STATISTICAL ANALYIS PLAN (SAP)

Study Details:

Protocol Version	2.0	Protocol Date	02-JUN-2021

SAP Version Control:

SAP Status	Final, Version 1.0, 10-Jun-2024
Previous Version Number, Date	

Name and Title	Signature and Date / DocuSign Stamp
Yueming Shi Statistician, Bayesian Statistics (Statistician)	DocuSigned by: Yueming Shi Signer Name: Yueming Shi Signing Reason: I am the author of this document Signing Time: 10-Jun-2024 | 14:12:37 BST 21B94D7EDD60456CBB43BBC136D9B217
Jay Jantz Director of Global Biostatistics & Data Sciences (Head of Global Biostatistics or designee)	DocuSigned by: Jay Jantz Signer Name: Jay Jantz Signing Reason: I approve this document Signing Time: 10-Jun-2024 | 19:38:36 BST B7D37248838E4CACAE11E7E55198A88D

Contents

1 LIST OF ABBREVIATIONS

Abbreviation	Definition
ADE	Adverse Device Effect(s)
AE	Adverse Event(s)
AVN	Avascular Necrosis
BCS	Bi-cruciate Stabilized
CRF	Case Report Form(s)
CR	Cruiciate Retaining
CRO	Contract Research Organization
CV	Curriculum Vitae
DA	Degenerative Arthriti
DD	Device Deficiency(ies)
DFMEA	Design Failure Mode Effects Analysis
FAS	Full Analysis Set Population
FDA	Food and Drug Administration
FU	Follow-Up
GCP	Good Clinical Practice
HTO	High Tibial Osteotomy
ICH	International Conference on Harmonisation
IEC	Independent Ethics Committee
IFU	Instructions for Use

Abbreviation	Definition
IP	Investigational Product
IRB	Institutional Review Board
ISF	Investigator Site File
ITT	Intention to Treat population
KSS	Knee Society Score
LOCF	Last Observation Carried Forward
NA or N/A	Not Applicable
N (or n)	Total Sample Size (or subgroup sample size)
NSAID	Nonsteroidal Anti-inflammatory Drug
OA	Osteoarthritis
OKS	Oxford Knee Score
PI	Principal Investigator
PMA	Pre-Market Authorization
PP	Per-protocol Population
PTA	Post Traumatic Arthritis
RA	Rheumatoid arthritis
RCT	Randomized Controlled Trial
ROM	Range of Motion
S+N	Smith&Nephew.Inc
SADE	Serious Adverse Device Effect(s)
SAE	Serious Adverse Event(s)
SAF	Safety population
SAP	Statistical Analysis Plan
SoC	Standard of care
THA	Total Hip Arthroplasty
TKA	Total Knee Arthroplasty

Abbreviation	Definition
USADE	Unanticipated Serious Adverse Device Effect(s)
WOMAC	Western Ontario and McMaster Universities Arthritis

2 INTRODUCTION

The following Statistical Analysis Plan (SAP) details the statistical considerations, including the data analysis methods, for the Study Protocol Journey II CR.2020.11, Version 2.0, 02Jun2021. Related documents to this SAP are the Study Protocol, Case Report Form (CRF), and Tables, Figures and Listings (TFL) Templates Shells.

3 STUDY DESIGN

This study is a prospective, non-randomized multi-centre study to evaluate the performance of JOURNEY II CR Total Knee System in APAC patient populations, also to evaluate impact of patella resurfacing on the outcomes of JOURNEY II TKA. Totally up to 15 sites will participate within India, China mainland, Hongkong, Singapore, Thailand and Japan. Totally up to 480 knees information will be collected.
The decision on whether to resurface patella or not will be made by investigator based on patient's conditions.

STUDY SCHEDULE

Schedule of Events[1]	Pre-Operative -28 to 0 days	Operation Day 1	Immediate Post - op to 6 weeks Post-Op (Day 1 to 6 weeks+14 days)	6 months Post-Op (± 14 days)	1 year Post-Op (± 30 days)	2 years Post-Op (± 60 days)	3 years Post-Op (± 60 days)
Informed Consent	x						
Inclusion/Exclusion	x						
Demographics/Medical History	x						
Operative Data Collection		x					
Discharge Data Collection			x				
OKS[2]	x			x	x	x	x
FJS[2]				x	x	x	x
KSS[2]	x			x	x	x	x
Patient Expectation	x						
Patient Satisfaction [2]				x	x	x	x
Quantitative Radiographic Assessment	x[3]		x[4]	x[5]	x[5]	x[5]	
Safety Assessment (AEs, DDs)		x	x	x	x	x	x
End of Study/Exit		x[6]	x[6]	x[6]	x[6]	x[6]	x

4 STUDY OBJECTIVES

The objectives of the study are listed as below,

1. Evaluate performance of JOURNEY II TKA in APAC patient populations
2. Evaluate impact of patella resurfacing on the outcomes of JOURNEY II TKA

5 STUDY ENDPOINTS

5.1 Primary Endpoint(s)

The primary endpoint for the study is defined as the OKS at 2 years.

