHg, > 20 mmHg

For each subject both the worst increase and worst decrease during the 2 year period (after first injection) will be used. If subject has no change (change=0), then it will not be counted in the categories of increase and decrease. The following summaries will be prepared for blood pressure as grouped above only for subjects who have post baseline values:

- Maximal shifts (change in categories [increase and decrease]), this will be done both relative to injections and relative to target days following either table 3a or table 3b.

Mean values will be plotted per timepoint. One symbol per treatment group will be used.

17.5 Physical Examination

Weight, height, and BMI will be summarized by treatment and timepoint as will absolute change from baseline.

17.6 ECG Evaluation

Standard 12-lead ECGs (reporting ventricular rate/HR, PR, QRS, QT, and QTc intervals) will be summarized by treatment and timepoint as will absolute change from baseline. Presentation of QTc intervals will also include the Bazett's and Fridericia's correction defined as

Bazett's Correction (QTcB) $$QTc_b = \frac{QT}{\sqrt{RR}}$$

Fridericia's Correction (QTcF) $$QTc_f = \frac{QT}{\sqrt[3]{RR}}$$

where: RR = RR-interval measured in seconds (RR = 1/HR).

Mean values of continuous parameters will be plotted per analysis timepoint. One symbol per treatment group will be used. All ECG parameters will be listed by subject and visit.

18 References

References available upon request.

1. Tubach F, Ravaud P, Baron G et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Annals of Rheumatic Diseases. 2005; 64:29-33.

2. Independent Data Monitoring Committee Charter version 2, July 2014.

3. Attur, M. et al. Radiographic severity of knee osteoarthritis is conditional on interleukin 1 receptor antagonist gene variations. Annals of Rheumatic Diseases. 2010; 69, 856–861.

4. Wu X et al. IL-1 receptor antagonist gene as a predictive biomarker of progression of knee osteoarthritis in a population cohort. Osteoarthritis and Cartilage 2013; 21, 930–938.

5. Siebuhr et al., Identification and characterization of osteoarthritis patients with inflammation derived tissue turnover. Osteoarthritis and Cartilage 2014; (22) 44-50.

6. Sharif et al., A 5-yr longitudinal study of type IIA collagen synthesis and total type II collagen degradation in patients with knee osteoarthritis- association with disease progression. Rheumatology 2007;46: 938–943.

7. Kraus et al., Application of biomarkers in the development of drugs intended for the treatment of osteoarthritis: Osteoarthritis and Cartilage 2011; (19) 515-542.

8. Gudmann et al., Cartilage turnover reflected by metabolic processing of type II Collagen: A novel marker of anabolic function in chondrocytes. International Journal of Molecular Sciences. 2014; 15:18789-18803.

9. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004; 12:389-99.

10. Medical Outcomes Study Short Form-36 General Health Survey User's Guide, version 2, 2000.

11. Measure of Intermittent and Constant Osteoarthritis Pain User's Guide, version 3, 2007.

12. Guermazi A, Roemer FW, Haugen IK, Crema MD and Hayashi D. MRI-based semi quantitative scoring of joint pathology in osteoarthritis. Nature Reviews Rheumatology. 2013; 9(4):236-51.

13. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007; 15 (Suppl A): A1–56.

19 Appendices

The shells for tables, figures and listings will be specified in separate document and will be considered integral part of this SAP. In addition, specification for the biomarker process and handling will be detailed in a separate document. These documents will be finalized and approved prior to breaking the blind of the study.

