Statistical Analysis Plan
APSS-66-00

Protocol Title:
A Two-Phase, Multicenter, Randomized Study Comparing Autologous Protein Solution with Hyaluronic Acid Intra Articular Injections in Patients with Knee Osteoarthritis

Zimmer Biomet, Inc.

Version 2

Date: September 24, 2019

Table of Contents

1. Study Details and Objectives

The nSTRIDE Autologous Protein Solution (APS) Kit is designed to safely and rapidly prepare APS from a small blood sample at the patient's point of care. The APS is to be injected intra-articularly for the treatment of knee osteoarthritis (OA) and associated symptoms. This study will evaluate the effectiveness and safety of use in a knee OA population.

The primary objective of this study is to determine whether nSTRIDE APS is superior to hyaluronic acid (HA) in improving mean Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) LK 3.1 pain. The metric used will be change from baseline to 12 months post-injection. The WOMAC LK 3.1 is a validated tool used for assessing knee pain, function, and stiffness. Thus, the WOMAC is comprised of three subscales. Of the three subscales comprising the WOMAC, the mean change in pain subscale score is the primary endpoint in this study. The hypothesis is that the APS group will demonstrate greater improvement than the HA group at 12 months post-injection. Further detail on the primary hypothesis is given in Section 4.2.

The impact of the usage of APS on function will be evaluated in support of the primary analysis; however, no formal statistical test for the superiority of function will be incorporated into the study success criteria. Instead, a qualitative assessment of the changes in WOMAC Function over time for APS and Hyaluronic Acid (HA) will be discussed in order to ensure that the impact of the treatment on function is neutral or positive. Further detail on this qualitative assessment is given in Section 4.2.

Secondary objectives of this study include determining whether nSTRIDE APS is superior to HA in improving WOMAC LK 3.1 function at 12 months (as evaluated using the percentage of subjects showing at least the minimal clinically important difference (MCID)), and WOMAC LK 3.1 pain at 12 months (as evaluated using the percentage of subjects showing at least the minimal clinically important difference (MCID)), OMERACT-OARSI responder rates at 12 months, analyzing WOMAC pain and function in only the KL-II Subgroup at 12 months, evaluating superiority of APS over HA in improving Numeric Rating Scale (NRS) pain at 12 months, assessment of the changes in WOMAC Pain scores over time (baseline through 12 months), and evaluation of the usage of rescue medication within 12 months. These will be tested in a fixed sequential order pending the rejection of the primary null hypothesis.

Exploratory objectives of this study include determining whether APS is superior to HA with regards to improvement in mean EQ-5D outcomes (12 months minus baseline) and repeated measures of WOMAC Function, WOMAC stiffness, and total WOMAC score. In addition, changes in joint morphology (determined with X-ray images) will be evaluated. Subgroup

analyses will be conducted to investigate whether the treatment effect varies depending on subjects' initial pain and baseline information. Analysis of exploratory endpoints will be done using alpha=0.05 with no adjustment for multiple comparisons.

A long-term follow-up phase will examine the superiority of nSTRIDE APS in the duration of the treatment effect, injection frequency, patient preferences, healthcare resource utilization, and associated costs. These endpoints will be treated as exploratory endpoints and analyses will be conducted using alpha = 0.05 with no adjustment for multiple comparisons.

Finally, the safety profile of nSTRIDE APS will be compared to HA by comparing adverse events and rates between treatment groups.

1.1 Background of the Study

Modalities presently available for treating early to moderate osteoarthritis are palliative, without exception [1]. The nSTRIDE APS Kit builds upon *in vitro* studies, animal studies, and a limited number of human studies designed to understand the causes of osteoarthritis. These research efforts suggest that osteoarthritis is associated with an imbalance in cytokines and growth factors [2]. This imbalance adversely affects cartilage, bone, and soft tissues.

Tissue remodeling is perpetually ongoing in the body. It is a cycle of tissue breakdown and rebuilding. In osteoarthritis, findings suggest that cytokines associated with cartilage breakdown are abnormally numerous when compared to the number of tissue-building cytokines; in particular Interleukin-1β (IL-1β) and Tumor Necrosis Factor alpha (TNFα)) [3-5].

The nSTRIDE APS Kit concentrates beneficial tissue-building cytokines and growth factors present in the patients' own blood in a way designed to block the activity of cytokines that break tissues down. In particular, APS concentrates Interleukin-1 receptor antagonist (IL-1ra), soluble Interleukin-1 Receptor II (sIL-1RII), and soluble Tumor Necrosis Factor Receptors I and II (sTNF-RI and sTNF-RII) [6]. This combination of concentrated beneficial factors is intended to act antagonistically to the pro-inflammatory factors that cause cartilage degradation and inflammation of the joint (IL-1β and TNFα [7, 8]. The blockade of inflammation by APS has been demonstrated *in vitro* [9, 10].

1.2 Study Design

The study is a two-phase, randomized, double-blind study with a planned enrollment of 246 subjects assigned to treatment groups on a 1:1 basis. It is designed to determine whether APS provides a more effective treatment for knee osteoarthritis than HA. The study will be conducted in two phases. Phase I is the randomized, double-blind portion of the study. After each subject completes all 12 month follow-up evaluations, only subjects will be blinded to the individual treatment allocation resulting in single-blind design following 12-month follow-up time point. Subjects from both groups will be able to request additional injections of their originally assigned treatment as frequently as needed, provided they did not experience any significant clinical concerns after previous treatment administrations and are benefiting from it, as determined by the investigator. During the long-term follow-up phase (12 – 60 months), subjects may also elect to cross over to the other treatment arm and receive additional injections of the other treatment as frequently as needed, provided they did not experience any significant clinical concerns after previous treatment administrations and are benefiting from it, as determined by the investigator. Subjects may only cross over from their originally assigned treatment group to the other treatment group one time during the study.

1.3 Randomization

Subjects will be randomized to either the APS or HA treatment group on a 1:1 basis. The subjects and investigators will both remain blinded throughout Phase I of the study. Assignment to treatment groups will be stratified by site and will use random block sizes. Specifically, designated unblinded personnel at each site will provide the masked syringe to the injecting health care provider. To protect the double-blind, the treating health care professional and evaluating health care professional will be different individuals. Assignment will not be revealed until after the blood has been drawn for nSTRIDE APS Kit processing.

The randomization plan will be produced using SAS v 9.4 or similar software. Balanced randomization with random block sizes (1:1, APS: HA) will be implemented. In the event that, post-randomization, no study treatment was given, randomization will not be reassigned; however, this case will not count toward the overall sample size. Randomization will continue with the next case enrolled until the minimum sample size is reached in both treatment groups. Randomization will be stratified by site, and each site will receive separate randomization plans using random predetermined block sizes that will remain undisclosed to the sites. The randomization file will be uploaded into the electronic data capture (EDC) system. Once the subject is enrolled into the EDC and has been identified as eligible for randomization, the randomization allocation will be visible within the EDC and viewable only by the unblinded research associate. The subject will be treated according to the contents of the displayed randomization allocation.

2. Analysis of Baseline Data

Analyses will be done to determine whether randomization succeeded in creating groups that were balanced with regards to key baseline characteristics. Planned tests aimed at determining this are shown in **Table *1***. All tests will be under a null hypothesis of no difference between treatment groups.

These tests are intended to be used along with the magnitude/variation as a guideline to see whether any significant baseline imbalances have occurred that would need to be taken into consideration in the analysis of outcomes.

Table 1. Randomization Verification

Variable	Test	Expected Outcome
Gender	Fisher's Exact	No Statistically Significant Difference
Age	T-Test	No Statistically Significant Difference
BMI	T-Test	No Statistically Significant Difference
Race	Likelihood Ratio Chi-Square	No Statistically Significant Difference
Baseline WOMAC Pain (Screening)	T-Test	No Statistically Significant Difference

A table showing major and minor protocol deviations will be generated comparing frequency of the occurrence of major and minor deviations between the two treatment groups.