5.2 Secondary Endpoint(s)

The following secondary endpoints have been defined for this study as below,

- OKS
- FJS
- KSS
- Patient Expectation
- Patient Satisfaction
- Radiographic Assessment

5.3 Safety Endpoint(s)

•All adverse events (AEs) including intra-operative adverse events and complications
•Device related re-intervention
•Device Deficiencies

6 STATISTICAL CONSIDERATIONS

6.1 Determination of Sample Size

The sample size is established primarily using a precision of estimates-based approach for resurfaced and unsurfaced Patella cohorts separately. As a secondary concern formal statistical hypothesis testing to assess differences between the cohorts will be a possibility.

Based on an underlying assumption of standard deviation of 7.4 in OKS reported in literature (Bohm et al 2012)., a minimum of 192 knees are required, per cohort, to ensure that there is an 80% probability that a 95% confidence interval constructed around the OKS score can be calculated to within ±1.2 units for each cohort of resurfaced and unresurfaced knees. To allow for up to 20% attrition during the data collection activity, this will be inflated to enrol a minimum of 240 knees per cohort.

6.2 Randomisation

No randomization has been planned for this study. The decision on whether to resurface patella or not will be made by investigator based on patient's conditions.

6.3 Interim Analysis

The protocol stated an interim analysis was planned after the 2 year visit. However due to premature termination of the study no interim analyses will be performed.

7 STATISTICAL ANALYSIS

7.1 General

Smith+Nephew's Global Biostatistics group will conduct the statistical analysis for this study. Unless otherwise stated, all significance tests and hypothesis testing will be two-sided, performed at the 5% significance level. Resulting p-values will be quoted and 95% two-sided confidence intervals will be generated where appropriate. All p-values will be rounded to three decimal places, p-values less than 0.001 will be presented as '<0.001' in all tables.

Where data summaries are specified, categorical and ordinal variables will be summarized with frequencies and percentages. Continuous variables will be summarized with the following summary statistics: number of observations, mean, median, standard deviation, minimum and maximum values. All analyses will be performed in SAS 9.4 (or later).

All table summaries will be provided overall, and by surgical method (resurfaced/un-resurfaced patella).

Example mock shells are provided below:

Table: Demographics by subject				
		Resurfaced Patella	Un-Resurfaced Patella	Overall
Age	Mean			
	Median			
	Std			
	Min			
	Max			
	Q1			
	Q3			
	Inter Quartile Range			
	N			
Age	<65			

	>=65			
	N			

PROM by knee					
	Assessment		Resurfaced Patella	Un-Resurfaced Patella	Overall
	Pre-Operative	Mean			
		Median			
		Std			
		Min			
		Max			
		Q1			
		Q3			
		Inter Quartile Range			
		N			
	6 Month				
	1 Year				
	2 Year				
	3 Year				

7.2 Analysis Populations

The following are the analysis populations included in the protocol:

- Safety Population (SAF): This is defined as all subjects who were enrolled and treated with the study devices.
- Intent-to-Treat Population (ITT): The ITT population is defined as subjects from SAF, who have at least one post-operative assessment on any of the effectiveness endpoints.
- Per-Protocol Population (PP): The PP population is a subset of subjects in the ITT population, who do not have major protocol deviations and who satisfied all enrolment eligibility criteria. All major protocol deviations will be formally classified on a case-by-case basis prior to the final study database lock.

However, due to premature termination of the study, all analyses will be performed with SAF population only.

7.3 Handling of Missing, Incomplete and Repeat Data

All data will be analysed as observed unless stated otherwise in analysis section.

7.4 Derived Data

Time to Revision:
Time to revision (in months) = (Revision Date – Operative Date)/30.4375.

2011 Knee Society Score (KSS)
The 2011 KSS consists of 34 questions and provides sub-scores across 5 dimensions as shown in Table 4.

