STATISTICAL ANALYIS PLAN (SAP)

Study Details:

ST Number	1028				
Protocol Reference	16-NPFS-11	**Protocol Version**	2.0	**Protocol Date**	05-JUN-2017
Study Title	A Retrospective Study of the NAVIO Robotic-assisted Surgical System				

SAP Version Control:

SAP Status	Final		
SAP Version Number	1.0	**SAP Date**	29-JAN-2018
Previous Version Number(s), Date(s)			

Signature Page:

Author(s) Name and job title(s):	Babajide Olayinka, Senior Biostatistician		
Role	**Approver Name and Job Title**	**Approver Signature**	**Date Signed (DD/MMM/YYYY)**
Head of Global Biostatistics/ Statistician	Alan Rossington		05/Feb/2018
Statistician	Babajide Olayinka		05/FEB/2018

Table of Contents

1. List of Abbreviations

Abbreviation	Definition
ADE	Adverse Device Effect(s)
AE	Adverse Event(s)
AEMB	Adverse Event Monitoring Board
CI	Confidence Interval
CRF	Case Report Form(s)
EC	Ethics Committee
FAS	Full Analysis Set
IRB	Institutional Review Board
LL	Lower Limit
NA or N/A	Not Applicable
KSS	Knee Society Score
MCS	Mental Component Score
N (or n)	Total Sample Size (or subgroup sample size)
PCS	Physical Component Score
QoL	Quality of Life
SADE	Serious Adverse Device Effect
SAE	Serious Adverse Event(s)
SAP	Statistical Analysis Plan
SF-12	Short Form 12 Health Survey Standard
TFL	Tables, Figures, and Listings
UADE	Unanticipated Adverse Device Effect(s)
UKR	Unicondylar Knee Replacement
VR-12	Veterans Rand 12 Item

2. Introduction

This statistical analysis plan (SAP) is written to specify the statistical analyses that would be carried out for study protocol 16-NPFS-11 in greater detail than the protocol. The statistical details provided herein would serve as guideline for the programming that would generate this study's tables, figures, and listings (TFLs).

The SAP is based on version 2.0 of the protocol dated 05 June 2017 and version 3.0 of the study's accompanying case report forms (CRFs).

3. Study Design

The NAVIO system is designed to assist surgeons in providing software-defined spatial boundaries for orientation and reference information to anatomical structures during orthopedic surgical procedures.

This study is a retrospective, multi-center, cohort study to be carried out on approximately 128 subjects enrolled at up to 10 investigational sites. These subjects would have previously undergone the NAVIO System-assisted Unicondylar Knee Replacement (UKR) for restoring either compartment of the knee that has been affected by any of the following conditions: (a) Non-inflammatory degenerative joint disease including osteoarthritis, traumatic arthritis, or avascular necrosis (b) Correction of functional deformity and (c) Treatment of fractures that are unmanageable using other techniques.

Subjects enrolled would be followed-up for evaluation of safety and effectiveness of the device. Subjects would retrospectively be evaluated pre-operatively and operatively and prospectively at subsequent available post-operative follow-up visits for up to 2 years (24 months) or beyond. In order to evaluate the effectiveness of the NAVIO system-assisted investigational device, information on the revision rate (or survivorship) from a currently used device from an appropriate registry would be used as historical control. The historical control's revision (or survivorship) rate which was obtained from an orthopedic registry is specified in future sections of the SAP.

The study schematic displayed in Table 1 provides a succinct overview of the parameters evaluated within the study design.

Table 1: Study schematic.

	Retrospective Data Collection[1]				**Prospective Data Collection[2]**
	Screening and Enrollment	Pre-operative	Operative / Discharge	Follow-up	2-year (-3months)or greater postoperative follow-up
Informed Consent	√[3]				
Inclusion/Exclusion	√				
Demographics/ Medical History		√			
Operative and Discharge Data Collection			√		
Implant Status[4]				√	√
Knee Society Score (KSS)		√		√	√
VR-12		√		√	√
Radiographic Assessment		√	√	√	√
Adverse Event Assessment			√	√	√
End of Study/Exit				√*	√

[1]Given the retrospective-prospective study design, data will be collected to the extent it is available.

