Statistical Plan for the NUsurface® Meniscus Implant IDE Study G120182

January 2014

Study Design and Hypothesis

The NUsurface® Meniscus Implant IDE Study is a multi-center, prospective, randomized pivotal clinical trial to test the hypothesis that implantation of the NUsurface device into the targeted patient population will provide results that are superior to patients treated with the non-surgical standard of care. More specifically, the following hypothesis of superiority will be tested: two years after treatment, recipients of the NUsurface® Meniscus Implant will have a higher probability of "Overall Success" vs. "Overall Failure" than recipients of the Non-Surgical Standard of Care. The null hypothesis is the converse: two years after treatment, the probability of "Overall Success" will not be greater in NUsurface Meniscus Implant recipients. (Campbell 1993) The pivotal study intends to enroll 62-64 subjects per group with a 1:1 allocation ratio for a total of 124-128 subjects.

Randomization Ratio, Method, and Blinding

Patients who meet the inclusion/exclusion criteria will be assigned by a 1:1 balanced randomization into one of two groups. Even though the study is not blinded, randomization should provide a comparable mix in two treatment groups in terms of patient characteristics such as gender, weight, treated knee side, and baseline pain. The use of a 1 to 1 randomization ratio will be stochastic, meaning that at the time of assignment to the two groups all of the patients will have an equal probability of being assigned to either group. (AAOS 2008) The use of a 1:1 balanced randomization not only minimizes the sample size necessary to have enough power to demonstrate superiority, but conveys the idea of "equipoise" to the patient in that both arms of the study will be viewed as having an equal chance of success.

Balance randomization at each site will be achieved by using blocks of 4 prepared in a 1:1 ratio of NUsurface device to non-surgical Control. In other words, for each site a set of 4 sequences will contain a randomized sequence of 2 NUsurface devices and 2 Controls. Each site will be told after informed consent has been obtained as to which arm the patient has been randomized. To prevent any possible bias, no one will be informed of the size of the randomization blocks. In the event that an unexpected out-of-balance allocation between the two arms of the study is detected for any reason, such as from surgery bailouts or patient dropouts after consent and before treatment, the randomization ratio may be adjusted so as that as near a perfect 1:1

randomization between the treatment arms is achieved. This event, if it occurs, would be a randomization re-balancing process.

Patients and surgeons cannot possibly be blinded about receiving surgical or non-surgical treatment since it will be obvious whether surgery was performed or not. Evaluators also cannot possibly be blinded since surgical scars will be visible. Thus, this study cannot and will not be blinded, either single or double.

Homogeneity of Demographic Parameters and Baseline Pain Scores Between the Two Arms of the Study

Baseline balance of the NUsurface Meniscus Implant and Non-Surgical Standard of Care Control arms will be checked. As appropriate, comparisons between the two treatment groups will be made using Student's t (e.g., for age and BMI at the time of treatment), chi-square (e.g., for gender and treated side), and Cochran-Mantel-Haenszel chi-square tests for variables, if any, with ordinal values such as "none, mild, moderate, severe." As mentioned, if during the course of the study it is noticed that the number of patients in each arm is severely imbalanced for any reason, some modification of the randomization ratio between future patients entered into the study may be carried out.

Poolability of Data Across Study Sites

The study sites will be compared for poolability of the data to identify any sites with unusual or significant heterogeneity in the NUsurface Meniscus Implant minus Standard of Care difference of the Overall Success outcome. Between-treatment *difference* of the outcomes will be examined for homogeneity across sites. So it is not problematic if individual sites have elevated or lower success rates for both study group arms. A similar analysis will be used to compare investigators with the FDA's definitions of "financial interests" to investigators without financial interests. If the results show these two groups of data are poolable, they will be pooled.

Statistical Method Employed to Determine Overall Success

As described in the protocol, the primary variable of the primary endpoint is KOOS (Knee injury Osteoarthritis Outcome Score) (Roos 1999). The methodology used to analyze the data will be similar to that described by Saris 2009 whereby the change from baseline for each patient will be examined over time. Justification for the threshold used for success (20 points improvement) is discussed in the Clinical Investigation Plan (CIP) or study protocol. At 24 months each individual observed patient will be determined to be either a study Overall Success or Overall Failure by first determining if a Provisional PRO KOOS success has occurred. Since the protocol

states that ≥20 point improvement in the 24 month KOOS Pain and KOOS Overall relative to baseline is required to be a Provisional PRO success, the exact details for determining when 20 points is achieved is needed. The reason details are needed in the Statistical Plan is that because the 20 point cut-off value is a whole interger and the KOOS subscale values are almost always fractional numbers. If the KOOS values are rounded to one decimal point, then it is possible in the KOOS Pain subscale, for example, that two differences of the same magnitude could be 19.4+ for one set of KOOS values and 19.5+ for another pair. Therefore, in order to make equal improvements on different locations of the KOOS scale ruled as a success for both arms of the study, a 19.4 or greater improvement in KOOS values will be used to determine a Provisional 20 point KOOS success.