Table 4 2011 Knee Society Score

#	Sub-Score	# Questions	Range	Completion by
1	Objective Knee Score	7	0-over 100	Surgeon
2	Subject Satisfaction Score	5	0-40	Subject
3	Subject Expectation	3	3-15	Subject
4	Functional Knee Score	19	0-100	Subject

The Knee Society Score (KSS) comprises information on Objective Knee Indicators (range 0 to over 100 points), Patient Satisfaction (range 0-40), Patient Expectation (range 0-15), and Functional Activities (range 0-100). The KSS endpoints are derived from a series of questions that relate to everyday activities. More elaborate information for deriving the KSS endpoints are described in the publication by Scuderi et al (2012) [1]. However, an abbreviated description of the KSS endpoints used in the study is described in this SAP.

- *Objective Knee Score (Completed by Physician/Surgeon)*

- The objective score allows for more than 100 points in patients with greater than 125° of flexion and a stable painless knee as outlined below.

- "Alignment" allows for a maximum of 25 points and is determined on a weight-bearing AP radiograph measuring the femoral-tibial (Anatomic) axis.
- "Instability" allows for a maximum of 25 points for a knee that is stable in the coronal and saggital axis.
- "Joint Motion" allows one point for each 5° of joint motion with a potential for greater than 25 points to be assigned for patients with greater than 125° of motion. There are deductions for flexion contracture and extension lag.
- "Instability" has a maximum allowable point allocation of 25+.
- "Symptoms" category contains two 10-level scales, ranging from "none" to "severe" for each patient to rate pain encountered when walking on level ground and on stairs/inclines. For this, a patient starts with 10 points on each scale for a painless knee with deductions of up to 10 points deductions as indicated by the patient's response on each pain scale. There is an additional question regarding how "normal" the knee feels to the patient. The maximum allowable point is 25.

- *Patient Expectations and Satisfaction (Completed by Patient)*

- "Patient Expectations" is a three-question fifteen-point scale that is collected pre-operatively and post-operatively. The pre-operative questions reflect a patient's opinion on the extent to which s/he expects that the operation will improve his/her knee pain, and the ability to perform activities of daily living and recreational activities.
- The post-operative patient expectations questions reflect the extent to which the outcome after the operation has met the patient's pre-operative expectations with respect to pain and function.

- "Patient Satisfaction" is a five-question 40-point scale that is collected preoperatively and at each follow-up visit.

- *Functional Score (Completed by Patient)*

- The functional score is composed of the four subgroups and has a maximum score of 100 as follows:
  - "Walking and Standing" has a maximum value of 30 points with deductions for the use of walking aids and supports.
  - "Standard Activities" has a maximum of 30 points and evaluates "standard" activities of daily living. Patients can also respond if they never participate in the activities. Patients who respond "I never do this" receive zero points for that activity.
  - "Advanced Activities" has a maximum of 25 points and evaluates function in performing more vigorous activities ranging from climbing a ladder or step-stool to running. Patients can also respond if they never participate in the activities. Patients who respond "I never do this" receive zero points for that activity.
  - "Discretionary Activities" has a maximum of 15 points and allows patients to select the three activities that they consider most important to them personally from a group of seventeen recreational and exercise activities. Patients who do not participate in any of the discretionary activities will have a functional knee score that is limited to 85 points. The discretionary activities do not need to be identical in the pre-operative and post-operative period.

Forgotten Joint Score (FJS) [2] :

The FJS-12 is a PROM that assesses joint awareness during the activities of daily living. The FJS-12 comprises of 12 questions with a 5-point Likert response format (0 to 4), where 0 = Never; 1 =

Almost Never; 2 = Seldom; 3 = Sometimes and; 4 = Mostly. The 12 questions that joint awareness is based on are as follows:

- Night Symptom
- Sitting
- Walking
- Bathing
- Travelling
- Stairs
- Walking uneven ground
- Rising
- Standing
- Housework/Gardening
- Walking/Hiking
- Favorite Sport

An overall score based on the sum of all 12 responses (minimum of 0 and maximum of 48) is computed and transformed linearly into a 0 to 100 score. The higher the transformed score the greater is the ability of the subject to forget about the affected/replaced knee during activities of daily living. If any answer to the above question is missing, the score should not be calculated.

Oxford Knee Score [3]:

In the Oxford Knee score, 12 questions are scored. Each of question is scored in the same way. All questions are laid out similarly with response categories denoting least (or no) symptoms to greatest severity. For example, for question 1:

During the past 4 weeks, How would you describe the pain you usually have from your knee?