[2]Visit may occur in an office or remotely. If completed remotely (e.g. telephone call, mailing of subject questionnaires and/or any other means), secondary endpoint data will be collected to the extent possible.

[3]Informed consent/waiver of informed consent from the IRB/EC must be obtained prior to retrospective and prospective data collection.

[4]Subjects who have undergone a revision procedure of the study knee will be considered terminated from the study from the date of the revision. Study related data will not be collected following the date of the revision.

*if applicable

4. Study Objectives

4.1 Primary Objective

The primary objective of this study is to evaluate the 2-year safety and effectiveness of the NAVIO System's assisted implanted UKR in subjects aged ≥ 18 years.

5. Study Endpoints

5.1 Primary Endpoint

The NAVIO System assisted UKR implant survivorship (absence of device revision) at 2+ years (96 weeks) post-surgical implantation.

5.2 Secondary Endpoints

Table 2 displays the visit windows for evaluation time points that would be used for summarizing the KSS and the VR-12 endpoints. Based on this, for situations in which two or more KSS or VR-12 observations present at a given visit, the observation closest to the upper range of the window would be selected for summary purposes for the visit.

Table 2. Summary of visit windows used for summarizing KSS and VR-12 endpoints

Month	Range of Days (windows) for Month Membership for follow-up post-surgical implantation
6 Months	≤ 182 days
12 Months	>182 to ≤ 365 days
18 Months	>365 to ≤ 548 days
24 Months	>548 to ≤ 730 days
30 Months	>730 to ≤ 913 days
36 Months	>913 to ≤ 1095 days

The secondary endpoints are as follows:

- Two-year (24 months) postoperative visit Knee Society Score (KSS).

- Two year (24 months) post-operative visit VR-12 domain and sub-domain scores. These scores would encompass the individual domains of physical component score (PCS), mental component score (MCS) as well as the sub-domain scores within the PCS (physical functioning, role physical, bodily pain, and general health) and those within the MCS (vitality, social functioning, role emotional, and mental health).
- Anterior/Posterior and lateral radiographic assessments which include but not only restricted to the following:
    - Radiographic findings
    - Component orientation
    - Radiolucencies
    - Migration
    - Osteolysis
    - Stress shielding
    - Subsidence

5.3 Exploratory Endpoints

N/A.

5.4 Safety Endpoints

Safety endpoints are as follows:

- Incidence of AEs,
- Incidence of AEs by severity,
- Incidence of AEs by relationship to knee implant (study device),
- Incidence of AEs by relationship to NAVIO System,
- Incidence of AEs that led to discontinuation,
- Incidence of AEs by outcome,
- Incidence of serious adverse events (SAEs),
- Incidence of adverse device effects (ADE),
- Incidence of serious adverse device effects (SADEs),
- Incidence of unanticipated adverse device effects (UADEs),
- AEs of special interest,
- Listing of revisions and details of components removed during revision.

6. Statistical Considerations

6.1 Determination of Sample Size

The sample size obtained for study is based on the non-inferiority of the percent of revision-free survival expected to be obtained at 2 years in the study, π to one obtained from existing literature, π_0. The Australian Orthopaedic Association National Joint Replacement Registry 2015 Annual Report (1) indicates that the revision-free survival at 2-years is 95.7%.

Using a non-inferiority margin, δ of 7%, a sample size of 115 subjects would provide at least 80% power at the 5% significance level using the exact binomial method. To allow for a 10% attrition and drop-out rate, a total of approximately 128 subjects would be enrolled into the study.

6.2 Randomisation

N/A.

6.3 Interim Analysis

N/A

7. Statistical Analysis

7.1 General

Smith & Nephew's Global Biostatistics group would carry out the statistical analysis for this study. Unless otherwise stated, all statistical significance tests and the hypothesis test would be two-sided, performed at the 5% significance level (i.e. $\alpha = 0.05$). Where appropriate, the resulting p-values would be quoted and if applicable, the corresponding 95% two-sided confidence intervals (CIs) would be generated. All p-values would be rounded to three decimal places; p-values less than 0.001 would be presented as '<0.001' in all tables.