Use of restricted medications will be summarized by type and compared between the two treatment groups.

3. Subject Disposition

A subject disposition table will be created showing the number of subjects in each treatment group that completed the study, with the number of 'in window' and 'all' visits indicated. Visits occurring outside of windows are not considered to be protocol deviations, but these windows can be used for the reporting of Adverse Events and Study Terminations, as well as for patients who return for follow-up outside of the protocol assigned visit window.

Visit windows are defined in **Table *2*** and **Table *3***.

Table 2. Visit Windows

Visit	1 Month	3 Month	6 Month	12 Month
In-Window	± 7 Days	± 14 Days	± 14 Days	± 28 Days
Days since Injection for In-Window Visits	23-37	77-105	169-197	337-393
Days since Injection for All Visits	1-61	62-137	138-274	275-456

After each subject completes all 12 month follow-up evaluations, individual treatment allocation will be unblinded. From this time-point on, only subjects will be blinded to the individual treatment allocation resulting in a single-blind design during the long-term follow-up period. Long-term follow-up assessment visits will be at the following intervals:

Table 3. Long-Term Visit Windows

Visit	18 Month	24 Month	30 Month	36 Month	42 Month	48 Month	54 Month	60 Month
In-Window	± 28 Days	± 28 Days	± 28 Days	± 28 Days	± 28 Days	± 28 Days	± 28 Days	± 28 Days
Days since Injection for In-Window Visits	520-576	703-759	885-941	1068-1124	1250-1306	1433-1489	1616-1672	1798-1854
Days since Injection for All Visits	457-639	640-821	822-1004	1005-1187	1188-1369	1370-1552	1553-1735	1736+

4. Effectiveness Outcomes

4.1 General Methods

The continuous variables included in the secondary and exploratory analyses will be tested for homoscedasticity using an F test and also graphically evaluated for normality (boxplots, histograms, and/or normal probability plots). If the p-value for the F-test is less than 0.05 or the plots indicate that the data are extremely non-normal, then the T-tests indicated in **Table *4*** and **Table *5*** may be replaced by an appropriate alternative (Satterthwaite T-test where unequal variances are found or Wilcoxon Mann-Whitney U-test (WMW) where evidence of extreme non-normality is found).

Table 4 and **Table 5** present a comprehensive list of all planned tests, excluding tests of imaging (which will be presented separately).

4.2 Primary Outcome

The primary objective of this study is to determine whether nSTRIDE APS is superior to HA with respect to the improvement in mean WOMAC LK 3.1 pain score (change from baseline to 12 months post-injection). The primary hypothesis described below will be tested along with the secondary hypotheses listed in Section 4.3 using a conventional fixed-sequence procedure [11, 12], constructed using a pre-specified order of hypotheses. These tests will be performed at the 0.05 level. If the first null hypothesis is rejected, the second test will be performed at the 0.05 level. If the first null

hypothesis is not rejected, the second test, and all subsequent tests, will be performed as exploratory analyses. Since the order of the tests is fixed a priori, and the second hypothesis is tested only if the previous hypothesis has been rejected, the principle of closed testing implies that no adjustment to control the familywise error rate is necessary. The Fixed-Sequence Method is described as an acceptable method for addressing the multiplicity problem in Section IV.C.5 of the FDA draft guidance entitled, "Guidance for Industry – Multiple Endpoints in Clinical Trials".

A two-tailed independent sample T-test will be used to test the primary endpoint.

The hypotheses will be:

H_0: $\mu_{APS} = \mu_{Control}$

Versus

H_A: $\mu_{APS} \neq \mu_{Control}$

A finding of nSTRIDE APS superiority on the WOMAC LK 3.1 Pain subscale will be considered evidence of nSTRIDE APS superior efficacy, and the device will be considered more efficacious than HA for the treatment of knee pain associated with OA at 12 months post-injection.

The impact of the usage of APS on function will be evaluated as part of the primary endpoint; however, no formal statistical test will be done as a part of the study success criteria. A qualitative assessment of the changes in WOMAC Function over time for APS and HA will be discussed, in order to ensure that the impact of the treatment on function is neutral or positive. The following descriptive analyses will be performed for this assessment of function:

(1) A graphical examination of the changes in WOMAC Function for APS and HA over time, including means and standard error bars.
(2) Descriptive statistics for the mean changes in WOMAC Function over time, separately for each treatment group. These will include mean, median, standard deviation, minimum, maximum, and 95% confidence intervals.

For each analysis, it is expected that the changes in WOMAC function will remain neutral or increase over time within the APS group, and also that the mean changes in WOMAC Function over time will be at least nominally better for the APS group than for HA.

Note that a repeated measures ANOVA will also be used to assess changes in WOMAC Function over time; this analysis will be exploratory in nature and is described in **Section 4.4.**

4.3 Secondary Outcomes

Secondary objectives of this study include determining whether nSTRIDE APS is superior to HA with regard to the endpoints in **Table *4***. These tests will be performed if the primary null hypotheses have been rejected. These secondary hypotheses will be tested using a fixed-sequence procedure, constructed using a pre-specified order of hypotheses. The order is shown in **Table *4***. These tests will be performed at the 0.05 level for each consecutive test until one hypothesis is tested not significantly, after which all subsequent tests will not be performed as secondary analyses, but instead will be performed as exploratory analyses only. Since the order of the tests is fixed a priori, and each subsequent hypothesis is tested only if the previous hypotheses have been rejected, the principle of closed testing implies that no adjustment to control the familywise error rate is necessary [11].

In particular, it is hypothesized that APS is superior to HA with regard to an improvement in both function and pain; thus these comprise the first two secondary endpoints tested. The subjects showing a minimum clinically important difference (MCID) in the WOMAC Function subscale will be referred to as “MCID Function responders.” The proportion of subjects showing a MCID in the WOMAC Pain subscale (i.e. “MCID Pain responders”) will be tested as the second-ranked secondary endpoint.

The development of the MCIDs for Function and Pain for which a patient is designated as a “function responder” or a “pain responder” is based on the data from the PROGRESS II pilot study. It is advantageous to use this data for the following reasons:

- The MCID is very much dependent on the choice of clinical score and the scale/version of the score and the follow up time point, so the best approach is to use pilot data where these variables are the same as in the current study.
- The Progress II data allows for development of an MCID that is context-specific, as it has the same or similar patient population, time point, baseline characteristics, and baseline symptom severity as this PROGRESS V study. These factors are important to consider in establishing an MCID [13].
- The Pilot study collected a transition question in which patients could rate their improvement as “Very Much Improved”, “Improved”, “Minimally Improved”, “No Change”, “Minimally Worse”, “Much Worse”,

and “Very Much Worse”. This question can be used as external criteria to define patients who have experienced a meaningful change in their condition [14].

- Other potential anchor questions are also available in the Pilot study, serving as a way to verify the MCIDs and make sure that the MCIDs based on these different possible anchor questions are consistent.
- The change that a patient rates as “clinically significant” is dependent on the expectations of the patient [15]. This can vary by treatment as well as by time period.

