STATISTICAL ANALYSIS PLAN (SAP)

Study Details:

Protocol Version	5.0	Protocol Date	18-Aug-2017

SAP Version Control:

SAP Status	Final, Version 1.0, 10-Mar-2023
Previous Version Number, Date	N/A

Name and Title	Signature and Date / DocuSign Stamp
Michael Robinson Senior Manager, Biostatistics & Data Sciences (Statistician)	DocuSigned by: Michael Robinson Signer Name: Michael Robinson Signing Reason: I approve this document Signing Time: 10-Mar-2023 | 11:47:08 GMT 0E27C33EB43441FB820C99CBC38E7E89
Michelle Foster Senior Statistician (Independent Statistician)	DocuSigned by: Michelle Foster Signer Name: Michelle Foster Signing Reason: I approve this document Signing Time: 10-Mar-2023 | 12:29:05 GMT 963D0337C7F9428089D4471F058C5A52

Contents

STATISTICAL ANALYSIS PLAN (SAP) 1
1 List of Abbreviations 3
2 Introduction 4
3 Study Design 4
4 Study Objectives 4
4.1 Primary Objective 4
4.2 Secondary Objectives 4
5 Study Endpoints 4
5.1 Primary Endpoint 4
5.2 Secondary Endpoints 4
5.3 Safety Endpoint(s) 5
6 Statistical Considerations 5
6.1 Determination of Sample Size 5
6.2 Randomisation 5
6.3 Interim Analysis 5
7 Statistical Analysis 5
7.1 General 5
7.2 Analysis Populations 6
7.3 Handling of Missing, Incomplete and Repeat Data 6
7.4 Derived Data 6
7.5 Baseline Data 11
7.6 Disposition Data 12
7.7 Protocol Deviations 12
7.8 Multiplicity 12
7.9 Analysis of Primary Endpoint(s) 12
7.10 Analysis of Secondary Endpoint(s). 13
7.11 Analysis of Exploratory Endpoint(s) 15
7.12 Analysis of Safety Endpoint(s) 15

7.13 Other Data Summaries 16
7.14 Changes in Analysis Methods Specified in the Protocol 16
8 References 16

1 LIST OF ABBREVIATIONS

Abbreviation	Definition
ADE	Adverse Device Effect(s)
ASES	American Shoulder and Elbow Surgeons Shoulder Score
ADL	Activities of Daily Living
AE	Adverse Event(s)
CRF	Case Report Form(s)
DF99	Datafax Safety Database
FU	Follow-Up
GCP	Good Clinical Practice
ITT	Intention to Treat Population
NA or N/A	Not Applicable
PP	Per-protocol Population
QuickDASH	The Disabilities of the Arm, Shoulder and Hand
SADE	Serious Adverse Device Effect(s)
SAE	Serious Adverse Event(s)
SAP	Statistical Analysis Plan
SAF	Safety Analysis Set
TFL	Tables, Figures and Listing
USADE	Unanticipated Serious Adverse Device Effect(s)
WOOS	Western Ontario Osteoarthritis of the Shoulder index

2 INTRODUCTION

The following Statistical Analysis Plan (SAP) details the statistical considerations, including the data analysis methods, for the Study Protocol version 5.0, 18-Aug-2017. Related documents to this SAP are the Study Protocol, Case Report Form (CRF), and Tables, Figures and Listings (TFL) Templates Shells.

3 STUDY DESIGN

Post-market, non-randomised, open-label, observational clinical study with retrospective and prospective enrolment.

4 STUDY OBJECTIVES

4.1 Primary Objective

The objective of this study is to evaluate the 2-year post implantation survivorship of the PyroTITAN HRA device following the implementation of a new proof test to identify and eliminate devices with sub-standard mechanical integrity. The results will be compared to data collected in a prior study conducted before implementing the new proof test.

4.2 Secondary Objectives

The secondary objective of the study is to evaluate any other adverse events and overall clinical outcomes of the PyroTITAN HRA device in the new cohort at 2-year and 5-year post implantation time-points.

