STATISTICAL ANALYSIS PLAN

A Prospective, Randomized, Controlled, Clinical Trial Evaluating the Non-Inferiority of Bridge-Enhanced ACL Repair (BEAR) to ACL Reconstruction with an Autologous Tendon Graft (ACLR)

Title: **B**ridge-**E**nhanced **A**CL **R**epair II Trial (BEAR II)

Protocol: P00012985

IDE Number: G150268

NCT02664545

Sponsor:
Martha M. Murray, MD
Professor, Department of Orthopaedic Surgery
Harvard Medical School
Boston Children's Hospital

Collaborator:
Miach Orthopaedics, Inc.
69 Milk St. Suite 100
Westborough MA, 01581

[redacted]

23May2019

Table of Contents

SIGNATURE APPROVAL PAGE

Sponsor Approval

Name: [redacted]

Approval: _____

Date: _____

Clinical Approva

Name: [redacted]

Approval: _____

Date: _____

Statistician Approval

Name: [redacted]

Approval: _____

Date: _____

1 ABBREVIATIONS

Abbreviation	Definition
ACL	Anterior Cruciate Ligament
ACL-RSI	ACL-Return to Sport after Injury Scale
BBA	Boston Biomedical Associates
BEAR	Bridge-Enhanced® ACL Repair Implant
BPTB	Bone Patellar Tendon Bone
CRF	Case Report Forms
CSR	Clinical Study Report
IKDC	International Knee Documentation Committee
KOOS	Knee Injury and Osteoarthritis Outcome Score
ITT	Intent-To-Treat Population
MITT	Modified Intent-To-Treat Population
MRI	Magnetic Resonance Imaging
PP	Per-Protocol Population
QOL	Quality of Life
SAP	Statistical Analysis Plan
SAE	Serious Adverse Event
VAS	Visual Analog Score for Pain

2 SUMMARY

TITLE	**B**ridge-**E**nhanced **A**CL **R**epair II Trial
PREFACE	This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for protocol P00012985 (A Prospective, Randomized, Controlled, Clinical Trial Evaluating the Non-Inferiority of Bridge-Enhanced ACL Repair (BEAR) to ACL Reconstruction with an Autologous Tendon Graft (ACLR)). This study is being completed to assess the safety and effectiveness of the BEAR® implant for the treatment of ACL repair in 14-35 year old individuals with a complete ACL tear. The following documents were reviewed in preparation of this SAP: • Clinical Research Protocol P00012985 version III issued 23JAN2017 • Case report forms (CRFs) issued 09FEB2018 for Protocol P00012985
PURPOSE	The purpose of this SAP is to outline the planned analyses in support of the Clinical Study Report (CSR) for protocol P00012985. Exploratory analyses not necessarily identified in this SAP may be performed to support the clinical development program. Any post-hoc, or unplanned, analyses not identified in this SAP will be clearly identified in the respective CSR.
STUDY OBJECTIVES	The overall objective is to determine the effectiveness of the BEAR™ implant and demonstrate its non-inferiority to the standard of ACL reconstruction in terms of knee stability and patient reported outcomes, as well as its superiority in terms of regaining muscle strength at 3 and 6 months after surgery. Additional objectives are to determine if safety outcomes including infection, graft rejection, and need for further surgical procedures are different between patients undergoing the BEAR procedure and those undergoing ACL reconstruction, as well as if markers of early osteoarthritis are clinically different in the two groups at the two year time point.
STUDY DESIGN	Single center, 2-arm, randomized, controlled clinical trial. 2:1 Randomization scheme.
ENDPOINTS	Primary: The study's **primary effectiveness endpoints** are International Knee Documentation Committee (IKDC) questionnaire score at 24 months post-surgery and AP laxity of the knee as measured by KT testing at 24 months post-surgery. For each score, a non-inferiority assessment of the mean for BEAR™ to the mean for the standard of ACL reconstruction will be carried out. The study will be considered a success if non-inferiority of BEAR™ to ACL reconstruction is met for both endpoints. Secondary Effectiveness – For Labeling Claims: • Hamstring strength, reported as percentage of the contralateral side, and as determined by hand-held dynamometer at 6 months post-surgery (superiority) • Hamstring strength, reported as percentage of the contralateral side, as determined by hand-held dynamometer at 12 months post-surgery (superiority) • Hamstring to quadricep ratio for the operated knee at 6 months post-surgery (superiority) • Hamstring to quadricep ratio for the operated knee at 12 months post-surgery (superiority) • ACL-RSI score at 6 months post-surgery (superiority) • Knee Injury and Osteoarthritis Outcome Score (KOOS) at 12 months post-

	surgery – Pain (non-inferiority) • KOOS at 12 months post-surgery – Symptoms (non-inferiority) • KOOS at 12 months post-surgery – Sports and Recreation (non-inferiority) • KOOS at 12 months post-surgery – Quality of Life (QOL) (non-inferiority) • KOOS at 12 months post-surgery – Activities of Daily Living (ADL) (non-inferiority) • KOOS at 12 months post-surgery – Pain (superiority) • KOOS at 12 months post-surgery – Symptoms (superiority) Secondary Effectiveness and Other – Not Intended for Labeling Claims: • IKDC at 6 and 12 months post-surgery • AP laxity of the knee at 6 and 12 months post-surgery • Hamstring strength, reported as percentage of the contralateral side, and as measured by hand-held dynamometer at 3 and 24 months post-surgery • Quadriceps and hip abductor strength for the operated knee at 3, 6, 12, and 24 months post-surgery • KOOS scores for Pain, Symptoms, Sports and Recreation, QOL, and ADL at 24 months post-surgery • SF-36 scores (all domains) at 12 and 24 months after surgery • X-ray imaging based on Kellgren and Lawrence grading system at 24 months post-surgery • Range of motion as measured via goniometer and standing active flexion angle at 3, 6, 12, and 24 months post-surgery • Time to return to full time work, school, and sports • ACL-RSI at 12 and 24 months post-surgery Secondary Safety: • Deep joint infection/incision and drainage of deep surgical site infection at any time point • Graft rejection/removal at any time point • Graft/repair failure • Adverse events of interest • Additional surgical procedures required • Presence of bovine type I gelatin IgE antibodies at 6 months post-surgery
INTERIM ANALYSES	Annual reporting per regulatory and IRB oversight was performed. No formal interim analyses were planned for this study.
FINAL ANALYSES	The data through 24 months will be locked and all final planned analyses of primary and secondary endpoints, identified in the protocol and this SAP, will be performed after the last patient completes the 24 month visit.

3 STUDY OBJECTIVES AND ENDPOINTS

3.1 STUDY OBJECTIVE

3.1.1 PRIMARY OBJECTIVE

The overall objective of this study is to determine the non-inferiority of the effectiveness of the BEAR Scaffold when compared with an ACL reconstruction with an autograft reconstruction (current gold standard). The outcomes for evaluating the primary objective will include a patient reported score on the International Knee Documentation Committee (IKDC) validated outcome measure at two years after surgery and a measure of AP knee laxity at two years after surgery.

3.1.2 SECONDARY OBJECTIVES

Secondary objectives are:

1) To determine if patients having ACL surgery with the BEAR scaffold recover their hamstring or quadriceps strength more quickly than patients undergoing ACL reconstruction.
2) To determine if markers of early osteoarthritic change are less prevalent in the BEAR patients than in patients undergoing ACL reconstruction at two years out from surgery.
3) To determine if the recorded patient reported outcomes (KOOS scores and SF-36 scores) are different in the two groups at any time point where they have been recorded.
4) To determine if there are any increased safety risks associated with use of the BEAR scaffold.
5) To determine if patients undergoing the BEAR procedure are able to get back to work and sports at the same rate as patients undergoing ACL reconstruction, and to determine if there is a difference in the rate of meeting the return-to-sport criteria for patients in the BEAR and ACL reconstruction groups.
6) To determine whether the patient reported outcomes or graft failure rates can be predicted by the surrogate prediction of maximum load and stiffness of the healing ACL using a novel MRI technique.
7) To determine whether the volume and orientation of the ACL after the BEAR procedure is different from the volume and orientation of the ACL graft after ACL reconstruction.