For data summaries, categorical and ordinal variables would be summarized using frequency and percent. Continuous variables would be summarized using characteristics such as number of observations, mean, median, standard deviation, minimum and maximum values. All analyses would be performed in SAS version 9.3 or a more recent version.

All demographic and baseline variables as well as effectiveness endpoints would be summarized using both the safety and per protocol populations unless otherwise specified. For the primary endpoint, the safety population would be used as the primary analysis population while the per protocol population would be used for sensitivity analysis of the analysis of the primary endpoint. The secondary endpoints would be summarized using only the safety population. All safety endpoints would be summarized using the safety population.

All statistical comparisons of the data would provide the test statistic used as well as its distributional assumptions. In general, an equivalent non-parametric test would be used as alternative to the parametric where means are markedly different from medians thus increasing the likelihood that violations of normality assumptions would occur.

Subject data listings would be created to provide supporting information for pertinent tables.

7.2 Analysis Populations

Safety Population

This is defined as the population of subjects who receive the investigational study device.

Per Protocol Population

The per protocol population is the subset of subjects in the safety population who are compliant with the protocol by satisfying the following criteria: (a) meeting all inclusion/exclusion criteria (b) attend follow-up assessments and do not have significant protocol deviations.

Subjects who have a device revision would retain membership in the per protocol population unless any of the above two specifications is met. The final decision on which deviations would be classified as significant would be finalized prior to database lock.

7.3 Handling of Missing, Incomplete and Repeat Data

Not all subjects enrolled into this study would complete the study. Subjects who die or discontinue without a revision would have their data censored on their date of death/discontinuation. Subjects who do not have a revision by the end of the study would have their data censored at their last known study visit date.

7.4 Derived Data

VR-12 domains/subscales

Table 3 displays the values from the scores from the original VR-12 instrument, an analogue of the SF-12 Quality of Life (QoL) (2) instrument, based on potential responses to

the questions. These values are transformed into scores for use in the different conceptual sub-domains (Table 3) that make up the PCS MCS domains.

The transformation of raw scale scores provided in Table 3 to transformed scores in Table 4 (0-100 transformed range) utilizes the following algorithm:

Transformed scale score =

{(Actual raw score – Lowest possible raw score)/Possible raw score range} x100.

Table 3: VR-12 scoring for individual questions

Item (Question) number	Response Choices	Likely scores on original response	Final Item Value *
Q1	Excellent	1	5.0
	Very Good	2	4.4
	Good	3	3.4
	Fair	4	2.0
	Poor	5	1.0
Q2a, Q2b	Yes, limited a lot	1	1
	Yes, limited a little	2	2
	No, not limited at all	3	3
Q3a, Q3b, Q4a, Q4b, Q6c, Q7	All of the time	1	1
	Most of the time	2	2
	Some of the time	3	3
	A little of the time	4	4
	None of the time	5	5
Q5	Not at all	1	5
	A little bit	2	4
	Moderately	3	3
	Quite a bit	4	2
	Extremely	5	1
Q6a, Q6b	All of the time	1	5
	Most of the time	2	4
	Some of the time	3	3
	A little of the time	4	2
	None of the time	5	1

* Note, the higher the score on the final item value, the more favorable the response.

Table 4: VR-12 Sub-domains of the PCS[1] and MCS[2].

Scale/Domain	Sum of Final Items from Table 3	Lowest and Highest Possible raw scores	Possible raw score range
Physical Functioning[1]	**Q2a, Q2b**	(2, 6)	4
Role Physical[1]	**Q3a, Q3b**	(2, 10)	8
Bodily Pain[1]	**Q5**	(1, 5)	4
General health[1]	**Q1**	(1, 5)	4
Vitality[2]	**Q6b**	(1, 5)	4
Social Functioning[2]	**Q7**	(1, 5)	4
Role Emotional[2]	**Q4a, Q4b**	(2, 10)	8
Mental Health[2]	**Q6a, Q6c**	(2, 10)	8

[1] Physical Component Score.
[2] Mental Component Score.

