STATISTICAL ANALYSIS PLAN
12 May 2023

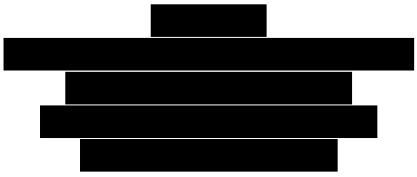
A Multicenter, Randomized, Double-blind, Sham-controlled, Comparative Study of SI-6603 in Subjects with Lumbar Disc Herniation (Phase 3)

PROTOCOL NUMBER
6603/1133

SPONSORED BY

SEIKAGAKU CORPORATION
6-1, Marunouchi 1-chome, Chiyoda-ku,
Tokyo 100-0005, Japan

PREPARED BY



Version: 2.0

DOCUMENT VERSION CONTROL

Version Number	Date	Comments/Changes
1.0	30Mar2018	Created
1.1	05DEC2022	Updated for COVID-19 pandemic Added clarifications on visit windowing, diary data averages, baseline record selection, worst back pain data handling Added sensitivity analyses Updated multiplicity to use serial gatekeeping Updated definition of recurrence of LDH to clarify treatment failure criteria.
2.0	12May2023	Changed odds ratios to differences in proportions Updated handling of missing responder-type endpoints Added efficacy subgroup analysis Edited for consistency on 30% and 50% improvement Added clarification around handling of baseline and follow-up averages when randomization and injection dates differ Updated randomization seeds to an appropriate number of digits Updated definition of recurrence of LDH

APPROVALS

[redacted] **Project Statistician Approval**	**Seikagaku Corporation Approval**
[redacted] Chief Strategist, Biostatistics & Protocol Design	Takayuki Seo, Biostatistics Lead
Signature and Date	Signature and Date
DocuSigned by: [redacted]	DocuSigned by: Takayuki Seo 署名者名: Takayuki Seo 署名理由: この文書を承認する 署名時刻： 12-May-2023 | 10:21:44 PM EDT BF16A1EAFAD546E5B11BF87B7717F1CC

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CGIC	Clinical Global Impression of Change
CI	Confidence Interval
COVID-19	COronaVIrus Disease of 2019
eCRF	electronic Case Report Form
EQ-5D-5L	EuroQol Group 5-Dimension Quality of Life
mITT	modified Intention-to-treat
LDH	Lumbar Disc Herniation
MAR	Missing at Random
MCAR	Missing Completely at Random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed Model for Repeated Measures
MNAR	Missing Not at Random
MRI	Magnetic Resonance Imaging
ODI	Oswestry Disability Index
PGIC	Patient Global Impression of Change
PP	Per-protocol
PT	Preferred Term
QOL	Quality of Life
SD	Standard Deviation
SLR	Straight Leg Raise
SF-36	36-item Short Form Health Survey
SOC	System Organ Class
TEAE	Treatment-emergent Adverse Event
U	Units

VAS	Visual Analogue Scale
WPAI	Work Productivity and Activity Impairment Questionnaire

1. PURPOSE OF THE ANALYSES

The purpose of the analyses described in this document is to compare the safety and efficacy of a single intervertebral disc injection of SI-6603 to a sham injection in subjects with lumbar disc herniation (LDH).

2. PROTOCOL SUMMARY

2.1 Study Objectives

The primary objective is to evaluate the efficacy of a single-dose intervertebral disc injection of SI-6603 at a dose of 1.25 units (U) compared to control in subjects with LDH by comparing the change in worst leg pain during the past 24 hours, as assessed by visual analogue scale (VAS), from baseline to Week 13 after injection of the investigational product.

The secondary objectives are to evaluate the efficacy of SI-6603 1.25 U for key secondary endpoints, and to demonstrate whether SI-6603 1.25 U is safe and well tolerated.

Key secondary endpoints include:

- Change from baseline to Week 13 in herniation volume
- Change from baseline to Week 13 in Oswestry Disability Index (ODI) score
- Change from baseline to Week 52 in average worst leg pain score during the past 24 hours over the previous 7 days, as assessed by 100 mm VAS.

Supportive endpoints may be found in section 7.2.

2.2 Overall Study Design and Plan

Up to three weeks following the screening, subjects will be randomized and injected at Week 0. Subjects will receive a single injection of SI-6603 or a sham injection and will be followed for 52 weeks for the efficacy and safety evaluations found in the schedule of events. Subjects will return for follow-up visits at Weeks 1, 2, 4, 6, 13, 26, 39 and 52.

2.3 Study Population

Subjects will meet the following key criteria:

- Ages 30 to 70 (inclusive)
- Have contained posterolateral LDH at either L4-L5 or L5-S1 (or L5-L6):
    - with the presence of demonstrable nerve root impingement, as assessed by Magnetic Resonance Imaging (MRI);
    - chief complaint of unilateral radiculopathy and/or radicular leg pain corresponding to the ipsilateral leg and distribution of the affected nerve root; and
    - positive result of Straight Leg Raise (SLR) test (≤70°) only on the ipsilateral leg having chief complaint of radiculopathy corresponding to the pain and distribution of the affected nerve root.
- Symptoms of radiculopathy and/or radicular leg pain only in the unilateral leg corresponding to the distribution of the affected nerve root for 6 weeks or more but 1 year or less, and which is still ongoing at the time of informed consent.

- Subjects whose worst leg pain (100 mm VAS) for the 7 consecutive days up to the day before randomization meets following conditions:
    - at least 5 worst leg pain scores on the days when not using any rescue medications or increasing or adding non-prohibited concomitant medications to treat LDH;
    - the mean of the worst leg pain scores is 50 to 90 mm; and
    - the range of fluctuation in worst leg pain scores over the 7 days (difference between minimum and maximum scores) is ≤ 25 mm.
- ≥30% on the ODI at the time of randomization

The full inclusion and exclusion criteria may be found in the protocol.

