Statistical Analysis Plan (SAP)

GLOBAL ICON Stemless Shoulder System Post Market Clinical Follow Up Study

Protocol Version: CT 1401 v3.0

SAP Revision: 1.0
SAP Revision Date: 11-MAY-2021

GLOBAL ICON Stemless Shoulder System Post Market Clinical Follow Up Study

Protocol Version: CT 1401 v3.0

The following individuals have reviewed this version of the Statistical Analysis Plan and are in agreement with the content:

Signature Page

Study Biostatistician:

PPD PPD

(Print) (Sign) Date

Head of Biostatistics:

PPD PPD

(Print) (Sign) Date

Franchise Clinical Study Lead:

PPD PPD

(Print) (Sign) Date

Franchise Platform Head of Clinical Research:

PPD PPD

(Print) (Sign) Date

Revision History

Revision Number	Revision Date (DD/MM/YYYY)	Reasons for Revision
0.0	14/08/2020	Original document – no revision
1.0	11/05/2021	Clarification of the radiographic endpoint (section 1.1) and clarification of the adjusted Constant-Murley Score reference (section 5 and references)

Table of Contents

List of Abbreviations

AE	Adverse Event
BMI	Body Mass Index
CIP	Clinical Investigational Plan
n	Non-missing data
PG	Performance Goal
PMCF	Post Market Clinical Follow-Up
SAP	Statistical Analysis Plan
SD	Standard Deviation

1. Study Design

This study is designed as a prospective, multi-centre, non-comparative, uncontrolled Post Market Clinical Follow-Up (PMCF) Study.

There will be 157 patients with the GLOBAL ICON Stemless Shoulder System in Total Shoulder Arthroplasty in up to 20 sites located in the United Kingdom, the Netherlands, Germany and Canada. A maximum of 30 subjects may be recruited from one site. More than one implanting surgeon may recruit subjects at each site. A complete description of the methods for determining sample size is contained in Section 7.

Patients will be followed after surgery at several times: immediate Post-Op, 3-, 12-, 24-, 60- and 120-months. The data up to and including 24-month follow-up visit will be used to determine the effectiveness and safety of the device. The primary, safety, secondary and tertiary objectives are presented below.

1.1 Primary Objective

The primary objective is to evaluate success of the device at the 24 month visit following implantation. Success is a composite outcome that includes both effectiveness and safety endpoints and will be established if all of the following conditions are met at the 24 month visit: 1) radiographs indicate that there is no continuous radiolucent line (RLL) around the humeral component, with a continuous RLL defined as a radiolucent line > 1mm in all five zones of both the AP and Axillary views, and 2) the adjusted Constant-Murley score is greater than 85, and 3) the Global ICON humeral component has not been removed for any reason, and 4) there has been no device-related serious adverse events.

1.2 Safety Objective

The safety objective is to evaluate the overall survivorship of the device as well as to characterize adverse events. For overall survivorship, a device is deemed to be surviving if no components have been removed for any reason. Survival is based on Kaplan-Meier methodology, with survival reported at the last day of the 24-, 60- and 120-months post-operative assessment window (see section 4). For adverse events (AE), the type and

frequency of all AEs in this study will be summarized, with distinction of serious AEs, operative and device related AEs.

1.3 Secondary Objectives

Secondary objectives include the assessment of clinical and radiographic performance post-operatively at 3-, 12-, 24-, 60- and 120-month visits. Assessments include: adjusted Constant-Murley Score, Oxford Shoulder Score, EQ-5D-5L and radiographic evidence of aseptic loosening of the GLOBAL ICON stemless humeral component. Of interest for the non-radiographic endpoints are mean values at the distinct time points. Based on radiographic data, time to aseptic loosening will be characterized by survival analysis using Kaplan-Meier methodology, where the event of interest is aseptic loosening. Survival will be reported at the last day of each assessment window.

1.4 Tertiary Objectives

Tertiary objectives are to evaluate levels of improvement, evaluated as a change from baseline for each post-operative assessment time point, for the adjusted Constant-Murley Score, Oxford Shoulder Score and EQ-5D-5L. This will be evaluated by examining change from baseline at 3-,12-, 24- , 60-, and 120- month visits. Lastly, survival of the GLOBAL ICON stemless humeral component for the reason of periprosthetic fracture will be reported using Kaplan-Meier methodology. Survival will be reported on the last day of the 24-, 60- and 120-month assessment windows.

2. Treatment Assignment

Treatment assignment is described in Section 1.

3. Randomization and Blinding

Neither randomization nor blinding will be implemented in this study.