3.2 STUDY ENDPOINTS

3.2.1 PRIMARY ENDPOINTS

The study's **primary effectiveness endpoints** are IKDC score at 24 months post-surgery and AP laxity of the knee as measured by KT testing at 24 months post-surgery.

3.2.2 SECONDARY ENDPOINTS – INTENDED FOR LABELING CLAIMS

- Hamstring strength, reported as percentage of the contralateral side, and as determined by hand-held dynamometer at 6 months post-surgery (superiority)
- Hamstring strength, reported as percentage of the contralateral side, as determined by hand-held dynamometer at 12 months post-surgery (superiority)
- Hamstring to quadricep ratio for the operated knee at 6 months post-surgery (superiority)
- Hamstring to quadricep ratio for the operated knee at 12 months post-surgery (superiority)
- RSI score at 6 months post-surgery (superiority)
- Knee Injury and Osteoarthritis Outcome Score (KOOS) at 12 months post-surgery – Pain (non-inferiority)
- KOOS at 12 months post-surgery – Symptoms (non-inferiority)
- KOOS at 12 months post-surgery – Sports and Recreation (non-inferiority)
- KOOS at 12 months post-surgery – Quality of Life (QOL) (non-inferiority)
- KOOS at 12 months post-surgery – Activities of Daily Living (ADL) (non-inferiority)

- KOOS at 12 months post-surgery – Pain (superiority)
- KOOS at 12 months post-surgery – Symptoms (superiority)

3.2.3 Secondary Endpoints – Not Intended for Labeling Claims

- IKDC at 6 and 12 months post-surgery
- AP laxity of the knee at 6 and 12 months post-surgery
- Hamstring strength, reported as percentage of the contralateral side, and as determined by hand-held dynamometer at 3 and 24 months post-surgery
- Quadricep and hip abductor strength for the operated knee at 3, 6, 12, and 24 months post-surgery
- KOOS score for Pain, Symptoms, Sports and Recreation, QOL, and ADL at 24 months post-surgery
- ACL-RSI at 12 and 24 months post-surgery
- SF-36 scores at 12 and 24 months after surgery
- X-ray imaging based on Kellgren and Lawrence grading system at 24 months post-surgery
- Range of motion as measured via goniometer and standing active flexion angle at 3, 6, 12, and 24 months post-surgery
- Time to return to full time work, school, and sports

3.2.4 Safety Endpoints

- Deep joint infection/incision and drainage of deep surgical site infection at any time point per CDC/NHSN surveillance definitions for specific types of infections
- Graft rejection/removal at any time point
- Graft/repair failure as determined by positive pivot shift exam, Lachman exam with >6mm side to side difference, absence of tissue in expected ACL location on MRI, MR evidence of graft or repair loss of continuity or symptomatic instability requiring revision ACL surgery.
- Any additional adverse events including (but not limited to) deep venous thrombosis, loss of function, need for prolonged parenteral pain medication, development of neurologic symptoms or additional trauma will also be recorded for all patients in the study.
- Any additional surgical procedures that the patient requires on the operative knee, as well as any surgical procedures required on the contralateral knee, will also be recorded and reported as an additional secondary outcome measure. These include (but are not limited to) additional surgery to address meniscal or cartilage pathology, scar tissue, arthrofibrosis, removal of symptomatic hardware or graft removal for any reason.
- Presence of bovine type I gelatin IgE antibodies at 6 months post-surgery

4 SAMPLE SIZE

Non-inferiority will be tested by constructing a 95% confidence interval for the difference between the intervention group means and observing whether or not the confidence interval overlaps a pre-specified non-inferiority margin for that endpoint.

Of the two primary outcomes, the knee laxity endpoint requires the larger sample size. For AP laxity, we assume a standard deviation (SD) of 2.7mm based on Fleming et al. [1] pooled standard deviations of the within-patient side-to-side difference scores for both the low tension and high tension groups (personal communication). Based on the Arneja and Leith [2] recommendation of a 2 to 3 mm threshold as the basis for a diagnostic test, we selected a clinically important difference to be 2.0 mm and hence are using 2.0 mm as the non-inferiority margin. I.e., if μ_1 is the mean side-to-side difference of the BEAR Scaffold and μ_2 is the mean side-to-side difference of the standard therapy (Control), then the null and alternative hypotheses are:

$$H_0: \mu_1 - \mu_2 \geq 2.0\text{mm}$$
$$H_1: \mu_1 - \mu_2 < 2.0\text{mm}$$

With a 2:1 randomization, a sample size of N=69 (46 BEAR, 23 Control) will provide 80% power at a one-sided 0.025 level of significance to reject the above H_0 in favor of H_1 (i.e., it will provide 80% power to detect non-inferiority of BEAR to Control as described above). Inflating the sample size to account for an anticipated 20% attrition brings the sample size goal to N=87.

Because of our special interest in the hamstring graft subgroup (but not as a primary or secondary analysis), we will recruit until we have met the target N=87 patients in the hamstring-preference stratum. Based on historical data at our institution, we expect ~85% of all randomized patients to be in the hamstring-preference subgroup so we project a total sample size, including patients in the bone patellar tendon bone (BPTB)-preference stratum, to be about N=100. The remainder of the power calculations will be based on the smaller number of N=87 and therefore power is slightly underestimated.

Our second primary endpoint is the IKDC score at 24 months post surgery. Irrgang [3] concluded that a change in the IKDC of 11.5 points was an optimal threshold with high sensitivity for distinguishing those who were or were not improved. Therefore, we used 11.5 as the non-inferiority margin. In a cohort of 69 patients reported by Reinke [4], the SD of IKDC scores at two years was 10.5. In a systematic literature review of reconstruction and non-surgical cohorts with mean 13.9 years follow-up, Chalmers [5] reported a mean difference in IKDC scores of 5.8 and an effect size of 0.73, which implies a SD of 7.95. We used the most conservative of these SD estimates, 10.5. With that SD and the clinically important difference of 11.5 as recommended by Irrgang [3], our sample size of 87 patients (minus 20% projected attrition) gives us 99% power to detect non-inferiority between the groups for this primary outcome measure. I.e. there is 99% power to reject the following null hypothesis in favor of its alternative at a one-sided 0.025 level of significance, where μ_1 is the mean IKDC score of the BEAR Scaffold and μ_2 is the mean IKDC score of the standard therapy:

$$H_0: \mu_1 - \mu_2 \leq -11.5$$
$$H_1: \mu_1 - \mu_2 > -11.5$$

To project statistical power for some additional secondary endpoints, we considered the KOOS score as a second patient-reported outcome measure. Roos and Lohmander [6] reported SDs for the Sports and Knee Related Quality of Life (krQOL) subscales at 6 months post ACL reconstruction of 15.8 and 10.1, respectively, and Roos and Lohmander [7] concluded that 10 points represented a clinically significant effect for the KOOS. Under these assumptions and using a non-inferiority margin of 10, our sample size of 87 patients will provide 69% and 97% power to detect non-inferiority for the Sports and krQOL subscales of the KOOS, assuming 20% dropout at the two year time point and using a one-sided 0.025 level of significance for the testing of each endpoint.

For hamstring strength deficits after ACL surgery, Landes et al. [8] reported a mean hamstring strength side-to-side difference of 17 Nm (a 23% deficit relative to the contralateral knee), with a SD of 14 Nm, 2 years after ACL reconstruction with hamstring tendon. Data obtained at Boston Children's Hospital for 235 patients undergoing autologous hamstring grafts showed a mean 29% deficit (SD = ±23%) of hamstring strength at six months post-operatively. From the initial BEAR study, the patients undergoing ACL reconstruction had an average 33% deficit of hamstring strength at the six month time point, while the BEAR patients had only a 13% deficit, an estimated 20-point difference. At 3 months there was a larger treatment effect observed, 25 points. Assuming a SD of ±23%, our sample size of 87 patients (minus 20% projected attrition) will have 91.9% power to detect an absolute 20% difference in percent hamstring strength deficit at 3 or 6 months at a two-sided 0.05 level of significance.