The MCIDs were derived using an anchor-based method with the patient transition question described above. This question is an appropriate anchor as it is easily clinically interpretable, as well as correlated with the WOMAC Function and Pain scores as shown in the table below. In the Pilot study, the mean improvement in WOMAC Function and Pain scores increases with each increasing patient rating:

TRANSITION QUESTION	N	WOMAC ADL Mean Improvement from Baseline to 12 months	WOMAC Pain Mean Improvement from Baseline to 12 months
Very Much Worse	1	-7.0	-1.0
Much Worse	1	4.5	1.5
Minimally Worse	5	3.8	3.2
No Change	6	7.4	4.2
Minimally Improved	9	15.8	5.8
Much Improved	18	25.3	8.4
Very Much Improved	4	37.8	10.9

Thus the anchor is appropriate as described in methodology in Guyatt, et al.[16] and the MCID is calculated as follows:

The AAOS published a guideline in which calculations of minimum clinically important improvement (MCII) were presented [17]. These calculations of MCII were based on patients with knee osteoarthritis whose final outcome of treatment was “good, satisfactory effect with occasional episodes of pain or stiffness.” The final response to treatment anchored by the baseline value was calculated for each patient. The determinations of clinical significance required patients in the included studies to achieve a change score comparable to that achieved by 75% of patients reporting good outcomes in the population.

In order to implement this method, the PGI-C response that most closely corresponds to a “good” result was determined. The closest category to the

AAOS definition was the “much improved” response (assuming that “minimally improved” corresponds to “fair”, and “much improved” corresponds to “good”). The 25th percentile of the distribution of WOMAC Pain scores (improvement from baseline) for these subjects corresponds to a score achieved by 75% of the patients reporting a “much improved (= good)” outcome.

In order to verify the MCID, other anchor question alternatives corresponding to a “good” outcome were explored:

- The SF-36 health transition question, using subjects who respond with “somewhat better”
- The symptom to benefit ratio, which takes into account the symptoms described by the AAOS definition “occasional episodes of pain or stiffness”. This includes subjects in the “Moderate/None” category.

The change from baseline in WOMAC function and pain scores for subjects within the appropriate category of each of the three measures were used to calculate and verify the MCID. The MCIDs (i.e. 25th percentile of the distribution of WOMAC Function and Pain Improvement scores) calculated using each of these three anchor questions are presented below.

Transition Question	**MCID for Improvement in WOMAC Function from Baseline to 12 Months**	**MCID for Improvement in WOMAC Pain from Baseline to 12 Months**
Based on 25th percentile of the PGI = “Much Improved” subjects	20.0 points	7.0 points
Based on 25th percentile of the SF-36 health transition question	20.5 points	7.0 points
Based on 25th percentile of the symptom to benefit ratio = “Moderate/None”	16.5 points	7.0 points

The 25th percentiles for each of the three measures were remarkably similar to each other.

To further assess the validity of the MCIDs, the percentage of subjects in each transition category who met the MCID for Function and Pain were calculated, and are as follows:

TRANSITION QUESTION	PERCENT OF SUBJECTS MEETING WOMAC Function MCID	PERCENT OF SUBJECTS MEETING WOMAC Pain MCID
Very Much Worse	0%	0%
Much Worse	0%	0%
Minimally Worse	0%	0%
No Change	16.7%	50%
Minimally Improved	44.4%	55.6%
Much Improved	77.8%	83.3%
Very Much Improved	100%	100%

Based on this data, the MCID is an appropriate differentiator of subjects who are Much improved or Very Much Improved.

Therefore, the MCID Function responder criterion is as follows:

MCID Function Responder:

A subject is considered an MCID Function responder if they show an absolute improvement of ≥ 20.0 points in WOMAC Function from baseline to 12 months.

A two-tailed Fisher's Exact will be used to test the first secondary endpoint. The hypotheses will be:

H_0: $p_{APS} = p_{Control}$

Versus

H_A: $p_{APS} \neq p_{Control}$

Where:

p_{APS} = proportion of MCID function responders in the APS group, and

$p_{control}$ = proportion of MCID function responders in the Control group

MCID Pain Responder:

A subject is considered an MCID Pain responder if they show an absolute improvement of ≥ 7.0 points in WOMAC Pain from baseline to 12 months.

A two-tailed Fisher's Exact will be used to test the first secondary endpoint. The hypotheses will be:

H_0: $p_{APS} = p_{Control}$

Versus

H_A: $p_{APS} \neq p_{Control}$

Where:

p_{APS} = proportion of MCID pain responders in the APS group, and

$p_{control}$ = proportion of MCID pain responders in the Control group

The next sequential secondary endpoint will be a comparison of the mean change in WOMAC function score change from baseline to 12 months post-injection.

These analyses of WOMAC function and pain were chosen as the first, second, and third-ranked secondary endpoints for analysis because increased pain and decreased function are the predominant clinical findings associated with OA, making both pain and functional improvement important aspects in the treatment of OA. It is anticipated that if treatment with APS decreases pain, it may also increase function.

Following the test of the mean change from baseline in WOMAC function, the OMERACT-OARSI Responder Criteria [18] will be applied to both treatment groups, categorizing each patient into one of two categories: responder and non-responder. These results will be tested to determine whether a difference between treatment groups exists. With respect to the OMERACT-OARSI classification, all measurements where a pain measurement is called for will employ NRS pain. For evaluation of function in the OMERACT-OARSI classification, all measurements will be based on WOMAC function scores. For the global assessment variable, the EQ-5D global assessment VAS scale will be used for OMERACT-OARSI classification. For the function score, absolute change must equal 20% of the total possible score to meet the criteria of an absolute change ≥ 20, and 10% of the total possible score to meet the criteria of an absolute change ≥ 10.

Responders will be defined as subjects who achieved a high degree of improvement in pain or in function (improvement of ≥ 50% and absolute change ≥ 20), or a moderate degree of improvement in 2 of the 3 response domains (pain, function, global assessment) as follows (Figure 1):

(1) NRS pain improvement of ≥ 20% and absolute improvement of ≥ 1 point
(2) WOMAC function improvement of ≥ 20% and absolute improvement of ≥

6.8 points (3) EQ-5D global assessment improvement of ≥ 20% and absolute improvement ≥ 10.

In the secondary analysis of OMERACT-OARSI responders at 12 months (**Table *4***), a subject who has recurring (two or more) documented uses of rescue medication within 48 hours of the 6 or 12 month visits for index knee OA, or recurring (two or more) documented uses of restricted medication for index knee OA between the 6 and 12 month visits will be classified as a "non-responder" at the 12 month visit. If rescue medication use is reported but number of uses is unclear, it will be assumed for purposes of this analysis that the usage is recurring.

The OMERACT-OARSI classification will be calculated as long as there is enough information to calculate per the definition.

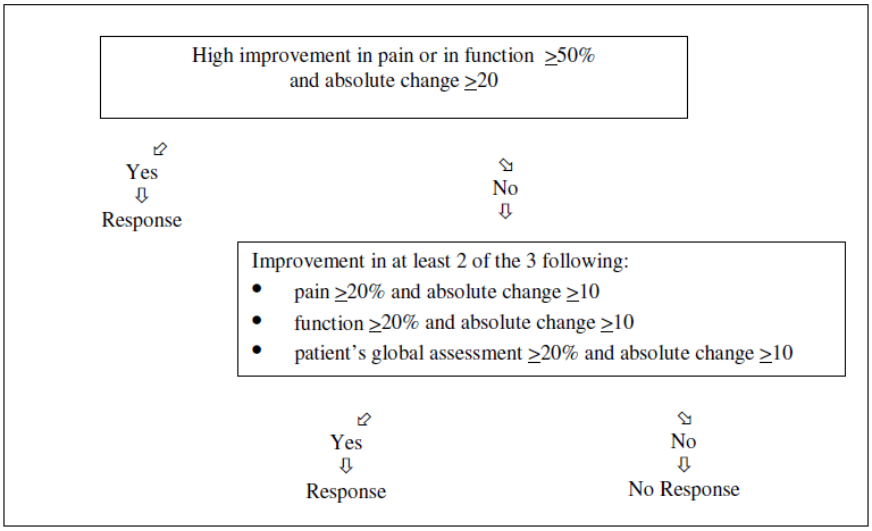

Figure 1. OMERACT-OARSI Responder Criteria [18]

In addition to the test for OMERACT-OARSI classification for the entire study population, the primary efficacy test will be repeated in the subgroup of subjects with Kellgren-Lawrence Grade II (K-L II) OA at baseline. There is supporting evidence in the literature that the lower grade OA patient population may be more responsive to injection therapy [19, 20].

