Protocol/CIP No. SM-2012-02

A Prospective, Double Blinded, Multi-Center, Randomized, Controlled Trial to Evaluate Mechanical Debridement vs. Radiofrequency-Based Debridement in the treatment of Articular Cartilage Lesions

Statistical Analysis Plan
Prepared for:
Smith & Nephew

Final Version 4.0, Date 10Mar2017

Prepared by:
Chiltern International, Inc.
King of Prussia, PA

Revision History

Version	Date	Revision Author	Comments
Draft 0.1	27Mar2013	Shashidhar J Savanur	Initial Draft
Draft 0.2	09Apr2013	Shashidhar J Savanur	Revisions
Draft 0.3	12Apr2013	Shashidhar J Savanur/Dan Pericak	Revisions
Draft 0.4	15Apr2013	Shashidhar J Savanur/Dan Pericak	Revisions
Draft 1.0	15Apr2013		Sponsor comments: MRI endpoints and subgroups, removal of sequential testing
Draft 1.1	15May2013	Shashidhar J Savanur/Dan Pericak	Additions of MRI endpoints and subgroups, removal of sequential testing.
Draft 1.2	22Jul2013	Shashidhar J Savanur/Dan Pericak	Subgroup Analysis section updated according to Sponsor suggestion,
Draft 1.3	24Jul2013	Shashidhar J Savanur/Dan Pericak	MRI endpoints
Draft 2.0	24Jul2013	Dan Pericak	MRI endpoints/minor edits
Final 1.0	12Sep2013		Sponsor comments (17Nov2014): Revisions to include KOOS subscale analysis as secondary efficacy endpoints
Draft 3.0	26Nov2014	Arijit Sarkar/Shashidhar J Savanur/Dan Pericak	Revisions to include KOOS subscale analysis as secondary efficacy endpoints
Draft 4.0	15Mar2016	Padmashree Jahagirdar/Dan Pericak	Revisions to include MMRM for analysis of primary end point and modifications for the SF-12 calculations
Draft 4.1	23Mar2016	Dan Pericak	Revisions to include MMRM for analysis of primary end point and secondary analyses
Draft 4.2	24Mar2016	Dan Pericak	Added details for repeated measures covariance model selection
Draft 4.3	29Mar2016	Padmashree Jahagirdar/ Dan Pericak	Revisions to covariance structures when non-convergence is observed, formatting of sections
Final 3.0	05April2016	Padmashree Jahagirdar	Sponsor comments/Abbreviation list updates
Draft 5.0	27Feb2017	Karisse M. Roman-Torres	Changes based on Protocol V4, description of new stratification for the randomization, change in the Non-inferiority margin for the Week 52 analysis, updates to SF-12 scoring

Draft 5.1	10Mar2017	Karisse M. Roman-Torres	Included 2nd interim analysis details, Analysis of EQ VAS
Final 4.0	10Mar2017	Karisse M. Roman-Torres	Sponsor comments/Sponsor titles

This document is the confidential information of Smith & Nephew. It may not be disclosed to parties not associated with the clinical investigation or used for any purpose without the prior written consent of Smith & Nephew.

SIGNATURE PAGE

Author:

*Name:*Karisse Roman-Torres, M.Sc. *Title:* Manager, Biometrics Chiltern International, Inc.	Date

Reviewed by:

Name: Dan Pericak, M.Math, P.Stat *Title:* Director, Biometrics Chiltern International, Inc.	Date

Approved by:

Name: Alan Rossington *Title:* Director Biostatistics and Data Management Smith & Nephew	Date

Table of Contents

1. LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Table 1: Abbreviations and Definitions of Terms

ACL	Anterior Cruciate Ligament
ACT	ArthroCare Cartilage Trial
ADE	Adverse Device Effect
AE(s)	Adverse Event(s)
AIC	Akaike's Information Criterion
ANCOVA	Analysis of Covariance
AR(1)	First order Autoregressive
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CFR	Code of Federal Regulations
CI	Confidence Interval
cm	Centimeters
CS	Compound Symmetry
CRF	Case Report Form
DICOM	Digital Imaging and Communications in Medicine
DSMB	Data Safety Monitoring Board
EuroQol	European Quality of Life
FOV	Field of Vision
IA	Interim Analysis
ICF	Informed Consent Form
ICRS	International Cartilage Repair Society
IFU	Instructions For Use
IKDC	International Knee Documentation Committee
ITT	Intent-To-Treat
KOOS	Knee Injury and Osteoarthritis Outcome Score

LOCF	Last Observation Carried Forward
LS	Least Squares
LSMEAN	Least Squares Mean
Max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities
MAR	Missing at Random
MCAR	Missing Completely at Random
MCS	Mental Component Summary
MMRM	Mixed Model Repeated Measures
Min	Minimum
MRI	Magnetic Resonance Image
MOAKS	MRI Osteoarthritis Knee Scores
N	Number of subjects
PCS	Physical Component Summary
PP	Per-Protocol
PT	Preferred Term
QOL	Quality of life
RCT	Randomized Clinical Trial
RF	Radiofrequency
SAE	Serious Adverse Event
SD	Standard deviation
SOC	System Organ Class
TEAE	Treatment-Emergent Adverse Event
UN	Unstructured
USADE	Unanticipated Serious Adverse Device Effect
VAS	Visual Analogue Scale
WHO	World Health Organization
WORMS	Whole-Organ Magnetic resonance Imaging Score

2. INTRODUCTION

Articular cartilage defects or chondral lesions are commonly detected during arthroscopy when treating knee pathology such as a torn meniscus or damaged anterior cruciate ligament (ACL). Three recent large studies consisting of 190, 1,000, and 25,124 patients each have shown that between 60% and 70% of patients having arthroscopic knee surgery have concomitant chondral lesions (1-3). These lesions were found to occur most often in patients undergoing meniscal tear repair or ACL repair and were usually found on the medial femoral condyle.

Focal chondral lesions observed during arthroscopy are usually addressed using any one of several different treatment options that are selected using a clearly defined algorithm (4). The treatment selected depends on the severity of the lesion, which is graded using a scheme such as the International Cartilage Research Society (ICRS) classification system (5). When chondral lesions are treated surgically, first line methods generally consist of debriding loose or worn articular cartilage, with the goal of stabilizing the lesion to prevent further degeneration. One method of treatment includes radiofrequency (RF)-based debridement.

A prospective study by Voloshin et al. followed one hundred ninety-three patients that underwent bipolar RF–based chondroplasty over 38 months. Of these patients, fifteen with a total of twenty-five defects were re-examined when undergoing repeat arthroscopy for recurrent or new injuries. Of the twenty-five lesions, only three demonstrated further deterioration of the cartilage defects after treatment with RF (6). Limited clinical data exists, however, evaluating the use of RF-based debridement for the treatment of chondral lesions compared to another commonly used treatment method, mechanical debridement. Owens and Busconi reported that women treated for isolated patellofemoral chondral lesions using a bipolar RF-based device tended to have better joint evaluation scores and demonstrated less incidence of crepitus at one and two years post-operatively than those receiving mechanical shaver-based treatment (7). In patients undergoing partial meniscectomy and having concomitant medial femoral condyle chondral lesions treated using mechanical debridement alone or mechanical and RF-based

debridement, Barber and colleagues observed no significant differences between treatment groups in clinical outcomes through one year (8). Spahn et al. reported contrary findings to the Barber group when studying a similar patient cohort in Germany (9). In their prospective, randomized, controlled trial they found significantly better clinical results, as measured using the Knee Injury and Osteoarthritis Outcome Score (KOOS), Tegner scores, and qualitative recovery measures, in patients receiving mechanical and RF-based debridement compared to those receiving mechanical debridement alone.

Currently, it is not known if clinical outcomes following treatment of chondral lesions by mechanical debridement (i.e. mechanical shaver) or RF-based debridement may equal one another.

The primary aim of this study is to evaluate clinical outcomes following RF-based debridement or mechanical debridement for subjects requiring treatment of a single medial femoral chondral lesion plus partial medial meniscectomy procedure. The secondary aims include evaluating imaging and additional clinical outcomes.

The Quantum 2™ System and WEREWOLF COBLATION System are FDA cleared bipolar, RF electrosurgical systems designed for use in orthopaedic/arthroscopic surgical procedures.

The Quantum 2 Controller System consists of the following components:

1. A bipolar radiofrequency Controller;
2. A reusable, non-sterile Foot Control;

2a. A reusable, non-sterile wireless Foot Control (optional);

3. A reusable, non-sterile Power Cord;
4. A reusable, non-sterile Patient Cable (optional); and
5. A disposable, sterile ARTHROWAND Wand (sold separately).

The WEREWOLF™ Controller and FLOW 50 Wand system consists of the following components:

1. A bipolar radiofrequency Controller with Integrated Fluid Outflow Regulator
2. A non-sterile, reusable wired Foot Control and Power Cord
3. Sterile, disposable, single-use COBLATION Wands

3. STUDY OBJECTIVES

The primary objective is to evaluate clinical outcomes following RF-based debridement or mechanical debridement for subjects requiring treatment of a single medial femoral chondral lesion plus partial medial meniscectomy procedure. The secondary objectives are to evaluate imaging and additional clinical outcomes.

4. STUDY DESIGN

4.1 General Design

This is a non-inferiority, prospective, double blinded, multi-center, randomized, controlled, adaptive study design with enrollment of 82 randomized subjects at up to 13 study sites. Study duration will be until the last subject reaches 104 weeks post-operative evaluations.

The study will be comprised of two parts:

Part I: Part I will require all Investigators to perform 1 to 3 procedures using the Quantum 2 Controller plus Paragon T2 ICW Wand or the WEREWOLF Controller plus FLOW 50 Wand. The purpose of this Part I will be to minimize variability with the recommended directions for use established in the instructions for use (IFU). Part I subjects will be followed per the protocol follow-up requirements, and will be included in the safety population only. These subjects will be additive (to the safety population), that is, in addition to the 82 randomized subjects planned as part of the primary evaluation in Part II.

Part II: Part II will include 82 study subjects. Each Investigator may initiate enrollment of subjects in this part of the study following completion of Part I requirements. The randomization will be by-site randomization to assure a balanced number of subjects between the study devices and control device group at each investigational site. In addition, the randomization will be stratified by lesion grade 3A or grade 2 (see section 4.3 for further details). The assignment will be based on a 1:1 ratio (Group I: Group II) where the treatment groups will be defined as:

- Group I: RF-based debridement **(study devices)**
- Group II: Mechanical debridement **(control device)**

If randomized to RF-Based Debridement, the subject will receive the Quantum 2 Controller plus Paragon T2 ICW Wand or the WEREWOLF Controller plus FLOW 50 Wand.

An interim analysis is planned to occur when 50% of subjects are randomized and have completed the Week 24 post-operative follow up evaluations. This interim analysis will be primarily for sample size re-assessment of the Part II study. There will be no reduction in sample size as a result of the interim analysis; however, the sample size may be increased either to establish the non-inferiority and/or may be increased sufficiently to establish superiority depending on the results of the interim analysis. The study will not be stopped due to the effectiveness results, and no hypotheses testing will be conducted to assess the differences between the two groups.

Study duration will be until the last subject enrolled reaches 104 weeks post-operative treatment. Subjects will be assessed pre-operatively and return post-operatively at Day 10, Weeks 6, 12, 24, 36, 52, and 104. At each follow-up visit, subject questionnaires will be administered and AEs and concomitant medication will be reviewed, if applicable. At Day 10, Weeks 52 and 104 post-operatively, the subjects will complete a Magnetic Resonance Image (MRI) of the treated knee.