2.4 Treatment Regimens

Subjects will be randomized to receive a single dose injection of SI-6603 1.25 U into an intervertebral disc or a sham injection.

2.5 Randomization

All eligible subjects will be randomized on a 1:1 basis to SI-6603 or control based on a computer-generated randomization schedule. The randomization will be balanced by randomly permuted blocks and stratified by site.

2.6 Sample Size Determination

It is planned to enroll approximately 320 subjects, 160 in the SI-6603 group and 160 in the control group.

Sample size estimation is based on using the mean difference in the worst leg pain score change from baseline in the primary efficacy analysis. The following assumptions were made to compute the sample size:

- The 2-sample t-test comparing the group means of the change from baseline was used. (The 2-sample t-test approximates the test of the null hypothesis based on the repeated measures model that will be used in the primary efficacy analysis.)
- Treatment difference between SI-6603 and control groups of 12 mm
- Common standard deviation (SD) of 30 mm
- Dropout rate of 15%
- Power of 90%
- Two-sided significance level of 5%

3. GENERAL ANALYSIS AND REPORTING CONVENTIONS

All efficacy analyses will use a two-sided alpha = 0.05 test unless otherwise stated. P-values will be rounded to four decimal places. If a p-value is less than 0.0001, it will be reported as "<0.0001."

Continuous data will be described using descriptive statistics: sample size, mean, SD, median, Q1 and Q3, minimum, and maximum. Minimums and maximums will be reported with the same precision as the raw values; medians, Q1, Q3 and means will show precision to one decimal place greater than the raw values; SDs will show precision to two decimal places greater than the raw values. For example, if height is recorded in whole centimeters, min and max will be displayed as whole centimeters, means will be shown as xxx.x and SD will be shown as xxx.xx. Data will be displayed in all listings sorted by treatment group and subject identification number. Subjects will be identified in the listings by the subject identification number concatenated with the investigator number.

When count data are presented, the percentage will be suppressed when the count is zero in order to draw attention to the non-zero counts. The denominator for all percentages will be the number of subjects with non-missing data in that treatment group within the population of interest.

For the duration of leg pain covariate, if partial dates are provided, then missing dates will be imputed as the 15th for a missing day and July for a missing month, consistent with prior protocols.

3.1 Baseline values

For visit-based measures, baseline will be the last non-missing value prior to injection. For diary-based measures, the baseline will be the average of the scores for the 7 consecutive days up to the day prior to randomization (independently for worst leg pain and worst back pain). Per the inclusion criteria, subjects should have a minimum of 5 non-missing values; however, if a subject is inadvertently randomized with fewer values, then the baseline will be the mean of the available values from that 7 day period. Note that post-baseline visits were scheduled per the protocol with the injection date as day 1; in rare instances, this may be after the randomization date. In these cases, the baseline will be anchored around the randomization visit. Additionally, in the instance that a subject is inadvertently entered into the randomization system and assigned a kit on a day other than their visit 2A, the anchor date will be the day that they were confirmed to have qualified for the trial (typically the day of visit 2A).

4. ANALYSIS POPULATIONS

Three analysis populations as described below will be utilized in reporting the study results. All efficacy analyses will be performed on the modified intention-to-treat (mITT) population, with some analyses repeated on the per-protocol (PP) population.

All safety analyses will be performed on the safety population.

4.1 Modified Intention-to-Treat

The mITT population is defined as all randomized subjects who received the study injection, analyzed according to the assigned treatment.

4.2 Per-Protocol

The PP population is defined as all mITT subjects who had no major protocol deviations that could affect the primary efficacy assessment. The details of major protocol deviations are described in section 5.2, and included subjects will be determined prior to unblinding.

4.3 Safety

The safety population is defined as all randomized subjects who received the study injection, analyzed according to the treatments subjects received.

5. STUDY SUBJECTS

5.1 Disposition of Subjects

A disposition table of subjects will include the number and percentage of subjects for the following categories: subjects screened, subjects randomized, subjects treated (safety sample), subjects in the mITT sample, subjects in the PP sample, subjects completing through Week 13, subjects completed study, and subjects discontinued from the study. A separate summary of discontinuations through Week 13 will also be presented. All percentages will be based on the number of subjects randomized (with the exception of subjects screened). Additionally, the number of subjects excluded from each population and reason for exclusion will be presented overall and by site.

The reasons for study discontinuation will also be summarized in this table. The reason for discontinuation may include any of the following: lack of efficacy, death, adverse event, lost to follow-up, failure to meet randomization criteria, protocol violation, physician decision, withdrawal by subject, pregnancy, study terminated by sponsor, and other. Subjects that do not complete the study will be flagged if they had early termination due to confirmed COVID-19, early termination at site discretion due to COVID-19, or early termination at subject discretion due to COVID-19; this will summarized with the number and percentage of subjects in each category. Only the primary reason for study discontinuation will be recorded for a given subject.

The number and percentage of subjects attending each visit will be summarized. Due to the COVID-19 pandemic, visits could be in-person, remote or a combination; the reason for a remote visit could be due to the subject having confirmed COVID-19, at site discretion due to COVID-19, at subject discretion due to COVID-19 or other reasons. Likewise visits may have been missed or canceled due to the subject having confirmed COVID-19, at site discretion due to COVID-19, at subject discretion due to COVID-19 or other reasons. Additionally, individual assessments may have been omitted due to COVID-19; this is recorded on the CRF.