The Kellgren-Lawrence subgroup analyses are ranked below the OMERACT-OARSI classification for the overall study population because the outcomes of the general OA population have greater utility than subgroup analyses.

An analysis of the usage of rescue and restricted medication is considered an important endpoint as it can affect the assessment of pain; thus, this is also a key secondary endpoint. This endpoint is also incorporated as covariate in an ANOVA model to determine if the usage of pain medication has any effect on the mean change in WOMAC Pain from baseline to 12 months.

Table 4 - Planned Secondary Efficacy Tests

Order of Testing	Secondary Outcomes	Test	Expected Outcome
1	WOMAC MCID Function Responder rate	Fisher's Exact	APS Superior
2	WOMAC MCID Pain Responder rate	Fisher's Exact	APS Superior
3	Mean WOMAC Function Δ (12 Month minus Baseline)	T-Test or WMW	T-Test or WMW
4	OMERACT-OARSI Responder / Non-responder (12 Month)	Fisher's Exact	APS Superior
5	Mean WOMAC Pain Δ (12 Month minus Baseline) in K-L II Subgroup	T-Test or WMW	APS Superior
6	Mean WOMAC Function Δ (12 Month minus Baseline) in K-L II Subgroup	T-Test or WMW	APS Superior
7	Use of rescue medication (acetaminophen) use (for index knee OA) over time	Fisher's exact test	APS Superior
8	Mean WOMAC Pain Δ (12 Month minus Baseline) with Usage of rescue medication within 48 hours of the 12 month visit for knee OA as a covariate	ANOVA	Exploratory

Order of Testing	Secondary Outcomes	Test	Expected Outcome
9	Mean WOMAC Pain Δ (12 Month minus Baseline) with Usage of rescue medication within 48 hours of the 12 month visit for any reason as a covariate	ANOVA	Exploratory
10	Mean WOMAC Pain Δ (12 Month minus Baseline) with Usage of restricted medication within 48 hours of the 12 month visit for knee OA as a covariate	ANOVA	Exploratory
11	Mean WOMAC Pain Δ (12 Month minus Baseline) with Usage of restricted medication within 48 hours of the 12 month visit for any reason as a covariate	ANOVA	Exploratory
12	Mean NRS Pain Δ (12 Month minus Baseline)	T-Test or WMW	APS Superior
13	Mean WOMAC Pain changes over time within treatment	Repeated measures ANOVA	Significant change from baseline

4.4 Exploratory Outcomes (Phase I)

4.4.1 Questionnaire Data

Exploratory objectives of this study include determining whether nSTRIDE APS is superior to HA with regard to the endpoints in **Table *5***. Where repeated measures tests are statistically significant, tests at individual time

points will be performed. All tests will be conducted at α = 0.05 and will not be adjusted for multiple comparisons as these are exploratory outcomes.

Table 5 – Planned Exploratory Analyses, Questionnaire Data

Exploratory Outcomes	Test	Expected Outcome
Mean NRS Pain Δ (1, 3, 6 Months)	Repeated Measures ANOVA	Trend APS Superior
Mean WOMAC Pain Δ (Percent change from baseline to12 Month)	T-Test or WMW	APS Superior
Mean WOMAC Function Δ (Percent change from baseline to12 Month)	T-Test or WMW	APS Superior
Mean WOMAC Function changes over time within treatment	Repeated measures ANOVA	Significant change from baseline
Mean EQ-5D Δ Global Assessment VAS (12 Month minus Baseline)	T-Test or WMW	APS Superior
Mean EQ-5D Δ Global Assessment VAS (Percent change from baseline to12 Month	T-Test or WMW	APS Superior
Mean EQ-5D Global Assessment VAS changes over time within treatment	Repeated measures ANOVA	Significant change from baseline
Mean EQ-5D *Δ Single Index Value (12 Month minus Baseline)	T-Test or WMW	APS Superior
Mean EQ-5D Δ Single Index Value (Percent change from baseline to12 Month	T-Test or WMW	APS Superior
Mean EQ-5D Single Index Value changes over time within treatment	Repeated measures ANOVA	Significant change from baseline

Exploratory Outcomes	Test	Expected Outcome
EQ-5D Dimensions (Mobility, Self-Care, Usual Activities, Pain/Discomfort, Anxiety/Depression) (12 Month)	Likelihood Ratio Chi-Square	APS Superior
Mean WOMAC Pain Δ (1, 3, & 6 Months)	Repeated Measures ANOVA or Friedman Test	Trend APS Superior
Mean WOMAC Function Δ (1, 3, & 6 Months)	Repeated Measures ANOVA or Friedman Test	Trend APS Superior
MCID WOMAC Function Responder (1, 3, & 6 Months)	Fisher's Exact Family α = 0.05	Trend APS Superior
MCID WOMAC Pain Responder (1, 3, & 6 Months)	Fisher's Exact Family α = 0.05	Trend APS Superior
OMERACT-OARSI Responder / Non-responder** (1, 3, 6 & 12 Months) **Traditional published criteria; i.e. not including multiple restricted/ rescue medication usage between 6-12 months	Fisher's Exact Family α = 0.05	Trend APS Superior
Total Mean WOMAC Δ (1, 3, & 6 and 12 Months)	Repeated Measures ANOVA or Friedman Test	Trend APS Superior
Mean WOMAC Stiffness Δ (1, 3, & 6 and 12 Months)	Repeated Measures ANOVA or Friedman Test	Trend APS Superior
Mean EQ-5D Δ Global Assessment VAS (1, 3, & 6 Months)	Repeated Measures ANOVA or Friedman Test	Trend APS Superior

Exploratory Outcomes	Test	Expected Outcome
EQ-5D Dimensions (Mobility, Self-Care, Usual Activities, Pain/Discomfort, Anxiety/Depression) (1, 3, & 6 Months)	Likelihood Ratio Chi-Square	Trend APS Superior
Mean WOMAC Pain Δ (12 Month minus Baseline) by Treatment and Site	ANOVA	No Site Effect
Mean WOMAC Function Δ (12 Month minus Baseline) by Treatment and Site	ANOVA	No Site Effect
Mean WOMAC Function Δ (12 Month minus Baseline) with Usage of rescue medication within 48 hours of the 12 month visit for knee OA as a covariate	ANOVA	Exploratory
Mean WOMAC Function Δ (12 Month minus Baseline) with Usage of rescue medication within 48 hours of the 12 month visit for any reason as a covariate	ANOVA	Exploratory
Mean WOMAC Function Δ (12 Month minus Baseline) with Usage of restricted medication within 48 hours of the 12 month visit for knee OA as a covariate	ANOVA	Exploratory

Exploratory Outcomes	Test	Expected Outcome
Mean WOMAC Function Δ (12 Month minus Baseline) with Usage of restricted medication within 48 hours of the 12 month visit for any reason as a covariate	ANOVA	Exploratory

4.4.2 Exploratory Outcomes (Imaging)

X-rays are evaluated for Kellgren-Lawrence (K-L) grade and absence of severe osteoarthritis by the central laboratory for confirmation of eligibility.