4.2 Discussion of Study Design

This study has the following design characteristics:

Non-inferiority: In this study the treatments groups will be compared to test the null hypothesis that there is an important difference between the treatments and reject the null hypothesis in favor of the alternative hypothesis that the new treatment (RF-based debridement) is not inferior to control treatment (mechanical debridement) with a certain non-inferiority margin. Non-inferiority is established by showing that the upper bound of the one-sided 97.5% confidence interval (CI) for (new treatment - control) is less than the pre-specified non-inferiority margin. After review of the literature (10), as pre-specified in the protocol, a non-inferiority margin difference of 8 points in change in KOOS from baseline to Week 52 between the two groups has been chosen as being clinically significant. A decision of modifying the non-inferiority margin from 8 (as specified in the protocol) to 10 points for the analysis at Week 52 is based on literature which provides a range of 8-10 points as the maximum acceptable extent of clinical non-inferiority of the study device and control group [10].

Double Blind: In order to eliminate assessment bias both the study subjects and study radiologists reviewing the MRI scans will be blinded to the randomly assigned treatment group.

Multicenter: The study is designed as a multicenter study in order to recruit the necessary number of subjects in a shorter timeframe and to allow for the generalizability of results across different regions.

Treatment duration: The study duration is 104 weeks post-operative treatment and is considered as an adequate evaluation period to assess clinical outcomes of RF-based debridement or mechanical debridement in the treatment of articular cartilage lesions.

Randomization: Randomization is employed to minimize systematic differences between the treatment groups and to provide a sound basis for statistical inference.

4.3 Method of Assignment of Subjects to Treatment Groups

The randomization will be in 1:1 ratio (i.e., study device: control device). In each investigational site, subjects who are eligible for randomization will be randomized in a double-blind manner to a device. Subject randomization will be stratified by site. As per protocol version 4.0, the inclusion criteria of lesion grade has been expanded, to allow, in addition to lesion Grade 3A, subjects with lesion Grade 2 with widely displaceable fibrillation or flaps. As a consequence, the

randomization schedule was modified to include lesion grade as an additioal stratification factor (either Grade 3A or Grade 2). Accordingly, subjects randomized prior to protocol version 4.0 entered the study with lesion grade of 3A and therefore will be considered as already randomized under this strata. Subjects randomized after protocol version 4.0 will be further stratified at the site level to either lesion grade 3A or 2.

4.4 Blinding

Subjects will be blinded to treatment assignment until they complete the course of the study. All efforts will be made to keep the subject blinded. All post-operative MRI scans will be transferred to the study radiologists in Digital Imaging and Communications in Medicine (DICOM) format. All subject information will be masked to ensure blinding. The study radiologists reviewing the scans will also be blinded to the treatment received by the subject. MRI scans will be evaluated at Day 10, Weeks 52 and 104 post-operative treatment. All efforts will be made to keep the study radiologists blinded to treatment assignment information by restricting access to related information.

4.5 Determination of Sample Size

This study will be conducted in a two parts. There is no power calculation for Part I. In Part II, study subjects will be randomized to one of the two treatment groups in a 1:1 ratio.

A total of 82 subjects (41 in the study device group and 41 in the control device group) will be randomized in an effort to assure that 70 subjects (i.e., 35 subjects per group) complete the study. PASS version 2011, is used for this sample size and power calculation. The calculation for sample size is based on the assessment of non-inferiority of the two treatment groups with respect to their effect on change from baseline in KOOS at Week 52. The requirement of 70 subjects (35 active and 35 control) is based on achieving 80% statistical power to detect a non-inferiority margin difference of 8 points (10) between the two groups at 5 % level significance. The discontinuation rate is predicted to be no more than 15%. To accommodate the potential discontinuations, 82 randomized subjects (41 in the study device group and 41 in the control device group) are needed for this study.

Per protocol, an interim analysis is planned to occur when 50% of subjects are randomized and have completed the Week 24 post-operative follow up evaluations. Re-estimation of the sample size and conditional power will be computed based on the interim analysis results – the details of this analysis is provided in Section 7.5.

5. CHANGES IN THE CONDUCT OF THE STUDY OR PLANNED ANALYSES

5.1 Changes in the Conduct of the Study

There were no changes in the conduct of the study at the time of preparing this statistical analysis plan.

5.2 Changes from the Analyses Planned in the Protocol/CIP

The following analyses noted in the protocol were clarified:

- The "average" KOOS score will be utilized for primary.

The following analyses noted in the protocol were added:

- KOOS subscales will be presented at each time point as secondary endpoint.

The following analyses noted are described in the SAP:

- Change in the non-inferiority margin of 8 to 10 for the final efficacy analysis for the Week 52 analysis.
- Statistical efficacy models will be also adjusted by baseline lesion grade (Grade 3A, Grade 2).
- As requested by the Sponsor, a second interim analysis will be performed and include all Part 2 subjects' 52 week to re-evaluate the sample size and conditional power using the data available at the time of analysis.

6. BASELINE, EFFECTIVENESS AND SAFETY EVALUATIONS

6.1 Schedule of Evaluations

Table 2: Schedule of Evaluations

Procedure	Screening/ Baseline	Surgery	Post-Operative Follow-up Evaluation						
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9
	Day -21 to -1	Day 0	Day 10 ±3 Days	Wk 6 ±5 Days	Wk 12 ±7 Days	Wk 24 ± 14 Days	Wk 36 ±14 Days	Wk 52 ±28 Days	Wk 104 ±56 Days
Informed Consent	X[1]								
Medical History & Demographics	X								
Subject Eligibility Verification	X	X							
Physical Assessment – Knee	X		X	X	X	X	X	X	X
Magnetic Resonance Imaging (MRI)	X[2]		X					X	X
Weight bearing AP / Lateral / Merchant View	X								
Weight bearing posteroanterior radiograph	X								
Standing Long Leg Alignment	X							X	
International Cartilage Repair Society classification (ICRS)		X[3]	X[4]					X[4]	X[4]
International Knee Documentation Committee (IKDC) [5]	X			X[6]	X	X	X	X	X
Visual Analogue Scale (VAS), knee pain	X		X	X	X	X	X	X	X
Knee Injury and Osteoarthritis Outcome Score (KOOS)	X			X	X	X	X	X	X
SF-12	X			X	X	X	X	X	X
EQ-5D-5L	X		X	X	X	X	X	X	X
Subject Satisfaction								X	X
Adverse Events		X	X	X	X	X	X	X	X
Post-operative Rehabilitation[7]			X	X					
Concomitant Medications	X	X	X	X	X	X	X	X	X
Protocol Deviations	X	X	X	X	X	X	X	X	X

[1] Must occur prior to any study–specific procedures
[2] MRI must occur within 9 months prior to surgery confirming presence of a single medial femoral chondral lesion and medial meniscal tear requiring a partial meniscectomy

[3] International Cartilage Repair Society (ICRS) classification will be used for intra-operative arthroscopic confirmation of the lesion grade
[4] International Cartilage Repair Society (ICRS) classification will be used for determination of lesion grade by MRI
[5] International Knee Documentation Committee (IKDC) Subject Knee Evaluation and Knee Examination Forms will be used
[6] IKDC Knee Examination Form - Subject dependent
[7]Post-operative Rehabilitation per Investigator standard of care

6.2 Time Point Algorithms

6.2.1 Relative Day

The date of debridement will be considered to be relative day 0, and the day before debridement will be relative day -1. Relative days will be calculated as follows and only when the full assessment date is known, i.e., partial dates will have missing relative days, unless otherwise noted:

Relative. Day =
- Date of Assessment – Date of debridement. For days on or after debridement
- Date of Assessment – Date of debridement For days prior to debridement

6.2.2 Visit Windows

The time schedule described in the protocol for each scheduled activity will be followed as closely as possible. The analysis visit windows in Table 3 below will be utilized for analysis purposes. All scheduled and unscheduled visits will be windowed according to the table below.

Table 3: Visit Windows

Protocol Visit	Target Day	Visit window for Analysis (days)
Baseline	-1	Last observation prior to debridement.
Visit 2 (Surgery)	0	0
Visit 3 (Day 10 ± 3 Days)	10	1-26
Visit 4 (Week 6, Day 42 ± 5 Days)	42	27-63
Visit 5 (Week 12, Day 84 ± 7 Days)	84	64-126

Visit 6 (Week 24, Day 168 ± 14 Days)	168	127-210
Visit 7 (Week 36, Day 252 ±14 Days)	252	211-308
Visit 8 (Week 52, Day 364 ± 28 Days)	364	309-546
Visit 9 (Week 104, Day 728 ± 56 Days)	728	>= 547

If a subject has more than one assessment occurring in the same analysis visit window, the data from the visit closest to the scheduled study day will be used for summaries. If two assessments have the same distance from the scheduled study day, the assessment after the scheduled study day will be used.

If two assessments both occur on the protocol-specified day, and both are valid, then the later one will be summarized. All assessments will be provided in the data listings.

6.3 Baseline Assessments

The following baseline/screening assessments will be conducted:

- Confirmation of written informed consent
- Inclusion/exclusion criteria
- Demographics (e.g., age, gender, height, weight, BMI, work status and nicotine use)
- Concomitant medication (if applicable)
- Protocol deviation (if applicable)
- Medical and surgical history
- Physical Assessment – Knee
- IKDC Knee Examination Form
- Mechanism of injury
- Screening number assignment
- Pre-operative imaging (Imaging Assessment)
    - Weight-bearing Anterior / Posterior view
    - Weight-bearing Lateral view
    - Merchant view

- Weight-bearing posteroanterior radiograph
- Standing Long Leg Alignment
- MRI (within 9 months of surgery)

- Self-reported questionnaires:
  - KOOS
  - IKDC Subjective Knee Evaluation Form
  - SF-12
  - EQ-5D-5L
  - VAS knee pain

6.4 Effectiveness Variables

For all effectiveness evaluations, the baseline measurement is defined as most recent assessment prior to surgery.

6.4.1 Primary Effectiveness Variable(s)

6.4.1.1 Change from baseline in Knee and Osteoarthritis Outcomes Scores (KOOS) at Week 52 post-operative

The primary effectiveness variable is the change from baseline in Knee and Osteoarthritis Outcomes Scores (KOOS) at Week 52 post-operative. The baseline KOOS is the score prior to the debridement surgery. The change from baseline in KOOS at Week 52 will be computed as follows:

Change from baseline in KOOS at Week 52 = KOOS at Week 52 – KOOS at baseline

KOOS consists of 5 subscales: Pain, other Symptoms, Function in daily living (ADL), Function in Sport and Recreation (Sport/Rec) and knee related Quality of Life (QOL). A 1-week recall is taken into consideration when answering the questions. Each subscale response is based on a 5-point Likert system with each response score ranging from 0 (No problems) to 4 (Extreme problems) (12). Each subscale score is calculated independently.

The individual five KOOS subscale scores (i.e., Pain, ther Symptoms, Function in daily living (ADL), Function in Sport and Recreation (Sport/Rec) and Knee related Quality of Life (QOL)) will be summarized and analyzed to enable clinical interpretation (See section 6.4.2.1.3). Traditionally, in orthopedics, a score of 100 indicates no problems and a score of 0 indicates extreme problems. Each subscale score is normalized and transformed to meet this standard. The KOOS subscale score is computed as follows:

$$KOOS\ Subscale\ score = 100 - \frac{\text{mean of the observed items within the subscale x 100}}{4}$$

A total score has not been validated and is not recommended (13). For the purposes of this trial, the average of KOOS subscale scores will be considered as the primary endpoint.

Missing and improper response to questions in the KOOS questionnaire will be handled as follows:

1. As long as at least 50% of the subscale questions are answered for each subscale, a mean score will be calculated.
2. If more than 50% of the subscale questions are omitted, the response is considered invalid and no subscale score will be calculated.

6.4.2 Secondary Effectiveness Variables

The individual five KOOS subscale scores will be summarized and analyzed to enable clinical interpretation.

The secondary effectiveness variables for this study are classified in two sets: Clinical Endpoints and Imaging Endpoints. These are discussed below in Sections 6.4.2.1 and 6.4.2.2.