After excluding patients in both arms who have died, who are Lost-to-Follow-up, or who have withdrawn from the study before 24 months (a number expected to be 15% or less for each arm of the study), and after excluding those patients that are automatic study failures as defined in the protocol, each of the remaining patients will be determined to be either an Overall Success or Overall Failure using the Provisional PRO KOOS success criteria described above for KOOS Pain and KOOS Overall. Then the percent of observed patients at 24 months who are an Overall Success in each arm will be calculated by taking the number of patients with Overall Success and dividing by that number plus the number of Overall Failures. Further details and explanations of the Overall Success/Overall Failure definition are provided in the IDE study protocol or CIP. In brief, the definition of Overall Success for the NUsurface arm also requires that the NUsurface patient not be revised for infection or a device-related reason and not have a negative image on MRI. The non-surgical Control arm patient must not receive surgery on the index compartment of the knee during the study duration in order to be an Overall Success.

Once calculated, the ratio of Overall Success at 24 months for the two arms of the study will then be compared by a one degree of freedom chi-square test with significance defined by a two-sided p-value of less than 0.05 with part of the 0.05 spent in the interim analysis and the remainder at the end of the study, as described below.

Interim Analysis of the Data

Since the IDE study is not blinded, nor could it be, and no code needs to be broken to analyze the data, the Sponsor intends to perform many interim analyses throughout the course of the study, such as for annual reports for the FDA, for investor presentations, for when all patients reach 6 or 12 month follow-up, for publication purposes, etc. However, these types of interim analyses will only be performed for business or planning purposes and not for study decision

making. For these types of interim analyses no adjustment to the significance levels are needed. In other words, no p-value penalty is applicable for any non-decision making interim analysis.

Only one planned interim analysis for a statistical decision making purpose will occur when ~~half~~ the first 16 patients in each arm of the study reach 24 months. Since any decision making interim analysis does have the potential for an elevated risk of Type I error, the Sponsor will employ the Haybittle-Peto rule and spend only a fraction of the total possible p = 0.05 for the study in the decision making interim analysis. Namely, only 0.001 of the total 0.05 possible p in the study will be allocated to this decision making interim analysis. The identical means of determining Overall Success/Failure at the end of the study, as described in the protocol, and accounting for patients will be used for the decision-making interim analysis.

This single planned interim analysis using this higher significance threshold to check for superiority will have one of five possible decision making outcomes for the study sample size of each arm of the pivotal study (32-33, 33-62, 62-64, >64, and >>>64) a described below:

1) <u>The first 16 patients in each arm reach 24 months and the NUsurface arm is found to be superior at the p level of 0.001 level of confidence</u>. This finding would mean that the original sample size for the IDE study was unnecessarily large for demonstrating the effectiveness of the NUsurface treatment over the standard of care for the target population. For example, if the 2 year decision making interim analysis had the following rates: 75% of the NUsurface patients are an Overall Success and 15% of the Controls are an Overall Success, then only 25 patients in each group at 2 years would be needed to demonstrate superiority of the NUsurface treatment over non-operative Controls. If such a possible outcome occurs at this decision making interim analysis, a Pre-Market Approval (PMA) application for the NUsurface® Meniscus Implant will then be filed using the Stage I patients (32-33 in each arm).

2) <u>The first 16 patients in each arm reach 24 months and the interim analysis finds that the NUsurface arm is not yet superior to the Control arm at an alpha level of 0.001.</u> Even though superiority has not yet been demonstrated, the drop-out rates, failure rates, infection rates, and Overall Success means and standard deviations for the two arms of the study would be known with confidence. These parameters can then be used to re-calculate the sample size needed to demonstrate superiority of the NUsurface arm at a p level = 0.049. As such, the decision making interim analysis may trigger a sample size re-estimation of the exact sample size required to demonstrate superiority of the NUsurface® Meniscus Implant arm over the non-operative conservative care arm. The new re-estimation of sample size may be between

half the study and the original size of the pivotal study. If so, then the originally planned study over-estimated the sample size needed and the study will be then modified so that a marketing application can be submitted before all the originally planned patients in the study reach 24 months.

3) The re-estimation or re-calculation of the sample size from this one time interim decision making analysis concludes that the study should proceed as originally planned until all the patients in both arms (Stage I and II) reach the 24 month time point. If this outcome occurs, then the study protocol will not be amended.