Post-injection X-rays taken at 12 months will receive K-L grades that will be compared to baseline. For each subject, results will be categorized as (1) worse—a higher Kellgren-Lawrence grade at 12 months compared to baseline, (2) no change in grade from baseline to 12 months or (3) Improvement—a decrease in grade from baseline to 12 months. An analysis of the percentage of subjects in each treatment group who fall into each of these three categories will be done using a Likelihood Ratio Chi-square test. This test will also be repeated within subgroups defined by the baseline K-L grade.

Measurements of medial and lateral Joint Space Width (JSW) will be performed to assess narrowing over time in the medial and lateral compartments of the treated joint. Joint Space Narrowing (JSN) will be calculated for each compartment as the change in JSW between the 12 Month visit and the Baseline visit.

A list of planned statistical tests on imaging data is presented in **Table *6***. **.**All tests will be conducted at α = 0.05 and will not be adjusted for multiple comparisons as these are exploratory outcomes. In addition to these tests, logistic regression will be used to determine whether imaging results are associated with OMERACT-OARSI categorization.

Table 6 - Planned Statistical Tests for X-ray Results

Variable	Test
X-ray	
Change in K-L	Likelihood Ratio Chi-Square
Change in K-L by baseline K-L grade	Likelihood Ratio Chi-Square for each subgroup

Mean JSW at 12 Months	T-Test or WMW
Mean JSN (12 months minus baseline)	T-Test or WMW

4.5 Exploratory Outcomes, Phase II

Tests planned for Phase II are shown in **Table *7*** and **Table *8***. With regard to the analyses with outcome variable "Treatment Success", Treatment Success will be defined as follows:

Treatment Success from Original Injection

(1) No additional injections
(2) No cross-over or study exit for other invasive treatment
(3) Clinically meaningful improvement in WOMAC Pain score (>20% from baseline)

Treatment Success from Original Treatment Assignment

(1) No cross-over or study exit for other invasive treatment
(2) Clinically meaningful improvement in WOMAC Pain score (>20% from baseline)

The items in **Table *7*** will be analyzed according to the original treatment assignment (APS or HA). The items in **Table *8*** will use Treatment / Crossover group as the predictor variable: (1) randomized to APS and not crossed over to HA, (2) randomized to APS and crossed over to HA, (3) randomized to HA and crossed over to APS and (4) randomized to HA and not crossed over to APS. For all tests where the Treatment / Crossover Group is the predictor variable, patients progressing to other invasive treatment for their OA will be included as a Non-Responder (where OMERACT-OARSI responder is the outcome variable) or using the most recently collected outcome score (where pain, function or EQ-5D is the outcome variable).

Table 7. Planned Phase II Statistical Analysis-Originally Assigned Treatment

Predictor Variable	Outcome Variable(s)	Statistic	Expected Outcome
Treatment Group	Election to Cross Over	Fisher's Exact	Statistically Significant, More HA than APS Cross Over
Treatment Group	Percentage of patients achieving	Fisher's Exact	APS Superior

	treatment success from initial injection		
Treatment Group	Percentage of patients achieving treatment success from originally assigned treatment	Fisher's Exact	APS Superior
Treatment Group	Percentage of patients achieving success in each treatment arm (i.e. patients with clinically meaningful improvement in WOMAC Pain score, inclusive of originally assigned and cross-over patients)	Fisher's Exact	APS Superior
Treatment Group	Average number of cumulative interventions per patient	Likelihood Ratio Chi-Square	Trend APS Superior
Treatment Group	Time from the initial injection (nSTRIDE APS or HA) to first subsequent injection, cross-over, or study exit for other invasive treatment	Kaplan Meier – Time to first subsequent injection/cross-over/other invasive treatment	Trend APS Superior

Table 8 - Planned Phase II Statistical Analysis – Treatment/Crossover Group

Predictor Variable	Outcome Variable(s)	Statistic	Expected Outcome
Treatment / Crossover Group	Progression to other invasive OA treatment	Chi-Square	Trend APS Superior

Treatment / Crossover Group	OMERACT-ORSI Responder / Non-Responder	Chi-Square	Trend APS Superior
Treatment / Crossover Group	WOMAC Pain, 6 Mo Intervals	One-Way Anova	Trend APS Superior
Treatment / Crossover Group	WOMAC Function, 6 Mo Intervals	One-Way Anova	Trend APS Superior
Treatment / Crossover Group	NRS Pain, 6 Mo Intervals	One-Way Anova	Trend APS Superior
Treatment / Crossover Group	EQ-5D Global Assessment 6 Mo Intervals	One-Way Anova	Trend APS Superior
Treatment / Cross Over Group	Patient Preference	Chi-Square	Trend APS Superior
Treatment / Cross Over Group	WOMAC Pain Δ: Baseline-12M	Summary Statistics and ANCOVA Model*	Trend APS Superior
Treatment / Cross Over Group	WOMAC Pain Δ: Baseline-6M	Summary Statistics and ANCOVA Model*	Trend APS Superior
Treatment / Cross Over Group	Percent of Subjects in each Treatment/Cross Over Category	Chi-Square	Trend APS Superior

*Analysis of Covariance models will be used in order to determine if changes in WOMAC Pain score from Baseline to12 months (first model) or from 6 to 12 months (second model) differ according to what type of 2nd injection they chose to receive (β_3), and whether this difference varies for APS and HA ($\beta_2 * \beta_3$). The model will be as follows:

$Y=\beta_0+ \beta_1+ \beta_2+ \beta_3+ \beta_2*\beta_3$

Where

β_0= Intercept

β_1= Baseline WOMAC Pain score

β_2 = Treatment Group

β_3= 2nd Injection (None, Same, Crossover)

Y= Change in WOMAC Pain

In addition to these tests, results of the patient preference questionnaire will be tabulated by crossover group.

4.6 Health Economic Outcomes

Additional exploratory analyses related to resource cost and utilization will be completed. Specifically, the analyses listed in **Table *9*** will be conducted. Note that resource cost data will not be collected during Phase 1 of the study, so the analyses which relate to cost will only be conducted using Phase II data.

Table 9 - Planned Phase II Statistical Analysis – Health Economic Outcomes

Predictor Variable	Outcome Variable(s)	Statistic	Expected Outcome	Time points
Treatment / Crossover Group	Difference in Quality-Adjusted Life Years (QALY)	One-Way Anova	Trend APS Superior	Time of initial treatment until "failure" or end of study (Phases I and II)
Treatment / Crossover Group	Mean monthly OA-related cost for help and/or transportation	One-Way Anova	Trend APS Superior	Phase II
Treatment / Crossover Group	Mean cost per 6 months for OA-related cost for medical and/or workplace-related cost	One-Way Anova	Trend APS Superior	Phase II
Treatment / Crossover Group	Improvement in Utility Score as measured by EQ-5D Δ Single Index Value	Repeated Measures ANOVA or Friedman Test	Trend APS Superior	Time of initial treatment until "failure" or end of study (Phases I and II)
Treatment/Crossover	Mean cost per 6 months for	One-Way Anova	Trend APS Superior	Phase I

Group	OA-related cost for restricted medications			

At each follow-up visit, the quality-adjusted life years (QALYs) will be calculated by multiplying the duration of time to failure (i.e. rescue medication or procedure) by a health-related quality of life (HRQoL) weight (i.e. utility score) associated with that health state. Therefore, the two key elements—HRQoL and survival—are incorporated. The EQ-5D general health state rating (scored on a scale of 0 (worst imaginable health state) to 100 (best imaginable health state)) will be used as the HRQoL weights.

In addition, the analyses will include economic cost-benefit/cost-effectiveness analyses to measure and calculate the direct and indirect cost comparison between nSTRIDE and alternative treatment options. Zimmer Biomet will rely on publicly-available, standardized datasets, such as those available from the U.S. Department of Health & Human Services, the National Center for Health Statistics and the Bureau of Labor Statistics, when conducting these analyses to help ensure outcome validity and reproducibility.