4. Interval Windows

Data collected throughout the study will be assessed for compliance with the protocol-specified visit schedule.

Table Seven windows are defined based on the number of days prior to or after surgery (Day 0): baseline, Immediate post-op, 3-months post-op, 12-months post-op, 24-months post-op, 60-months post-op and 120-months post-op. If multiple visits fall into the same interval, the result latest in the visit window with complete data will be used in the analysis.

Table 1. Interval Windows

Analysis Visit	Study Visit	Study Interval (Days from Surgery)	Visit Window
Baseline	**Pre-Op**	-180 days to day of surgery (day 0)	-180 to 0 days
Surgery	**Surgery**	0	
Immediate Post-Op	**Immediate Post-Op**	0-10	**5** (-5 to +5 days)
3 Months	**3 Months**	35 – 126	**80** ± 45 days
12 Months	**12 Months**	281 – 449	**365** ± 84 days
24 Months	**24 Months**	646 - 1644	**730** - 84 days **730** + 914 days
60 Months	**60 Months**	1645 - 2005	**1825** ± 180 days
120 Months	**120 Months**	3470 – 3830	**3650** ± 180 days

5. Description of Scales and Scoring

The Constant-Murley score (CMS) is a 100-points scale composed of a number of individual parameters. These parameters define the level of pain and the ability to carry out the normal daily activities of the patient. The Constant-Murley score was used to determine the functionality after the treatment of a shoulder injury. It has four subscales: pain (15 points), activities of daily living (20 points), strength (25 points) and range of motion: forward elevation, external rotation, abduction and internal rotation of the shoulder (40 points). The higher the score, the higher the quality of the function. Once calculated, the scores are adjusted or normalized for gender and age using Constant's normative data. This normative data is reported by both Constant [1] and Katolik [2]. We will follow the original method described by Constant as Katolik incorrectly identifies categories of age.

The Oxford Shoulder Score (OSS) is a patient-based questionnaire used to assess shoulder pain after surgery. It consists of 12 questionnaire items with 5 ordinal response options for each question. Each response is scored from 0 to 4 points (4 = best/least problems). All item scores across 12 questions are summed to produce scale 0–48 (48 = best/least problems).

The EQ-5D-5L consists of the EQ-5D-5L descriptive system and the EQ Visual Analogue scale (EQ VAS). The descriptive system comprises 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). And each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. When a summary score is used across dimensions, each combination is mapped onto a value score. A specific mapping algorithm is used with country-specific values (U.S. values will be used for Canadian patients) [3].

6. Levels of Significance

Only the primary objective for the endpoint of device success is prospectively powered and will be conducted with a one-sided test of $\alpha = 0.025$, or equivalently, whether the lower bound of the 95% confidence intervals is above the maximum value under the null hypothesis. Confidence intervals (two-sided 95%) and p-values may be provided for analyses involving the primary, safety, secondary and tertiary endpoints. For safety endpoints no hypotheses will be tested, and the secondary/tertiary objectives are deemed exploratory. In light of these considerations, there will be no adjustment of significance levels because of testing multiple hypotheses. No labeling claims will therefore be made based on the analysis of safety, secondary and tertiary endpoints.

7. Analysis Sets

Analysis set definitions from the CIP appear below.

Consented/Enrolled Population: The Consented/Enrolled Population will consist of all subjects who were consented and enrolled in the study based on preliminary subject eligibility.

Safety Population: The Safety Population will consist of all subjects who were enrolled and were either implanted with the GLOBAL ICON stemless humeral component or an attempt was made to do so. Demographic data and analysis of the primary, safety, secondary and tertiary endpoints will be based on the Safety Population.

Per Protocol Population: The Per Protocol Population will consist of the subset of the Safety Population without specific protocol deviations. Based on review of subject data, the subset will exclude those patients that: 1) fail to satisfy all inclusion/exclusion criteria, 2) received a non-Global ICON humeral component or did not receive an approved DePuy Synthes glenoid system, and 3) failed to adhere to critical protocol-required restrictions and prohibitions, receipt of prohibited concomitant procedures or therapies, and severe non-compliance to protocol-specific procedures. The final Per Protocol Set will be determined prior to database hard lock and in agreement with clinical research and clinical operations. Analyses using the Per Protocol Population will be used to complement analyses based on the safety population.