For imaging outcomes, we don't anticipate any significant changes in joint space on plain radiography at two years, so while we will collect these images at time zero and year two, this will be simply for baseline recording in the event the results of this proposed study warrant longer term study of these patients at time points when we might anticipate seeing joint space narrowing, such as at 6 or 10 years after injury as has been done for prior ACL reconstruction cohort studies.

Finally, with the 2:1 randomization, our sample size of N=87 will yield 58 BEAR patients, 46 after 20% attrition, for observing rare safety events such as graft removal due to an immune response. Forty-six evaluable patients will provide a 90% chance of observing at least one event if the true event rate is 0.049. However, such events are likely to occur early in follow up, before potential dropout, so the number of evaluable patients could be as high as 58. In addition, if we consider the 10 BEAR patients in the pilot study, we could have a maximum potential of 68 evaluable patients in which to observe rare safety events. Therefore, we will plan to consider the safety outcomes for the patients in the initial BEAR I trial (IDE G140151) in addition to those in this current IDE study to improve our power to detect any differences between the ACL reconstruction and BEAR groups for these relatively rare safety events. With 58 or 68 evaluable patients, we will have a 90% chance of observing at least one event even for events that are more rare (event rates 0.039 or 0.033, respectively).

5 Sequence of Planned Analyses

5.1 Interim Analyses

Annual reporting per regulatory and IRB oversight was completed. No formal interim analyses were planned for this study.

5.2 Final Analyses and Reporting

All final, planned, analyses identified in the protocol and in this SAP will be performed only after the last patient has completed the final follow-up visit (the month 24 visit). Key statistics and study results will be made available to the trial sponsor following database lock. Any post-hoc, exploratory analyses completed to support planned study analyses, which were not identified in this SAP, will be documented and reported as necessary. Any results from these unplanned analyses will also be clearly identified as post-hoc analyses.

6 Analysis Populations

6.1 Intent to Treat Population (ITT)

The intent-to-treat (ITT) population for this study includes all enrolled and randomized patients. Patients are considered enrolled in the trial after they have signed the informed consent form and are randomized for treatment assignment. Patients are analyzed under the treatment to which they were randomized.

6.2 As-Treated (AT)

The as-treated population (AT) includes all patients who have the procedure attempted and are analyzed under the treatment to which they received. The AT population is the primary analysis population for safety.

6.3 Modified Intent to Treat Population (mITT)

The modified intent-to-treat population (mITT) includes all ITT patients who have the procedure attempted. Patients in the mITT population will be followed a minimum of 24 months. The mITT population is the primary analysis population for effectiveness. Patients are analyzed under the treatment to which they were randomized.

6.4 Per-Protocol Population (PP)

The per-protocol population (PP) will include all ITT patients who additionally meet all study eligibility criteria, have available study data for the study endpoint and do not have a major protocol violation that affects primary effectiveness. Patients in the PP population will be evaluated based upon the type of surgery that they receive. Major protocol violations are described in Section 7.4.

7 General Issues for Statistical Analysis

Descriptive statistics (mean, standard deviation, frequencies, etc.) for baseline patient characteristics, patient disposition and other relevant study parameters will be reported. Results will be generated by treatment group.

7.1 Analysis Software

Analysis data sets, statistical analyses and associated output generated by BBA will be generated using SAS® Software version 9.4.

7.2 Disposition of Patients and Withdrawals

The number and percent of patients in each analysis population will be presented, with percentages based on the ITT population.

All patients who provide written informed consent will be accounted for. The frequency of patients who completed each scheduled assessment will be presented in a flow chart. The number and percentage of ITT patients prematurely withdrawing will be presented overall and by reason of discontinuation.

7.3 Methods for Withdrawals and Missing Data

All practical monitoring and follow-up steps will be taken to ensure complete and accurate data collection. Since the primary effectiveness endpoints are assessed 24 months following randomization, it is anticipated that there will be some missing data. In the case of missing data for a scheduled visit, if there is an unscheduled visit within the respective scheduled visit range, the data from the unscheduled visit will be used for analysis. Note: Data from an unscheduled visit will not replace data that is provided for the scheduled visit.

The primary effectiveness endpoint analysis will be carried out on the mITT population with a supporting analysis on the PP population. In the mITT population, patients may have missing information on the primary endpoints of IKDC and AP laxity of the knee, primarily due to premature withdrawal from the study. Additionally, patients may have subsequent interventions that could impact their IKDC and AP laxity of the knee at 24 months. Therefore, several analyses are planned to assess the impact of missing data and subsequent interventions on the primary effectiveness endpoint:

1. Multiple imputation for missing data where post-intervention observations are not set to missing (Primary). Patients who are missing IKDC at 24 months will be considered as "missing data patients" for this endpoint. Missing IKDC at 24 months will be imputed using a monotone linear regression multiple imputation approach for continuous outcome data. However, assuming the missing data pattern will not be completely monotone at first, then a Markov Chain Monte Carlo (MCMC) imputation will first be carried out using IKDC scores at all time points where it is collected to make a monotone missing data pattern. There will be 50 datasets generated in this manner to create 50 datasets with a monotone missing data pattern. For each of these 50 data sets, missing IKDC at 24 months will then be imputed once from a monotone multiple imputation linear regression model with independent variables of age, gender, and other covariates including: randomized treatment group, baseline body mass index (BMI), baseline MARX score, hamstring : quadricep ratio for the injured knee, knee flexion angle for the injured knee, tibial slope, and results from each of the five components of the baseline KOOS questionnaire. Also included will be baseline and post-baseline non-missing IKDC scores. The one-sided, two-sample t-test for non-inferiority will then be carried out on each of the resulting 50 complete datasets, with the t-test results being combined across the 50 complete datasets using standard multiple imputation theory to obtain one overall p-value comparing the two treatments on IKDC after accounting for missing data.

   Similarly, patients who are missing AP laxity of the knee at 24 months will be considered as "missing data patients" for this endpoint. Missing AP laxity of the knee at 24 months will be imputed using the same approach as described above for missing IKDC, with independent variables of age, gender, and other covariates including: randomized treatment group, baseline BMI, baseline MARX score, hamstring : quadricep ratio for the injured knee, IKDC, tibial slope, and results from each of the five components of the KOOS questionnaire. Also included will be baseline and post-baseline non-missing AP laxity of the knee. The one-sided, two-sample t-test for non-inferiority will then be carried out on each of the resulting 50 complete datasets, with the t-test results being combined across the 50 complete datasets using standard multiple imputation theory to obtain one overall p-value comparing the two treatments on AP laxity of the knee after accounting for missing data.

2. Multiple imputation for missing data where post-intervention observations are first set to missing (Sensitivity). The above multiple imputation approach for missing data will be repeated, but where post-intervention observations are first set to missing.

3. Available data, removing patients who are missing primary endpoint data and ignoring subsequent intervention. (Sensitivity). For this sensitivity analysis, the analysis will be run on available data, as described below in Section 9.1. Any patient with a subsequent intervention will have their post-intervention data used, as available, without censoring.

4. Available data, removing patients who are missing primary endpoint data and censoring data for patients who have a subsequent intervention (Sensitivity). For this sensitivity analysis, any patient who has a subsequent intervention prior to the 24 month endpoint will have their IKDC and AP laxity of the knee data set to missing after the intervention. The analysis will then be run on available data, as described below in Section 9.1.