The type of visit (in-person, remote, or combination), the reason for remote visits, the reason visits were missed and assessments missed due to COVID-19 will be summarized with counts and percentages for each visit and treatment group.

A listing will present data concerning subject disposition and other reasons for discontinuation. Additionally, a listing will show the visit attendance, type of visit, reason for remote/missed visit, and assessments missed due to COVID-19 for each subject.

5.2 Protocol Deviations

Protocol deviations will be categorized as major or minor prior to database lock and unblinding according to Protocol Deviation Plan. Additionally, deviations are categorized whether they are related to the COVID-19 pandemic. All recorded protocol deviations and their categorization will be presented in a listing.

6. DEMOGRAPHICS, BASELINE CHARACTERISTICS, CONCOMITANT MEDICATIONS, MEDICAL HISTORY

6.1 Demographics and Other Baseline Characteristics

Demographics and baseline characteristics will be summarized by treatment for the mITT, safety, and PP populations. With the exception of the worst leg pain scores, Week 0 will be presented unless an outcome is not collected, in which case the screening visit will be used. For VAS pain scores, the baseline will be the average of the 7 daily worst pain leading up to the randomization visit. Outcomes presented will include:

- age
- sex
- race
- ethnicity
- height
- baseline weight
- body mass index (BMI)
- smoking history
- occupation
- days since diagnosis
- herniation site
- location of herniation
- location of radicular leg pain
- days since onset of current radicular leg pain
- ongoing low back pain caused by herniation
- VAS pain score for worst leg pain
- VAS pain score for worst back pain
- ODI score

All outcomes except the VAS pain scores and ODI scores will appear in a listing of baseline characteristics; VAS pain scores and ODI scores will appear with all other visits in the efficacy listings.

Note that 36-item Short Form Health Survey (SF-36), EQ-5D-5L, and Work Productivity and Activity Impairment Questionnaire (WPAI) assessments are collected at baseline; however, baseline values of these will be presented only on their associated displays.

6.2 Prior and Concomitant Medications/Therapies

Concomitant medications/therapies include all medications taken at any point following the injection of the study therapy; in case of partial dates, a medication will be considered to be concomitant unless the partial dates clearly exclude the post-injection period.

Concomitant medications will be coded using the WHODrug Global using the current version at the time coding is first initiated and will be presented with counts and percentages by treatment for Anatomical Therapeutic Chemical (ATC) class and standardized drug name. Subjects will be counted at most once for a drug and class; this will be presented for the safety population. Concomitant medications will also be presented in a listing.

Prior medications include any medications recorded on the electronic case report form (eCRF) that were taken at any time prior to injection of study therapy. Prior medications will include medications that are ongoing at time of injection of study therapy. Prior medications will be presented in a manner parallel to the concomitant medications.

Additionally, prohibited and restricted medications will be summarized as above and presented in a listing. Restricted medications include medications for LDH taken at stable dose and regimen as described in protocol section 6.6.2. A separate summary for concomitant medication excluding those taken for LDH will also be provided.

Prior and concomitant therapies will be presented in a listing; prohibited therapies and block procedures will be flagged in this listing.

6.3 Medical History

All medical history for the safety population will be presented in a listing.

The number and percentage of subjects having received nerve block, epidural injection, and Facet block will be summarized for each treatment group.

7. EFFICACY EVALUATION

7.1 Overview of Efficacy Analysis Issues

7.1.1 Handling of Dropouts or Missing Data

For the efficacy analysis based on repeated-measures models of continuous endpoints, missing data will be implicitly handled via a mixed-effect model, without explicit imputation. Inferences based on this approach are unbiased under the assumption of missing at random (MAR), which is a weaker and more generally true assumption than missing completely at random (MCAR).

For the primary efficacy analysis, sensitivity analysis based on multiple imputation (MI) methods that changes imputation methods by missing pattern such as discontinuations due to AEs, lack of efficacy, and other will be performed for the mITT population.

For the analysis of percentages of positive responders by composite definition, no imputation will be performed. However, subjects will only be considered responders if they meet all criteria, thus, subjects with missing data and dropouts will implicitly be treated as non-responders.

No replacement of any missing data will be made for the safety analyses.

Additionally, see Section 7.3.2.1 for data handling regarding the VAS pain scores for the definition of a non-missing week and data censoring rules.

7.1.2 Assessment Time Windows

All data collected during follow-up will be displayed and analyzed according to the actual visit that the data are assigned in the eCRF.

Study discontinuation visits will be assigned to the corresponding visit if it falls within the protocol window for that visit. For example, a discontinuation visit on day 100 would be assigned to Week 13 since it falls within the -7/+14 day window for Week 13. However, if scheduled visit data are present, then the discontinuation visit data will not be used for that visit. This will be done on a parameter level and discontinuation visit values will not be assigned to visits where the assessment was not scheduled to be taken (For example, Week 4 does not have MRI, but the discontinuation visit does; even if the discontinuation visit falls in the Week 4 window, we would not assign the MRI findings to a Week 4 visit).

Unscheduled assessments (laboratory data, imaging findings or vital signs associated with non-protocol clinical visits or obtained in investigating or managing AEs) will be included in listings but not summaries of the data. Likewise, discontinuation visit data that do not fall within a protocol-defined window or fall within a window where

scheduled visit data are already present will not be summarized, but will appear in listings.

7.2 Efficacy Variables

The primary efficacy outcome is the change from baseline in the weekly averages of the VAS worst leg pain assessed at Weeks 1, 2, 4, 6, 13, 26, 39, and 52. Baseline is defined as the average of the 7 daily worst pain measures leading up to the randomization visit.

Key secondary outcomes include:

- Change from baseline to Week 13 in herniation volume
- Change from baseline to Week 13 in ODI score
- Change from baseline to Week 52 in average worst leg pain score during the past 24 hours over the previous 7 days, as assessed by 100 mm VAS.