6.4.2.1 Clinical Endpoints

6.4.2.1.1 Change from Baseline in International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form scores

The IKDC Subjective Knee Evaluation Form scores will be assessed at baseline, Weeks 6, 12, 24, 36, 52 and 104. At each timepoint assessed, summary statistics will be presented for the each subgroup (i.e. symptoms, function, and sports activities and IKDC score along with the associated change from baseline). The change from baseline in IKDC score will be computed at Weeks 6, 12, 24, 36, 52 and 104 as follows:

Change in IKDC score from baseline at Week (i) = IKDC score at Week (i) – IKDC score at baseline

Details of IKDC scoring are provided in Appendix 2: Statistical Analysis and Programming Details of the SAP.

6.4.2.1.2 International Knee Documentation Committee (IKDC) Knee Examination Form scores

The IKDC -Knee Examination Form scores will be assessed at baseline, Weeks 6, 12, 24, 36, 52 and 104. At each timepoint assessed, the group grades and final grade will be summarized as frequency and percentage. The change from baseline in IKDC Knee Examination Form score will be computed at Weeks 6, 12, 24, 36, 52 and 104.

Details of IKDC Knee Examination Form scoring are provided in Appendix 2: Statistical Analysis and Programming Details of the SAP.

6.4.2.1.3 KOOS Subscale Score

As described in section 6.4.1.1 individual KOOS subscale scores will be computed. Each subscale score is calculated seperately for baseline and Weeks 6, 12, 24, 36, 52 and 104.

6.4.2.1.4 Change from baseline in Average KOOS for Weeks 6, 12, 24, 36 and 104

As described in Section 6.4.1.1, the change from baseline in average KOOS will be calculated for Weeks 6, 12, 24, 36 and 104.

6.4.2.1.5 Change from baseline in KOOS subscale scores for Weeks 6, 12, 24, 36, 52 and 104

For each subscale the change in KOOS subscale score from the corresponding KOOS subscale score at baseline will be calculated for weeks 6, 12, 24, 36 52 and 104.

6.4.2.1.6 Change in Visual Analogue Scale (VAS) Knee Pain Score from baseline

VAS knee pain will be assessed at baseline, Day 10, Weeks 6, 12, 24, 36, 52 and 104.
At each timepoint assessed, summary statistics will be presented for the VAS knee pain score and the change from baseline. The change from baseline in VAS knee pain score will be computed at Day 10, Weeks 6, 12, 24, 36, 52 and 104 as follows:

Change in VAS knee pain score from baseline at Week (i) or Day (i) = VAS knee pain at Week (i) or Day (i) – VAS knee pain at baseline

6.4.2.1.7 Change in SF-12 scores from baseline

The SF-12 will be assessed at baseline, Weeks 6, 12, 24, 36, 52 and 104. SF-12 Health Survey includes 8 domains commonly represented in health surveys: physical functioning, role functioning physical, bodily pain, general health, vitality, social functioning, role functioning emotional, and mental health. Two summary measures will be computed at each timepoint: Physical Component Summary (PCS) and the Mental Component Summary (MCS). The change from baseline in SF-12 domain scores, PCS and MCS scores will be computed at Weeks 6, 12, 24, 36, 52 and 104 as follows:

Change in the SF-12 domain score from baseline at Week (i) = domain score at Week (i) – domain score at baseline

Details of the SF-12 scoring are provided in Appendix 2: Statistical Analysis and Programming Details of the SAP.

6.4.2.1.8 Change in EQ-5D-5L scores from baseline

EQ-5D-5L and EQ-VAS will be assessed at baseline, Day 10, Weeks 6, 12, 24, 36, 52 and 104. Summary descriptive statistics of the 5 dimensions (i.e., mobility, self-care, usual activities, pain/discomfort and anxiety/depression) will be provided for each response level as well as the change in index-based values ("utilities") from baseline.

The change from baseline in EQ-5D-5L utility score will be computed at Day 10, Weeks 6, 12, 24, 36, 52 and 104 as follows:
Change in utility score from baseline at Week (i) or Day (i) = utility score at Week (i) or Day (i) – utility score at baseline
The EQ-VAS scale ranges from 0 to 100 with higher scores meaning better health and lower score representing worst health. The observed EQ-VAS values will be summarized using continuous statistics at each follow-up visit.

In addition, the change from baseline in the EQ VAS score will be computed as:
Change from baseline in the EQ-VAS at Week (i) or Day (i) = EQ-VAS score at Week (i) or Day (i) – EQ-VAS score at baseline.

Details of EQ-5D-5L scoring are provided in Appendix 2: Statistical Analysis and Programming Details of the SAP.

6.4.2.1.9 Subject Satisfaction at week 52 and 104

Subject Satisfaction with Study Treatment is evaluated at Week 52 (Visit 8) and Week 104 (Visit 9) by asking one single question with responses on a 4-point Likert scale ranging from 1 (very

satisfied) to 4 (very dissatisfied). The satisfaction rating will be considered as a continuous variable (14).

6.4.2.2 Imaging Endpoints

MRI will be assessed at Day 10, Weeks 52 and 104 post-operatively. Images will be scored with respect to 8 independent articular features: cartilage signal and morphology, subarticular bone marrow abnormality, subarticular cysts, subarticular bone attrition, marginal osteophytes, synovial thickening & joint effusion and loose bodies. Readings will be taken independently by two blinded radiologists. Readers will use all images to evaluate each feature.

Five of the features to be examined (cartilage signal and morphology, subarticular bone marrow abnormality, subarticular cysts, subarticular bone attrition, marginal osteophytes) relate to the articular surfaces. These features are to be evaluated in 15 different regions: Medial Femur – Anterior, Medial Femur – Central, Medial Femur – Posterior, Lateral Femur – Anterior, Lateral Femur – Central, Lateral Femur – Posterior, Medial Tibia – Anterior, Medial Tibia – Central, Medial Tibia – Posterior, Lateral Tibia – Anterior, Lateral Tibia – Central, Lateral Tibia – Posterior, Medial Patella, Lateral Patella and Tibial Subspinous.
These assessments will have the following possible responses:

Absent:	If all subregional surface areas have either no cartilage loss or above grading criteria not met.
Present:	Above grading Criteria met. If present the region meeting the grading criteria will be presented
Indeterminate:	If assessment cannot be made from the available images due to technical/confounding factors (such as obscured anatomy, poor contrast or imaging artifact).
Unable to assess:	If assessment cannot be made due to missing images or inadequate field of vision (FOV).
Not applicable:	Assessment is not applicable due to prerequisite conditions not met.

6.4.2.2.1 Cartilage Signal and Morphology

Cartilage signal and morphology will be evaluated in each of the 15 articular-surface regions. The responses will be graded using MRI Osteoarthritis Knee Scores (MOAKS) based percent cartilage loss in one or more subregional surface area.

Table 4: Cartilage Signal and Morphology Grade (MOAKS)

Grade(MOAKS)	Description
Grade 1	One or more subregional surface areas have cartilage loss < 10% of region.
Grade 2	One or more subregional surface areas have cartilage loss ranging between 10% to 75% of region.
Grade 3	One or more subregional surface areas have cartilage loss > 75% of region.
Indeterminate	Assessment cannot be made from the available images due to technical/confounding factors (such as obscured anatomy, poor contrast or imaging artifact).
UA - Unable to assess	Assessment cannot be made due to missing images or inadequate FOV.

6.4.2.2.2 Subarticular Bone Marrow Abnormality

The subarticular bone marrow abnormality will be graded using MOAKS based on percent of subregion volume spanned by subarticular bone marrow lesion.

Table 5: Subarticular Bone Marrow Abnormality Grade (MOAKS)

Grade(MOAKS)	Description
Grade 1	One or more subregional surface areas have bone marrow lesion spanning < 33% of subregional volume.
Grade 2	One or more subregional surface areas have bone marrow lesion spanning 33% to 66% of subregional volume.
Grade 3	One or more subregional surface areas have bone marrow lesion spanning > 66% of subregional volume.
Indeterminate	Assessment cannot be made from the available images due to technical/confounding factors (such as obscured anatomy, poor contrast or imaging artifact).

Unable to assess	Assessment cannot be made due to missing images or inadequate FOV.

6.4.2.2.3 Subarticular Cysts

The subarticular cysts will be graded using MOAKS based on percent of subregion volume occupied by cysts in one or more subregions.

Table 6: Subarticular Cysts Grade (MOAKS)

Grade(MOAKS)	Description
Grade 1	One or more subregional surface areas have cysts occupying < 33% of subregional volume.
Grade 2	One or more subregional surface areas have cysts occupying 33% to 66% of subregional volume.
Grade 3	One or more subregional surface areas have cysts occupying > 66% of subregional volume.
Indeterminate	Assessment cannot be made from the available images due to technical/confounding factors (such as obscured anatomy, poor contrast or imaging artifact).
Unable to assess	Assessment cannot be made due to missing images or inadequate FOV.

6.4.2.2.4 Subarticular Bone Attrition

The subarticular bone attrition will be graded using Whole-Organ Magnetic resonance Imaging Score (WORMS).

Table 7: Subarticular Bone Attrition (WORMS)

Grade(WORMS)	Description
Grade 1	One or more subregional surface areas show flattening of the (normally) convex osseous articular surface
Grade 2	One or more subregional surface areas show slight concavity of the (normally) convex osseous articular surfaces
Grade 3	One or more subregional surface areas show marked concavity of the (normally) convex osseous articular surfaces.
Indeterminate	Assessment cannot be made from the available images due to technical/confounding factors (such as obscured anatomy, poor contrast or imaging artifact).
Unable to assess	Assessment cannot be made due to missing images or inadequate FOV.

6.4.2.2.5 Marginal Osteophytes

The marginal osteophytes will be graded using Whole-Organ Magnetic resonance Imaging Score (WORMS).

Table 8: Marginal Osteophytes (WORMS)

Grade(WORMS)	Description
Grade 1	One or more subregional surface areas have equivocal osteophytes.
Grade 2	One or more subregional surface areas have small osteophytes.
Grade 3	One or more subregional surface areas have small-moderate osteophytes.
Grade 4	One or more subregional surface areas have moderate osteophytes.
Grade 5	One or more subregional surface areas have moderate-large osteophytes.
Grade 6	One or more subregional surface areas have large osteophytes.
Grade 7	One or more subregional surface areas have very large osteophytes.

6.4.2.2.6 Synovial Thickening & Joint Effusion

Synovial thickening and joint effusion were not distinguished from each other, but graded collectively.

Table 9: Synovial Thickening & Joint Effusion Grade (MOAKS)

Assessment	Description
Absent	Appearance of normal physiological amount of fluid.
Small	Fluid continuous in the retropatellar space.
Medium	With slight convexity of the suprapatellar bursa.
Large	Evidence of capsular distention.
Indeterminate	Assessment cannot be made from the available images due to technical/confounding factors (such as obscured anatomy, poor contrast or imaging artifact).
Unable to assess	Assessment cannot be made due to missing images or inadequate FOV.

6.4.2.2.7 Loose Bodies

The presence of loose bodies in joint space will be determined using the following criteria.

Assessment	Description
0 - Absent	Appearance of normal physiological amount of fluid.
1 - Small	Fluid continuous in the retropatellar space.
2 - Medium	With slight convexity of the suprapatellar bursa.
3 - Large	Evidence of capsular distention.
Indeterminate	Assessment cannot be made from the available images due to technical/confounding factors (such as obscured anatomy, poor contrast or imaging artifact).
Unable to assess	Assessment cannot be made due to missing images or inadequate FOV.

6.4.2.3 International Cartilage Repair Score (ICRS) Grade

The Cartilage lesions will be graded using ICRS grades.

Table 10: International Cartilage Repair score (ICRS)

ICRS Grade	Description
Grade 0	Normal knee.
Grade 1	Nearly normal. Evidence of superficial lesions, soft indentations, and/or superficial fissures and cracks.
Grade 2	Abnormal. Lesions extending down to < 50% of cartilage depth.
Grade 3	Severely Abnormal. Cartilage defects extending down >50% of cartilage depth as well as down to calcified layer and down to , but not through the subchondral bone. Blisters are included in this Grade.
Grade 4	Severely Abnormal. Grade 3 defects continue into subchondral bone.
Indeterminate	Assessment cannot be made from the available images due to technical/confounding factors (such as obscured anatomy, poor contrast or imaging artifact).
Unable to assess	Assessment cannot be made due to missing images or inadequate FOV.