5. Tipping point analysis. (Sensitivity)
   This sensitivity analysis will be done via a 'tipping point' strategy for each endpoint, in which missing data are imputed under a range of assumed biases for the BEAR group until the p-value changes from significant to non-significant or vice versa. When analyzing a continuous endpoint, the tipping point analysis involves worsening the imputed values for the BEAR patients with missing data, and then re-running the analyses. The patients with subsequent interventions will have their data censored prior to the multiple imputation. For the comparison between groups for IKDC and AP knee laxity, the tipping point analysis will be conducted in SAS PROC MI, altering the shift parameter in the MNAR statement for the BEAR patients in order to yield worsening endpoint values. Specifically, the shift parameter adds a constant to the imputed values of the endpoint in the BEAR arm.

   For IKDC, the shift parameter will be negative (imputing lower IKDC scores for patients with missing data in the BEAR arm). For AP knee laxity, the shift parameter will be positive (imputing higher AP knee laxity scores for patients with missing data in the BEAR arm). Separate shift parameters will be applied to the patients with true missing data and patients with subsequent interventions, so that the shift parameter for patients with subsequent interventions will be twice as large as the shift parameter for patients with true missing data. The values imputed for the control arm under the MAR assumption will remain as is. The primary analysis for each endpoint will then be re-run to determine at what shift parameter the results become insignificant. This approach to the tipping point analysis involves seeing how extreme the imputed values for the BEAR patients with missing data would have to be in order to change the results of the study.

7.4 Protocol Violations

Protocol deviations will be listed and summarized in the CSR. The potential impact of protocol deviations on the study outcomes will be described. If necessary, a per protocol analysis will be conducted which excludes patients in whom a protocol deviation which has the potential to confound assessment of the study endpoints has occurred.

7.5 Multiple Comparisons and Multiplicity

The 12 secondary endpoints intended for labeling (Section 3.2.2) will be tested one at a time in a fixed hierarchical method in the order specified below to control the Type I error rate. The first endpoint (hamstring strength at 6 months post-surgery) will be compared between treatments in a superiority manner at a two-sided 0.05 level of significance. If the null hypothesis is rejected in favor of BEAR, then the BEAR scaffold will be considered statistically superior to the control on hamstring strength at 6 months post-surgery and treatment comparison analysis will proceed to the 2nd secondary hypothesis. This process will continue to the 12th secondary hypothesis as long as the prior null hypothesis is rejected at a two-sided 0.05 level of significance in favor of BEAR. These secondary endpoints/hypotheses will only be evaluated if the non-inferiority null hypothesis for both of the primary endpoints are rejected.

- Hamstring strength, reported as percentage of the contralateral side, and as determined by hand-held dynamometer at 6 months post-surgery (superiority)
- Hamstring strength, reported as percentage of the contralateral side, as determined by hand-held dynamometer at 12 months post-surgery (superiority)
- Hamstring to quadricep ratio for the operated knee at 6 months post-surgery (superiority)
- Hamstring to quadricep ratio for the operated knee at 12 months post-surgery (superiority)
- RSI score at 6 months post-surgery (superiority)
- KOOS at 12 months post-surgery – Pain (non-inferiority)
- KOOS at 12 months post-surgery – Symptoms (non-inferiority)
- KOOS at 12 months post-surgery – Sports and Recreation (non-inferiority)
- KOOS at 12 months post-surgery – QOL (non-inferiority)
- KOOS at 12 months post-surgery – ADL (non-inferiority)

- KOOS at 12 months post-surgery – Pain (superiority)
- KOOS at 12 months post-surgery – Symptoms (superiority)

7.6 Assessment of Homogeneity

As this study is being performed at a single investigative site, there will not be an analysis performed to assess homogeneity of treatment difference on the primary endpoint across investigative sites.

A poolability analysis will be performed to assess homogeneity of treatment difference on the primary and secondary endpoints intended for labeling across surgeons. Surgeons with fewer than 8 patients will be combined into a pooled surgeon for the purpose of this analysis. Pooling will be determined before the treatment blind is broken and prior to inspecting the outcome data.

The following will be carried out on the mITT (available data): Assessment of homogeneity of treatment difference on the primary and secondary endpoints intended for labeling will be carried out using an analysis of variance (ANOVA) model with treatment, surgeon, and treatment-by- surgeon interaction as the independent variables. Of interest is the significance of the treatment-by- surgeon interaction.
If the treatment-by- surgeon interaction effect is significant at a 0.10 level of significance, then analyses within surgeon will be further inspected. If the interaction is not significant or if it is significant but the direction of the effect is the same in all surgeons, then surgeons will be pooled for the final analysis. Otherwise, demographics and procedure characteristics within surgeon will be inspected to assess if differences in demographics or surgical procedure may be causing the interaction.

There will be no imputation of missing primary or secondary endpoint data for the poolability analysis.

8 Demographics and Other Baseline Characteristics

8.1 Demographics and Baseline Characteristics

Patient demographics for all analysis populations will be summarized in a table. Gender, ethnicity, and race will be summarized with frequency and percent. Age, height, weight, BMI, Marx Activity total score, and time from injury to surgery will be summarized with N, mean, standard deviation, median and minimum and maximum.

8.2 Baseline Physical Exam and Surgical Procedure

Patient baseline characteristics for the mITT population will be summarized in a table. Data related to the patient knee injury, surgery, intraoperative data, medial meniscal tear, and lateral meniscal tear will be summarized with N, mean, standard deviation, median and minimum and maximum for continuous variables and frequency and percentage of patients for categorical variables.

9 Effectiveness Analyses

The primary analysis population is the mITT, with multiple imputation for patients with missing primary endpoint data. Sensitivity analyses will be performed to assess the impact of missing data. The multiple imputation strategy and sensitivity analyses are described above in Section 7.3. Supporting analyses will be completed in the PP population.

9.1 Primary Effectiveness Variables

The primary effectiveness endpoints are IKDC score and AP laxity of the knee as measured by KT testing at the 24-month follow-up to determine if the BEAR Scaffold is noninferior to standard surgical reconstruction.

The difference in means between groups, (BEAR minus control) will be calculated for continuous effectiveness endpoints as a measure of relative treatment effectiveness. Assuming higher scores are better, a positive difference in means will indicate that BEAR is better than control and a negative difference that BEAR is worse. (If lower scores are better, this interpretation is reversed.) Two-sided 95% confidence intervals for the difference in means will be used to show a plausible range of relative treatment effectiveness. Analyses of anteroposterior knee laxity, range of motion and muscle strength measures that are performed on both knees will be based on within-patient side-to-side differences (involved minus contralateral knee) or percent deficit (side-to-side difference as a percent of the contralateral knee). The intervention groups will then be compared with respect to these side-to-side differences.

For the patient-reported functional primary endpoint, IKDC score at 24 months, non-inferiority for BEAR will be demonstrated if the lower bound of the two-sided 95% CI of the treatment group difference in means lies above the non-inferiority margin of -11.5. For the biomechanical primary endpoint, knee laxity at 24 months, a lower side-to-side difference (involved minus contralateral knee) is desirable so non-inferiority will be demonstrated if the upper limit of the two-sided 95% CI for the treatment group difference in means is below the non-inferiority margin of +2.0 mm.

If μ_1 is the mean IKDC score of the BEAR Scaffold and μ_2 is the mean IKDC score of the standard therapy, then the primary effectiveness hypothesis is:

$$H_0: \mu_1 - \mu_2 \leq -11.5$$
$$H_1: \mu_1 - \mu_2 > -11.5$$

If μ_1 is the mean side-to-side difference of the BEAR Scaffold and μ_2 is the mean side-to-side difference of the standard therapy, then the primary effectiveness hypothesis is:

$$H_0: \mu_1 - \mu_2 \geq 2.0\text{mm}$$
$$H_1: \mu_1 - \mu_2 < 2.0\text{mm}$$

Overall, demonstration of non-inferiority requires that both primary endpoints meet the CI criteria described above based on a two-sample t-test. Although the 95% CIs are two-sided intervals, only one side is relevant for demonstrating non-inferiority so the Type I error rate for each endpoint is α =0.025 and the combined statistical procedure is protected at an overall Type I error rate of $\alpha \leq 0.05$.