Supportive outcomes include:

- Change from baseline in worst leg pain score at each time point
- Change from baseline in worst back pain score at each time point
- Change from baseline in ODI score at each time point
- Percentage of subjects with negative SLR test at each time point
- Percentage of subjects without hypoesthesia, muscle weakness or hyporeflexia at each time point
- Patient Global Impression of Change (PGIC) score at each time point
- Clinical Global Impression of Change (CGIC) score at each time point
- Change from baseline in SF-36 scores at each time point
- EQ-5D-5L (EuroQol Group 5-Dimension Quality of Life instrument, 5-level version) quality of life (QOL) score and VAS score at Week 13 and Week 52
- WPAI score at Week 13 and Week 52
- Incidence and amount of rescue medication use over 13 weeks
- Change from baseline in amount of rescue medication use at each time point
- Cumulative distribution of percentage change from baseline in worst leg pain score at Week 13 and Week 52
- Cumulative distribution of percentage change from baseline in ODI score at Week 13 and Week 52
- Responder rate by composite definition at Week 13
- Change from baseline in intervertebral disc volume at Week 13 and Week 52
- Change from baseline in herniation volume at Week 13 and Week 52
- Incidence of post-treatment surgery for LDH at the same level of investigational product injection
- Number of subjects with recurrence of LDH at Week 52

7.3 Analysis Methods

7.3.1 Estimand

Population

The analysis population will be mITT subjects (randomized and injected) with LDH as defined by the protocol inclusion/exclusion criteria.

Variable

The primary endpoint will be the change from baseline to Week 13 in the weekly average of daily VAS worst leg pain scores. Three days of non-missing values are required for each weekly average.

Intercurrent Events

Observations following concomitant therapies of lumbar operation, lumbar percutaneous nucleotomy, or lumbar intradiscal therapies will be censored. Otherwise, no special handling will be employed for intercurrent events; once injected it is impossible for the subject to go off treatment and data will be analyzed as observed.

Population-Level Summary

The population-level summary will be the difference in least-square means between treatment arms (analyzed as randomized) at Week 13 from the primary analysis model. See Section 7.3.2 below for details.

7.3.2 Primary Efficacy Analyses

The change from baseline in the worst leg pain scores will be analyzed by a mixed model for repeated measures (MMRM) analysis for Week 1 through Week 52 in the mITT population. Records analyzed will be the weekly averages at Weeks 1, 2, 4, 6, 13, 26, 39, and 52 (see Section 7.3.2.1 for additional details).

The primary comparison will be the mean change from baseline of the SI-6603 group at Week 13 compared with the control group, estimated using this model. The model will include the baseline worst leg pain score, treatment, time, treatment-by-time interaction, and duration of leg pain as fixed effects.

An unstructured covariance will be used to model the covariance structure among repeated measures. Should the model fail to converge with the type = UN option, a compound symmetry structure will be used. Kenward-Roger method will be used for computing the denominator degrees of freedom. The missing data will be implicitly handled via a mixed-effect model, without explicit imputation.

The primary efficacy objective will be tested using the null hypothesis that there is no difference in the mean change from the pretreatment measure of pain relief (weekly average VAS worst leg pain) between SI-6603 and control at 13 weeks. This null hypothesis can be expressed as the hypothesis test

H_0: $\delta_{13} = 0$ vs. H_A: $\delta_{13} \neq 0$

where δ_{13} is the mean difference in the change from the pretreatment measure of pain relief (weekly average VAS worst leg pain) between SI-6603 and control at Week 13 from the model-based estimates of each treatment's change from pretreatment at those visits.

Worst leg pain score and change from baseline in worst leg pain score will be summarized by treatment group and time point. Mean change from baseline over time will be summarized in a line graph with error bars.

7.3.2.1 Data Handling of Worst Leg Pain

The mean worst leg pain score from 7 consecutive days prior to each visit will be targeted as the worst leg pain score for each time point; (the day of the visit itself will not be included but will serve as the anchor). If less than 3 days of worst leg pain scores are available in the day interval prior to a visit, alternative 7-day intervals within the protocol-specified visit window will be examined. Finally, for the Week 52 visit only, if the actual visit day falls outside of the protocol-defined window, intervals in that window will targeted first. Note that post-baseline visits were scheduled per the protocol with the injection date as day 1; in rare instances, this may be after the randomization date. For the purposes of the below algorithm, the injection date is considered to be day 1.

To calculate the 7-day average, first the worst leg pain score on any of the following days will be eliminated:

- Days after prohibited concomitant therapies of lumbar operation, lumbar percutaneous nucleotomy, or lumbar intradiscal therapies occurred, and
- Days before the previous visit (for example, the Week 2 visit may fall just 5 days after Week 1; thus, the days prior to the Week 1 visit will not be considered as eligible for inclusion in the 7-day Week 2 average).

Anchored by the visit date, the average of the consecutive days prior to each visit will be calculated. If less than 3 days of worst leg pain scores are available of the 7 consecutive days prior to a visit, the following approach will be utilized:

1. Using the visit target day, each day that would fall within the protocol-defined visit window will be identified as the anchor.
2. For each of those days, the corresponding 7 day interval anchored to that end date will be examined and all intervals with a valid average with at least 3 days of data will be identified
3. The interval with the anchor day that falls closest to the visit target day will be selected for the average. In the case of two equidistant valid intervals, the one with the later anchor day will be used.

If an in-person visit is missing altogether, the above process will be used with the corresponding visit target day. If no valid interval can be identified, worst leg pain score will be handled as missing at the time point.