5 STUDY ENDPOINTS

5.1 Primary Endpoint

Assessment of device survival (no device revisions or removal) at the two year time-point

5.2 Secondary Endpoints

- Absence of complications (device related Adverse Events), and post-op procedures on the affected joint including additional revision surgeries at 2-year and 5-year time-points.
- Assessment of the functionality of the PyroTITAN HRA Shoulder System through clinical assessments at 2-year and 5-year time-points.

5.3 Safety Endpoint(s)

- Adverse Event classifications (AE/ ADE/ SAE/ SADE/ USADE)

6 STATISTICAL CONSIDERATIONS

6.1 Determination of Sample Size

The non-inferiority hypothesis will be evaluated using the Blackwelder approach based on the Intent-to-treat population. The approach is based on 95% confidence interval by comparing the interval limit to the pre-specified non-inferiority margin. The non-inferiority will be claimed if the upper 95% confidence bound of (P_0 - P_T) is less than δ.

The success rate for the study device is unknown for this population but there is no reason to believe that it would be less than reference rate. Therefore, for a conservative success rate of 95.8% of the study device is used in the study. The non-inferiority margin δ of -0.045 is chosen in the study. 123 subjects are required to achieve 100(1 – β) % =80% power to detect non-inferiority at the Significance level of α = 0.05. With 10% lost to follow-up, the total sample size needed is 137.

6.2 Randomisation

No randomization has been planned for this study.

6.3 Interim Analysis

Data will be reviewed for safety in an ongoing manner, according to standard formal procedures at Smith + Nephew. No interim formal analysis is planned; but it is anticipated that a number of outcomes will be followed through time; for example, surgical technique, and return to function.

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).

7.2 Analysis Populations

The following subject populations will be used for the statistical analysis of this study.

- **Safety Population (SAF)**, including all subjects who are implanted with study device.
- **Intention to Treat population (ITT)**, following Intention to Treat principle including all subjects who are enrolled into the study.
- **Per-Protocol Population (PP)**, including all subjects in the full analysis set who have no significant protocol deviations and met the inclusion/exclusion criteria.

All safety and survisorship (primary endpoint) analyses will utilise the SAF Population. All other endpoints will be analysed using both the ITT and PP analysis populations.

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

<u>Analysis populations</u>

- Flag for inclusion in the SAF population
- Flag for inclusion in the ITT population
- Flag for inclusion in the PP population

<u>Visit dates</u>

- 6 month follow up within 30 day window
- 12 month follow up within 60 day window
- 24 month follow up within 60 day window
- 60 month follow up within 90 day window

<u>Surgery Details</u>

- Surgical Time = End time of Operation – Start time of operation

- Time to discharge = Date of discharge – Date of Surgery

VAS Pain and Satisfaction

Calculate the change from baseline VAS pain and VAS satisfaction scores to each visit (score at post-op visit minus score at baseline).

ASES[1]

The ASES questionnaire usually has a VAS Pain scale however as this study doesn't use this part of the questionnaire ASES score will be calculated using the activities of daily living functional questions only. The score is derived using the following equation:

*ASES score = (Cumulative functional questions score/ 30) * 100*

There are 10 functional questions: put on a coat, sleep on the affected side, wash back/do up bra, managing toileting, comb hair, reach a high shelf, lift 10lbs above shoulder, throw a ball overhand, usual work, usual sport.

Assign the response to each functional question a score as follows:

- Unable to do = 0
- Very difficult to do = 1
- Somewhat difficult = 2
- Not difficult = 3

To calculate the cumulative functional questions score, sum the responses to the 10 questions (each response is worth a maximum of 3 points, maximum possible functional score of 30).

Calculate the ASES score for each visit, as well as the change from baseline score to each visit (score at post-op visit minus score at baseline).

WOOS [4]

There are 19 questions, each question is answered on a visual analogue scale ranging from 0 to 100. The questions are divided into four domains: Physical symptoms (max score 600), sports and work (max score 500), lifestyle (max score 500) and emotions (max score 300). The overall score ranges from 0 to 1900, with 1900 being the worst. For ease of interpretation, scores are converted to a percentage of the maximum score for each domain and overall [(score/max score)*100].

Calculate the score for each domain and overall at each visit, as well as the change from baseline scores to each visit (score at post-op visit minus score at baseline).