6.5 Safety Assessments

6.5.1 Extent of Exposure

All randomized subjects will undergo surgery to receive one of the study devices or the control device on day 0.

A listing indicating details of surgery will be presented which includes (i) duration in operation room, (ii) duration and type of anesthesia, (iii) duration of procedure, (iv) device start and end time (wand), (v) Tourniquet time (initial and subsequent), (vi) Whether or not an intra-articular injection to joint space was administered, (vii) study device lot number and ablation setting used, and (viii) Arthrex measurement probe, 70 degree lot number.

The extent of exposure to study device will be computed as [(study completion date (or discontinuation date) – randomization date) + 1] reported in months and will be descriptively presented.

6.5.2 Adverse Events

An AE can be any unfavorable and unintended sign, symptom, or disease temporally associated with the use of a study device, whether or not considered related to the study device. All adverse event (AE) tables will include only treatment emergent AEs (TEAEs) unless otherwise noted. TEAEs are those AEs which start or worsen in severity after treatment with the study device. If it cannot be determined if the AE started or worsened after treatment with the study device, it will be assumed to be a TEAE. The investigator's verbatim term of each adverse event will be mapped to the system organ class (SOC) and preferred term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA) dictionary.

The expected adverse events include:

- Joint effusion
- Hematoma
- Adhesions or arthrofibrosis
- Hemarthrosis
- Reduced range of motion or gait status abnormality (temporary)

- Localized pain
- Sensation decrease at incision site
- Inflammation
- Infection
- Chondrolysis
- Fever
- Synovitis
- Deep Vein Thrombosis
- Treatment failure due to rehabilitation non-compliance
- Swelling and bruising
- Fracture
- Nerve injury
- Tendon injury
- Delayed wound healing
- Vascular injury
- Conversion to mini-open or open procedure
- Secondary surgical intervention to address complications associated with surgery or treatment
- General risks associated with surgery and anesthesia
- Prolonged surgery time due to device breakage or malfunction
- Patient burn
- Inadvertent ablation

An AE considered to be caused by or related to the device (both study and control) is an Adverse Device Effect (ADE).

Serious Adverse Events (SAEs) are AEs which meet any of the following criteria (in accordance with the recommendations of ICH [Federal Register, October 7, 1997, Vol. 62, No. 194, pp 52239-45]):

- Results in death,

- Is life-threatening (NOTE: the term "life-threatening" in the definition of "serious" refers to an event in which the subject was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe),
- Requires inpatient hospitalization or prolongation of existing hospitalization,
- Results in persistent or significant disability/incapacity,
- Is a congenital anomaly/birth defect

Additionally, events are classified as serious if they meet any of the following criteria:

- Requires intervention to prevent permanent impairment/damage, or
- Important medical events that may not result in death, be life-threatening, or require hospitalization may be considered a serious adverse event when, based upon appropriate medical judgment, they may jeopardize the subject and may require medical or surgical intervention to prevent one of the outcomes listed above.

An Unanticipated Serious Adverse Device Effect (USADE) is described as any serious adverse effect on health or safety or any life-threatening problem or death caused by or associated with a device, if that effect, problem, or death was not previously identified in nature, severity, or degree of incidence in the investigational plan or application (including a supplementary plan or application), or any other unanticipated serious problem associated with a device that relates to the rights, safety, or welfare of subjects (21 CFR 812.3(s)).

In addition to "serious" or "not serious" the intensity of each AE will be assessed using CTCAE v4.03 grading.

Table 11: CTCAE v4.03 General Guidelines

Grade	Description
Grade 1	Mild; asymptomatic or mild symptoms; clinical or diagnostic observation only; intervention not indicated.
Grade 2	Moderate minimal, local or noninvasive intervention indicated; limiting age-appropriate instrumental activities of daily living (ADL).*
Grade 3	Severe medically significant but not immediately life-threatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care ADL.†

Grade 4	Life-threatening consequences; urgent intervention indicated.
Grade 5	Death related to AE ‡

*Instrumental ADL refer to preparing meals, shopping for groceries or clothes, using the telephone, managing money, etc.
†Self-care ADL refer to bathing, dressing and undressing, feeding self, using the toilet, taking medications, and not bedridden
‡Unlike the AE outcome assessment (see protocol Section 13.3.2), a subject may have more than one Grade 5 event.

A missing CTCAE grade category will be presented as necessary to summarize any missing intensity.

Study device related AEs are those events that were rated by the investigator as related to study device. A missing category will be presented to summarize any missing relationship.

6.5.3 Clinical Laboratory Evaluations

This section is not applicable for this study.

6.5.4 Other Observations Related to Safety

6.5.4.1 Physical Assessment of Knee

At each visit the knee will be assessed for warmth, swelling and skin changes are graded using 4-point Likert scale ranging from 1 (none) to 4 (severe). In addition, the evidence of infection will be evaluated.

7 STATISTICAL METHODS

7.1 Treatment Group Descriptors for All displays

In all data displays the actual treatment received will be presented rather than the intended/planned treatment. Any departures from the planned treatment according to randomization will be documented in the Clinical Study Report.

For the purpose of all data displays the treatment groups will be labeled as presented in table below. The treatment will be presented in this order in the summary tables.

Table 12: Treatment Description for Mock Displays

Treatment Group	Description used for Mock Displays

Radiofrequency-Based debridement using the Quantum 2 Controller plus Paragon T2 ICW Wand, or, WEREWOLF Controller plus FLOW 50 Wand	RF-Based Debridement
Mechanical Debridement	Mechanical Debridement

7.2 General Methodology

All analyses will be performed after the database hard lock and the study treatment codes are unblinded.

All statistical tests will be two-sided with a significance level of α=0.05, unless otherwise specified, and will be performed using SAS® Version 9.1.3 or higher. Data will be summarized using descriptive statistics (number of subjects [N], mean, standard deviation [SD], median, range (minimum, maximum)).

One-sided 97.5% CI for treatment difference (Study – Control) will be used for determining non-inferiority/superiority of the study device.

All unscheduled visits will be included and nominal visits will be applied using analysis visit windows. Both the assigned analysis visit and the site reported nominal visit will be provided in the subject data listings.

Subject listings of all data from the case report forms (CRFs), as well as any derived variables, will be presented using the safety population.

7.3 Adjustments for Covariates

Pseudo-site (as required), baseline variable of the outcome measure being analyzed, the lesion grade at baseline and the interaction terms of treatment-by-site and treatment by lesion grade will be included as covariates in ANCOVA in the primary and secondary effectiveness models (unless otherwise specified). For the MMRM analysis, the treatment visit interaction will be added as a covariate along with the above mentioned covariates.

7.4 Handling of Dropouts or Missing Data

Dropout subjects (i.e., subjects who withdraw early from the study post-randomization) will not be replaced in this study.

Incomplete dates will not be imputed.

For the analysis involving missing data, the following procedures will be adopted.

a) Missing Baseline values

For any subject if the baseline value of analysis parameter is missing then that subject will not be included in the analysis of respective parameter.

b) Completely missing post-baseline values

For any subject where all the post-baseline values of an analysis parameter are completely missing then that subject will not be included in the analysis of respective parameter.

c) Missing value of effectiveness parameter at end of treatment.

For any subject, the missing post-operative assessments of effectiveness parameters will be imputed using last observed post baseline value (LOCF method). No data will be carried forward from baseline.

This approach assumes that the subject response remains constant from the last observed value to the end point of the trial. In other words, this method of imputation assumes that the missing data arose from completely missing random mechanism.

The method of LOCF will be used for subjects who are discontinued from the study prematurely.

d) MMRM method

The MMRM method assumes that the missing data arose from missing at random (MAR) mechanism (i.e., the observed response variable (dependent variable) is related to the probability of drop outs). The MAR assumption is often reasonable in clinical trials as the

observed data explain much of the missingness in many scenarios, particularly in well controlled studies such as clinical trials in which an extensive effort are made to observe all the outcomes and the factors that influence them. Moreover, MAR is valid in every case where missing completely at random (MCAR) is valid; however, the converse is not true.

In MMRM, the information from the observed data is used via the within-patient correlation structure to provide information about the unobserved data. The data from all the visits are simultaneously analyzed using restricted maximum likelihood method and no explicit imputation for missing value will be performed. Unstructured covariance matrix is assumed to explain within subject covariance in MMRM method in protecting the type I error rates. In the event if the assumption of unstructured covariance (UN) matrix results in non-convergence of the model, **only** then the following covariance structure will be considered in the order mentioned below and the covariance structure would be selected based on the AIC criteria (i.e., the covaraice structure corresponding to lowest AIC).

The order of covariance structure to be considered is

1. Compound Symmetry (CS)
2. First-order Autoregressive (AR(1)) and
3. Modified AR(1).

In the MMRM model, the time is considered as a factor variable and Treatment $*$ Time effects is considered as an unstructured interaction effect, instead of considering Treatment $*$ Time effect as the slope (rate of change) difference of treatment groups over the study time period. The advantage of considering the effect of Treatment $*$ Time as unstructured is that it provides the direct estimates and statistical test of least square mean (LSMEAN) differences of the treatment groups at the study endpoint, as well as at each scheduled study time point with respect to the primary efficacy measure. Since patients in clinical trials are often evaluated at a fixed number (relatively small) of time points, the MMRM modeling approach facilitates analyzing clinical trial data, considering Time as a factor variable in the model.

e) Missing covariates for Analysis of Covariance (ANCOVA) model.

Any subject with missing covariate will be excluded from analysis.

7.5 Interim Analyses and Data Monitoring

One interim analysis is planned to be conducted for this study. This interim analysis will be primarily planned for sample size recalculation for Part II of this study. The study will not be stopped due to the effectiveness results, and no hypotheses testing will be conducted in this interim analysis to assess the differences between the two treatment groups.

This interim analysis is planned to be conducted when 50% of subjects are randomized and have completed the Week 24 post-operative follow-up evaluations. The interim analysis will be conducted under the auspices of an independent Data Safety Monitoring Board (DSMB). There is no intention of decreasing the sample size, however, the sample size may be increased to either establish the non-inferiority and/or may be increased sufficiently to establish superiority depending on the results of the interim analysis.

The data to be used in the interim analysis and the treatment assignment of each randomized subject will be provided to the independent statistician. Using this data, the independent statistician will calculate the following metrics for the primary endpoint:

- difference in change in average KOOS between the treatment groups
- change in average KOOS from baseline in Group I and the observed number of subjects in the group
- change in average KOOS from baseline in Group II and the observed number of subjects in the group
- dropout rate at the time of the interim analysis
- statistical power of the study at the time of the Interim Analysis (using a conditional power approach)

- conditional power (CP) in this interim analysis will be calculated according to the following formula (Chen 2004) using primary endpoint data from the intent-to-treat (ITT) population (see Section 8.3.1).

$$CP(f_1,z_1) = \Phi \{ z_1 / \sqrt{f_1 (1 - f_1)} - z_\alpha / \sqrt{(1 - f_1)} \}$$

where:

CP(f1,z1) is the conditional power at the interim analysis

Φ{.} is the cumulative distribution function of a standard Normal distribution ($\mu=0$, $\sigma^2 = 1$)

f_1 is the fraction of subjects enrolled and used in the interim analysis before decision of increasing the sample size

z_α is the upper α quintile for standard Normal distribution [Z(alpha/2=0.025) = 1.96)]

z_1 is the standardized Normal

Non-inferiority margin will be specified as 8.

The sample size will be adjusted only if the conditional power at the time of the interim analysis is 50% or more and less than 80%. The sample size will be recalculated based on the observed difference between the treatment groups.