Finally, for the Week 52 visit, if the actual visit day falls outside of the protocol-defined window, the following approach will be utilized:

1. Each day from day 351 to day 379 will be identified as a possible anchor.
2. For each of those days, the corresponding 7 day interval anchored to that end date will be examined and all intervals with a valid average with at least 3 days of data will be identified
3. The interval with the anchor day that falls closest to day 365 will be selected for the average. In the case of two equidistant valid intervals, the one with the later anchor day will be used.

If no valid interval is identified anchoring from day 351 to day 379 (365 +/- 14 days), only then will the interval anchored by the out-of-window visit be used.

The data handling for the worst back pain will use rules identical to those above for worst leg pain.

7.3.2.2 Sensitivity analyses

As a sensitivity analysis, missing data for participants in the mITT population will be imputed via MI. Twenty repeats of the imputation will be performed using Markov Chain Monte Carlo (MCMC) assuming non-monotone missing. Where the subject was discontinued due to lack of efficacy, due to an AE (with the exception of COVID-19), the subject has met the criteria for treatment failure as described in Section 7.3.3.4, or the subject's values are censored due to a prohibited therapy as described in Section 7.3.2.1, missing values will be imputed drawing from the baseline values, conditioned on the non-missing post-baseline values, of all participants in the mITT population under the assumption that they are missing not at random (MNAR); duration of leg pain will be included as a covariate.

If the study medication was discontinued for any other reason, values will be imputed within treatment group using MI under the assumption that they are missing at random (MAR) with covariates for duration of leg pain, and the weekly average worst leg pain score recorded at each time point (including baseline). The MAR approach will also be applied for sporadic missing values (prior to discontinuation). Values missing due to COVID-19, whether the result of early termination or sporadic missing values, are assumed MAR and unrelated to the endpoint of the weekly average VAS worst leg pain.

Data will be imputed for the full set of visits as described above and the same output will be used as a sensitivity for both the primary at Week 13 and the secondary endpoint at Week 52.

The MMRM analysis as described for the primary will be repeated for each of the 20 sets of data and combined using SAS PROC MIANALYZE. See Appendix 13.2 for the list of random seeds.

7.3.2.3 Per Protocol Analyses

The primary analysis and the key secondary analyses will be repeated on the PP population.

7.3.2.4 Comparison of Results in Subpopulations

Analyses for the primary endpoint will be conducted for the mITT Population on the following subgroups:

- Age categories: 20-29 years, 30-49 years, 50-59 years and ≥ 60 years (only study 6603/1031 includes 20-29 group, but will be reported and discussed in the ISE text)
- Sex: Male, Female
- Race: White, Black or African American, Asian, and Other.
- Ethnicity: Hispanic or Latino, Not Hispanic or Latino (applies to Study 6603/1133 only)
- BMI categories: <18.5 kg/m^2, 18.5 to < 25.0 kg/m^2, 25.0 to < 35.0 kg/m^2, and ≥ 35.0 kg/m^2
- Height: <170 cm, ≥ 170 cm
- Disc level: L4/L5, L5/S1 (subjects that have the procedure on L6 will be grouped with the L5/S1subjects)

The goal of these subgroup analyses is to explore potential differences in treatment response by subgroup. The analysis and reporting will parallel the primary reporting and MMRM model where possible. If convergence issues arise due to insufficient sample size, then descriptive statistics only may be reported for a subgroup.

7.3.3 Secondary Efficacy Analyses

7.3.3.1 Key Secondary

A longitudinal analysis model will be used as described in the primary efficacy analysis for the analyses of change from baseline in the herniation volume at Week 13, change from baseline in the worst leg pain score at Week 52, and change from baseline in the ODI at Week 13. The baseline value of each outcome will be included as a covariate in these models in place of the baseline worst leg pain score (from the primary analysis). Note that herniation volume is collected only at Weeks 13 and 52; ODI assessment shares the same time points as the primary.

Key secondary endpoints will be summarized by treatment group and time point. Mean change from baseline in key secondary endpoints over time will be summarized in a line

graph with error bars. Estimated value in key secondary endpoints over time will be summarized in a line graph with 95% confidence interval (CI) displayed.

7.3.3.2 Sensitivity Analyses for Key Secondary endpoints

An MI approach identical to that used for the sensitivity analysis for the primary endpoint will be applied for each key secondary endpoint. As with the primary, a full set of data will be imputed for all visits including Week 52 and the analyses will report results for both the Week 13 and Week 52 time points.

7.3.3.3 Other Secondary

All endpoints, including the primary and key secondary endpoints, will be summarized by treatment group and by time point. The differences of other secondary endpoints between the SI-6603 and control groups will be analyzed at each time point they are collected as below.

- A similar longitudinal analysis model will be used as described in the primary efficacy analysis for the analysis of change from baseline in pain intensity scores, function scores, QOL scores, and volumes (intervertebral disc and herniation), by including the corresponding baseline value as the fixed effect instead of the baseline worst leg pain score.
- Percentage change from baseline in intervertebral disc and herniation volumes will be summarized in a line graph with error bars.
- Neurologic status as determined by neurological examinations will be compared between treatments with a difference in proportions Z test. In the case of low counts (<5 expected in a cell), a Fisher's exact test will be used.
  Outcomes will include SLR test, sensation, muscle strength, and deep tendon reflex. Counts and percentages in each category will be presented along with 95% confidence intervals, the difference in the percentages and its 95% Wald CI. Percentages will be reported out of non-missing values.
- For the global assessments of PGIC score and CGIC score, responder analysis with the two best categories as "responder" will be compared between treatments with a difference in proportions Z test. In the case of low counts (<5 expected in a cell), a Fisher's exact test will be used. Counts and percentages in each category for the PGIC and CGIC will be presented as well as the responder/non-responder counts and percentages along with 95% confidence intervals, the difference in the percentages of responders and its 95% Wald CI. Subjects with missing values will be considered non-responders.
- Scores on the WPAI will be analyzed with the analysis of covariance model including the baseline value and duration of leg pain as covariates.
- The incidence of rescue medication use will be compared between treatments with a difference in proportions Z test. In the case of low counts (<5 expected in a cell), a Fisher's exact test will be used. The amount of rescue medication use will be analyzed with the Wilcoxon rank sum test.