19.1 List of endpoints

Group	Endpoint	Treatment Phase	Follow-up (integrated)	Reference
		Baseline to LOCF1	***Baseline to LOCF2***	
Trial Subjects	Disposition	Descriptive	Descriptive	12
	Protocol Deviations	Descriptive	Descriptive	12.1
Demographics and other baseline characteristics	Demographics	Descriptive	Descriptive	13.1
	Medical History	Descriptive	Descriptive	13.2
	Baseline Characteristics	Descriptive	Descriptive	13.3
Concomitant Medication and procedures	Previous medication and procedures	Descriptive	Descriptive	14
	Concomitant medication and procedures	Descriptive	Descriptive	14
	Pain medication	Descriptive	Descriptive	14
Exposure and treatment compliance	Exposure	Descriptive	Descriptive	15
	Compliance	Descriptive	Descriptive	15
Efficacy	Change from baseline in average cartilage thickness in the total femorotibial joint as evaluated by MRI in the target knee	Primary: Descriptive Analysis Sensitivity analysis (PP) Model check	Exploratory: Descriptive Analysis Model check	16.1
	WOMAC total score in the target knee at baseline and post-baseline	Secondary endpoint Descriptive Analysis Sensitivity analysis (PP) Subgroup analysis	Exploratory: Descriptive Analysis Model check	16.2
	WOMAC sub scores in the target knee at baseline and post-baseline	Secondary endpoint Descriptive Analysis Sensitivity analysis (PP) Subgroup analysis	Exploratory: Descriptive Analysis Model check	16.2

Group	Endpoint	Treatment Phase	Follow-up (integrated)	Reference
		Baseline to LOCF1	***Baseline to LOCF2***	
	20-meter WALK test	Secondary endpoint Descriptive Analysis	Descriptive	16.2
	PGA	Secondary endpoint Descriptive	Descriptive	16.2
	JSW on X-ray in the target knee at screening or baseline – and post-treatment	Secondary endpoint Descriptive Analysis Sensitivity analysis (PP) Subgroup analysis	Exploratory: Descriptive Analysis Model check	16.2
	Regional Average Cartilage thickness on MRI	Secondary endpoints Descriptive Analysis	Descriptive	16.2
	Total and Regional Cartilage Volume	Secondary endpoint Descriptive Analysis	Descriptive	16.2
PK	Synovial fluid levels of sprifermin/FGF-18	Descriptive	NA	16.3
	Serum levels of sprifermin/FGF-18	Descriptive	NA	16.3
PK/Safety	Occurrence of antibodies	Descriptive	NA	16.3 and 17
Other endpoints	OMERACT-OARSI score for the target knee	Descriptive Analysis	Descriptive	16.5
	PGIC	Descriptive	Descriptive	16.5
	KOOS Qol Score	Descriptive	Descriptive	16.5
	KOOS Symptom index	Descriptive	Descriptive	16.5
	MOS SF-36 Score	Descriptive	Descriptive	16.5
	NRS Pain Score for the target and contralateral knee	Descriptive	Descriptive	16.5
	Pain in other joints	Descriptive	Descriptive	16.5

Group	Endpoint	Treatment Phase	Follow-up (integrated)	Reference
		Baseline to LOCF1	***Baseline to LOCF2***	
	ICOAP Score for the target knee	Descriptive	Descriptive	16.5
	Semi quantitative MRI endpoints (WORMS/BLOKS) for the target knee	Descriptive Analysis	Descriptive	16.5
	Biomarkers	Descriptive for PGX and biochemical biomarkers. Other biomarkers will be defined in the BSAP [to be finalized prior to unblinding]		16.4
	Modeling	Will be defined in Modeling SAP		
	Subgroup analysis (including biomarkers subgroups)	Descriptive Analysis		16.4 and 16.5
	Additional analyses (changes and in addition to those specified in the protocol)	Descriptive Analysis	Descriptive	16.5
Safety	Adverse Events	Descriptive	Descriptive	17.1
	Death	Descriptive	Descriptive	17.2
	AIR	Descriptive Analysis	NA	17.2.3
	Clinical laboratory parameters	Descriptive	Descriptive	17.3
	Vital signs	Descriptive	Descriptive	17.4
	Physical Examination	Descriptive	Descriptive	0
	ECG	Descriptive	Descriptive	17.6