If the conditional power is larger than 50%, then the sample size will be adjusted upward and no Type I error rate adjustment will be made to the final analysis.

If the conditional power is less than 50%, the Type I error rate will be inflated and statistical adjustment will be made to the final analysis, an adjustment to the final p-value will be made as follows:

$$\text{p-value} = 2 [1 - \boldsymbol{\Phi}\{ | f_2^{\frac{1}{2}} z_1 + (1-f_2)^{\frac{1}{2}} z_2 | \}]$$

where:

P-value is the adjusted target p-value at the end of the study reflecting the adjustment for increasing the study size at the interim analysis

n_1, is the sample size at the time of the interim analysis ($n_1 = n_{gp1} + n_{gp2}$)

n_2 is the re-calculated sample size to be enrolled after the interim analysis

$f_2 = n_1 / (n_1+n_2)$ is the fraction of the newly planned sample size at the interim analysis

z_1 is the observed z score at the time of the interim analysis

z_2 is the observed z value of the z score only based on the data collected after the interim analysis

The 1st interim analysis was performed and reviewied by the DSMB on 20APR2016. After reviewing the interim analysis of the data evaluating the KOOS at Week 24 the DSMB recommended increasing the sample size based on the conditional power results.

The sample size increase was reviewed but not actioned by the sponsor at the time of the DSMB recommendation in Protocol Version 4.0.

In lieu, a review into the study design by an independent project team at the Sponsor highlighted two potential sources of imprecision in the 1st interim analysis;

a) The interim analysis and subsequent sample size re-estimation comprised of data on the difference in change in average KOOS between the treatment groups at week 24 post-operatively, while the primary endpoint extends to week 52.

b) The non-inferiority margin chosen as part of the study design is overly conservative based on supporting literature [10].

Therefore, the interim analysis previously performed will be repeated using the following modifications:

- All part 2 subjects with week 52 data available up to the time of the IA analysis (including those subjects who discontinued from the study).
- The noninferiority margin will be defined as a 10 point boundary in the KOOS score at Week 52. The decision of modifying the NI margin is based on literature which provides a range of 8-10 points as the the maximum acceptable extent of clinical noninferiority of the study device and reference group [10].

The methodology for sample size adjustment and adjustment to the final p-value for the second interim analysis the will follow the details noted above in section 7.5.

The second interim analysis will be performed by an independent unblinded Chiltern statistician. Chiltern's Unblinded Statistician will provide a recommendation that will not include any unblinding information. These details of the sample size re-estimation and conditional power will appear only in an unblinded report until the study is completed.
Once the second interim analysis is completed the independent unblinded Chiltern statistician will provide the recomendation in a written form to the Sponsor's statistician and clinical study manager.

7.6 Data Safety Monitoring Board

The Data and Safety Monitoring Board (DSMB) will consist of a group of individuals, appointed by the Sponsor or its designee, with pertinent expertise that will review accumulated data at the interim analysis from the study. The DSMB will advise the sponsor regarding the continuing safety of subjects and those yet to be voluntarily recruited to the study, as well as the continuing validity and scientific merit of the study. Unblinded data reviewed by the DSMB will be kept confidential and protected from inadvertent or inappropriate access by the sponsor or its designee. Following review of data generated from the interim analysis, the DSMB may advise the Sponsor to continue, redesign, or stop the study.

A DSMB will meet twice, once to establish the charter, and once to review the interim analysis. This meeting will be held after 50% of the subjects are randomized and have completed the week 24 visit.

7.7 Study Stopping Rules

The Sponsor may terminate the study at any study site, at any time, for any of the following reasons:

- Non-compliance to Good Clinical Practice (GCP) or protocol
- Failure to enroll subjects
- Major protocol deviations
- Inaccurate or incomplete data
- Unsafe or unethical practices

- Safety or performance considerations
- Recommendation made by the DSMB

7.8 Multi-center Studies and Pooling of Centers

A sufficient number of subjects will be required within each investigative site to determine if there are any outlying sites. If some of the investigative sites are unable to enroll a sufficient number of subjects, it may be necessary to combine smaller sites (where enrollment is too low to detect differences) with larger "similar" sites. If a site has fewer than 5 subjects, then the site will be considered to have an insufficient number of subjects and will be considered eligible for pooling. Eligible sites within the same state or region of the country will be pooled sequentially until a sufficient number of subjects are reached (at least 10). Any required pooling will also be applied to the per-protocol population to allow a valid comparison of the study results.

In this process, sites with fewer than 5 subjects will be ordered numerically in ascending order, based on the investigator number. The median number of subjects in the sites where the number of subjects >= 5 subjects will be estimated. Sites with 4 or fewer subjects will be sequentially pooled in ascending order of site number until the resulting pseudo-site has at least as many subjects as the median size of the other sites. A new pseudo-site to account for additional low-enrolling sites will be started upon achieving the median number of subjects. The process will be repeated until there are fewer than the median number of subjects in all remaining sites that require pooling.

7.9 Multiple Comparisons/Multiplicity

No multiplicity adjustments will be implemented for this study. All secondary endpoints are exploratory in nature.

7.10 Use of an "Effectiveness Subset" of Subjects

Subjects randomized to one of the study devices and who do not have major protocol deviations will form the Per Protocol (PP) Population. The major protocol deviations will be evaluated at

the time of evaluability evaluation, the time between the database soft lock and hard lock before unblinding (as per Chiltern CI409v07 SOP).

The relevant protocol deviations/violations are as follows:

- Violations of inclusion criteria
- Violations of exclusion criteria
- All visit window deviation
- Surgical technique, Procedure,/or Physiotherapy deviation
- Subject questionnaire deviations

Smith & Nephew will review protocol deviations/violations to determine if a subject completed the study according to the protocol. A listing of protocol deviations/protocol violations will be provided.

7.11 Active-Control Studies Intended to Show Equivalence

For the purposes of non-inferiority testing of the primary effectiveness endpoint (change from baseline in KOOS at Week 52), the one-sided 97.5% confidence interval for the treatment difference between control and study device will be calculated to evaluate whether the upper limit of the confidence interval is less than 8 points as pre-specified in the protocol. A decision of modifying the non-inferiority margin from 8 (as pre-specified in the protocol) to 10 points for the analysis at Week 52 is based on literature which provides a range of 8-10 points as the maximum acceptable extent of clinical non-inferiority of the study device and control group [10].

7.12 Examination of Subgroups

Exploratory subgroup analyses will be performed on the primary effectiveness variable to determine if the effectiveness is consistent across certain subgroup levels. The following subgroups will be evaluated:

- BMI (<30 kg/m^2, >=30 kg/m^2)
- Age (<40, >=40 years)
- Total Pre-Debridement Lesion Size (<=2 cm^2, >2-4 cm^2)
- Gender (M, F)

- Nicotine Use (Currently, Stopped/Never)
- Duration since onset of symptoms for index knee (Acute: ≤30 days, Subacute: 31 days – 6 months, Chronic: > 6 months)
- Knee Alignment (Varus malalignment, Valgus malalignment)
- Narrowest width of meniscal rim post-operatively [Radial Measurement of Meniscus post resection] (0-3°, > 3°)

8 STATISTICAL ANALYSIS

A SAP will be developed and approved before the database is locked. The SAP will present the detailed statistical methodology to be used in analyzing the effectiveness and safety data from this study. No inferential statistics will be performed on the baseline and safety data summaries.

8.1 Disposition of Subjects

The disposition of all subjects who sign an informed consent form (ICF) will be provided. The numbers of subjects screened, randomized, completed, and discontinued during the study, as well as the reasons for all post-randomization discontinuations will be summarized by treatment group. Disposition and reason for study discontinuation will also be provided as a by-subject listing.

8.2 Protocol Deviations

The deviations occurring during the clinical study will be summarized by treatment groups.

8.3 Analysis Populations

8.3.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) population is defined as all randomized subjects. The Intent-to-Treat population will be the primary population for the analysis of the primary and secondary endpoints. For the ITT population, no data will be excluded from these analyses due to protocol violations.

8.3.2 Per Protocol Population

The Per Protocol (PP) population is defined as all randomized subjects who were not associated with a major protocol violation.

8.3.3 Safety Population

The Safety population is defined as any subject receiving the device/treatment. This population will be used for the analysis of safety parameters. This population includes the subjects from both Parts I and II of the study (refer to Section 4.1).

8.4 Demographic and Other Baseline Characteristics

All baseline summaries will be displayed by treatment group based on the safety population unless otherwise specified.

8.4.1 Demographics

Age (years), height (cm), weight (kg) and calculated body mass index (BMI in kg/m^2) will be summarized using descriptive statistics (n, mean, SD, median, min, and max). Age group (<40, >=40 years), gender, race, ethnicity, BMI group (<30 kg/m^2, >=30 kg/m^2), nicotine history (currently, stopped/never), work status (Employed full time, Employed part time, Semi-retired, Retired, Unemployed, Disabled, Student), living environment (Home (alone), Home (with spouse, family, friend aid, being cared for by family/other and other), Total Pre-Debridement Lesion Size (<=2 cm^2, >2 cm^2), Duration since onset of symptoms for index knee (acute: ≤30 days, subacute: 31 days – 6 months, chronic: > 6 months), Knee alignment (Varus malalignment, Valgus malalignment) and Narrowest width of meniscal rim post-operatively (0-3°, > 3°) will be summarized using number and percentage. Number of cigarettes, cigar or pipes consumed per day will be presented in the data listing. The lesion grade (3A and 2) frequency distribution will be presented using number and percentage.

8.4.2 History of symptoms

Index knee (left, right), bilateral knee pain, symptoms for index knee (gradual, acute), activity at injury (activity of daily living, sports, work, traffic and other), sports (contact, non-contact), type of work and type of vehicle will be summarized using number and percentage. Duration since onset of symptoms for index knee will be summarized using descriptive statistics (n, mean, SD, median, min, and max).

8.4.3 Pre-Surgery Imaging

Weight-bearing Anterior/Posterior, Weight-bearing Lateral, Merchant, Weight-bearing Posteroanterior Radiogrpah, Standing Long Leg Alignment, anatomic alignment (AP x-ray) degrees varus and valgus will be summarized using number and percentage.

8.4.4 Medical History

The number and percent of subjects with concurrent disease will be summarized according to organ class.

8.4.5 International Cartilage Repair Society (ICRS) Knee history registration-Previous surgery

The number and percent of subjects with previous surgical history of the index knee will be summarized for following:

Ligament surgery

- ACL reconstruction (intra-articular, extra-articular)
- PCL reconstruction (intra-articular, extra-articular)
- Collateral ligament reconstruction (medial, lateral)
- Type of graft
    - Patella tendon (ipsilateral, contralateral)
    - Single hamstring
    - 2 bundle hamstrings

- 4 bundle hamstrings
- quadriceps
- allograft
- other

Extensor mechanism surgery

- Tibial tubercle medialization
- Patella tendon repair
- Anteromedialization
- Quadriceps tendon repair and
- Lateral release

Patellofemora surgery

- Soft tissue realignment (medial imbrication, lateral release)
- Bone realignment
    - Tibial tubercle transfer (proximal, distal, medial, lateral, anterior anteromedial)
    - Trochlear plasty
    - Patellectomy

Osteotomy

- Proximal tibia
- Distal femur
- Varus
- Valgus

8.4.6 Intra-Operative Subject verification on Day 0 (Visit 2: surgery day)

Prior to randomization the following criteria will be evaluated to ensure the subject continues to meet eligibility:

- Single, treatable chondral lesion, localized to the medial femoral condyle
- ICRS Grade 2 with widely displaceable fibrillation or flaps or ICRS Grade 3A,
- $< 4cm^2$ in size

A listing to this effect will be produced for all subjects enrolled in the study.

8.4.7 Baseline Knee evaluation and Classification

Medial meniscus evaluation, location of tear, medial femoral chondral lesion evaluation, lateral meniscus evaluation and ICRS lesion classification at baseline will be summarized as number and percent of subjects in each treatment group.