The answer and corresponding score are as following:

- None (score as 4)
- Very mild (score as 3)
- Mild (score as 2)
- Moderate (score as 1)
- Severe (score as 0).

The overall score will be calculated by simply summing the scores received for individual questions. This results in a continuous score ranging from 0 (most severe symptoms) to 48 (least symptoms). If there are only one or two questions with no answer, the mean value of all of other responses will be used to imputed. If more than two questions are unanswered, overall score will not be calculated.

Adverse Events

Investigator events do not receive a classification and so S+N classify events as follows:

- An event will be classified as an ADE if the relationship to the study device and/or the relationship to the study procedure is possibly/definitely
- An event will be classified as an SAE if the event is serious
- An event will be classified as an SADE if it is an ADE which is also serious
- An event will be classified as a USADE if it is an SADE which is also unanticipated
- Otherwise, the event will be a non-device, non-procedure related, non-serious AE

For adverse events that have been resolved, duration of adverse event is calculated as end date minus start date

Due to system updates, checks need to be performed to ensure the most stringent classification is being reported. When comparing Investigator assessments (classifications as above) with AEMB classifications, the most stringent classification should be reported as follows:

Investigator assessment	AEMB classification	Most stringent
AE	SAE	SAE
SAE	AE	SAE
AE	ADE	ADE
ADE	AE	ADE
ADE	SADE	SADE
SADE	ADE	SADE

SAE	SADE	SADE
SADE	SAE	SADE
SADE	USADE	USADE
USADE	SADE	USADE

7.5 Baseline Data

All observations available prior to the operative date will be defined as baseline (or preoperative) data. All demographic and pre-operative characteristics data will be summarized at baseline. All demographic (including age categorised whether "<65" and "≥65") and baseline characteristics will be summarized by surgical method (with or without Patella Resurfacing) using the SAF analysis populations.

All data under below forms will be summarized for screening:
Demographics
Vital Signs
Subject Pregnancy

Whether a subject has relevant medical history and/or diseases or past surgery will be summarised, and the average number per subject summarised. Details provided in a listing of medical history term, start/end date and ongoing.

All data under below forms will be summarized for Operative Details:
Operative Data

Device IDs will be provided in a listing.

7.6 Disposition Data

A table will be provided summarising number of subjects.

Subject (and knee) disposition will be summarized after completion of the study using frequency (n) and percentage (%) of subjects (or knee) who completed and who did not complete the study. For those who did not complete the study, the primary reason for discontinuation will additionally be summarized.

The number of knees that enter the study, and the number of knees with follow up information will be provided by visit:

- Theoretically due: number enrolled and implanted with study device
- Deaths: cumulative number of subjects that died prior to the study visit
- Revisions / failures: cumulative number of subjects that revised/failed prior to the study visit
- Cumulative terminations: terminated from study other than from revision or death
- Expected = theoretical due – cumulative (deaths + revisions + termination)
- Actual = actual number of subjects at follow up visit
- Follow up = actual / expected x 100.

The number of subjects that have discontinued from the study and reasons for withdrawal will be summarised.

Study duration and the duration between surgery and each visit will be summarised.

Screening and enrolment summary: For each centre/site, dates that the first subject was screened/enrolled and the last subject completed will be provided, duration (site, 1st enrolled – final subject exit), together with number of subjects enrolled and number of subjects completed 10 year visit.

7.7 Protocol Deviations

The frequency of protocol deviations will be summarized along with the number of subjects experiencing each.

7.8 Measurement of Treatment Compliance

Not applicable.

7.9 Multiplicity

No adjustments for multiplicity are planned for this study.

7.10 Analysis of Primary Endpoint(s)

The protocol states:

The primary endpoint of this study is the OKS at 2 years. If we assume the mean of OKS at 2 years for the subjects with implant JOURNEY II TKA with and without Patella Resurfaced as μ1 and μ2. The hypothesis test will be as following:

H0: μ1 = μ2 vs
H1: μ1≠ μ2

The mean difference (μ1-μ2) and 95% CI will be estimated. If zero is included in 95% CI of mean difference, then null hypothesis will be accepted, which means patella resurfacing has no impact on the outcomes of JOURNEY II TKA. Otherwise if zero is not included in 95% CI of mean difference, then alternative hypothesis (H1) will be accepted, which means patella resurfacing has significant impact on the outcomes of JOURNEY II TKA. If the lower limit of 95% CI is greater than 0, which means that the subject implant with Patella Resurfaced has higher OKS at 2 years. If the upper limit of 95% CI is less than 0, which means that the subject implant with Patella Resurfaced has lower OKS at 2 years.
In addition, the mean and 95% CI of OKS at 2 years for subjects implant with and without Patella Resurfaced will be estimated separately.