19.2 List of permitted pain medication

Analgesic Washout List_05JUN13.pdf

19.3 Handling of medication timing

	CMstart Year/Month/day	**CMstart Year/Month**	**CMstart Year**	**CMstart missing**
CMend Year/Month/day	(1) if CMstart Year/month/day > DoLD year/month/day => CM is Post; (2) else if CMend Year/month/day < DoFD year/month/day => CM is Previous; (3) else CM is Concomitant	(1) if CMstart Year/month > DoLD year/month => CM is Post; (2) else if CMend Year/month/day < DoFD year/month/day => CM is Previous; (3) else CM is Concomitant	(1) if CMstart Year > DoLD year => CM is Post; (2) else if CMend Year/month/day < DoFD year/month/day => CM is Previous; (3) else CM is Concomitant	(1) if CMend Year/month/day < DoFD year/month/day => CM is Previous; (2) else if CMend Year/month/day >= DoFD year/month/day => CM is Concomitant
CMend Year/Month	(1) if CMstart Year/month/day > DoLD year/month/day => CM is Post; (2) else if CMend Year/month < DoFD year/month => CM is Previous; (3) else CM is Concomitant	(1) if CMstart Year/month > DoLD year/month => CM is Post; (2) else if CMend Year/month < DoFD year/month => CM is Previous; (3) else CM is Concomitant	(1) if CMstart Year > DoLD year => CM is Post; (2) else if CMend Year/month < DoFD year/month => CM is Previous; (3) else CM is Concomitant	(1) if CMend Year/month < DoFD year/month => CM is Previous; (2) else if CMend Year/month>= DoFD year/month => CM is Concomitant
CMend Year	(1) if CMstart Year/month/day > DoLD year/month/day => CM is Post; (2) else if CMend Year < DoFD year => CM is Previous; (3) else CM is Concomitant	(1) if CMstart Year/month > DoLD year/month => CM is Post; (2) else if CMend Year < DoFD year => CM is Previous; (3) else CM is Concomitant	(1) if CMstart Year > DoLD year => CM is Post; (2) else if CMend Year < DoFD year => CM is Previous; (3) else CM is Concomitant	(1) if CMend Year < DoFD year => CM is Previous; (2) if CMend Year>= DoFD year => CM is Concomitant
CMend missing	(1) if CMstart Year/month/day > DoLD year/month/day => CM is Post; (2) else if CMstart Year/month/day < =DoLD	(1) if CMstart Year/month > DoLD year/month => CM is Post; (2) else if CMstart Year/month < =DoLD year/month =>	(1) if CMstart Year > DoLD year => CM is Post; (2) else if CMstart Year < =DoLD year => CM is Concomitant	Concomitant

	CMstart Year/Month/day	CMstart Year/Month	CMstart Year	CMstart missing
	year/month/day => CM is Concomitant	CM is Concomitant		

If subject is not treated, medication flag will be always "Previous" and use planned treatment for ITT/mITT Analysis sets in TLFs. Assumption is Day of First Dose (DoFD) and Day of Last Dose (DoLD) are not missing and complete for all treated subjects.

19.4 Handling of TEAE timing

	AEstart Year/Month/day	AEstart Year/Month	AEstart Year	AEstart missing
AEend Year/Month/day*	(1) if AEstart Year/month/day > DoLD year/month/day +30 => AE is Post; (2) else if AEend Year/month/day < DoFD year/month/ day => AE is Prior; (3) else AE is TEAE	(1) if AEstart Year/month > DoLD year/ month +1 => AE is Post; (2) else if AEend Year/month/day < DoFD year/month/day => AE is Prior; (3) else AE is TEAE	(1) if AEstart Year > DoLD year => AE is Post; (2) else if AEend Year/month/day < DoFD year/month/day => AE is Prior; (3) else AE is TEAE	(1) if AEend Year/month/day < DoFD year/month/day => AE is Prior; (2) else if AEend Year/month/day >= DoFD year/month/day => AE is TEAE
AEend Year/Month	(1) if AEstart Year/month/day > DoLD year/month/day +30 => AE is Post; (2) else if AEend Year/month < DoFD year/month => AE is Prior; (3) else AE is TEAE	(1) if AEstart Year/month > DoLD year/ month +1 => AE is Post; (2) else if AEend Year/month < DoFD year/month => AE is Prior; (3) else AE is TEAE	(1) if AEstart Year > DoLD year => AE is Post; (2) else if AEend Year/month < DoFD year/month => AE is Prior; (3) else AE is TEAE	(1) if AEend Year/month < DoFD year/month => AE is Prior; (2) else if AEend Year/month>= DoFD year/month => AE is TEAE