The IKDC scores and AP laxity of the knee will be summarized by treatment group and time point.

9.2 Secondary Effectiveness Variables – Intended for Labeling Claims

Secondary endpoints will be tested in the order specified below to control the Type I error rate and adjust for multiple testing. The primary analysis of secondary endpoints intended for labeling will be performed on the mITT population with multiple imputation for patients with missing secondary endpoint data. The multiple imputation approach for secondary endpoints will be carried out in the same manner as described for the primary endpoint in Section 7.3. A sensitivity analysis for the secondary endpoints intended for labeling will be performed on the available data without censoring for subsequent interventions.

9.2.1 Prone Hamstring Strength at 6 Months Post-Surgery

The strength of the hamstring as measured by hand-held dynamometer will be determined at 6 months post-surgery. The average strength percentage [100*(Injured knee/Non-injured knee)] between the two treatment groups will be compared. The null hypothesis is that the true means are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test with a two-sided significance level of 0.05.

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

A significant p-value combined with a higher mean in the BEAR group would indicate superiority of the BEAR group over control group.

9.2.2 Prone Hamstring Strength at 12 Months Post-Surgery

The strength of the hamstring as measured by hand-held dynamometer will be determined at 12 months post-surgery. The average strength percentage [100*(Injured knee/Non-injured knee)] between the two treatment groups will be compared. The null hypothesis is that the true means are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test with a two-sided significance level of 0.05.

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

A significant p-value combined with a higher mean in the BEAR group would indicate superiority of the BEAR group over control group.

9.2.3 Hamstring to Quadricep Ratio at 6 Months Post-Surgery

The ratio of the hamstring to quadricep strength on the operated knee will be determined at 6 months post-surgery. The average ratio between the two treatment groups will be compared. The null hypothesis is that the true means are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test with a two-sided significance level of 0.05.

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

A significant p-value combined with a higher mean in the BEAR group would indicate superiority of the BEAR group over control group.

9.2.4 Hamstring to Quadricep Ratio at 12 Months Post-Surgery

The ratio of the hamstring to quadricep strength on the operated knee will be determined at 12 months post-surgery. The average ratio between the two treatment groups will be compared. The null hypothesis is that the true means are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test with a two-sided significance level of 0.05.

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

A significant p-value combined with a higher mean in the BEAR group would indicate superiority of the BEAR group over control group.

9.2.5 ACL Return to Sports After Injury at 6 Months Post-Surgery

The ACL-RSI score will be determined at 6 months post-surgery. The average score between the two treatment groups will be compared. The null hypothesis is that the true means are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test with a two-sided significance level of 0.05.

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

A significant p-value combined with a higher mean in the BEAR group would indicate superiority of the BEAR group over control group.

9.2.6 Pain Score (non-inferiority)

The pain domain from the KOOS score will be collected at 12 months post-surgery and compared between groups. A 10 point difference between groups will be considered statistically significant when a two-sided, two-sample t-test is used based on prior studies validating the KOOS score [6] and is declared significant at a two-sided 0.05 level of significance.

$$H_0: \mu_1 - \mu_2 \geq 10$$
$$H_1: \mu_1 - \mu_2 < 10$$

Although the 95% CIs are two-sided intervals, only one side is relevant for demonstrating non-inferiority so the Type I error rate for each endpoint is α =0.025.

9.2.7 Symptoms Score (non-inferiority)

The symptoms domain from the KOOS score will be collected at 12 months post-surgery and compared between groups. A 10 point difference between groups will be considered statistically significant when a two-sided, two-sample t-test is used based on prior studies validating the KOOS score [6] and is declared significant at a two-sided 0.05 level of significance.

$$H_0: \mu_1 - \mu_2 \geq 10$$
$$H_1: \mu_1 - \mu_2 < 10$$

Although the 95% CIs are two-sided intervals, only one side is relevant for demonstrating non-inferiority so the Type I error rate for each endpoint is α =0.025.

9.2.8 Sports and Recreation Score (non-inferiority)

The sports and recreation domain from the KOOS score will be collected at 12 months post-surgery and compared between groups. A 10 point difference between groups will be considered statistically significant when a two-sided, two-sample t-test is used based on prior studies validating the KOOS score [6] and is declared significant at a two-sided 0.05 level of significance.

$$H_0: \mu_1 - \mu_2 \geq 10$$
$$H_1: \mu_1 - \mu_2 < 10$$

Although the 95% CIs are two-sided intervals, only one side is relevant for demonstrating non-inferiority so the Type I error rate for each endpoint is α =0.025.

9.2.9 Knee Related Quality of Life Score (non-inferiority)

The knee related QOL domain from the KOOS score will be collected at 12 months post-surgery and compared between groups. A 10 point difference between groups will be considered statistically significant when a two-sided, two-sample t-test is used based on prior studies validating the KOOS score [6] and is declared significant at a two-sided 0.05 level of significance.

$$H_0: \mu_1 - \mu_2 \geq 10$$
$$H_1: \mu_1 - \mu_2 < 10$$

Although the 95% CIs are two-sided intervals, only one side is relevant for demonstrating non-inferiority so the Type I error rate for each endpoint is α =0.025.

9.2.10 Activities of Daily Living Score (non-inferiority)

The ADL domain from the KOOS score will be collected at 12 months post-surgery and compared between groups. A 10 point difference between groups will be considered statistically significant when a two-sided, two-sample t-test is used based on prior studies validating the KOOS score [6] and is declared significant at a two-sided 0.05 level of significance.

$$H_0: \mu_1 - \mu_2 \geq 10$$
$$H_1: \mu_1 - \mu_2 < 10$$

Although the 95% CIs are two-sided intervals, only one side is relevant for demonstrating non-inferiority so the Type I error rate for each endpoint is α =0.025.

9.2.11 Pain Score (superiority)

The pain domain from the KOOS score will be collected at 12 months post-surgery and compared between groups. In addition to showing non-inferiority in Section 9.2.6 above, we hypothesize superiority of BEAR over standard therapy for the KOOS score in the pain domain. BEAR will be considered superior when a two-sided, two-sample t-test is used and is declared significant at a two-sided 0.05 level of significance combined with a higher mean score for the BEAR group than the standard therapy group.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

A significant p-value combined with a higher mean in the BEAR group would indicate superiority of the BEAR group over control group.

9.2.12 Symptoms Score (superiority)

The symptoms domain from the KOOS score will be collected at 12 months post-surgery and compared between groups. In addition to showing non-inferiority in Section 9.2.7 above, we hypothesize superiority of BEAR over standard therapy for the KOOS score in the symptoms domain. BEAR will be considered superior when a two-sided, two-sample t-test is used and is declared significant at a two-sided 0.05 level of significance combined with a higher mean score for the BEAR group than the standard therapy group.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

A significant p-value combined with a higher mean in the BEAR group would indicate superiority of the BEAR group over control group.

9.3 Secondary Effectiveness and Other Variables – Not Intended for Labeling Claims

The summary and analysis of the secondary endpoints not intended for labeling will be performed on available data without censoring for subsequent intervention. No multiple imputation for missing data is planned for these endpoints.

9.3.1 IKDC

For the patient-reported functional primary endpoint, IKDC score at 3, 6, and 12 months post-surgery, non-inferiority for BEAR will be demonstrated if the lower bound of the two-sided 95% CI of the treatment group difference in means lies above the non-inferiority margin of -11.5.

If μ_1 is the mean IKDC score of the BEAR Scaffold and μ_2 is the mean IKDC score of the standard therapy, then the hypothesis is:

$$H_0: \mu_1 - \mu_2 \leq -11.5$$
$$H_1: \mu_1 - \mu_2 > -11.5$$

9.3.2 AP Laxity of the Knee

For the biomechanical primary endpoint, knee laxity at 6 and 12 months post-surgery, a lower side-to-side difference (involved minus contralateral knee) is desirable so non-inferiority will be demonstrated if the upper limit of the two-sided 95% CI for the treatment group difference in means is below the non-inferiority margin of +2.0 mm.