7.5 Baseline Data

The following demographic and baseline variables whose information is available either at the pre-operative or the operative visit would be summarized using descriptive characteristics for continuous[1] or categorical[2] data as follows:

- Age[1] (in years),
- Sex[2] (males or females),
- Race[2],
- Weight[1] (in kg),
- Height[1] (in cm),
- Body Mass Index[1],
- Primary diagnosis[2],
- Smoking status at time of surgery[2]
- Duration of surgery (in hours)[1],
- Type of surgical approach[2],
- Type of knee system implanted (Smith & Nephew Inc. Unicondylar Knee System or Other Smith & Nephew Inc. Unicondylar System)[2],
- Type of UKR used (medial or lateral)[2],
- Intra-operative complications[2] and,
- Blood loss prevention method[2].

Age would further be summarized categorically by stratifying into: < 60 years versus ≥ 60 years. BMI would also be further summarized categorically by stratifying into: < 30 kg/m^2 versus ≥ 30 kg/m^2.

7.6 Disposition of Patients

- Subject disposition would be summarized as frequency (n) and percentage (%) of subjects who complete and who do not complete the study. For those who do not complete the study, the primary reason for discontinuation would additionally be summarized.

- Subject accountability information would also be presented. This would be summarized by visit for the number of subjects that are theoretically due, subject deaths, subjects who had a revision and, subjects who terminated from the study for any reason. Additional information that would be summarized on the subject accountability table would include but not limited to the number of subjects presenting at each visit, number of subjects expected at each visit.

7.7 Protocol Deviations

A listing of deviations encountered during the study would be listed. Deviations that would lead to the exclusion of subjects in the per protocol population would be appropriately identified and flagged.

7.8 Measurement of Treatment Compliance

N/A.

7.9 Multiplicity

N/A.

7.10 Analysis of Primary Endpoint

For the primary endpoint, cumulative proportion (or percentage) of subjects with implant survivorship at 2+ years (96 weeks), the following is the hypothesis would be performed:

H_0 (null): $S(t) = \pi \leq (\pi_0 - \delta)$ versus

H_a (alternate): $S(t) = \pi > (\pi_0 - \delta)$,

where π = expected percentage of subjects with implant survivorship at 2 years, π_0 = historical control implant survivorship (95.7%) and δ = margin of non-inferiority (7%). Because not all subjects would be fully evaluated through study completion, the Kaplan-Meier product limit survival estimate would be used to estimate implant survivorship, S(t) where S(t) is the proportion of subjects at time t, without a device revision. Time to implant revision would be the endpoint of interest in the determination of implant survivorship. Cumulative proportions of subjects with implant survivorship at 2 years would be calculated and displayed accordingly. Subjects who complete the study without a revision would be censored at their last known date in the study. A subject who prematurely discontinues from study as a result of death or for any other reason would be censored at the date this event occurred. A subject who is lost to follow-up would be censored at the last known contact date. The Kaplan-Meier estimate for implant survivorship would be presented with the corresponding two-sided 95% Confidence Intervals (CIs). This study's hypothesis test would only be carried out at the end of the study at 2+ years (96 weeks). From this test, if the lower limit (LL) of the 95% two-sided CI for $S(t = 2 \text{ years}) > (\pi_0 - \delta) = 88.7\%$, then H_0 will be rejected in favor of H_a (i.e. demonstration of non-inferiority of the investigational device) otherwise H_0 would not be rejected. Kaplan-Meier estimates for implant survivorship would also be non-inferentially estimated at the intermediate time points of 6 months (24 weeks), 1 year (48 weeks), and at 18 months (72 weeks) as well as subsequently at 30 months (120 weeks).

The primary endpoint would be analyzed using the safety population as the primary analysis population while the per protocol population would be used for sensitivity analysis.

7.11 Analysis of Secondary Endpoints

- The KSS endpoints would be evaluated by deriving the scores for objective knee indicators, patient satisfaction, patient expectation and functional activities. Each of these derived scores would be summarized using descriptive summary characteristics for continuous variables at the 2+ year (24 months) follow-up visit.