	AEstart Year/Month/day	AEstart Year/Month	AEstart Year	AEstart missing
AEend Year	(1) if AEstart Year/month/day > DoLD year/month/day +30 => AE is Post; (2) else if AEend Year < DoFD year => AE is Prior; (3) else AE is TEAE	(1) if AEstart Year/month > DoLD year/ month +1 => AE is Post; (2) else if AEend Year < DoFD year => AE is Prior; (3) else AE is TEAE	(1) if AEstart Year > DoLD year => AE is Post; (2) else if AEend Year < DoFD year => AE is Prior; (3) else AE is TEAE	(1) if AEend Year < DoFD year => AE is Prior; (2) if AEend Year>= DoFD year => AE is TEAE
AEend missing	(1) if AEstart Year/month/day > DoLD year/month/day +30 => AE is Post; (2) else AE is TEAE	(1) if AEstart Year/month > DoLD year/ month +1 => AE is Post; (2) else AE is TEAE	(1) if AEstart Year > DoLD year => AE is Post; (2) else AE is TEAE	TEAE

*When end time of AE is collected both end date and end time of AE will be compared to date and time of first dose and if the AE ends before first dose then AE is Prior, else AE is TEAE.

If subject is not treated, then flags are blank. Assumption is DoFD and DoLD are not missing and complete for all treated subjects.

19.5 List of key statistical outputs

19.5.1 General Information Tables

1. Disposition of Patients by Treatment - ITT Analysis Set
2. Major and Minor Protocol Deviations - ITT Analysis Set
3. Demographics by Treatment - ITT Analysis Set
4. Baseline Target and Contralateral Knee Characteristics by Treatment - ITT Analysis Set
5. Treatment Compliance by Treatment - ITT Analysis Set
6. Exposure to Sprifermin - ITT Analysis Set

19.5.2 Efficacy Tables

MRI Cartilage Thickness:

1. Repeated Measurement ANCOVA on Absolute 2 Years Change from Baseline Average Cartilage Thickness [mm] in the Total Femorotibial Joint - mITT Analysis Set
2. Repeated Measurement ANCOVA on Absolute 2 Years Change from Baseline Average Cartilage Thickness [mm] in the Medial Femorotibial Compartment - mITT Analysis Set
3. Repeated Measurement ANCOVA on Absolute 2 Years Change from Baseline Average Cartilage Thickness [mm] in the Lateral Femorotibial Compartment - mITT Analysis Set

4. Repeated Measurement ANCOVA on Absolute 2 Years Change from Baseline Average Cartilage Thickness [mm] in the Total Femorotibial Joint Transformed as Ranked Data - mITT Analysis Set
5. Repeated Measurement ANCOVA on Absolute 2 Years Change from Baseline Average Cartilage Thickness [mm] in the Medial Femorotibial Compartment Transformed as Ranked Data - mITT Analysis Set
6. Repeated Measurement ANCOVA on Absolute 2 Years Change from Baseline Average Cartilage Thickness [mm] in the Lateral Femorotibial Compartment Transformed as Ranked Data - mITT Analysis Set
7. Summary Statistics by Visit, Relative and Absolute Change from Baseline in Cartilage Thickness [mm] of the Total Femorotibial Joint, by Treatment - mITT Analysis Set
8. Summary Statistics by Visit, Relative and Absolute Change from Baseline in Cartilage Thickness [mm] of the Medial Femorotibial Compartment, by Treatment - mITT Analysis Set
9. Summary Statistics by Visit, Relative and Absolute Change from Baseline in Cartilage Thickness [mm] of the Lateral Femorotibial Compartment, by Treatment - mITT Analysis Set
10. Repeat 1-3 and 7-9 for PP Analysis set.