4.7 Subgroup Analysis

Clinically relevant differences in treatment effect are not anticipated across age, gender, race, or other subgroups, and the primary analysis will not be stratified by any subgroups. However, subgroup analyses of primary and secondary efficacy and safety endpoints will be performed in an exploratory fashion as specified below. These analyses will be performed for each of the following subgroups:

- Age (treated as continuous)
- Age (< Median vs. ≥ Median)
- Gender (Male/Female)
- Race (White (Hispanic), White (non-Hispanic), African-American, Native American, Asian or Pacific Islander, Other, or Not Specified).
- Baseline KL Grade
- Presence of Bone Marrow Lesions at baseline
- Site
- Baseline WOMAC Pain (< Median vs. ≥ Median)
- Usage of rescue or restricted medications prior to the 12M visit
- Presence of Contralateral Knee Pain prior to the 12M visit
- KL improvement status (improved, worsened, unchanged)

The statistical models in **Table *10*** will be performed for each subgroup. If the Treatment*Subgroup interaction is statistically significant ($p < 0.05$) and/or clinically meaningful, further analysis will be performed to determine the particular subgroup(s) in which the treatment effect differs, and the impact of this will be assessed and described. The least squared means of each treatment*subgroup combination will be output from the model.

Table 10. Subgroup Analysis

Dependent Variable(s)	Statistical Test
Mean WOMAC Pain Δ (12 Month minus Baseline)	Analysis of Covariance Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction
WOMAC MCID Function Responder rate	Logistic Regression Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction
Mean WOMAC Function Δ (12 Month minus Baseline)	Analysis of Covariance Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction
OMERACT-OARSI Responder / Non-responder (12 Month)	Logistic Regression Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction
Mean WOMAC Pain Δ (12 Month minus Baseline) in K-L II Subgroup	Analysis of Covariance Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction
Mean WOMAC Function Δ (12 Month minus Baseline) in K-L II Subgroup	Analysis of Covariance Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction
Mean NRS Pain Δ (12 Month minus Baseline)	Analysis of Covariance Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction

Dependent Variable(s)	Statistical Test
Occurrence of one or more SAE (Yes/No)	Logistic Regression Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction
Occurrence of one or more Device-Related AE (Yes/No)	Logistic Regression Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction
Rescue Medication Use for Index Knee OA Pain (Yes/No)	Logistic Regression Independent variables: Treatment, Subgroup, Treatment*Subgroup Interaction

5. Safety Endpoints

During Phase I, all Adverse Events will be collected regardless of relationship to the device. During Phase II, only AEs that are potentially related to the device/procedure will be collected.

Adverse events will be designated in MedDRA categories by a qualified reviewer that is blinded with respect to treatment assignment. Specific adverse events are expected to occur at low frequencies, and therefore no significant differences between groups are expected. Accordingly, the *a priori* tests planned will determine whether the overall rate of adverse events differs between treatment groups, and whether differences between groups exist with respect to device relatedness and severity. Results for each category will be tabulated by treatment group. In this tabulation the number of subjects with one or more adverse events fitting into an individual adverse event (AE) category will be displayed beside the count of subjects with no AEs fitting into that category. A Fisher's Exact test will be performed to determine whether treatment groups differ. In addition, the number of AEs and the corresponding percentage of total AEs within each category of severity and relatedness to the device will be summarized by treatment. A listing of all AEs will be provided, and a narrative of each serious AE will be generated.

In addition, a time-course distribution of adverse events will be created based on contiguous intervals between planned visits, as shown in ***Table 11***, **Table *12***, **Table *13***, and **Table *14***. In this table, one AE per subject, category, and time interval will be counted.

Table 11. Time Course Distribution of AEs - APS

	Injection	1 Month	3 Month	6 Month	12 Month
Interval	Injection to Day 15	Day 16 to Day 61	Day 62 to Day 137	Day 138 to Day 274	Day 275 to Day 547
AEs					
AE Type 1					
AE Type 2 …					
Serious AE					
AE Type 1					
AE Type 2 …					
Device Related AE					
AE Type 1					
AE Type 2 …					

Table 12 - Time Course Distribution of AEs - HA

	Injection	1 Month	3 Month	6 Month	12 Month
Interval	Injection to Day 15	Day 16 to Day 61	Day 62 to Day 137	Day 138 to Day 274	Day 275 to Day 547
AEs					
AE Type 1					
AE Type 2 …					
Serious AE					
AE Type 1					

AE Type 2 …					
Device Related AE					
AE Type 1					
AE Type 2 …					

Table 13 - Time Course Distribution of AEs – Post-Crossover APS to HA

	2nd Injection	1 Month	3 Month	6 Month	12 Month
Interval	Injection to Day 15	Day 16 to Day 61	Day 62 to Day 137	Day 138 to Day 274	Day 275 to Day 547
AEs					
AE Type 1					
AE Type 2 …					
Serious AE					
AE Type 1					
AE Type 2 …					
Device Related AE					
AE Type 1					
AE Type 2 …					

Table 14 - Time Course Distribution of AEs – Post-Crossover HA to APS

	2nd Injection	1 Month	3 Month	6 Month	12 Month

Interval	Injection to Day 15	Day 16 to Day 61	Day 62 to Day 137	Day 138 to Day 274	Day 275 to Day 547
AEs					
AE Type 1					
AE Type 2 …					
Serious AE					
AE Type 1					
AE Type 2 …					
Device Related AE					
AE Type 1					
AE Type 2 …					

A time-course tabulation showing rescue medication use (for index knee OA) by treatment group will also be generated. Adverse Event and Medication Use will be tabulated for all Phase I intervals.

6. Study Populations

A list of protocol violations will be evaluated to determine subjects who (1) violated one or more eligibility criteria, and (2) subjects with major protocol deviations that impact patient safety or the scientific validity of the study comparisons, or have the potential to do so. This will be determined by a blinded review of the protocol deviations.

Results will be tabulated. For the primary endpoint, two analyses will be performed: (1) intent-to-treat (ITT) and (2) per protocol (PP), as defined below. The ITT analysis will be considered the primary analysis of the study, and the PP analysis will be considered as sensitivity analysis.

Secondary endpoints will be analyzed using the per protocol group, with no imputation for missing data. Safety evaluations will be performed on the intent-to-treat group, with no imputation for missing data. Further detail regarding these analysis groups is given below.

Intent-to-Treat (ITT)

The ITT population includes all randomized subjects, regardless of whether they received an injection or not. This will be the population used for any data listings, for the primary study analysis, for safety analyses, and for demographic data summarized according to **Table *1***.

A table showing major and minor protocol deviations will be generated, comparing major and minor deviation frequency between the two treatment groups.

Per Protocol (PP)

The PP population includes all subjects from the mITT population who do not have major protocol deviations that impact patient safety or the scientific validity of the study outcomes, or the potential to do so. Analyses which use the PP population will use all available data on these subjects, with no imputation of missing data. This is the population that will be used in the analysis of secondary and exploratory efficacy endpoints, and as a sensitivity analysis of the primary endpoint.

6.1 Sample Size

Data used to estimate sample size are from Zimmer Biomet study APSS-33-00, a randomized study comparing APS to saline in human subjects. This calculation assumes that hyaluronic acid will perform similarly to saline. Estimates of expected difference and pooled standard deviation from this study are shown in **Table *15***.

Table 15- Results from APSS-33-00

Group	n	Mean Improvement in WOMAC Pain	St. Dev.	Mean Diff.	Pooled St. Dev.	Sample Size Each Group	Combined Plus 20% Attrition
APS	30	7.20	3.25				
Saline	15	4.93	4.73	2.27	4.06	93	246

The calculation results in 93 per group. Results are based on a two-tailed alpha of 0.049 and statistical power of 90%.