8. Sample Size Justification

The sample size was determined based on the primary endpoint using the one-sample exact-test of a proportion as implemented in PROC POWER in SAS, version 9.3 or higher. For the calculation, we assume the composite success proportion at 24 months after surgery is 85% and implement a one-sided test at the 0.025 alpha-level to determine the minimal N which provides more than 80% power; this sample size is 133. To accommodate potential 15% attrition, the sample size will be increased to 157.
The sample size applies to patients enrolled and actually treated. Some patients who were enrolled in a study may not be ultimately treated for various reasons. Patient enrolment will continue until the proposed sample size for treated patients is complete.

NOTE: Although the sample size was originally determined based on the anticipated use of a one-sample exact-test of a proportion, it has been determined after enrolment was

complete that a one-sample Z-test of a proportion using a normal approximation would be utilized in the final primary endpoint analysis to facilitate data imputation methods.

9. Analyses to be Conducted

9.1 Descriptive Analyses

Study data will be tabulated for all subjects in the Safety Population using SAS version 9.4 or higher. Planned tabulations are described below and table, figure, and listing shells are provided separately (see Appendix).

Standard descriptive summaries for continuous data include the number of subjects with nonmissing data (n), mean, standard deviation (SD), median, minimum, and maximum values. For categorical data, the count and percentage will be provided. Percentages will be based on the number of subjects without missing data.

9.2 Disposition of Study Subjects

An overall summary of the number of subjects who were (or had): Enrolled, Enrolled but not treated, in the Safety Set, in the Per Protocol Set, withdrew before study completion, and who completed the study will be tabulated for all sites combined. A listing will be created for completion status and will include columns for all the items included in the summary table.

9.3 Demographic and Baseline Characteristics

Descriptive statistics will be displayed using the subjects in the Safety and the Per Protocol Analysis Populations for:

- Age at consent (in years)
- Gender
- Race
- Ethnicity
- Height (cm)
- Weight (kg)

- Body Mass Index (BMI) calculated as (Weight in kilograms)/(Height in meters)^2
- Primary Diagnosis

9.4 Hypothesis Testing of Primary Endpoint

The primary endpoint analysis will seek to demonstrate that at the 24 month visit after surgery, the composite success (p), which is based on four study outcomes (radiograph demonstrating no continuous radiolucent line around the humeral component, adjusted Constant-Murley Score > 85, no removal of the humeral component, no serious adverse events related to the device), is significantly greater than a performance goal (PG) of 75%. The null and alternative hypotheses for this endpoint are as follows:

Null hypothesis H_{o1}: $p \leq 75\%$

Alternative hypothesis H_{A1}: $p > 75\%$

The null hypothesis will be rejected and the alternative hypothesis will be accepted if the lower bound of a two-sided 95% confidence interval is greater than 75%. The performance goal of 75% used in the hypothesis test described above is based on the percentage of successful stemmed implants at the 24 month visit reported elsewhere using a similar definition of success [4]. Therefore, the hypothesis test implies a comparison to performance of stemmed implants. In a prior study, non-inferiority of the Simpliciti stemless humeral device to stemmed devices was demonstrated using a PG of 65% (75%-10% non-inferiority margin) [4]. Therefore, the current study sets a higher bar to establish successful performance of Global Icon when compared to the study evaluating the Simpliciti device.

Another objective using the same endpoint is to establish non-inferiority of Global Icon to Simpliciti conditional on rejection of H_{o1}. To do so a PG or non-inferiority margin of 78.74% (Simpliciti success percentage of 88.74% [5] -10%) is selected. The null and alternative hypotheses for this objective are as follows:

Null hypothesis H_{o2}: $p \leq 78.74\%$

Alternative hypothesis H_{A2}: $p > 78.74\%$

The null hypothesis will be rejected and the alternative hypothesis will be accepted if the lower bound of a two-sided 95% confidence interval is greater than 78.74%. Notably,

simultaneous testing of these nested null hypotheses (H_{o1}, H_{o2}) does not inflate the Type I error rate [6,7].

9.5 Modeling of the Primary Endpoint

Tests of hypotheses for a one-sample proportion described in 9.4 will calculate the variance of the proportion using a normal approximation to the binomial distribution with PROC FREQ. As noted in 9.8.4, the analysis is repeated in each imputed data set and thereafter aggregated. Once aggregated, inference will be based on the lower bound of the 95% confidence interval.

9.5.1 Adequacy of the Normal Approximation

A reasonable concern is that the success proportion will be close to 1 in this study, and it is unclear how well the normal approximation performs in this case. Using Monte Carlo simulation, performance of the normal approximation was evaluated. We generated 100,000 replications with either no missing data or 15% missing data (under a missing completely at random mechanism) with values that were subsequently imputed (10 imputations) using methods described elsewhere [8]. The population success proportion was set at 0.90 with 157 observations in each replication. Both in the absence of missing data and in the presence of missing data, the normal approximation yielded slight under-coverage relative the nominal level of 95%, 94.5% and 94.0%, respectively. By comparison, the Wilson method yielded slight over-coverage, 95.6% and 95.2%. Given these results, the normal approximation was chosen.