X-ray mJSW:

11. Absolute Change (ANCOVA) from Baseline up to 2 years in mJSW [mm] in the Medial Femorotibial Compartment - ITT Analysis Set
12. Absolute Change (ANCOVA) from Baseline up to 2 years in mJSW [mm] in the Lateral Femorotibial Compartment - ITT Analysis Set
13. Summary Statistics by Visit, Relative and Absolute Change from Baseline in mJSW [mm] in the Medial Femorotibial Compartment, by Treatment - ITT Analysis Set
14. Summary Statistics by Visit, Relative and Absolute Change from Baseline in mJSW [mm] in the Lateral Femorotibial Compartment, by Treatment - ITT Analysis Set
15. Repeat 11-14 for PP Analysis set

WOMAC:

16. Summary Statistics by Visit and Change in WOMAC Total Score from Baseline to 2 Years, by Treatment - ITT Analysis Set
17. Summary Statistics by Visit and Change in WOMAC Pain Sub Score from Baseline to 2 Years, by Treatment - ITT Analysis Set
18. Summary Statistics by Visit and Change in WOMAC Function Sub Score from Baseline to 2 Years, by Treatment - ITT Analysis Set
19. Absolute Change (ANCOVA) from Baseline up to 2 years in WOMAC Total Score - ITT Analysis Set
20. Absolute Change (ANCOVA) from Baseline up to 2 years in WOMAC Pain Sub Score - ITT Analysis Set
21. Absolute Change (ANCOVA) from Baseline up to 2 years in WOMAC Function Sub Score - ITT Analysis Set
22. Repeat 16-21 for PP Analysis set
23. Concomitant Pain Medication by Treatment - ITT Analysis Set

19.5.3 Safety Tables

1. Overview of AEs, by Treatment – Safety Analysis Set
2. Summary of TEAEs Leading to Discontinuation of Treatment – Safety Analysis Set
3. Summary of SAEs – Safety Analysis Set
4. Summary of Most Frequent TEAEs (at Least 5% in Any Treatment Group) – Safety Analysis Set
5. Summary of Treatment Emergent Injection Site Reactions – Safety Analysis Set
6. Summary of Patient Reported Acute Inflammatory Reactions by Cycle and Overall, by Treatment – Safety Analysis Set

19.5.4 Figures

1. Mean and 95% CI in Absolute Change from Baseline Over Visits by Treatment, Cartilage Thickness [mm] in the Total Femorotibial Joint - mITT Analysis Set
2. Mean and 95% CI in Absolute Change from Baseline Over Visits by Treatment, Cartilage Thickness [mm] in the Medial Femorotibial Compartment - mITT Analysis Set
3. Mean and 95% CI in Absolute Change from Baseline Over Visits by Treatment, Cartilage Thickness [mm] in the Lateral Femorotibial Compartment - mITT Analysis Set
4. Mean and 95% CI in Absolute Change from Baseline Over Visits by Treatment, mJSW [mm] in the Medial Femorotibial Compartment - ITT Analysis Set
5. Mean and 95% CI in Absolute Change from Baseline Over Visits by Treatment, mJSW [mm] in the Lateral Femorotibial Compartment - ITT Analysis Set
6. Mean and 95% CI in Absolute Change from Baseline Over Visits by Treatment, WOMAC Total Score in the Target Knee - ITT Analysis Set
7. Mean and 95% CI in Absolute Change from Baseline Over Visits by Treatment, WOMAC Pain Sub Score in the Target Knee - ITT Analysis Set
8. Mean and 95% CI in Absolute Change from Baseline Over Visits by Treatment, WOMAC Function Sub Score in the Target Knee - ITT Analysis

19.5.5 Data Listings

1. Adverse Events Leading to Discontinuation of Treatment - Safety Analysis Set
2. Serious Adverse Events - Safety Analysis Set
3. AIR Reported in Subject Diary - Safety Analysis Set