Planned enrollment was increased to 123 per group to protect against attrition of approximately 25% in either treatment group[1].

[1] Under the original sample size calculations, the sample size of 196 randomized (98 per group) was adjusted to accommodate potential dropout rate of approximately 20%, resulting in a total sample size of 246. This sample size

A maximum of 30 sites will participate in the study. Enrollment will be stopped at any site reaching a limit of 30% of planned study enrollment.

6.2 Interim Analysis Specification

An interim analysis is planned to formally assess the sample size calculation for this study. A designated unblinded statistician, independent from the study, will conduct the interim analysis when approximately 50% of the patients have reached the 12 month time point. The observed effect size at this interim time point will be estimated for analysis, spending 0.001 of the overall alpha-level of 0.05. This leaves 0.049 alpha for the final analysis. This is appropriate, as the study Sponsor is not expecting statistical significance, nor planning to stop the study, only to re-assess the sample size calculation. Furthermore, the effect of sample size adjustment through interim analysis has been shown to have negligible effect on Type I error rates [21].

The observed effect size at the interim time point will be used by the independent, unblinded statistician to re-calculate the sample size for the study. The effect size will not be shared with the Sponsor, only whether or not an increase in the sample size is warranted. The independent, unblinded statistician will only report to the Sponsor the sample size required for 80% and 90% power given the data to that point.

If the initial sample size is large enough to provide a minimum of 80% power (i.e. $n \geq n^*$), the trial will continue until all the planned number of subjects (n) are recruited. Otherwise, if $n < n^*$, we will increase the sample size and the trial will continue until enough patients (n^*) have been recruited that the desired power is achieved. The final new sample size for is

$$n_{new} = Max(n, n^*).$$

Therefore, if an increase in sample size is warranted by the re-calculation of the sample size requirements at the interim time point, it will only be reported how many more subjects are required in order to meet the study endpoint at

---

assumed a 2-sided alpha of 0.025 and power of 0.95. Note that, in accordance with the primary hypothesis, the sample size has been re-calculated under a 2-sided alpha=0.049 (instead of 2-sided alpha = 0.025) and a power of 0.90. This results in a sample size of 93 per group, or 117 when including 20% attrition and 124 when including 25% attrition. Thus the original sample size of 246 is sufficient for testing the primary endpoint with 90% power, and accommodate almost 25% attrition.

sufficient power. The sample size will not be reduced as a result of the sample size re-calculation. The sample size will only be increased, if the reassessment is so indicated. If the reassessment does not indicate an increase in sample size, then the study will continue as originally planned.

6.3 Missing Data and Sensitivity Analyses

6.3.1 Imputation of Missing Data in the Primary Analysis

The primary analysis will be performed on the ITT population. In the primary analysis, missing WOMAC Pain scores for subjects who are withdrawn due to reasons related to the index knee OA ("non-responders") will be considered Missing Not At Random (MNAR), and the subjects' baseline score will be carried forward in the analysis of WOMAC Pain. Scores missing for all other reasons will be considered Missing At Random and will be imputed using a Multiple Imputation method, which will replace each missing value with a set of plausible values that represent the variability around the choice of which value to impute. The monotone[2] regression method will be used [22], and will include variables from **Table *1*** (i.e. gender, age, BMI, race, and preop WOMAC Pain) as well as treatment group. As suggested in Bodner and White et al [23, 24], the number of imputations will reflect the percentage of incomplete cases. Approximately 15% attrition is estimated in the sample size calculation and thus approximately fifteen imputed datasets will be incorporated, and a seed of 20190516 will be used in generating the random numbers.

No imputation of missing values is planned for secondary or exploratory outcome variables.

6.3.2 Sensitivity Analyses

Scores for subjects who are withdrawn due to reasons related to the index knee OA are considered Missing Not At Random. In all of the ITT analyses described below, these subjects will be considered "non-responders", and the subjects' baseline score will be used in the analysis of WOMAC Pain.

The primary study analysis assumes that the all other missing data is Missing At Random (MAR).

In order to investigate the assumption that the missing data (excluding the non-responders who are withdrawn for reasons related to the index knee OA)

[2] If data are severely non-normal based on a graphical review, then the predictive mean matching method will be used instead of the monotone regression method.

is MAR, sensitivity analysis of the primary study endpoint will be performed under assumption of other missing data mechanisms, as follows:

1. A complete case analysis on the ITT population (Missing Completely at Random (MCAR))
2. A complete case analysis on the Per Protocol population (Missing Completely at Random (MCAR))
3. A tipping point analysis on the ITT population, performed under the assumption that the data is not MAR or MCAR by searching for a tipping point that reverses the study conclusion. This analysis will examine the possibility that the distribution of missing responses will have a different expected value than that of the corresponding distribution of the observed responses. Thus, the analysis will generate multiple imputed data sets with a specified sequence of shift parameters that adjust the imputed values for observations in the treatment group (Missing Not At Random (MNAR)).

In the event the conclusion from one or more of the sensitivity analyses disagrees with the conclusion from the primary analysis, the sources of differences between them will be investigated and subjected to explicit discussion and interpretation.

Prior results indicate that the efficacy of APS and the saline control differs over time, but this difference is small until some point after the 6 month post-injection time point. Therefore, although carrying data forward from an earlier time is commonly applied as a sensitivity analysis, in this case it is likely to produce misleading results if applied to primary endpoints and so is not planned.

7. Data Collection Time Points

Data will be collected at the screening visit, the injection visit, and at 1, 3, 6 and 12 months following the injection. After each subject completes all 12 month follow-up evaluations (Phase I), individual treatment allocation will be unblinded to the investigators. From this time-point on, only subjects will be blinded to study treatments resulting in a single-blind design during the long-term follow-up period (Phase II). During Phase II of the study, the subject may opt to get additional injections of their originally assigned treatment, or they may opt to crossover to the other treatment group and may receive multiple injections of the crossover treatment. Subjects may only crossover one time during the study. In Phase II, follow-up visits will occur at six month intervals until five years following the initial injection.

The follow-up visits will preferably be completed during an office visit to ensure subject compliance, but this is not mandatory and follow-up visits may consist of a structured telephone interview or electronic self-report with

exception of annual visits at which radiograph is to be taken. The investigative center will contact the subject to complete the Follow-up and AE forms. The subject will be asked to complete a Patient Questionnaire.

A table showing the data collected at each time interval is presented in the study protocol.

8. Summary of Data Analyses

Table 16 summarizes the analyses that are laid out in this report.

Table 16 – Summary of Data Analyses

Table Number	Table Title
Table 1	Randomization Verification
Table 2	Visit Windows
Table 3	Long-Term Visit Windows
Table 4	Planned Secondary Efficacy Test
Table 5	Planned Exploratory Efficacy Tests, Questionnaire Data
Table 6	Planned Statistical Tests for X-ray Results
Table 7	Planned Phase II Statistical Analysis-Originally Assigned Treatment
Table 8	Planned Phase II Statistical Analysis – Treatment/Crossover Group
Table 9	Planned Phase II Statistical Analysis – Health Economic Outcomes
Table 10	Subgroup Analysis
Table 11	Time course distribution of Adverse Events - APS
Table 12	Time course distribution of Adverse Events – HA
Table 13	Time course distribution of Adverse Events – Post Crossover APS to HA
Table 14	Time course distribution of Adverse Events – Post Crossover HA to APS
Table 15	Results from APSS-33-00