9.5.2 Sensitivity Analyses Incorporating Site Variation

In this study, implants are implanted within study sites and therefore may not be independent. Therefore, as a sensitivity analysis, a beta-binomial model will be used for incorporating possible between-site variability on implantation success [9]. In this model the number of implant successes (X_i) within each site follows a binomial distribution, $X_i|\ p_i \sim Binomial\ (n_i, p_i)$, with the proportion of successes (p_i) across sites following a beta distribution, $p_i \sim Beta\ (\mu, \theta)$ with $\mu = \alpha/(\alpha + \beta)$ and $\theta = 1/(\alpha + \beta)$ (where α and β are positive-valued parameters of the beta distribution). This is a random-effects model with a

mean success rate across sites (μ) that allows for variation by site (θ). The model is estimated using NLMIXED via the BETABIN SAS macro. In the absence of variation in the success rate across sites (i.e., $\gamma < .000001, \text{where } \gamma = \theta/(1+\theta)$), the normal approximation to the binomial will be used. As noted in 9.8.4, the analysis is repeated in each imputed data set and thereafter aggregated.

9.6 Reporting of the Safety Endpoints

Survivorship of the GLOBAL ICON device at 24-, 60- and 120- months post-operative will be reported, where the device is deemed to be surviving if no humeral components (anchor plate, humeral head) have been removed for any reason. An estimate of device survivorship of the GLOBAL ICON implant will be presented using Kaplan-Meier methods. Removal of any component, defined as revision for any reason, is the event of interest. Cases not revised will be censored at their date of last follow-up, death (if a death occurred) or study end date. Survivorship point estimates and 95% confidence intervals will be presented for a time point if at least 40 subjects remain at risk. The conventional Greenwood estimate will be used to calculate the variance and the complementary log-log transformation will be used to construct 95% confidence intervals using PROC LIFETEST.

The type and frequency of all adverse events (AEs) in this study will be summarized, with distinction of serious AEs, operative and device related AEs. The type and frequency of AEs through 120 months post-operative will be presented in table form and via a listing. An overall summary of AEs will be provided, including the number and the percent of subjects with all AEs, all serious AEs, all related AEs (device and procedure related), and all AEs by severity. In addition, all adverse events, serious, and non-serious adverse events will be summarized and tabulated by preferred term, both overall and by severity, and by time period of onset.

9.7 Modeling of the Secondary and Tertiary Endpoints

Confidence intervals (95%) and p-values will be based on means at each time point using a one sample t-test; change from baseline summaries will also be provided in this fashion. In addition, for analyses which involve change from baseline (tertiary endpoints), each

post-surgical measurement will be represented as a change from baseline, with baseline measurement and other possible factors of interest included as covariates in a linear regression model. These analyses are deemed exploratory, with no adjustment for multiplicity and no labeling claims based on the results. PROC TTEST and PROC REG will be used for the analyses. As noted in 9.8.4, the analysis is repeated in each imputed data set and thereafter aggregated. Aggregated results will be reported.

The Kaplan-Meier methods will be used to report results for survival endpoints. Survivorship point estimates and 95% confidence intervals will be presented for a time point if at least 40 subjects remain at risk. The conventional Greenwood estimate will be used to calculate the variance and the complementary log-log transformation will be used to construct 95% confidence intervals using PROC LIFETEST.

9.8 Missing Data

9.8.1 Primary endpoint

As detailed in the CIP, participants who withdraw prior to the 24-month follow-up visit (i.e., lost to follow-up) and did not experience a continuous RLL around the humeral component, a serious adverse device-related event or a revision surgery in which the humeral component was removed, would have missing data on their clinical outcomes, which will be addressed using imputation. Imputed values will be used to determine whether patients who withdraw prior to the 24-month follow-up visit are categorized as successes or failures on the primary endpoint. Specifically, determining success or failure on the primary endpoint for each patient withdrawing prior to the 24-month follow-up visit will be based on imputed values of clinical outcomes that make up the primary endpoint. Patients will be scored as a failure on the primary endpoint if their imputed values indicate continuous RLL around the humeral prosthesis or a revision in which the humeral component was removed or a SADE took place or the adjusted Constant-Murley score $\leq$ 85.