8.5 Prior and Concomitant Therapy

All prior and concomitant medications recorded in the Case Report Form will be coded using the most recent version of WHO Drug dictionary. Descriptive summaries, by device/treatment group, will be provided using the coded term. All prior and concomitant medications recorded in the Case Report Form will be listed.

8.6 Analysis of Effectiveness Parameters

The primary effectiveness endpoint, the change from baseline in KOOS at Week 52 will be analyzed using LOCF and MMRM method. The analyses will be performed for both ITT and PP population.

In addition, parallel analyses will be performed on (i) the observed data only and (ii) imputation using worst observation carried forward to impute missing KOOS at week 52 to assess the sensitivity of LOCF imputation.

The secondary effectiveness endpoints will be analyzed using the LOCF method to impute missing responses. These analyses will be performed using the ITT population.

8.6.1 Analysis of Primary Effectiveness Variable

8.6.1.1 Change from baseline in KOOS at Week 52

The scores of 5 subscales of KOOS (i.e., Pain, other Symptoms, Function in daily living (ADL), Function in Sport and Recreation (Sport/Rec) and knee related Quality of Life (QOL)) at baseline, Week 52 and change from baseline will be summarized descriptively by treatment group. The average of KOOS subscale scores will be considered as the primary endpoint. At each timepoint assessed, summary statistics will be presented for the KOOS and the change from baseline.

For any subject if the value of effectiveness parameter is missing at Week 52 then it will be imputed by last observed post baseline value (LOCF method).

An ANCOVA model will be used to compare the difference in the devices for change from baseline in the KOOS at Week 52.

For each treatment group, the "adjusted" mean change from baseline value (i.e., SAS® least-squares (LS) means) will be calculated based upon an analysis of covariance model, with treatment ,study site, and lesion grade as factors, and the baseline KOOS value as the covariate. In addition, a treatment-by-site and treatment-by-lesion grade interaction terms will be evaluated in the model. In general, there are two types of interaction: qualitative interactions which occur in those situations where the magnitudes of the between-treatment difference varies across an effect, but are always in favor of one treatment group (e.g., across sites) and quantitative interactions which occur in those situations where the direction of the between-treatment difference varies across an effect. In the presence of qualitative interactions, interpretation of the treatment effect is still valid, whereas, in the presence of quantitative interactions, the interpretation of the treatment effect is typically not valid. In this trial, the presence of a quantitative interaction in the primary effectiveness analysis may require further discussions regarding interpretation of the results. The covariates and treatment- by covariate interaction terms will be evaluated on separate models and if deemed significant at the 0.05 alpha then will

be included into a full model. If the interaction terms are non-significant at the 0.05 alpha level then these will dropped from the model.

The analysis will be presented as n, least square (LS) means, SE, and associated two sided 95% CI for each treatment and for treatment difference.

The underlying assumptions of normality and homogeneity of variance will be examined. If the assumptions are seriously violated then a suitable transformation for the response will be considered.

Non-inferiority will be tested using the ITT population. The null hypothesis is that the difference between the two randomized devices (study device – control device) in the change from baseline in KOOS is less than 10 points. In statistical notation the hypothesis to be tested is:

$$H_o : (\mu_{\text{STUDY DEVICE}} - \mu_{\text{CONTROL DEVICE}}) \geq -|\mathbf{d}|$$

vs

$$H_a: (\mu_{\text{STUDY DEVICE}} - \mu_{\text{CONTROL DEVICE}}) < -|\mathbf{d}|$$

Where:

- $\mu_{\text{CONTROL DEVICE}}$ = mean change from baseline in KOOS at Week 52 for control device
- $\mu_{\text{STUDY DEVICE}}$ = mean change from baseline in KOOS at Week 52 for study device
- d is the non-inferiority margin. In this study d = 10
- The hypothesis will be tested at an effective one-sided alpha level of 0.025

The one-sided 97.5% CI for the difference between devices (for the change from baseline in KOOS at Week 52) will be calculated. If the lower limit of the confidence interval is greater than 10 points, the null hypothesis will be rejected and non-inferiority will be concluded. If the point estimate of treatment difference (study device – control device) and the lower limit of confidence interval is greater than 0, then both superiority and non-inferiority of study device will be concluded.

To assess the consistency of the primary effectiveness analysis results a supportive analysis will be conducted using the PP population.

8.6.1.2 Sensitivity analyses of the Change from baseline in KOOS at Week 52

A sensitivity analsyis of primary efficacy analysis on the KOOS will be performed using a MMRM on the ITT & PP population. For the KOOS score, the statistical model will be a MMRM analyzing change from baseline in KOOS score at Week 52 as the dependent variable. Analysis will include terms for treatment group, site, treatment*site, lesion grade, treatment*lesion grade, visit, visit*treatment, baseline KOOS, and subject as a random effect. Descriptive summary statistics including number of subjects, mean, standard error, and LS means along with the p-values will be provided. In addition, the difference in LS means, the corresponding two-sided 95% CI and the p-value will be provided. Additionally, the p-value testing treatment-by-visit interaction will be provided to assess the consistency of treatment effects over different visits. The site and lesion grade by treatment interaction terms will be evaluated separately, and if deemed significant, then will be included in the full model. If the interaction terms are non-significant at the 0.05 level then these will be dropped from the statistical model.

The repeated measures analysis will be based on the restricted maximum likelihood method. Unstructured covariance matrix is assumed to explain within subject covariance in MMRM method in protecting the type I error rates.

The covariance matrix for the repeated measures to model the within-subject errors will be determined using the following procedure if the assumption of unstructured covariance (UN) matrix results in non-convergence of the model:

1) The following covariance structure will be explored in the following order: Compound Symmetry (CS), first-order Autoregressive (AR(1)), and modified AR(1).
2) Obtain the Akaike's Information Criterion (AIC) value from the PROC MIXED output for each of the covariance structures using the model specified above. The covariance

structure would be selected based on the AIC criteria (i.e., the covaraice structure corresponding to lowest AIC).

3) Obtain diagnostic plots of the residuals from the same PROC MIXED model #2 above for each of the covariance structures.

A Kenward-Roger approximation will be used for the denominator degrees of freedom. With a mixed effects model as the primary analysis model, assuming data are missing at random, no imputation of missing data will be done.

In addtition, sensitivity analyses will be performed using (i) observed data only and (ii) imputation using worst observation carry forward to impute missing KOOS at week 52. The sensitivity analysis will be performed using ITT population.

8.6.2 Analysis of Secondary Effectiveness Variables

Statistical analyses will be performed on Secondary Effectiveness Variables at Weeks 52 and 104. These analyses will be exploratory in nature.

8.6.2.1 Change from baseline in IKDC Subjective Knee Evaluation Form scores

The scores of three domains of IKDC Subjective Knee Evaluation Form(i.e., 1) symptoms, including pain, stiffness, swelling, locking/catching, and giving way; 2) sports and daily activities; and 3) current knee function) at baseline, each of scheduled post-operative visits and changes from baseline will be summarized descriptively by treatment group.

The change in IKDC score from baseline at each scheduled post-operative visit will be analyzed using an ANCOVA model with change in IKDC score from baseline as the response variable and baseline IKDC score, site, treatment, treatment-by-site interaction, lesion grade and treatment-by-lesion grade interaction as independent variables. The site and lesion grade by treatment interaction terms will be evaluated separately and if deemed significant then will be included in the full model. If the interaction terms are non-significant at the 0.05 level then these

will be dropped from the statistical model. The analysis will be presented as n, LS means, SE, and associated two sided 95% CI for each treatment and for treatment difference.

8.6.2.2 Change from baseline IKDC Knee Examination Form scores

Knee examination with respect to generalized laxity (tight, normal, Lax), alignment (obvious varus, normal, obvious valgus), patella position (obvious baja, normal, obvious alta), patella subluxation/ dislocation (centered, subluxable, subluxed and dislocated) will be summarized as number and percent of subjects in treatment group for baseline and post-operative follow-up visits. The treatment comparison of each examination for each post-operative visit will be performed using Chi-square test.

Range of motion (Ext/Flex) will be descriptively summarized for passive and active component within index side and opposite side by treatment group. T-test will be used to compare mean of range of motion between treatment groups

The shift in grade from baseline at each of the scheduled post-operative visits will be presented for final evaluation scores and each of the seven domains: effusion, passive motion deficit, ligament examination, compartment findings, harvest site pathology, X-ray findings and functional test by treatment group. The p-value from chi-square test will be presented to compare IKDC knee examination form score between treatment groups.

8.6.2.3 Change in average KOOS from baseline at Weeks 6, 12, 24, 36 and 104

An ANCOVA model will be used to compare the difference in the devices for change from baseline in the average KOOS at each of scheduled post-operative visits. The model will have change in KOOS as the response variable and treatment, baseline KOOS, site, treatment-by-site interaction, lesion grade and treatment-by-lesion grade interaction as independent variables. These interaction terms will be evaluated separately and if deemed significant then will be included in the full model. If the interaction terms are non-significant at the 0.05 level then these will be dropped from the statistical model.

The analysis will be presented as n, LS means, SE, and associated two sided 95% CI for each treatment and for treatment difference.

In addition, the MMRM method will be performed as described in section 8.6.2.1.3 to compare the difference in the devices for change from baseline in the KOOS at each of scheduled post-operative visits.

8.6.2.4 Change in KOOS subscale score from baseline at Weeks 6, 12, 24, 36, 52 and 104

An ANCOVA model as described in section 8.6.2.1.3 will be used to compare the difference in the devices for change from baseline in the KOOS subscale score at each of scheduled post-operative visits. The KOOS subscale score and change in KOOS subscale from baseline will be summarized descriptively for each scheduled post-operative visits. The above analysis will be present separately for each of 5 subscales (i.e., Pain, other Symptoms, Function in daily living (ADL), Function in Sport and Recreation (Sport/Rec) and knee related Quality of Life (QOL)).

In addition, the MMRM method will be performed as described in section 8.6.2.1.3 to compare the difference in the devices for change from baseline in the KOOS subscale score at each of scheduled post-operative visits.

8.6.2.5 Visual Analogue Scale (VAS) Knee Pain score

The VAS knee pain score at baseline, post-operative visits (Day 10, Weeks 6, 12, 24, 36, 52 and 104) and change from baseline will be summarized descriptively by treatment group.

An ANCOVA model will be used to compare the difference in the devices for change from baseline in the VAS knee pain score at each of scheduled post-operative visits. The model will have change in VAS knee pain as the response variable and treatment, baseline VAS knee pain, site, treatment-by-site interaction, lesion-grade, and lesion-grade interaction as independent variables. These interaction terms will be evaluated separately and if deemed significant then will be included in the full model. If the interaction terms are non-significant at the 0.05 level then these will be dropped from the statistical model. The analysis will be presented as n, LS means, SE, and associated two sided 95% CI for each treatment and for treatment difference.

8.2.6.6 Change in SF-12 scores

SF-12 Physical Component Summary (PCS) score and SF-12 Mental Component Summary (MCS) scores at baseline, post-operative visits (Weeks 6, 12, 24, 36, 52 and 104) and changes from baseline will be summarized descriptively by treatment group.

An ANCOVA model will be used to compare the difference in the devices for change from baseline in the SF-12 PCS scores at each of the scheduled post-operative visits. The model will have change in SF-12 PCS score as the response variable and treatment, baseline SF-12 PCS score, site, treatment-by-site interaction, lesion-grade and treatment-by-lesion grade interaction as independent variables. These interaction terms will be evaluated separately and if deemed significant then will be included in the full model. If the interaction terms are non-significant at the 0.05 level then these will be dropped from the statistical model. The analysis will be presented as n, LS means, SE, and associated two sided 95% CI for each treatment and for treatment difference.

Similar analysis will be performed for the SF-12 Mental Component Summary scores at each of scheduled post-operative visits.