4) The re-estimation or re-calculation of the sample size from the interim analysis concludes that the original sample size for the entire study (Stage I and II) was under-estimated and more patients need to be added. If this outcome occurs, then the IDE study will be modified through an IDE Amendment.

5) The decision making interim analysis demonstrates that the study should be stopped for futility. The non-operative Control arm may be found to be performing better than the NUsurface device. Alternatively, the sample size needed to demonstrate superiority of the NUsurface device may be calculated to be so large that carrying out such a study would be financially prohibited. If this outcome occurs, further enrollment in the study will be stopped and the IDE halted.

As mentioned, to avoid elevating the total Type I error, the remaining portion of the p = 0.05 (0.049) will be “spent” on the final analysis when all the patients (the size of which is determined or re-determined by the interim analysis) reach 24 months. If the Overall Success primary endpoint variable at 24 months is superior for the NUsurface arm, then the secondary endpoints will be analyzed for superiority by a hierarchal rank order closed procedure statistical methodology until one variable no longer demonstrates superiority. All the secondary outcome results after the cut-off of superiority including any cost or reimbursement analyses will be described with descriptive statistics (i.e., no p values). See the Tabular Summary of the Statistical Plan to Analyze for Superiority that is attached to the end of this statistical plan.

Study Outcomes:

Primary Safety/Effectiveness Outcomes

The primary outcome variable is Overall Success which is defined in the study protocol or CIP. In brief, Overall Success is made of the following effectiveness and safety components:

-Effectiveness: The KOOS Pain Subscale and KOOS Overall at 24 months relative to baseline.

-NUsurface Device Related Complications: The device related complications of the NUsurface Meniscus Implant during the 24 month post-operative period that leads to a secondary surgical intervention or negative MRI images as defined in the protocol.

-NUsurface Patients Revised for Infection: Patients who become infected after surgery that result in device removal at 24 months or less.

-Non-Surgical Control Post-Intervention Surgical Procedures: The treatment failures of the Control group during the first 24 month period.

Secondary Safety/Effectiveness Outcomes

Effectiveness: The effectiveness of the NUsurface Meniscus Implant relative to the Control non-surgical standard of care will be evaluated as secondary endpoints at the pre-op, 1.5, 6, and 12 months post-operative periods using the KOOS Pain Subscale and KOOS Overall Scale. The effectiveness of the IKDC and VAS outcome will be evaluated at 1.5, 12, and 24 months relative to baseline.

-Safety: Adverse events in either arm occurring at a rate of 5% or more at any post intervention follow-up period will be compared between study groups. Examples of safety measurements are:

1) Incidence rate of all adverse events through the 1.5, 6, 12, and 24 month post-intervention period.

2) Incidence rate of secondary (or primary for the non-surgical Controls patients) surgical interventions through the 1.5, 6, 12, and 24 month post-intervention period.

3) Device-related complication is defined as a complication found to be caused by the device and device malfunctions (including severity) as assessed at each post-op period.

Thresholds for stopping the study for some of these safety elements are described in the protocol (CIP). Definitions for many of the terms used in this Statistical Plan are also in the study protocol.

Statistical Analyses of Study Objectives

The incidence of adverse events, secondary surgical interventions, and device related complications will be compared between both study arms using standard statistical methods for data on counts and rates. Any patients with a post-intervention surgical procedure in either arm of the study will also be followed for adverse events.

Additional Secondary Statistical Analyses

Additional secondary statistical analyses of the data obtained in this study will be performed. For example, the sub-group of the Medicare age population (age ~65-75 inclusive) will be statistically analyzed to determine whether Medicare age patients benefited from the intervention of the investigational device and whether their results were poolable with, and therefore not statistically different from, the study group as a whole. Also, radiographic analysis or measurements taken from the MRI's taken at 1.5, 12, and 24 months are planned. Spearman's Rho used to correlate the self-reported and objectively assessed outcomes may also be computed. The preference-based (multi-attribute) health status classification instrument, the EuroQol EQ-5D™, that is taken at pre-op or baseline and at 1.5, 12 and 24 months will be analyzed using the methods provided by the makers of the EQ-5D form. Secondary analyses of costs and reimbursement issues will also be performed. A Kaplan-Meier survivorship analysis of the data may also be performed. An analysis of the results as a function of the duration of symptoms, as well as number of previous surgeries, will also be performed. Any other secondary analyses requested by the FDA during the PMA review will also be performed.