At time of writing the SAP, this study has been early terminated, the study no longer has sufficient power for this hypothesis test to be relevant and it will therefore not be performed. Summary statistics, 95% CI for μ1 and μ2 would be presented.
The primary endpoint will be summarized using the SAF population only due to premature termination.

7.11 Analysis of Secondary Endpoint(s)

All secondary endpoints will be summarized using the SAF analysis population. Changes from preoperative to postoperative will not be analysed for statistical evidence of difference in PROMs

due to early termination of the study and resultant lack of power. Summary statistics for observed PROMs score would be presented.

Knee Society Score (KSS), Oxford Knee Score (OKS) and Forgotten Joint Score (FJS)

The KSS score, OKS score and FJS score will be summarized at the pre-operative and postoperative visits and stratified by surgical method (resurfaced/un-resurfaced) using continuous summary characteristics.

For each PROM, the absolute change from pre-operative assessment (within subject) will also be summarised.

Standard Radiographic Assessment

The presence of radiographic observations: implant loosening, implant migration, and osteolysis and all other pertinent radiographic findings (i.e. all data under form 'Quantitative Radiographic Assessment') will be summarized as counts (n) and percentages (%) by visit. These would additionally be stratified by surgical method (resurfaced/un-resurfaced). Fisher Exact or Chi-Square test will not be performed due to premature termination.

Radiolucent lines (mm) by zone and visit will be summarized using descriptive statistics for continuous variables. These would additionally be stratified by surgical method (resurfaced/un-resurfaced).

Patient expectation and patient satisfaction

The patient expectation and patient satisfaction will be summarized as counts (n) and percentages (%) by visit. These would additionally be stratified by surgical method (resurfaced/un-resurfaced). Fisher Exact or Chi-Square test will not be performed due to premature termination.

Time to Revision

The protocol stated that: "Kaplan Knee implant revision rate will be summarized as proportions with 95% CIs estimated using exact binomial methods for two surgical methods separately."

Any revisions, and detail of, will be provided in a listing.

If feasible based on data available at premature termination, Kaplan Meier estimates and 95% confidence intervals will be provided by surgical method. Knees implanted without an on-study revision will be censored on the date of their last known follow-up visit. Any premature knee discontinuation from study due to a subject's death or for any other reason, excluding lost-to-follow-up, will be censored on the date of this discontinuation. Cumulative implant survivorship also be graphically illustrated.

7.12 Analysis of Exploratory Endpoint(s)

Not Applicable

7.13 Analysis of Safety Endpoint(s)

All safety data will be summarized overall, and by the different surgical methods (resurfaced/un-resurfaced).

Adverse event table summaries will be provided using the most stringent classification from Investigator and Adverse Event Monitoring Board (AEMB). Both Investigator and AEMB classifications will be provided alongside the most stringent classification in the Listings.

The number of events and the number of subjects reporting: adverse events (AE) and serious adverse events (SAE), adverse device effects (ADE), serious adverse device (SADE) and unanticipated serious adverse device effects (USADE) will be summarised.

In addition, AEs will be summarised by: frequency, severity, anticipated/unanticipated, relationship to device and procedure; treatment and surgical intervention details (type, reason), outcome and duration of adverse events at trial discontinuation will be summarised.

All safety analysis will be performed using the SAF analysis population.

Device Deficiencies

The number of device deficiencies, and the number with potential to cause an SAE will be summarised.

7.14 Other Data Summaries

Not applicable

7.15 Changes in Analysis Methods Specified in the Protocol

Statistical testing will not be performed for the endpoints due to premature termination of the study. Descriptive summary statistics will be presented for the endpoints only.

8 REFERENCES

1. Scuderi GR, Bourne RB, Noble PC, Benjamin JB, Lonner JH, Scott WN, The New Knee Society Knee Scoring System, Clin Orthop Relat Res. 2012 Jan;470(1):3-19. doi: 10.1007/s11999-011-2135-0.
2. Behrend, H., et al. The Forgotten Joint Score Website. 2016 [cited 2016 May 5]; Available from: http://www.forgotten-joint-score.info/.
3. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg Br. 1998;80-B(1):63-69. doi:10.1302/0301-620X.80B1.0800063