- The post-operative 24 month visit of the PCS domain, MCS domain as well as sub-domains of the PCS and MCS would be summarized using descriptive summary characteristics for continuous variables.

- The following are the descriptive summary of the anterior/posterior and lateral radiographic assessments by visit of the assessment:

    - Presence of heterotropic ossification, patellar subluxation, patellar dislocation would be summarized as count (n) and percent (%).

    - For component orientation, femur/tibia angle, tibia component, posterior slope of tibia tray would be summarized using descriptive characteristics for continuous variables.

    - For radiolucencies, overall summary as n and % of subjects who have radiolucent lines would be summarized. Furthermore, the distribution of the number of radiolucent lines and corresponding dimensions (in mm) by visit within the femoral zone, Tibial AP zone, and Tibial ML zone would also be categorically summarized.

    - For migration, categorical summary characteristics as n and % by visit would be used to summarize the presence of tibial migration/subsidence and femoral migration/subsidence.

- For osteolysis, categorical summary characteristics by visit using n and %.

All secondary endpoints would be summarized using the safety population. All data for the secondary endpoints would be analyzed as observed cases, i.e. there would be no imputation for missing values.

7.12 Analysis of Exploratory Endpoints

N/A.

7.13 Analysis of Safety Endpoints

- An overall summary AE table that would summarize as number (n) and percentages (%), the overall incidence according to subjects with at least one AE; subjects with at least one AE by worst severity (mild, moderate, or severe); subjects with at least one AE by worst outcome; subjects with an AE that led to study discontinuation); subjects with at least one AE by worst relatedness to investigational device; subjects with AEs by worst relationship to the NAVIO device; SAEs; ADEs; SADEs; UADEs and; AEs of special interest would be summarized. A Smith and Nephew Adverse Event Monitoring Board (AEMB) may be convened to review the list of and subsequently classify AEs into special interest AEs for additional AE summaries.

- Tables that summarize the incidence of AEs (each based on criteria previously specified in the overall summary table) would be generated. Each table would include the numbers (n) and percentages (%) of subjects with each AE as well as the cumulative events/episodes per AE.

All summaries and listings for the safety endpoints would use the safety population as the analysis population.

7.14 Other Data Summaries

- The type of analysis carried out on the primary endpoint, the Kaplan-Meier estimates would be summarized non-inferentially for the following subgroups using the safety population:
  - Age (< 60 years versus ≥ 60 years),
  - BMI (< 30 kg/m^2 versus ≥ 30 kg/m^2)

- Sex (males versus females), and
- Region (Europe versus USA).

For the subgroup analysis by region, investigational sites within Europe would be pooled into one virtual site and all investigational sites within the USA would be pooled into another virtual site.

- Other data summaries that would be performed using the safety population, are as follows:
  - Shift from baseline to 2 years (24 months) in overall physical health,
  - Shift from baseline to 2 years (24 months) in overall emotional problems.
- Subject discharge data would be summarized descriptively using the safety population for length of hospital stay prior to discharge (days) as a continuous variable and using categorical summary statistics (n and %) for place discharged to and primary ambulatory support prior to discharge.
- Listing of medical history.
- Summary of KSS and VR-12 derived endpoints at 6 months, 12 months, 18 months, 30 months and 36 months.

7.15 Changes in Analysis Methods Specified in the Protocol

- The safety population and the full analysis set (FAS) population as defined in the protocol are exactly the same. As a result thus, in this SAP, there was no further reference to the FAS population.
- The change from baseline to 24 month endpoints for KSS and VR-12 has been modified to be the summary characteristics for the endpoints at 24 months.

These modifications were implemented because all 27 subjects analyzed for the interim analyses had missing baseline KSS and VR-12 data.

8. References

1. Australian Orthopaedic Association National Joint Replacement Registry. Annual Report. Adelaide: AOA; 2015.
2. Ware JE, Kosinski M, Keller SD. SF-12: How to score the SF-12 Physical and Mental Health Summary Scales. Boston, MA: The Health Institute, New England Medical Center, Second Edition, 1995.