8.2.6.7 EQ-5D-5L scores

Each of the 5 dimensions will be summarized descriptively by treatment groups over the five levels using counts and percentages for the ITT population. Index-based values ("utilities") will be summarized descriptively by treatment group. An ANCOVA model will be used to compare the difference in the devices for change from baseline in utilities scores at each of scheduled post-operative visits. The model will have change in EQ-5D-5L utility score as the response variable and treatment, baseline EQ-5D-5L utility score, site, treatment-by-site interaction, lesion grade and treatment-by-lesion grade interaction as independent variables. These interaction terms will be evaluated separately and if deemed significant then will be included in the full model. If the interaction terms are non-significant at the 0.05 level then these will be dropped from the

statistical model. The analysis will be presented as n, LS means, SE, and associated two sided 95% CI for each treatment and for treatment difference.

The EQ-VAS score will be summarized using descriptive statistics by treatment group at each study visit. An ANCOVA model will be used to compare the difference in the devices for change from baseline in the EQ-VAS scores at each of scheduled post-operative visits. The model will have change in EQ-VAS score as the response variable and treatment, baseline EQ-VAS score, site, treatment-by-site interaction, lesion grade and treatment-by-lesion grade interaction as independent variables. These interaction terms will be evaluated separately and if deemed significant then will be included in the full model. If the interaction terms are non-significant at the 0.05 level then these will be dropped from the statistical model. The analysis will be presented as n, LS means, SE, and associated two sided 95% CI for each treatment and for treatment difference.

8.2.6.8 Subject Satisfaction

Subject satisfaction scores will be summarized using descriptive statistics by treatment groups Week 52 and Week 104. Mean scores of two treatment groups will be compared using t-test.

8.2.6.9 Imaging Endpoint

The grade at each of the scheduled post-operative visits (Day 10, Week 52 and Week 104) will be summarized descriptively by treatment group for

- Cartilage Signal and Morphology (MOAKS)
- Subarticular Bone Marrow Abnormality (MOAKS)
- Subarticular Cysts (MOAKS)
- Subarticular Bone Attrition (WORMS)
- Marginal Osteophytes (WORMS)

The chi-square test will be used to compare the the two treatment groups.

Synovial thickening and joint effusion and presence/absence of loose bodies in joint space will be summarized descriptively by treatment groups and chi-square test will be used for treatment comparisons.

8.2.6.10 ICRS Lesion Grades

ICRS lesion grading will be summarized descriptively by treatment groups at Day 10, Week 52 and Week 104. The chi-square test will be used for treatment comparisons.

8.6.3 Subgroup Analyses

Exploratory subgroup analyses will be performed on the primary effectiveness variable to determine if the effectiveness is consistent across certain subgroup levels. Trial comparison for the primary endpoint will be carried out in each subgroup provided the sample size in each of the subgroup is at least 10% of the total population. The analysis method will be the same as described in Section 8.6.1. The subgroups to be examined will be as described in section 7.12.

8.6.4 Exploratory Analyses

All secondary endpoint analyses will be exploratory in nature.

8.7 Analysis of Safety

All safety analyses will be performed using the safety population.

8.7.1 Extent of Exposure and Compliance to Study Treatment

See Section 6.6.1

8.7.2 Adverse Events

Adverse events will be summarized by system organ class and preferred term. A subject will only be counted once per system organ class and per preferred term within a treatment.

An overall summary of adverse events will be presented and will include:

- at least one TEAE,
- any severe TEAE,
- any treatment related TEAE/ADE,
- any serious TEAE,
- any TEAE resulting in death, and
- any TEAE resulting in study discontinuation.

The following summary tables will be presented by SOC and PT:

- All TEAEs,
- All TEAEs by severity,
- All TEAEs/ADEs by relationship to study treatment,
- All SAEs

The following listings will be presented by subject:

- All adverse events,
- SAEs,
- AEs leading to discontinuation from the study, and
- Deaths

For the summary of adverse events by severity, if a subject has multiple events occurring in the same body system or same preferred term, then the event with the maximum severity will be counted. For treatment related TEAEs, if a subject has multiple events occurring in the same body system or same preferred term, the event with the highest association will be summarized. Listings will be presented by subject for all AEs as well as for SAEs, for adverse events resulting in death and adverse events leading to discontinuation from the study.

The AE onset date will be compared to the study intervention administration date in order to determine if the AE occurred prior to the administration of the study intervention. Any symptoms or AEs recorded before administration of the study interventions will only be presented in listings.

8.7.3 Clinical Laboratory Evaluations

Not applicable.

8.7.4 Other Observations Related to Safety

8.7.4.1 Physical Assessment of Knee

For each post-operative visit, the shift in grading from baseline for warmth, swelling, skin changes and evidence of infection will be summarized as counts and percentage of subjects in each treatment group.

8.7.4.2 Clinical Follow-up

For each post-operative visit, information whether subject retuned to work and date of returning work is collected. A listing to this effect will be produced.

9 COMPUTER SOFTWARE

All analyses will be performed by Chiltern International, Inc. using Version 9.1.3 or later of SAS® software. All summary tables and data listings will be prepared utilizing SAS® software.

For continuous variables, descriptive statistics (number of subjects [n], mean, standard deviation [SD], median, minimum, and maximum) will be generated. For discrete/categorical variables, the number and proportion of subjects will be generated. The standard operating procedures (SOPs) of Chiltern International, Inc. will be followed in the creation and quality control of all data displays and analyses.

REFERENCES

1. Figueroa D, Calvo R, Vaisman A, Carrasco MA, Moraga C, Delgado I. Knee chondral lesions: incidence and correlation between arthroscopic and magnetic resonance findings. *Arthroscopy* 2007; 23(3):312-315.
2. Hjelle K, Solheim E, Strand T, Muri R, Brittberg M. Articular cartilage defects in 1,000 knee arthroscopies. *Arthroscopy* 2002; 18(7):730-734.
3. Widuchowski W, Widuchowski J, Trzaska T. Articular cartilage defects: study of 25,124 knee arthroscopies. *Knee* 2007; 14(3):177-182.
4. Sgaglione NA, Miniaci A, Gillogly SD, Carter TR. Update on advanced surgical techniques in the treatment of traumatic focal articular cartilage lesions in the knee. *Arthroscopy* 2002; 18(2 Suppl 1):9-32.
5. Brittberg M, Winalski CS. Evaluation of cartilage injuries and repair. *J Bone Joint Surg Am* 2003; 85-A Suppl 2:58-69.
6. Voloshin I, Morse KR, Allred CD, Bissell SA, Maloney MD, DeHaven KE. Arthroscopic evaluation of radiofrequency chondroplasty of the knee. *Am J Sports Med* 2007; 35(10):1702-1707.
7. Owens BD, Stickles BJ, Balikian P, Busconi BD. Prospective analysis of radiofrequency versus mechanical debridement of isolated patellar chondral lesions. *Arthroscopy* 2002; 18(2):151-155.
8. Barber FA, Iwasko NG. Treatment of grade III femoral chondral lesions: mechanical chondroplasty versus monopolar radiofrequency probe. *Arthroscopy* 2006; 22(12):1312-1317.
9. Spahn G, Klinger HM, Mückley T, Hofmann GO. Four-year results from a randomized controlled study of knee chondroplasty with concomitant medial meniscectomy: mechanical debridement versus radiofrequency chondroplasty. *Arthroscopy* 2010; 26(9 Suppl):S73-80.
10. Roos M. Ewa, Lohamander Stephan L, The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. BioMed Central 2003; 1:64, page 5 of 8

11. Chen JYH, Demets DL, and Lan KKG ‘ Increasing the Sample Size when unblinded interim results is promising,’ Statistics in Medicine 23 (2004) 1023- 1038.
12. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther. 1998 Aug;28(2):88-96.
13. The 2012 User's Guide to Knee injury and Osteoarthritis Outcome Score. Version 1.1. (http://www.koos.nu/KOOSGuide2003.pdf)
14. Heeren, T. and D’Agostino, R. (1987). Robustness of the two-independent samples t-test when applied to ordinal scaled data. *Statistics in Medicine* **6**, 79-90.
15. Avron Spiro III, William H. Rogers, Shirley Qian, Lewis E. Kazis: (2004). Imputing Physical and Mental Summary Scores (PCS and MCS) for the Veterans SF-12 Health Survey in the Context of Missing Data, *Technical Report* by The Health Outcomes Technologies Program, Health Services Department, Boston University School of Public Health, Boston, MA and The Institute for Health Outcomes and Policy, Center for Health Quality, Outcomes and conomic Research, Veterans Affairs Medical Center, Bedford, MA
16. Ware, John, Mark Kosinski and Susan Keller. 1995. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales, 2nd edition. Boston: The Health Institute, New England Medical Center.

APPENDICES
APPENDIX 1: VARIABLE DEFINITIONS

Subject Age:

Age is computed with reference to inform consent date as follows:

Age (yrs) = Integer part of ((Date of Informed consent – Date of Birth)/365.25)

Body Mass Index (BMI (kg/m^2))

BMI is calculated as:

BMI (kg/m^2) = weight (kg) / ((height [m])^2).

APPENDIX 2: STATISTICAL ANALYSIS AND PROGRAMMING DETAILS

Scoring for International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form

The International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form was developed to detect improvement or deterioration in symptoms, function, and sports activities due to knee impairment. The IKDC Subjective Knee Evaluation Form is comprised of three domains: 1) symptoms, including pain, stiffness, swelling, locking/catching, and giving way; 2) sports and daily activities; and 3) current knee function and knee function prior to knee injury (not included in the total score). In all there are 18 questions: 7 questions in the symptoms domain, 10 questions in the sport and daily activities domain (1 question on sport participation and 9 questions on daily activities), and 2 questions relating to the current knee function domain. However, the question regarding knee function prior to knee injury in current knee function domain is not included in computation of the total score.

Response options vary for each question. Question 6 dichotomizes response into yes/no; Questions 1, 4, 5, 7, 8, and 9 use 5-point likert scales; and Question 2, 3, and 10 use 11-point numerical rating scales.

The IKDC Subjective Knee Evaluation Form is scored by summing the scores for the individual questions and then transforming the score to a scale that ranges from 0 to 100. The response to Question10(i.e., "Function Prior to Knee Injury") is not included in the overall score. The steps to score the IKDC Subjective Knee Evaluation Form are as follows:

1. Assign a score to the individual's response for each question, such that lowest score represents the lowest level of function or highest level of symptoms.
2. Calculate the raw score by summing the responses to all questions with the exception of the response to Question 10 "Function Prior to Your Knee Injury"
3. Transform the raw score to a 0 to 100 scale score as follows:

$$IKDC\ score = \left[\frac{\text{Raw Score - Lowest Possible Score}}{\text{Range of Scores}} \right] \times 100$$

The transformed score is interpreted as a measure of function such that higher scores represent higher levels of function and lower levels of symptoms. A score of 100 is interpreted to mean no limitation with activities of daily living or sports activities and the absence of symptoms.

The IKDC Subjective Knee Form score will be calculated if missing data occurs, provided that responses to at least 90% of the questions (i.e. responses have been provided for at least 16 questions) are available. To calculate the raw IKDC Subjective Knee Form score when there are missing data, the average score of the questions that have been answered are substituted for the missing score(s). Once the raw IKDC Subjective Knee Form score has been calculated, it is transformed to the IKDC Subjective Knee Form score as described above.

Scoring for SF-12 v2®

SF-12 v2® includes 8 concepts commonly represented in health surveys: physical functioning, role functioning physical, bodily pain, general health, vitality, social functioning, role functioning emotional, and mental health. Results are expressed as the Physical Component Summary (PCS) and the Mental Component Summary (MCS).

Scoring the SF-12 v2® will be performed via the Health Outcomes Scoring Software 4.5 offered by QualityMetric. The software uses all the 12 items to produce scores for the PCS-12 and the MCS-12 and applies a norm-based scoring algorithm empirically derived from the data of a US general population survey. It has been recommended that the US-derived summary scores, that assume a mean of 50 and a standard deviation (SD) of 10, be used in order to facilitate cross-cultural comparison of results. In theory the possible scores for the PCS-12 and the MCS-12 could be ranged from 0 (the worst) to 100 (the best).