Constant-Murley[2]

The Constant-Murley score is derived using the following equation:

Constant-Murley score = [VAS pain score (0-15) + Cumulative ADL score (0-20) + Cumulative mobility score (0-40) + Maximum strength score (0-25)]

The VAS Pain score refers to the question "Indicate the highest pain level you have experienced in your shoulder during ordinary activities within the last 24 hours" and ranges from 0 to 15.

Calculate cumulative ADL score as follows (maximum 20 points):

- Is your occupation or daily living limited by your shoulder?
  - No (4 points)
  - Moderate limitation (2 point)
  - Severe limitation (0 points)
- Is your leisure and recreational activities limited by your shoulder?
  - No (4 points)
  - Moderate limitation (2 point)
  - Severe limitation (0 points)
- Is your night sleep disturbed by your shoulder?
  - No (2 points)
  - Sometimes (1 point)
  - Yes (0 points)
- State to what level you can use your arm for painless, reasonable activities?
  - Waist (2 points)
  - Xiphoid (4 points)
  - Neck (6 points)
  - Head (8 points)
  - Above head (10 points)

Calculate cumulative mobility score as follows (maximum 40 points):

- Forward flexion
  - 0-30 (0 points)
  - 31-60 (2 points)
  - 61-90 (4 points)
  - 91-120 (6 points)
  - 121-150 (8 points)
  - ≥151 (10 points)
- Abduction
  - 0-30 (0 points)
  - 31-60 (2 points)
  - 61-90 (4 points)

  - 91-120 (6 points)
  - 121-150 (8 points)
  - ≥151 (10 points)
- External rotation – check all that apply
  - Hand behind head, elbow forward (2 points)
  - Hand behind head, elbow back (4 points)
  - Hand to top of head, elbow forward (6 points)
  - Hand to top of head, elbow back (8 points)
  - Full elevation (10 points)
- Internal rotation
  - Lateral thigh (0 points)
  - Behind the buttock (2 points)
  - Sacroiliac joint (4 points)
  - Waist (6 points)
  - 12th thoracic vertebra (8 points)
  - Interscapular (10 points)

Calculate maximum strength score as follows (maximum 25 points):
If strength has been measured in kilograms, convert it to pounds by multiplying by 2.2.
Maximum strength score (in pounds) = max(1st attempt, 2nd attempt, 3rd attempt)

Calculate the CM score for each visit, as well as the change from baseline score to each visit (score at post-op visit minus score at baseline).

QuickDASH[3]
QuickDASH is an 11 item self-reported questionnaire in which the response options are presented as 5-point Likert scales with scores ranging from 1 to 5. At least 10 of the 11 items must be completed for a score to be calculated. The total score is then transformed into a score out of 100 with 0 indicating no disability and 100 being the most severe disability. The formula needed to transform the score is:

$\left[\left(\frac{Sum\ of\ responses}{n}\right) - 1\right] * 25$,where n=number of responses

Calculate scores as follows:

- Please rate your ability to do the following activities in the last week: open jar, chores, carry shopping, wash back, use knife to cut food, recreational activities (6 separate questions):
  - No difficulty (1 point)
  - Mild difficulty (2 points)
  - Moderate difficulty (3 points)
  - Severe difficulty (4 points)

  - Unable (5 points)

- During your past week, to what extent has your arm, shoulder or hand problem interfered with your normal social activities with family, friends, neighbors or groups?
  - Not at all (1 point)
  - Slightly (2 points)
  - Moderately (3 points)
  - Quite a bit (4 points)
  - Extremely (5 points)

- During the past week, were you limited in your work or other regular daily activities as a result of your arm, shoulder or hand problem?
  - Not at all (1 point)
  - Slightly limited (2 points)
  - Moderately limited (3 points)
  - Very limited (4 points)
  - Unable (5 points)

- Please rate the severity of the following symptoms in the last week: Arm shoulder hand pain, tingling (2 separate questions)
  - None (1 point)
  - Mild (2 points)
  - Moderate (3 points)
  - Severe (4 points)
  - Extreme (5 points)

- During the past week, how much difficulty have you had sleeping because of the pain in your arm, shoulder or hand?
  - None (1 point)
  - Mild (2 points)
  - Moderate (3 points)
  - Severe (4 points)
  - Extreme (5 points)

Calculate the score for each visit, as well as the change from baseline score to each visit (score at post-op visit minus score at baseline).