- Cumulative distribution of percentage change from baseline in worst leg pain and ODI will be generated, and differences between the SI-6603 and control groups in the percentages of subjects experiencing a ≥30% and ≥50% improvement in worst leg pain score will be compared between treatments with a difference in proportions Z test. In the case of low counts (<5 expected in a cell), a Fisher's exact test will be used. Additionally, the difference in the percentages of responders and its 95% Wald CI will be reported. Subjects with missing values will be considered non-responders/not having improvement.
- For the analysis of the percentages of responders by composite definition, differences between the SI-6603 and control groups in the percentages of positive responders will be compared between treatments with a difference in proportions Z test. In the case of low counts (<5 expected in a cell), a Fisher's exact test will be used. Counts and percentages of responders will be presented along with 95% confidence intervals, the difference in the percentages and its 95% Wald CI. See section 7.3.3.4 for the definition of composite responder.
- The number of subjects with recurrence of LDH, percentages and differences between the SI-6603 and control groups in the percentages will be reported. See section 7.3.3.5 for the definition of recurrence of LDH.
- Time to post-treatment surgery for LDH will be evaluated by survival analysis using Kaplan-Meier methodology and the log-rank test. Counts and percentages of subjects with post-treatment surgery will be reported as well as medians and quartiles of the time to surgery (if they are defined). Additionally, the number and percentage of subjects with post-treatment surgery for LDH will be reported and compared between treatment groups with a Fisher's exact test. The difference in the percentages of responders and its 95% Wald CI will also be reported.
- Time to post-treatment surgery for LDH will be summarized in a Kaplan-Meier plot.

7.3.3.4 Composite responder

A composite responder is a subject who meets all 4 criteria below (A, B, C, and D) as assessed at Week 13:

a. A reduction of ≥30% in worst leg pain score during the past 24 hours assessed by VAS from baseline to Week 13;

b. Improvement of ≥30% in ODI score from baseline to Week 13;

c. Maintenance or improvement of neurologic status (motor, sensory, reflexes) from baseline to Week 13; and

d. No "Treatment Failure"

"Treatment Failure" is defined as a subject to whom any of the conditions shown below become applicable until Week 13 after investigational product injection:

i. Use of additional medication/therapy to treat treatment-emergent adverse events (TEAEs) or complications associated with LDH
   - New prohibited medications used for more than 7 days;
   - Increase of restricted medications used for more than 7 days;
   - Use of new prohibited therapies; or
   - Increase in frequency or intensity of restricted therapies.

ii. Poor response to investigational product that leads the investigator to judge it necessary to perform a surgical intervention for back pain/leg pain or withdraw from the study

iii. Increase of measurement value on either vertebral body angle formed by flexion or vertebral body translation from baseline to values:
   - Vertebral body angle formed by flexion of ≥5º, or
   - Vertebral body translation of ≥3 mm

iv. Occurrence of treatment related AEs that lead the investigator to judge it necessary for the subject to have surgical intervention or withdraw from the study

7.3.3.5 Recurrence of LDH

Recurrence of LDH is defined as below:

A subject who has all of the following conditions:

a. The subject shows improvement according to all of the following clinical symptoms at Week 13:
   - Reduction of ≥30% in worst leg pain score from baseline,
   - Improvement of ≥30% in ODI score from baseline, and
   - Maintenance or improvement of neurologic status (motor, sensory, reflexes) from baseline.

b. At week 52:
   - Subject does NOT have reduction of ≥30% in worst leg pain score from baseline,

- Subject does NOT have improvement of ≥30% in ODI score from baseline, and
- Maintenance of or experiences worsening of neurologic status (motor, sensory, reflexes) from baseline.

Percentages will be reported out of subjects with any Week 52 data. If a subject has missing values, they will be imputed as negative outcomes (for example, missing leg pain will be interpreted as not having 30% reduction).

7.3.3.6 Cumulative Distribution Plots

Cumulative distribution plots of percentage change from baseline at Week 13 and Week 52 will be presented for the worst leg pain and the ODI score. Percentages will be out of all subjects, thus subjects with missing data will be counted as not having achieved a given percentage of improvement. Percentage of subjects will be on the Y axis and percentage of improvement will be on the X axis. Reference lines will mark 30% and 50% improvement.

7.3.4 Multiple Comparisons

To control family-wise error rate for multiple tests of the primary and key secondary endpoints, serial gatekeeping testing algorithm will be used.

1. The treatment effect on the primary endpoint will be evaluated at $\alpha = 0.05$.
2. If the effect on the primary endpoint is significant, the treatment effect on the key secondary endpoint of worst leg pain score at Week 52 will be evaluated using $\alpha = 0.05$.
3. If worst leg pain score at Week 52 is significant $\alpha = 0.05$, then the herniation volume at Week 13 will be tested at $\alpha = 0.05$.
4. If the herniation volume at Week 13 is significant $\alpha = 0.05$, then the ODI at Week 13 will be tested at $\alpha = 0.05$.

If at any point in the sequence an endpoint fails to meet significance at $\alpha = 0.05$, then the algorithm will be halted and the remaining tests declared non-significant regardless of their p-values. Endpoints and time points other than those above may have p-values reported, but none of these will be considered to be statistically significant.