Modified Intent-to-Treat (ITT) Analysis and Handling of Missing Data

The primary effectiveness analysis will be performed using the modified Intent-To-Treat (ITT) sample. A pure ITT analysis examines every patient enrolled in the study and counts them towards the group into which they were randomized, regardless of whether or not the patient ever received treatment or actually received the treatment that they were randomized to receive. A modified ITT only examines the results of patients who actually received treatment. Patients who were randomized but not treated will be documented. While the primary effectiveness analysis will be based on observed data only, various sensitivity analyses will be performed to investigate the impact of missing data in KOOS Pain and KOOS Overall scores. In the first sensitivity analysis, missing post-randomization KOOS Pain or KOOS Overall data will be handled using last observation carried forward (LOCF). The next two sensitivity analyses will impute missing values to the best possible outcome (optimistic imputation) and the worst possible outcome (pessimistic

imputation), respectively. A more rigorous sensitivity analysis will be based on the method of Multiple Imputation or MI (Rubin 1987). MI is a regression-based method in which missing values are imputed using information that is available for a patient. Implemented in SAS/STAT PROC MI, each missing outcome value will be imputed five times and results of the five data sets (i.e., five artificially complete data sets with no missing values) were pooled using PROC MIANALYZE to reflect their consensus and the differences among the five sets of regression coefficients.

Possible Modifications to the Statistical Plan and Number of Patients in the Study

As mentioned, since the NUsurface Meniscus Implant pivotal IDE study is in two stages and an interim analysis will be performed, the Sponsor may submit an IDE Supplement to refine the sample size requested in Stage II of the pivotal study. Additionally, since long-term Stage I data will be available from some patients, the Sponsor, depending upon the enrollment rate and other reasons, may submit to the FDA a Bayesian statistical study plan for the Stage II portion of the pivotal study so that the short-term study results of the Stage II patients could be projected to 24 months using the Stage I data at 24 and earlier months. The clinical data from both stages will be statistically integrated to make one pivotal study of NUsurface device safety and effectiveness.

Any changes in the clinical study, continued access, and post-market study design that would affect the Statistical Plan will be provided in future IDE Supplements. If the clinical investigation includes any combination of US and non-US populations, as mentioned in the draft November 2011 and June 2013 FDA IDE Guidance, the Sponsor will perform statistical analyses demonstrating that the US and non-US data can be pooled.

References:

American Academy of Orthopaedic Surgeons Clinical Practice Guideline on the Treatment of Osteoarthritis of the Knee (Non-Arthroplasty). Rosemont (IL): American Academy of Orthopaedic Surgeons (AAOS); 2008.

Campbell MJ, Machin D. Medical statistics: a commonsense approach. John Wiley & Sons, Chichester, 1993.

Roos E. Knee Injury and Knee Osteoarthritis. Development, evaluation and clinical application of patient-relevant questionnaires. Studentlitteratur AB, Sweden, 1999.

Rubin, D.B., *Multiple Imputation for Nonresponse in Surveys*. 1987, New York: John Wiley & Sons.

Saris DBF, Vanlauwe J, Victor J, Almqvist KF, Verdonk R, Bellemans J, Luyten FP. Treatment of Symptomatic Cartilage Defects of the Knee. Am J Sports Med 2009 37: 10S.

Statistical Plan for the NUsurface IDE Randomized Study

Start of Study	→ Half Patients at 24 Months	→	Final Sample Size of Patients at 24 Months
Total Type I Error 5%	**$p < 0.001$**		**$p < 0.049$**
	Interim Analysis for Early Stop of Study,		**Hierarcheal Rank Order for Superiority Tests:**
	Stopping for Futility, or Sample Size Re-Estimation	**1**	**Overall Success of All* Patients at 24 Months**
		2	24 Month VAS vs Baseline
	Overall Success for 1/2 Patients at 24 Months	3	24 month MRI vs Baseline of Cartilage Condition In Medial Compartment
		4	24 Month IKDC SKEF Score vs Baseline
		5	24 Month QLYA Score vs Baseline (using EQ-5D)
		6	24 Month KOOS Pain
		7	24 Month KOOS Overall
		8	12 Month KOOS Pain
		9	12 Month KOOS Pain vs Baseline
		10	12 Month VAS vs Baseline
		11	12 Month KOOS Overall vs Baseline
		12	12 month MRI vs Baseline Cartilage Thickness at Center of Medial Tibial Plateau
		13	12 Month IKDC SKEF Score vs Baseline
		14	12 Month QLYA Score vs Baseline (using EQ-5D)
		15	24 Month Return to Work
		16	6 Month KOOS Pain
		17	6 Month VAS vs Baseline
		18	6 Month IKDC SKEF Score vs Baseline
		19	6 Month KOOS Overall
		20	6 Month QLYA Score vs Baseline (using EQ-5D)

Note: Variables in Black are Primary Endpoints,
Variables in Blue are Secondary Endpoints
By the Hierarcheal Testing Procedure Each Variable at the End of the Study Must Show Superiority at $p < 0.049$ Before Proceeding to the Next Variable

***All" patients is defined as ≥ 85% of final pivotal study sample size determined after interim analysis**