<u>Adverse events</u>

Adverse events are recorded by the Investigator and then reviewed by the Adverse Events Monitoring Board (AEMB) at Smith + Nephew and documented in DF99. The most stringent classification will be presented and reported. Investigator assessments will be provided in listings.

Investigators do not classify events, S+N use the Investigator assessment to classify the event as follows:

- An event will be classified as an ADE if the event is related to the study device and/or procedure
- 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

The following data is to be summarised at baseline. If any baseline data is to be summarised together with later visit data then the analysis will be described in the relevant endpoint section and not in this section.

Screening

- Baseline demographics: age (years); gender and dominant side
- Baseline Vital Signs: height (cm); weight (kg); BMI(kg/m2)
- Pre-Op Evaluation: Presenting symptoms, Primary Dignosis
- Surgical History (operative/ non-operative shoulder)
- Concomitant Medications/ Medical Conditions

Operative Details

- Surgical details: Surgical time, cement use, device stable, surgical approach, treatment of biceps tendon, glenoid status, bone quality
- Concurrent surgical procedures
- Intra-operative complications
- Time to discharge

7.6 Disposition Data

- The number of subjects screened.
- The number of subjects attending each study visit will be summarised with the dates of first subject first visit and last subject last visit inserted into the footnote.
- Reason for study completion

7.7 Protocol Deviations

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

7.8 Multiplicity

No adjustments for multiplicity are planned for this study

7.9 Analysis of Primary Endpoint(s)

The primary endpoint is the device survival at 2-years.
The study hypothesis is to test whether the 2-year survival (i.e., no revisions) rate in the investigational PyroTITAN HRA group is non-inferior to the reference rate at 95.8%. This study will be considered successful if the upper bound of the two-sided 95% confidence interval for the success rate difference (reference – study survival rate) is less than the non-inferiority margin of 4.5%.

The primary non-inferiority hypothesis is formulated as:

$\mathbf{H_0: P_0 - \delta \geq P_T}$

$\mathbf{H_a: P_0 - \delta < P_T}$

The variables are defined as follows:

P_T = 24-month success rate in the PyroTitan HRA group

P_0 = reference 24-month success rate from CP-HRA-002 study

H_0 = Null hypothesis

H_a = alternative hypothesis that that the success rate in the investigational group is non-inferior to the reference rate

δ = non-inferiority margin pre-specified to be 4.5%.

The survival rate (no device revisions or device removal) will be calculated and the associated 95% exact confidence interval presented. If the upper 95% confidence limit is greater than the reference rate minus the non-inferiority margin (95.8 - 4.5 = 91.3%) then non-inferiority of the observed rate to the reference rate to within a 4.5% margin can be concluded.

Time-to-event curves for device survival will be estimated using the Kaplan- Meier technique. For analysis of device survival, those lost to follow up or withdrawn from the study for reasons other than device revision or removal will be regarded as censored observations.

7.10 Analysis of Secondary Endpoint(s).

Implant survivorship

The proportion of implant survivorship at 5 years will be estimated (with exact 95% confidence intervals) and displayed accordingly.

Percentage of device complications (device related Adverse Events) and post-op procedures will also be summarised with associated 95% exact confidence interval presented

ASES

ASES Scores range from 0 to 100 with a score of 0 indicating a worse shoulder condition and 100 indicating a better shoulder condition

Although the ASES standardised form contains both physician assessment and patient self-evaluation section, only the 10 functional questions are used to calculate the ASES score.

ASES questionnaires that are incomplete for the questions used for scoring at the baseline visit will not be included in the analyses.

The ASES score will be derived for each patient at each visit and summary statistics will be presented. Change from baseline score to each post-operative visit will also be presented.
A paired sample t-test will be used to test changes in the scores from baseline to each post-operative visit. If the t-test assumptions are not met then a Wilcoxon Signed Rank test will be considered.

VAS Pain

A higher VAS score indicates a higher level of pain.
VAS Pain will be summarised at each study visit and the absolute change in VAS scores, from baseline to each study visit will also be summarised.
A paired sample t-test will be used to test changes in the scores from baseline to each post-operative visit.