Statistical Analysis Plan Revision History

SAP Version	Description of Change	Sections
V 1.0	Original SAP (Dated 2019-08-05)	
V 2.0	Added an additional analysis of the change in K-L grade by baseline K-L grade	Section 4.4 Exploratory Outcomes (Imaging)
	Added additional analyses by Treatment / Crossover group and specified an ANCOVA model to analyze trends in Δ WOMAC Pain for treatment / Crossover groups.	Section 4.5 Exploratory Outcomes, Phase II
	Added additional subgroups for analysis; specified that LS Means would be examined for each subgroup.	Section 4.7 Subgroup Analysis

9. References

1. Brzusek, D. and D. Petron, *Treating knee osteoarthritis with intra-articular hyaluronans.* Curr Med Res Opin, 2008. **24**(12): p. 3307-3322.
2. Goldring, S.R. and M.B. Goldring, *The role of cytokines in cartilage matrix degeneration in osteoarthritis.* Clin Orthop Relat Res, 2004(427 Suppl): p. S27-S36.
3. Goldring, M.B., *Anticytokine therapy for osteoarthritis.* Expert Opin. Biol Ther, 2001. **1**(5): p. 817-829.
4. Goldring, M.B., et al. *Defining the roles of inflammatory and anabolic cytokines in cartilage metabolism*. in *Annals of the Rheumatic Diseases*. 2008.
5. Goldring, M.B., *Update on the biology of the chondrocyte and new approaches to treating cartilage diseases.* Best Pract. Res Clin Rheumatol, 2006. **20**(5): p. 1003-1025.
6. O'shaughnessey, K., et al., *Autologous protein solution prepared from the blood of osteoarthritic patients contains an enhanced profile of anti-inflammatory cytokines and anabolic growth factors.* Journal of Orthopaedic Research, 2014. **32**(10): p. 1349-1355.
7. Dinarello, C.A., *The role of the interleukin-1-receptor antagonist in blocking inflammation mediated by interleukin-1.* New England Journal of Medicine, 2000. **343**(10): p. 732-734.
8. Van Zee, K.J., et al., *Tumor necrosis factor soluble receptors circulate during experimental and clinical inflammation and can protect against excessive tumor necrosis factor alpha in vitro and in vivo.* Proc Natl. Acad Sci U. S. A, 1992. **89**(11): p. 4845-4849.
9. O'Shaughnessey, K.M., A. Panitch, and J.E. Woodell-May, *Blood-derived anti-inflammatory protein solution blocks the effect of IL-1b on human macrophages in vitro.* Inflammation Research, 2011. **60**(10): p. 929-936.
10. Matuska, A., et al., *Autologous solution protects bovine cartilage explants from IL-1α-and TNFα-induced cartilage degradation.* Journal of Orthopaedic Research, 2013. **31**(12): p. 1929-1935.
11. Dmitrienko, A., et al., *Analysis of clinical trials using SAS: A practical guide*. The Power To Know. 2005, Cary, NC 27513: SAS Institute.
12. Maurer, W., *Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypothesis.* Biomed. Chem.-Pharm. Ind., 1995. **6**: p. 3-18.
13. Bannuru, R.R., E.E. Vaysbrot, and L.F. McIntyre, *Did the American Academy of Orthopaedic Surgeons osteoarthritis guidelines miss the mark?* Arthroscopy: The Journal of Arthroscopic & Related Surgery, 2014. **30**(1): p. 86-89.
14. Administration, U.S.F.D. *Methods to Identify What is Important to Patients & Select, Develop or Modify Fit-for-Purpose Clinical Outcomes Assessments.* 2018; Available from: https://www.fda.gov/media/116276/download
15. Bellamy, N., et al., *Towards a definition of" difference" in osteoarthritis.* The Journal of rheumatology, 2001. **28**(2): p. 427-430.
16. Guyatt, G.H., et al. *Methods to explain the clinical significance of health status measures*. in *Mayo Clinic Proceedings*. 2002. Elsevier.
17. Jevsevar, D.S., *Treatment of osteoarthritis of the knee: evidence-based guideline.* JAAOS-Journal of the American Academy of Orthopaedic Surgeons, 2013. **21**(9): p. 571-576.

18. Pham, T., et al., *Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited.* Osteoarthritis Cartilage, 2004. **12**: p. 389-399.
19. Jang, S.-J., J.-D. Kim, and S.-S. Cha, *Platelet-rich plasma (PRP) injections as an effective treatment for early osteoarthritis.* European Journal of Orthopaedic Surgery & Traumatology, 2013. **23**(5): p. 573-580.
20. Cerza, F., et al., *Comparison between hyaluronic acid and platelet-rich plasma, intra-articular infiltration in the treatment of gonarthrosis.* The American journal of sports medicine, 2012. **40**(12): p. 2822-2827.
21. Herson, J. and J. Wittes, *The use of interim analysis for sample size adjustment.* Drug Information Journal, 1993. **27**(3): p. 753-760.
22. Rubin, D.B., *Multiple imputation for nonresponse in surveys*. Vol. 81. 2004: John Wiley & Sons.
23. Bodner, T.E., *What improves with increased missing data imputations?* Structural Equation Modeling: A Multidisciplinary Journal, 2008. **15**(4): p. 651-675.
24. White, I.R., P. Royston, and A.M. Wood, *Multiple imputation using chained equations: issues and guidance for practice.* Statistics in medicine, 2011. **30**(4): p. 377-399.

Signature Page

Statistical Analysis Plan
For APSS-66-00 Study (EU)

Protocol Title:

A Two-Phase, Multicenter, Randomized Study Comparing a Single Intra-Articular Injection of Autologous Protein Solution with a Single Injection Hyaluronic Acid in Patients with Knee Osteoarthritis

Signature Page

Cahit Akbas
Cahit Akbas (Oct 1, 2019)

Cahit Akbas
Scientific Affairs Principal
Zimmer Biomet, Inc.
Sponsor Representative

Date

Katie Miller
Katie Miller (Oct 1, 2019)

Katie Miller
M Squared Associates, Inc.
Study Statistician

Date

nSTRIDE PROGRESS V Statistical Analysis Plan V2-20190924 Clean

Final Audit Report 2019-10-01

Created:	2019-10-01
By:	Ann Blanton (ann.blanton@zimmerbiomet.com)
Status:	Signed
Transaction ID:	CBJCHBCAABAAIazRPLeP3NfUaIBi6Oc_SPqLWreiY1gr

"nSTRIDE PROGRESS V Statistical Analysis Plan V2-20190924 Clean" History

Document created by Ann Blanton (ann.blanton@zimmerbiomet.com)
2019-10-01 - 8:38:48 PM GMT- IP address: 71.222.230.137

Document emailed to Cahit Akbas (cahit.akbas@zimmerbiomet.com) for signature
2019-10-01 - 8:39:36 PM GMT

Email viewed by Cahit Akbas (cahit.akbas@zimmerbiomet.com)
2019-10-01 - 8:51:54 PM GMT- IP address: 84.80.19.13

Document e-signed by Cahit Akbas (cahit.akbas@zimmerbiomet.com)
Signature Date: 2019-10-01 - 8:52:14 PM GMT - Time Source: server- IP address: 84.80.19.13

Document emailed to Katie Miller (kmiller@msquaredassociates.com) for signature
2019-10-01 - 8:52:17 PM GMT

Email viewed by Katie Miller (kmiller@msquaredassociates.com)
2019-10-01 - 9:35:17 PM GMT- IP address: 174.221.7.181

Document e-signed by Katie Miller (kmiller@msquaredassociates.com)
Signature Date: 2019-10-01 - 9:35:46 PM GMT - Time Source: server- IP address: 174.221.7.181

Signed document emailed to Ann Blanton (ann.blanton@zimmerbiomet.com), Katie Miller (kmiller@msquaredassociates.com) and Cahit Akbas (cahit.akbas@zimmerbiomet.com)
2019-10-01 - 9:35:46 PM GMT