8. SAFETY EVALUATION

8.1 Overview of Safety Analysis Methods

All safety analyses will be performed on the safety population (any subject that received an injection) and will be presented by the treatment subjects actually received. No imputation of missing values will be performed.

8.2 Adverse Events

All AEs will be coded using Medical Dictionary for Regulatory Activities (MedDRA) using the current version at the time coding is first initiated. TEAEs will be defined as any AEs that were recorded during or following the investigational product injection. Significant TEAEs will include TEAEs that result in study discontinuation. TEAEs of special interest will include TEAEs associated with hypersensitivity, imaging findings, leg pain, and back pain.

An overall display will show the number and percentage of subjects (with exact 95% CI) with TEAEs or treatment-related TEAEs that fall into the following categories: any TEAE; TEAE leading to death; serious TEAE; significant TEAE; TEAE associated with hypersensitivity; TEAE associated with imaging findings; TEAE associated with leg pain; TEAE associated with back pain. TEAEs related to study procedure will be summarized. In addition the number of subjects with TEAEs or treatment-related TEAEs will be summarized by the following:

Relationship to Study Drug	**Severity**	**Time of First Occurrence (after injection)**
Related AEs	Mild AEs	≤ 1 day
Not Related AEs	Moderate AEs	2 – 7 days
	Severe AEs	8 days – 13 weeks
		>13 weeks

Subjects experiencing AEs in more than one category will be counted in the most severe or most related group.

Detailed listings of all AEs will include severity and relationship to treatment, as well as action taken, subject outcome, and whether the AE is diagnosed or suspected COVID-19. Separate listings will be provided of subjects who experience serious TEAEs, significant TEAEs, and special interest TEAEs. Additionally, any information recorded for AEs of diagnosed or suspected COVID-19 regarding descriptions of signs/symptoms, treatments, diagnostic test results, and the resolution will be presented in a listing.

8.2.1 Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

The total number of TEAEs and the number and percentage of subjects with events will be presented by system organ class (SOC) and preferred term (PT). Subjects are counted once in each SOC and PT. Percentages will be based on the total number of treated subjects in the treatment group. SOC will be listed in the Internationally Agreed Order (found in MedDRA Data Retrieval and Presentation: Points to Consider, Release 3.2) and PTs within a SOC will be listed alphabetically.

This will be repeated for treatment-related AEs.

8.2.2 Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Relationship/Severity

The total number of TEAEs and the number and percentage of subjects with events will be presented by SOC and PT as described above by relationship to study drug (related, not related). At each level (SOC, PT, overall), subjects will be counted as related if they have any TEAE within that level that was deemed related.

Likewise, this presentation will be repeated by severity. Subjects will be counted at the highest severity that they experience with a given level of summarization (SOC, PT, overall).

Additionally, TEAEs related to the study procedure will be displayed as above as well as broken down by severity.

8.2.3 Serious AEs, AEs leading to death, Special Interest and Significant Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

The total number of TEAEs and the number and percentage of subjects with events will be presented by SOC and PT as described above for serious AEs, TEAEs leading to death, special interest TEAEs, and significant TEAEs.

TEAEs of special interest will include TEAEs associated with hypersensitivity, imaging findings, leg pain, and back pain; significant TEAEs will include TEAEs that result in study discontinuation. Specific PTs related to each of the special interest groupings will be described in the define file for the datasets.

The number of TEAEs of special interest and percentage of subjects with events will be summarized by severity and time of occurrence. Time of first occurrence will be based on time from injection and will include the following categories: ≤ 1 day; 2 – 7 days; 8 days to 13 weeks; > 13 weeks.

All the above will be repeated for treatment-related AEs only.

8.3 Vital Signs, Labs, Imaging Findings, and Other Observations Related to Safety

8.3.1 Vital Signs

Vital signs at each visit and changes from baseline will be summarized descriptively using the conventions for continuous variables. A listing of vital signs will be presented by treatment group, subject, vital sign, and visit.

8.3.2 Laboratory Tests

Clinical laboratory tests are collected at screening, Weeks 1, 6, 13 and at discontinuation. Changes from screening will be summarized descriptively using the conventions for continuous variables.

A listing of laboratory results will be presented by treatment group, subject, lab test, and visit. The results of pregnancy tests and urine drug screens will be presented in separate listings.

8.3.3 Imaging Findings

Counts and percentages of significant findings from the MRI and X-ray imaging will be presented by treatment for screening, Week 13 and Week 52; results from discontinuation visit and last visit for all subjects will be presented separately.

Findings to be presented will include:

- Disc height (disc index): decrease in disc height ≥30% compared to baseline value
- Vertebral posterior angle: vertebral body angle formed by flexion of ≥5°
- Vertebral body translation: vertebral body translation of ≥3 mm
- Changes of vertebral body endplates and adjacent bone marrow as visible by MRI. The criteria are as follows: Modic Type 1, Type 2, or Type 3 change in vertebral body endplates and adjacent bone marrow.

Disc height, vertebral posterior angle, and vertebral body translation will be summarized by visit for the observed values, the change from baseline, and the percent change from baseline (disc height only). Percent change from baseline in disc height, observed value of vertebral posterior angle and vertebral body translation will also be summarized in a line graph over time.

Summaries will use the adjudicated values, but all reader data will appear in the listings.

8.3.4 Post Treatment Lumbar Surgery

Counts and percentages of subjects with post-treatment lumbar surgery by Week 52 will be presented by treatment. Percentages will be out of subjects with successful contact.

This will be summarized for any surgery, LDH at a non-target level and surgery not for LDH.