If μ_1 is the mean side-to-side difference of the BEAR Scaffold and μ_2 is the mean side-to-side difference of the standard therapy, then the hypothesis is:

$$H_0: \mu_1 - \mu_2 \geq 2.0\text{mm}$$
$$H_1: \mu_1 - \mu_2 < 2.0\text{mm}$$

9.3.3 Prone Hamstring Strength

The strength of the hamstring as measured by hand-held dynamometer will be determined at 3 and 24 months post-surgery. The average strength percentage [100*(Injured knee/Non-injured knee)] between the two treatment groups will be compared. The tested null hypothesis is that the true means are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

9.3.4 Prone Hip/Seated Quadriceps Strength

The strength and torque of the quadriceps and hip abductor musculature will be determined at 3, 6, 12 and 24 months after surgery. The average strength percentage [100*(Injured knee/Non-injured knee)] between the two treatment groups will be compared. The tested null hypothesis is that the true means are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

9.3.5 KOOS

The KOOS score is another validated patient outcome measure with five domains: Pain, Symptoms, Sports and Recreation, Knee Related QOL and ADL. We will collect data for all five domains at 24 months after surgery. A 10 point difference between groups will be considered statistically significant when a two-sided, two-sample t-test is used based on prior studies validating the KOOS score [6] for each domain.

$$H_0: \mu_1 - \mu_2 \geq 10$$
$$H_1: \mu_1 - \mu_2 < 10$$

9.3.6 ACL Return to Sports After Injury Scale

The ACL-RSI score will be collected at 12 and 24 months post-surgery. The average score between the two treatment groups will be compared at each time point. The null hypothesis is that the true means are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

9.3.7 SF-36

The SF-36 instrument is another validated patient outcome measure with eight domains and two additional scores. All domains and scores at 12 and 24 months after surgery will be compared between the BEAR and ACLR groups. Results will be summarized descriptively by time point. The null hypothesis is that the true means are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

9.3.8 X-Ray Imaging Based on Kellgren and Lawrence Grading System

X-ray imaging will be completed at 24 months post-surgery. The Kellgren and Lawrence system will be used to classify the severity of knee osteoarthritis using five grades:

- **Grade 0**: no radiographic features of OA are present
- **Grade 1**: doubtful joint space narrowing (JSN) and possible osteophytic lipping
- **Grade 2**: definite osteophytes and possible JSN on anteroposterior weight-bearing radiograph
- **Grade 3**: multiple osteophytes, definite JSN, sclerosis, possible bony deformity
- **Grade 4**: large osteophytes, marked JSN, severe sclerosis and definite bony deformity

The frequency and percent of patients in each grade will be summarized by treatment group. The grade will also be summarized numerically within each treatment group. The tested null hypothesis is that the true mean differences between groups are equal compared to the alternative hypothesis that they are not based on a two-sided Wilcoxon Rank Sum Test.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

9.3.9 ROM Active Flexion/Extension

Active and passive range of motion will be measured using a goniometer pre-operatively, intra-operatively and at 2 weeks, 6 weeks, 3 months, 6 months and 1 and 2 years after surgery. Thigh circumference will additionally be measured at these time points. The tested null hypothesis is that the true mean differences between injured and uninjured knees between groups are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

9.3.10 Lachman Test at 25^o Flexion

The Lachman test will be conducted at baseline, 3, 6, 12, and 24 months post-surgery. The tested null hypothesis is that the true mean differences between injured and uninjured knees between groups are equal compared to the alternative hypothesis that they are not based on a two-sided, two-sample t-test.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

9.3.11 Time to Return to Work, School, Sports

Time for patients to return to full time work or school, as well as sports, will be recorded for all patients in the study. Failure to return to work or sports will be recorded as a secondary outcome, and if the failure is due to the operative knee, that will also be recorded. In addition, the length of time for patients to meet return to sport criteria, including 90% strength on the operative side for quadriceps and hamstring strength and achievement of a hamstring to quadriceps ratio of 0.60, will also be recorded for all patients. The tested null hypothesis is that the true proportions who have returned to work/school/sports are equal between treatment groups compared to the alternative hypothesis that they are not based on a two-sided Fisher's Exact Test.

$$H_0: p_1 = p_2$$
$$H_1: p_1 \neq p_2$$

In addition, a Kaplan-Meier plot and analysis will be generated to compare the time to event across groups for each endpoint above.

9.4 Additional Effectiveness Variables

The summary and analysis of the additional variables collected that were not pre-specified as endpoints will be performed on available data without censoring for subsequent intervention. No multiple imputation for missing data is planned for these endpoints.

9.4.1 Hop Test

Hop testing will be performed at 6, 12 and 24 months post-surgery if the patients have progressed to hopping as part of their rehabilitation with their physical therapist and if the patient feels comfortable doing the test. We will compare the results between the BEAR and ACLR groups. Results will be summarized descriptively by time point.

9.4.2 Static and Dynamic Balance

Static and dynamic testing will be performed at 6, 12 and 24 months post-surgery. The results between the BEAR and ACLR groups will be compared. Results will be summarized descriptively by time point.

9.4.3 Biodex Testing

Biodex testing will be performed at 6, 12 and 24 months post-surgery. The results between the BEAR and ACLR groups will be compared. Results will be summarized descriptively by time point.

9.4.4 Pivot Shift

A pivot test will be conducted as part of a knee examination at 3, 6, 12, and 24 months post-surgery. The results between the BEAR and ACLR groups will be compared. Results will be summarized descriptively by time point.

9.4.5 Effusion Grade

An effusion grade will be assigned as part of a knee examination at 3, 6, 12, and 24 months post-surgery to help identify an infection. The results between the BEAR and ACLR groups will be compared. Results will be summarized descriptively by time point.

9.4.6 VAS Pain

VAS pain scores will be collected at 2 weeks, 6 weeks, 3 months, 6 months, 12 months and 24 months post-surgery. The results between the BEAR and ACLR groups will be compared. Results will be summarized descriptively by time point.

9.4.7 Follow-Up Activity

Patients are asked questions regarding their current level of activity (return to work/school and participation in sports) at the 2 week, 6 week, 3 month, 6 month, 12 month and 24 month follow-up visits. We will compare the results between the BEAR and ACLR groups. Results will be summarized descriptively by time point.

9.4.8 Marx Activity Scale

Patients are asked to complete the Marx Activity level at baseline, 12 month and at the 24 month follow-up visits. The results between the BEAR and ACLR groups will be compared. Results will be summarized descriptively by time point.

9.4.9 Surgeon Evaluation

The surgeon will complete an evaluation of complication history since last visit at the 2 week, 6 week, 3 month, 6 month, 12 month and 24 month follow-up visits. Results will be summarized descriptively by time point.

10 Safety Analyses

Safety analysis will be conducted based on the AT Population.

10.1 Safety Variables

10.1.1 Deep Joint Infection/Incision and Drainage of Deep Surgical Site Infection

Patients will be monitored for any signs of a post-operative infection. If there is clinical suspicion for a possible deep joint infection (fever greater than 101 degrees Fahrenheit, increasing pain in the knee, presence of an effusion, drainage from the knee), a knee arthrocentesis will be performed and if organisms are cultured from the joint fluid, the patient will be classified as having a deep joint infection (according to CDC/NHSN Surveillance Definitions for Specific Types of Infections) and treated accordingly. Any patient diagnosed with a deep joint infection or who undergoes incision and drainage of a deep surgical site infection will have the event recorded as an adverse event. Summary statistics will be provided, and a Fisher's Exact test will be used for comparison between the treatment groups.