VAS Satisfaction

A higher VAS score indicates a higher level of satisfaction.
VAS Satisfaction will be summarised at each study visit. As there is no baseline assessment of VAS satisfaction no analysis for change over time will be planned.

QuickDASH

QuickDASH is an 11 item self-reported questionnaire in which the response options are presented as 5-point Likert scales with scores ranging from 1 to 5. At least 10 of the 11 items must be completed for a score to be calculated. The total score is then transformed into a score out of 100 with 0 indicating no disability and 100 being the most severe disability. The formula needed to transform the score is:

$$\left[\left(\frac{Sum\ of\ responses}{n}\right) - 1\right] * 25$$,where n=number of responses

The QuickDASH score will be derived for each patient at each visit and summary statistics will be presented. Change from baseline score to each post-operative visit will also be presented. A paired sample t-test will be used to test changes in the scores from baseline to each post-operative visit. If the t-test assumptions are not met then a Wilcoxon Signed Rank test will be considered.

WOOS

There are 19 questions, each question is answered on a visual analogue scale ranging from 0 to 100. The questions are divided into four domains: Physical symptoms (max score 600), sports and work (max score 500), lifestyle (max score 500) and emotions (max score 300). The overall score ranges from 0 to 1900, with 1900 being the worst. For ease of interpretation, scores are converted to a percentage of the maximum score for each domain and overall [(score/max score)*100].

The WOOS score for each domain and overall will be derived for each patient at each visit and summary statistics will be presented. Change from baseline scores per domain and overall to each post-operative visit will also be presented.
A paired sample t-test on the overall WOOS score will be used to test changes in the scores from baseline to each post-operative visit. If the t-test assumptions are not met then a Wilcoxon Signed Rank test will be considered.

Constant Murley Score
The CM score consists of 4 domains: pain (15 points), activities of daily living (ADL) (20 points), movement (40 points) and power/strength (25 points). Strength score is calculated as the highest score of the 3 attempts.

Constant-Murley score questionnaires that are incomplete for the questions used for scoring at the baseline visit will not be included in the analyses.

The total score is calculated as the sum of each score from the 4 domains and ranges from 0 to 100 points, with a higher score indicating better shoulder function.

The CM score will be derived for each patient at each visit and summary statistics will be presented. Change from baseline score to each post-operative visit will also be presented. A paired sample t-test will be used to test changes in the scores from baseline to each post-operative visit. If the t-test assumptions are not met then a Wilcoxon Signed Rank test will be considered.

7.11 Analysis of Exploratory Endpoint(s)

Not Applicable

7.12 Analysis of Safety Endpoint(s)

Adverse events
The number of events and the number of subjects reporting: adverse events (AEs), serious adverse events (SAEs), adverse device effects (ADEs), serious adverse device effects (SADEs), and unanticipated serious adverse device effects (USADEs) will be summarised.

Events will be summarised by the AE/SAE/ADE/SADE/USADE classifications derived from the Investigator assessment of the event. In the case of disagreements, a separate summary of event classifications made by the Sponsor will also be provided.

In addition, AEs will be summarised by; severity (intensity), relationship to study device, relationship to procedure, outcome and duration of adverse event at trial discontinuation.

A listing will be provided which details subject number, AE description, start date, end date/ongoing, classification, severity, seriousness, relationship to device, relationship to procedure, anticipation, actions taken, outcome and whether study participation was discontinued due to the event.

7.13 Other Data Summaries

Not Applicable

7.14 Changes in Analysis Methods Specified in the Protocol

- Safety Population added to Analysis populations to define which subjects will be summarised for Safety and Survival endpoints.

8 REFERENCES

ASES Scoring:[1] https://www.codetechnology.com/blog/ases-shouldertool/#:~:text=Calculation%20of%20the%20ASES%20score,maximum%20functional%20sco re%20of%2030.

Constant Score: [2] https://orthotoolkit.com/constant-shoulder/static/media/Constant-MurleyScore.e77fd97c.pdf%208.%20Kukkonen%20J,%20Kauko%20T,%20Vah

QuickDASH scoring: [3] https://dash.iwh.on.ca/sites/dash/files/downloads/quickdash_info_2010.pdf

WOOS scoring: [4] https://core.ac.uk/download/pdf/82064477.pdf