9. OTHER ANALYSES

The association between imaging findings and clinical symptoms will be assessed. The following dichotomous imaging findings will be evaluated:

- Disc height (disc index): decrease in disc height ≥30% compared to baseline value
- Vertebral posterior angle: vertebral body angle formed by flexion of ≥5°
- Vertebral body translation: vertebral body translation of ≥3 mm
- Changes of vertebral body endplates and adjacent bone marrow as visible by MRI: Change of Modic Type 1, Type 2 and Type 3

For each outcome, treatment and level (decrease in disc height ≥30% and decrease in disc height <30%, for example) the following will be reported:

- Subjects with TEAEs associated with leg pain (number and percentage)
- Subjects with TEAEs associated with back pain (number and percentage)
- Subjects with worsening of worst leg pain (number and percentage)
- Subjects with worsening of worst back pain (number and percentage)
- Continuous summary statistics of worst leg pain
- Continuous summary statistics of worst back pain

These will be reported through Week 13 and Week 52.

10. INTERIM ANALYSES

No interim analyses are planned. All final analyses will be performed following database lock and unblinding.

11. CHANGES TO THE ANALYSES PLANNED IN THE PROTOCOL

The protocol-specified MI sensitivity analysis to the primary did not include consideration for subject data that is set to missing following a prohibited procedure; this has been included among the study outcomes that use the baseline distribution for imputation.

The multiple comparison algorithm has been updated to do a more straightforward serial gatekeeping approach.

The recurrence of LDH endpoint in the protocol references one of the treatment failure criteria; because this criteria was nested in other requirements that would conflict with the recurrence of LDH criteria, the text was updated to focus on only the applicable items. Additionally, surgical intervention was removed as a criteria and the symptom criteria were clarified.

12. REFERENCES

1. Dmitrienko A, Tamhane A. C, Wiens B. L General multistage gatekeeping procedures: *Biometrical Journal*. 2008 Oct 50(5):667-77.

13. APPENDICES

13.1 Schedule of Activities

	Screening				Follow up									
Visit	1	2			3	4	5	6	7	8	9	10	DC	
Week	-3	0			1	2	4	6	13	26	39	52	≤ 13	>13
Day	-21	1			8	15	29	43	92	183	274	365		
		Before Injection		After Injection										
Window (days)		-1			±1	±3	±3	±7	-7, +14	±14	±14	±14		
Informed consent	X													
Inclusion/exclusion criteria	X	X												
Demographics, medical history	X													
Concomitant medications/ rescue medication use	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Vital signs	X	X		X	X	X	X	X	X	X	X	X	X	X
Adverse events	→	→	→	→	→	→	→	→	→	→	→	→	→	→
Urine drug test	X	X												
Neurological examination	X	X			X	X	X	X	X	X	X	X	X	X
Laboratory tests[a]	X				X			X	X				X	
Pregnancy test[b]	X	X												
Imaging (X-ray, MRI)[c]	X								X			X	X	X
Central eligibility assessment for imaging	X													
Confirmation of wash out		X												
Daily diary (VAS pain assessment and rescue medication use)	→	→	→	→	→	→	→	→	→	→	→	→	→	→
Subject education[d]	X													
Distribution of rescue medication[e]	X			X	X	X	X	X	X	X	X			
ODI		X			X	X	X	X	X	X	X	X	X	X
SF-36		X					X	X	X	X	X	X	X	X
EQ-5D-5L		X							X			X	X	X
WPAI		X							X			X	X	X
Randomization[f]		X												
Injection[g]			X											
PGIC						X	X	X	X	X	X	X	X	X
CGIC						X	X	X	X	X	X	X	X	X
Telephone call before visit[h]									X	X	X	X		
Recurrence of LDH												X	X	X
Occurrence of lumbar surgery[i]												X	X	X
Collect remaining rescue medication												X	X	X

X=Essential examination; Solid line means assessment will be made throughout the period, and dotted line means the assessment will be made during the 14 consecutive days prior to the visit.

CGIC = Clinical Global Impression of Change; DC = Discontinuation; EQ-5D-5L = EuroQol Group 5-Dimension Quality of Life instrument, 5-level version; LDH = Lumbar Disc Herniation; MRI = Magnetic Resonance Imaging; ODI = Oswestry Disability Index; PGIC = Patient Global Impression of Change; SF-36 = 36-Item Short-Form Health Survey; VAS = Visual Analogue Scale; WPAI = Work Productivity and Activity Impairment Questionnaire.

a Laboratory tests are provided as listed in Table 2 of the study protocol.

b Women of childbearing potential only. Blood test will be performed at Visit 1, and urine test will be performed at Visit 2.

c If repeat imaging is needed, the window for imaging will be extended by a maximum of an additional 7 days.

d Subjects will be educated on appropriate expectations around their participation in a clinical study and the importance of reliably consistently and accurately reporting their pain throughout the study.

e If a subject has enough remaining acetaminophen, additional medication will not be provided.

f Subjects may be randomized on the day before (Day -1) or the day of (Day 1) investigational product administration.

g Treatment should be administered within 3 weeks after informed consent. If repeat imaging is needed, the window for screening will be extended by a maximum of an additional 7 days.

h Site staff should contact a subject by telephone call and/or other method by a minimum of 14 days before scheduled visit.
i Subjects who discontinued the study will be followed until the Week 52 Visit.

13.2 Randomization Seeds

The following is the list of seeds to be used for any imputations; they will be used in the order that they appear in the production dataset. If the list below is insufficient for the number of seeds required, the list will repeat with values 1 greater than those given.

514895564
763527595
221543962
855875309
919252181
252975352
524264245
800238255
734425865
366652284

14. ATTACHMENTS