9.8.2 Survival Endpoints

Safety, secondary and tertiary endpoints include survival outcomes. Patients with missing data resulting from withdrawal will be treated as censored, with their censoring time based on the date of last assessment.

9.8.3 Continuous Secondary and Tertiary Endpoints

Continuous secondary and tertiary endpoints include the adjusted Constant Murley Score, Oxford Shoulder score and the EQ-5D-5L scores. Missing data that could impact the proposed analysis would result from participants withdrawing from the study as well as missed intermediate assessments. In either case missing data will be handled using imputation.

9.8.4 Method of Imputation

Multiple imputation will be performed using a fully conditional specification [10], given the need to impute values for binary variables. Fifty imputations will be performed. In order to combine results across imputed datasets, the parameter estimate will be calculated using an average of estimates obtained across datasets and the variance calculated using Rubin's method [11]. All calculations will be performed in SAS and imputations will be performed using PROC MI and aggregated using PROC MIANALYZE.

9.8.5 Imputation Model

A single imputation model will be used to impute values for the primary, secondary and tertiary endpoints (not including survival endpoints). The variables used to define the primary endpoint are presence/absence of a: continuous RLL around the humeral component, a serious adverse device-related event, revision surgery of the humeral component and the adjusted Constant-Murley score. For the adjusted Constant-Murley score, we will create separate variables for scores obtained at each assessment to include in the imputation model. For the other variables that make up the primary endpoint, partitioning by time could produce computational problems, therefore we will consider a model in which the presence/absence of the outcomes will be imputed irrespective of their timing. Even with this simplification, computational problems may arise for various reasons, such as no events for one of the clinical outcomes or severe multicollinearity. In

these cases, we will document the problem and fit a further simplified imputation model that pools one or more of these clinical outcomes together. Another potential modification to the imputation model is the use of individual items rather than total adjusted Constant-Murley score if for some patients there were missing items for a particular assessment.

Incorporating additional variables in the imputation model served two objectives. First, they can improve prediction of clinical outcomes used to construct the primary endpoint. Second, they can be used to address missing data on the secondary and tertiary endpoints. Among the variables included in the imputation model to support these objectives are the Oxford Shoulder score, EQ-5D-5L summary score and demographic variables (age, gender BMI and primary diagnoses). For the Oxford Shoulder score and EQ-5D-5L summary score separate variables for scores obtained at each assessment will be created and included in the imputation model.

9.9 Sensitivity Analyses

The analysis of the primary, safety, secondary and tertiary endpoints will be repeated using all subjects in the Per Protocol Population to determine the robustness of the study results.

10. Data Monitoring Committee (DMC)

No DMC is required for this study.

References

1. Constant CR, Gerber C, Emery RJ, Søjbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. *Journal of Shoulder Elbow Surgery* 2008 Mar-Apr;17(2):355-61. DOI: 10.1016/j.jse.2007.06.022.
2. Katolik, L. I., Romeo, A. A., Cole, B. J., Verma, N. N., Hayden, J. K., & Bach, B. R. (2005). Normalization of the Constant score. *Journal of Shoulder and Elbow Surgery, 14*(3), 279-285. DOI: 10.1016/j.jse.2004.10.009
3. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the 5Q-5D-5L to 5Q-5D-3L value sets. *Value Health.* 2012;15(5):708-715. DOI:10.1016/j.jval.2012.02.008
4. Churchill R., et al. Clinical and radiographic outcomes of the SIMPLICITI canal-sparing shoulder arthroplasty system: a prospective two-year multicenter study. *Journal of Bone and Joint Surgery America* 2016; 98:552-60.
5. Simpliciti study supporting 510(k): K143552
6. European Agency for the Evaluation of Medicinal Products, Committee for Proprietary Medicinal Products (CPMP). Points to Consider on Switching Between Superiority and Non-inferiority. EMEA: London 2000,
7. Morikawa T, Yoshida, M. A useful testing strategy in phase III trials: Combined test of superiority and test of equivalence. *Journal of Biopharmaceutical Statistics* 1995; 5:297–306.
8. Lott A, Reiter JP. Wilson confidence intervals for binomial proportions with multiple imputation for missing data. *The American Statistician* 2018, DOI 10.1080/00031305.2018.1473796
9. Young-Xu Y, Chan KA. Pooling overdispersed binomial data to estimate event rate. *BMC Medical Research Methodology* 2008; 8:58.
10. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research* 2007;16:219–242.
11. Rubin, D. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons. 1987.

Appendix: Tables, Listings and Graph Shells

The document that contains the tables, listings and graphs shells is entitled Study DPO_CT1401_ICON Table Figure and Listing Shells.

End of Document