The algorithm used to score the data is detailed below:

First, the data are checked for out-of range values. Out of range values are any values that are outside the range of acceptable item responses values for the SF-12v2® Health Survey. Out-of-range values will be converted to missing values. Next, four items (GH01, BP02, MH03, VT02) are reversed scored. Reverse scoring of these items is required so that a higher item response value indicates better health for all SF-12v2® Health Survey items and summary measures.

Step 1: Standardization of the SF-12v2® Health Survey Scales

The first step in scoring the component summary measures consists of standardizing each SF-12v2® Health Survey scale using a z-score transformation. The z-score is computed by subtracting the mean 0-100 score observed in the 1998 general U.S. population from each SF-12v2® Health Survey scale score (0-100) scale and dividing the difference by the corresponding scale standard deviation observed in the 1998 general U.S. population.

Step 2: Aggregation of the Scale Scores

After a z-score has been computed for each SF-12v2® Health Survey scale, the second step involves computation of aggregate scores for the physical and mental summaries using weights (factor score coefficients) from the 1990 general U.S. An aggregate physical score is computed by multiplying the z-score of each SF-12v2® Health Survey scale by its associated physical factor score coefficient and summing the eight products. If any of the scale scores are missing, then the aggregate physical score is not computed. An aggregate mental score is computed by multiplying the z-score of each SF-12v2® Health Survey scale by its associated mental factor score coefficient and summing the products. If any of the scale scores are missing, then the aggregate mental score is not computed.

Step 3: Transformation of Summary Scores

The third step involves transforming the aggregate physical and mental summary scores to the T-score Based (50, 10) scoring. This is done by multiplying each aggregate summary score obtained from Step 2 by 10 and adding the resulting product to 50.

Source: Ware JE, Kosinski M, Turner-Bowker DM, Gandek B: How to score version 2 of the SF-12 HEALTH Survey. Lincoln, RI: Quality Metric Incorporated; 2002.

Scoring for EQ-5D-5L

EQ-5D is a standardized measure of health status developed by the European Quality of Life Scale (EuroQol) Group in order to provide a simple, generic measure of health for clinical and economic appraisal[1]. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care as well as in population health surveys. EQ-5D is designed for self-completion by respondents and is ideally suited for use in postal surveys, in clinics, and in face-to-face interviews.

The EQ-5D-5L descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The responses for 5 dimensions can be combined in a 5-digit number describing the respondent's health state. Index-based values ('utilities') are a major feature of the EQ-5D instrument and can be computed using the Crosswalk value sets for the EQ-5D-5L that can be downloaded from: http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html. The EQ-VAS score scale represents the subject's health status (good or bad) on the day of the assessment. The scale ranges from 0 to 100 in which 100 represents the best health and 0 represents the worst health. Subjects will mark with an X on the scale to indicate the status of their health. Once the subject have marked the scale the number noted on the scale will be recorded.

A scientific publication describing the methodology is expected to be published in the Value in Health Journal.

[1]EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208

ANCOVA Model Specification

The code below will be used to carry out the analyses as detailed in the specific sections pertaining to each endpoint.

```
Proc MIXED   Data = <Input data set> Method = reml;
     Class  TRT  SITE  ;
      Model  CHG = BASE SITE TRT  TRT*SITE/DDFM = KR S;
      LSMEANS  TRT  TRT* SITE;
 Run;
```

MMRM Model Specification

The code below will be used to carry out the analyses as detailed in the specific sections pertaining to each endpoint.

```
PROC MIXED DATA=<Input data set>;
        WHERE PARAMCD=<Value>;
        Class SUBJID TRT AVISIT;
        Model CHG=TRT AVISIT BASE TRT*AVISIT SITE TRT*SITE LESION
        LESION*TRT /SOLUTION DDFM=KENWARDROGER OUT=PRED;
        REPEATED AVISIT/SUB=SUBJID TYPE=UN;
        LSMEANS TRT*AVISIT TRT*SITE/SLICE=AVISIT DIFF;
        ODS OUTPUT FitStatistics=FitUn(rename=(value=UN))
                        FitStatistics=FitUNp
                        Dimensions=ParmUN(rename=(value=NumUN));
RUN;

PROC PRINT DATA=FitUN;
RUN;

GOOPTIONS RESET=ALL;
PROC GPLOT DATA=PRED;
        PLOT RESID*PRED;
RUN;
QUIT;
```

APPENDIX 3: PRESENTATION CONVENTIONS

Categories that are empty across all treatment groups will not be displayed in the respective tables.

Percentages are calculated based on all non-missing data unless noted otherwise.

APPENDIX 4: TABLE SHELLS SPECIFICATIONS

Table Number	**Table Title**	**Population**
Table 14.1.1.1	Summary of subjects randomized by site and by treatment group	Intent-to-treat population
Table 14.1.1.2	Summary of Subject Disposition	Enrolled Population
Table 14.1.1.3	Summary of Protocol Deviations	Enrolled Population
Table 14.1.2	Summary of Demographic Characteristics	Safety Population
Table 14.1.3.1	Summary of Symptom History Requiring Treatment	Safety Population
Table 14.1.3.2	Summary of Pre-surgery Imaging	Safety Population
Table 14.1.3.3	Summary of Medical History	Safety Population
Table 14.1.3.4	Summary of International Cartilage Research Society Knee History Registration - Previous Surgery	Safety Population
Table 14.1.3.5	Summary of Baseline Knee Evaluation	Safety Population
Table 14.1.3.6	Summary of International Cartilage Research Society Lesion Classification at Baseline	Safety Population
Table 14.1.4	Summary of Concomitant Medication	Safety Population
Table 14.1.5	Exposure to Study Treatment	Safety Population
Table 14.1.4	Summary of Concomitant Medication	Safety Population
Table 14.2.1.1.1	Primary Endpoint - Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Score at Week 52 using LOCF	Intent-to-Treat Population
Table 14.2.1.1.2	Primary Endpoint - Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Score at Week 52 using MMRM	Intent-to-Treat Population

Table 14.2.1.2.1	Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 using LOCF	Per-Protocol Population
Table 14.2.1.2.2	Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 using MMRM	Per-Protocol Population
Table 14.2.1.3	Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Score at Week 52 (Sensitivity Analysis – With No Imputation for Missing Score)	Intent-to-Treat Population
Table 14.2.1.4	Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Score at Week 52 (Sensitivity Analysis – Missing Value Imputed by Worst Score)	Intent-to-Treat Population
Table 14.2.1.5	Subgroup Analysis: Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 by BMI	Intent-to-Treat Population
Table 14.2.1.6	Subgroup Analysis: Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 by Age Group	Intent-to-Treat Population
Table 14.2.1.7	Subgroup Analysis: Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 by Total Pre-Debridement Lesion Size	Intent-to-Treat Population
Table 14.2.1.8	Subgroup Analysis: Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 by Gender	Intent-to-Treat Population
Table 14.2.1.9	Subgroup Analysis: Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 by Nicotine Use	Intent-to-Treat Population
Table 14.2.1.10	Subgroup Analysis: Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 by Duration since Onset of Symptoms for Index Knee	Intent-to-Treat Population
Table 14.2.1.11	Subgroup Analysis: Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 by Knee Alignment	Intent-to-Treat Population
Table 14.2.1.12	Subgroup Analysis: Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Scores at Week 52 by Narrowest Width of Meniscal Rim Post-Operatively	Intent-to-Treat Population
Table 14.2.1.13	Secondary Endpoint - Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Score Subscales at scheduled post-operative visits using LOCF	Intent-to-Treat Population
Table 14.2.1.14	Secondary Endpoint - Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Score Subscales at scheduled post-operative visits using MMRM	Intent-to-Treat Population
Table 14.2.2.1	Summary of Change from Baseline in International Knee Documentation Committee Subjective Knee Evaluation Form Score by Visit	Intent-to-Treat Population
Table 14.2.2.2	Analysis of Change from Baseline in International Knee Documentation Committee Score	Intent-to-Treat Population

Table 14.2.3.1	Summary of International Knee Documentation Committee Knee Examination – General Laxity, Alignment, Patella Position and Patella Subluxation	Intent-to-Treat Population
Table 14.2.3.2	Summary of 2000 International Knee Documentation Committee Knee Examination -Range of Motion	Intent-to-Treat Population
Table 14.2.3.3	Shifts in International Knee Documentation Committee Knee Examination Scores from Baseline	Intent-to-Treat Population
Table 14.2.4.1	Summary of Change from Baseline in Normalized Knee and Osteoarthritis Outcomes Score by Visit	Intent-to-Treat Population
Table 14.2.4.2	Analysis of Change from Baseline in Knee and Osteoarthritis Outcomes Score by visit	Intent-to-Treat Population
Table 14.2.5.1	Summary of Visual Analogue Scale Knee Pain Score by Visit	Intent-to-Treat Population
Table 14.2.5.2	Analysis of Change from Baseline in Visual Analogue Knee Pain Scale Score	Intent-to-Treat Population
Table 14.2.6.1.1	Summary of SF-12 physical component scale score by Visit	Intent-to-Treat Population
Table 14.2.6.1.2	Analysis of Change from Baseline in SF-12 PCS Score	Intent-to-Treat Population
Table 14.2.6.2.1	Summary of SF-12 Mental Component Scale Score by Visit	Intent-to-Treat Population
Table 14.2.6.2.2	Analysis of Change from Baseline in SF-12 MCS Score	Intent-to-Treat Population
Table 14.2.7.1	EQ-5D-5L by Visit	Intent-to-Treat Population
Table 14.2.7.2	Summary of Change in EQ-5D-5L Utility Score by Visit	Intent-to-Treat Population
Table 14.2.7.3	Analysis of Change from Baseline in EQ-5D-5L Utility Score	Intent-to-Treat Population
Table 14.2.8	Summary of Subject Satisfaction with Study Treatment for Knee Pain	Intent-to-Treat Population
Table 14.2.9.1	Summary of Change from Baseline in Assessment: Cartilage Signal & Morphology by Visit	Intent-to-Treat Population
Table 14.2.9.2	Summary of Change from Baseline in Assessment: Subarticular Bone Marrow Abnormality by Visit	Intent-to-Treat Population
Table 14.2.9.3	Summary of Change from Baseline in Assessment: Subarticular Cysts by Visit	Intent-to-Treat Population
Table 14.2.9.4	Summary of Change from Baseline in Assessment: Subarticular Bone Attrition by Visit	Intent-to-Treat Population
Table 14.2.9.5	Summary of Change from Baseline in Assessment: Marginal Osteophytes by Visit	Intent-to-Treat Population
Table 14.2.9.6	Summary of Change from Baseline in Assessment: Synovial Thickening & Joint Effusion by Visit	Intent-to-Treat Population
Table 14.2.9.7	Summary of Change from Baseline in Assessment: Loose Bodies by Visit	Intent-to-Treat Population

Table 14.2.9.8	Summary of Change from Baseline in Assessment: International Cartilage Repair Score (ICRS) by Visit	Intent-to-Treat Population
Table 14.3.1.1	Summary of Overall Treatment Emergent Adverse Events	Safety Population
Table 14.3.1.2	Summary of Treatment Emergent Adverse Events by System Organ Class and Preferred Term	Safety Population
Table 14.3.1.3	Summary of Treatment Emergent Adverse Events by System Organ Class, Preferred Term and Severity	Safety Population
Table 14.3.1.4	Summary of Treatment Emergent Adverse Events By System Organ Class, Preferred Term And Relationship To Study Device	Safety Population
Table 14.3.2.1	Summary of Serious Treatment Emergent Adverse Events by System Organ Class and Preferred Term	Safety Population
Table 14.3.6.1	Summary of Shift from Baseline in the Physical Assessment of the Index Knee	Safety Population