10.1.2 Evidence of Graft Rejection

If a patient presents with a swollen, warm knee and there is clinical suspicion of marked inflammation versus septic joint, an arthrocentesis will be performed. If the synovial fluid culture is negative for organisms, the patient will be classified as having a marked inflammatory reaction and treated accordingly. Bovine Type I collagen antibodies, ANA, CBC with differential, CRP and ESR lab tests will be performed as well for all symptomatic patients. In addition, a urinalysis will be performed and in patients having protein in the urinalysis, a protein electrophoresis will also be performed. Evidence of graft rejection, clinically or by serology, will be recorded as a secondary outcome measure as well graft removal for any reason. Summary statistics will be provided, and a Fisher's Exact test will be used for comparison between the treatment groups.

10.1.3 Graft or Repair Failure

A patient shall be noted to have had a graft or repair failure when one or more of the following criteria are met: positive pivot shift exam, Lachman exam with greater than 6 mm side to side difference, absence of tissue in the expected ACL location on MRI imaging, MR evidence of graft or repair loss of continuity or symptomatic instability requiring revision of the ACL surgery. Relevant summary statistics will be provided, and a Fisher's Exact test will be used for comparison between the treatment groups.

10.1.4 Adverse Events of Interest

Incidence of the following adverse events will be of particular interest and will be recorded for all patients in the study: deep venous thrombosis, loss of function, need for prolonged parenteral pain medication, development of neurologic symptoms or additional trauma. Relevant summary statistics will be provided, and a Fisher's Exact test will be used for comparison between the treatment groups.

10.1.5 Additional Surgical Procedures Required

Any additional surgical procedures that the patient requires on the operative knee, as well as any surgical procedures required on the contralateral knee, will also be recorded and reported as an additional secondary outcome measure. These include (but are not limited to) additional surgery to address meniscal or cartilage pathology, scar tissue, arthrofibrosis, removal of symptomatic hardware or graft removal for any reason. Relevant summary statistics will be provided, and a Fisher's Exact test will be used for comparison between the treatment groups.

10.1.6 Bovine Type I Gelatin Antibodies

After 6 months post-procedure, patients will be tested for IgE bovine gelatin antibodies. The number and percentage of patients who test positive will be presented, and the bovine antibody levels will be summarized descriptively. A Fisher's Exact test will be used for comparison of the presence of bovine type I collagen antibodies between the treatment groups.

$$H_0: p_1 = p_2$$
$$H_1: p_1 \neq p_2$$

A two-sample t-test will be used to compare the mean bovine antibody level between the treatment groups.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Data collected at baseline will also be presented for comparison.

11 ADVERSE EVENTS

11.1 ALL ADVERSE EVENTS

Summaries of incidence rates of individual AEs overall and by System Organ Class and Preferred Term will be prepared. Only treatment emergent AEs will be analyzed (a treatment emergent adverse event is one that started or worsened in severity at or after start of randomized treatment). Because a patient may experience more than one AE, summaries will provide both the number of patients experiencing at least one event and the number of events within a reporting period. Percentages provided will be the percent of patients experiencing one or more adverse events. In addition, incidence of AEs will be presented by severity (mild, moderate, severe, life-threatening, fatal) and by relationship to investigational product or procedure. Patients experiencing an event more than once will be counted under the maximum severity/relationship experienced.

A listing of all adverse events will include the patient number, AE number, days since index procedure, the AE name, the severity of AE, whether or not the AE is classified as serious (SAE), the relationship of the AE to the investigational device or procedure, the action taken, and the outcome.

11.2 ADVERSE EVENTS LEADING TO WITHDRAWAL

A summary of incidence rates (frequencies and percentages) of AEs leading to study withdrawal by System Organ Class and Preferred Term will be prepared for the safety Population. A data listing of AEs leading to withdrawal will also be provided, displaying details of the event(s) captured on the CRF.

11.3 SERIOUS ADVERSE EVENTS

Summaries of incidence rates and relationship to the investigational device/procedure of individual SAEs by System Organ Class and Preferred Term will be prepared. Summaries will provide both the number of patients and the number of events within a reporting period. Percentages provided will be the percent of patients experiencing one or more serious adverse events. A data listing of SAEs will also be provided, displaying details of the event(s) captured on the CRF.

11.4 DEVICE OR PROCEDURE RELATED ADVERSE EVENTS

Summaries of incidence rates of device and procedure related AEs by System Organ Class and Preferred Term will be prepared. Summaries will provide both the number of patients and the number of events within a reporting period. Percentages provided will be the percent of patients experiencing one or more device or procedure related adverse events. Data listings of device and procedure related AEs will also be provided, displaying details of the event(s) captured on the CRF.

11.5 POST RE-INTERVENTION

The incidence of AEs resulting in surgical re-intervention or re-treatment will be summarized. AEs and SAEs occurring post re-intervention or re-treatment will also be summarized according to the treatment provided during the re-intervention. Summaries will provide both the number of patients and the number of events.

11.6 DEATHS

Should any patients die during the course of the BEAR II trial, relevant information will be supplied in a data listing.

12 Other Planned Analyses

12.1 Planned Subgroup Analyses

12.1.1 Hamstring-preference Stratum

In addition to conducting the non-inferiority analyses for all patients, the hamstring-preference stratum is a subgroup of special interest and we will conduct the analyses within that group as a pre-specified subgroup analysis. If non-inferiority is demonstrated in the hamstring-preference subgroup, then we will test for superiority for the hamstring strength endpoint at 6 and 12 months in that subgroup. We hypothesize superiority of BEAR over hamstring graft at these early time points.

12.1.2 Age

Analyses of primary and secondary endpoints intended for labeling will be repeated by age group. A comparison of outcomes by age group ≤ 18 years old and > 18 years old will be generated.

12.1.3 Sex

Analyses of primary and secondary endpoints intended for labeling will be repeated by sex.

12.1.4 BMI

Analyses of primary and secondary endpoints intended for labeling will be repeated by BMI subgroup (defined as normal, overweight or obese).

12.2 Exploratory Analyses

12.2.1 Predicted Yield and Maximum Load of the Repaired ACL/ACL graft as per MRI

MRI will be used at six months, 1 and 2 years with the same specific sequence previously validated in the porcine model to predict the yield and maximum load of the repaired ACL or ACL graft (see Appendix B to the protocol). The sequence has not yet been validated in humans. In addition, MR imaging will be used to measure the volume and orientation of both the ACL repaired using the BEAR technique and the ACL reconstruction as well as that of the contralateral knee. During the course of this study, we will determine if the values for maximum load, yield load, stiffness, volume or orientation of the healing ACL or graft, as well as the relative value of these parameters compared with the contralateral side, are predictive of patient outcomes, including patient reported outcomes on the IKDC or KOOS testing, graft or ACL re-rupture, rate of return to sport/work and muscle strength.

12.2.2 Historic Control for BPTB Analysis

Given that both hamstring autograft and BPTB autograft are standard of care options, an analysis will be done on the available outcomes comparing the BEAR arm to the BPTB autograft from a historical control group. The analysis and discussion of the results in the BEAR II study will be considered in light of data from the MOON ACL reconstruction cohort, which includes sufficient line by line data to allow the application of the BEAR II Inclusion/Exclusion criteria, so comparison can be made for several of the key outcomes. This includes the primary patient reported outcomes (IKDC, KOOS, Marx) and KT-1000 testing results, for both hamstring and BPTB grafts. This analysis will allow for a more robust comparison of results across hamstring autograft and BPTB autograft, as the number of BPTB cases in BEAR II is expected to be limited. Endpoint analysis will be performed as described for each available endpoint in this SAP.

12.2.3 Rate of Re-intervention

An additional analysis will be performed to determine if factors known to influence the incidence of ACL reconstruction failure also influence the incidence of graft or repair failure in this cohort. Analyses and modeling will be conducted to assess the impact of covariates such as age, gender, randomized treatment group, baseline body mass index (BMI), baseline MARX score, hamstring : quadricep ratio for the injured knee, knee flexion angle for the injured knee, and tibial slope on the rate of re-intervention.

12.2.4 Ad Hoc Analyses

All pre-planned statistical analyses are included in the SAP. Additional unplanned analyses may occur at the time of the CSR submission or in support of future publications. Ad hoc analyses will be described and labeled as such.

13 Reporting Conventions

All reporting will meet the standards of BBA SOP BS002 and its associated work instructions.

15 LIST OF TABLES, LISTINGS AND FIGURES

Table	Title	Population(s)	Subgroup
1	Patient Accountability	Consented Patients	
2	Demographics and Baseline Characteristics	ITT, AT, mITT, PP	
3	Knee Injury	mITT	
4	Surgery	mITT	
5	Intraoperative Data	mITT	
6	Medial Meniscal Tear	mITT	
7	Lateral Meniscal Tear	mITT	
8	Primary Endpoints	mITT, PP	Hamstring preference, age, sex, BMI, surgeon
9	Secondary Endpoints Intended for Labeling	mITT, PP	Hamstring preference, age, sex, BMI, surgeon
10	Secondary Endpoints Not Intended for Labeling	mITT	
11	Safety Endpoints	mITT	
12	IKDC Patient Reported Score	mITT, PP	
13	Functional Testing Examination	mITT, PP	
14	KOOS Scores	mITT, PP	
15	Physical Exam: Knee Range of Motion (ROM) and Thigh Circumference	mITT, PP	
16	Physical Exam: Dynamometer Testing at 3 Months	mITT, PP	
17	Physical Exam: Lachman and Single Leg Squat at 3 Months	mITT, PP	
18	Physical Exam: Lachman at 6 Months, 1 Year, and 2 Years	mITT, PP	
19	Physical Exam: Pivot Shift	mITT, PP	
20	Physical Exam: Effusion Grade	mITT, PP	
21	Follow-up History	mITT, PP	
22	SF-36	mITT	
23	Marx Activity Scale	mITT	
24	Safety Summary at 2 Years	mITT	
25	Evidence of Graft Rejection/Graft Removal for Any Reason/Graft or Repair Failure/Reoperation	mITT	
26	Summary of AEs	AT	
27	AEs by SOC and PT	AT	
28	SAEs by SOC and PT	AT	
29	AEs Leading to Withdrawal	AT	
30	AEs by Relatedness	AT	
31	AEs Leading to Surgical Re-Intervention or Re-Treatment	AT	
32	AEs Occurring Post Surgical Re-Intervention or Re-Treatment	AT	
33	Surgeon Evaluation: Complication History Since Last Visit	mITT	
34	Laboratory Results at 6 Months	mITT	
35	Post-Operative X-Ray Findings at 2 Years	mITT	

16 CHANGES FROM PROTOCOL

The following table provides a list of changes from the protocol to the SAP, and the justification for each change.

Section	Description	Justification
3.2.3 Secondary Endpoints – Not for Labeling Claim	Effectiveness endpoints that were identified as secondary endpoints in the protocol but that are not intended for labeling have been included in the secondary endpoints not for labeling claim section.	Secondary endpoints identified in the protocol have been divided into more specific subgroups. SF-36 has additionally been identified as a secondary endpoint.
3.2.4 Safety Endpoints	Safety related endpoints that were identified as secondary endpoints in the protocol but that are not intended for labeling have been included in the secondary safety endpoints section.	Secondary endpoints identified in the protocol have been divided into more specific subgroups.
4. Sample Size	There were 2 typos identified in the sample size justification section of the protocol that have now been corrected in the SAP.	Corrections made when authoring SAP.
6. Analysis Populations	The SAP specifies an ITT, mITT, AT, and PP analysis population. The protocol only discusses the ITT and PP populations.	In FDA correspondence (Q180185/S001 – De Novo Pre-Sub Supplement, dated 25Feb2019), the agency recommended that safety data be presented on the AT population. In this regard, the mITT is also more appropriate for the effectiveness data presentation as there were several subjects that were removed from the study post randomization and prior to treatment attempt.
7.3 Methods for Withdrawals and Missing Data	The SAP specifies that mITT with multiple imputation will be the primary analysis.	In FDA correspondence (G150268 – BEAR II IDE approval letter, dated 07Jan2016), the agency recommended that MI be considered.
7.3 Methods for Withdrawals and Missing Data	The SAP outlines several sensitivity analyses.	In FDA correspondence (G150268 – BEAR II IDE approval letter, dated 07Jan2016), the agency recommended that sensitivity analyses for missing data be added in their study design considerations.
7.5 Multiple Comparisons and Multiplicity	The SAP specifies several secondary endpoints that are being sought for labeling claims. The protocol specifies secondary endpoints, but does not specify the endpoints for labeling.	The BEAR II submission is being entered for approval in support of marketing. While the protocol was first submitted under a more academic construct, the sponsor wishes to add endpoints for marketing consideration. FDA also recommended this approach (G150268 – BEAR II IDE approval letter, dated 07Jan2016).

Section	Description	Justification
7.5 Multiple Comparisons and Multiplicity	The SAP specifies a gatekeeping approach will be used to test the secondary endpoints that are being sought for labeling claims to control the Type I error rate.	In FDA correspondence (G150268 – BEAR II IDE approval letter, dated 07Jan2016 and G150268/S001 – BEAR II IDE approval letter, dated 28Apr2016), the agency recommended that if it there are intentions to make claims in the labeling for any secondary endpoints that the sponsor would need to show that they have controlled statistical multiplicity.
9.2.6 – 9.2.12 KOOS Testing	The protocol identifies 3 domains that will be tested: Pain, Quality of Life and Sports; in the SAP we indicate that all 5 domains will be tested.	Data is being collected on all 5 domains, and it was the intention that they would all be analyzed. It is hypothesized that pain and symptoms will be superior in the BEAR group compared to the control group.
10.1.6 Bovine Type I Gelatin Antibodies	The protocol states that bovine type I collagen antibodies were collected, however it was actually bovine type I gelatin antibodies that were collected. The SAP correctly identifies these as gelatin antibodies.	Corrections made when authoring SAP.
10.1.6 Bovine Type I Gelatin Antibodies	The protocol states that presence of IgG antibodies will be collected as part of the secondary endpoint of bovine Type I antibodies, however we have removed it as part of the secondary endpoint in the SAP.	It was determined after the study started that IgE antibodies are more appropriate than IgG antibodies to use as a screen for hypersensitive individuals, so the SAP identifies IgE antibodies as the secondary endpoint.
12.2.1 Predicted Yield and Maximum Load of the Repaired ACL/ACL Graft as per MRI	This endpoint was originally listed as a secondary endpoint, however it has been moved to an exploratory analysis. Language has been added to clarify that this MRI sequence has not been validated in humans.	In FDA correspondence (Q180185/S001 – De Novo Pre-Sub Supplement, dated 25Feb2019), the agency requested this exploratory analysis be addressed in the SAP.

Section	Description	Justification
12.2.2 Historic Control for BPTB Analysis	This exploratory analysis was not mentioned in the protocol, but was added to the SAP.	In the Q180185 submission (De Novo Pre-Sub, dated 25Apr2018), it was also noted that the BEAR I and BEAR II studies primarily used hamstring tendon as the autograft for the ACLR control group. FDA expressed their awareness that autograft tissue selection may be related to the individual preferences of each surgeon, and noted that this study did not specify what autograft source tissue to use in the ACLR control groups. FDA also expressed awareness that both hamstring autograft and bone-patellar tendon-bone (BPTB) autograft are standard of care options in ACLR, and recommended that our data analysis and discussion include a comparison to outcomes from BPTB autograft from a historical control group if assessment of comparable patient populations and outcomes is available.
12.2.3 Rate of Re-intervention	This exploratory analysis was not mentioned in the protocol but was added to the SAP.	In the Q180185 submission (De Novo Pre-Sub, dated 25Apr2018), FDA noted that the analysis evaluating the factors that predispose patients towards re-injury should be performed on the ACLR patients as well.