A Prospective, Multi-Center Study of the IlluminOss® Photodynamic Bone Stabilization System for the Treatment of Impending and Actual Pathological Fractures in the Humerus from Metastatic Bone Disease

NCT02338492

Statistical Analysis Plan (SAP)

November 15, 2016

IlluminOss Medical, Inc.

Protocol Number: 14-03-PATHOLHUM-02

A Prospective, Multi-Center Study of the IlluminOss® Photodynamic Bone Stabilization System for the Treatment of Impending and Actual Pathological Fractures in the Humerus from Metastatic Bone Disease

STATISTICAL ANALYSIS PLAN

Version: V3.1

Date: 15 November 2016

ICON Study Number 3330/0002

Statistical Analysis Plan Signature Page

Final Version 3.1, Dated 15NOV2016

Sponsor:	IlluminOss Medical, Inc.
Study Title:	A Prospective, Multi-Center Study of the IlluminOss® Photodynamic Bone Stabilization System for the Treatment of Impending and Actual Pathological Fractures in the Humerus from Metastatic Bone Disease
Protocol Number:	14-03-PATHOLHUM-02
Protocol Version & Date:	Version 5.0: 6 January 2016

Prepared at ICON Clinical Research by:

Lin Pan
15 Nov 2016 19:25:009+0000
I approve this document.

Lin Pan, 7e94b5d9-3b48-4abe-b5cc-c2d4f376d1cc Date
Senior Biostatistician I, Biostatistics

Reviewed at ICON Clinical Research by:

Jill Stankowski
16 Nov 2016 16:11:012+0000
I approve this document.

Jill Stankowski, 906d67fe-fac1-41d0-912c-ccc680d6400c Date
Director, Biostatistics

Reviewed and Approved at IlluminOss Medical, Inc by:

Amy Berman
15 Nov 2016 19:58:057+0000
I approve this document.

Amy Berman c9a3bd2e-6e49-46d8-834a-52846eb29c20 Date
Vice President, Clinical Affairs

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Term
aCRF	Annotated Case Report Form
ADL	Activities of Daily Living
AE	Adverse Event
CDL	Chen-DeMets-Lan
CI	Confidence Interval
CM	Concomitant Medication
CRF	Case Report Form
eCRF	Electronic Case Report Form
FAS	Full Analysis Set
FDA	Food and Drug Administration
FI	Functional Interference
IA	Interim Analysis
ICH	International Conference on Harmonisation
ICF	Informed Consent Form
IRB	Internal Review Board
ITT	Intent-to-treat Population
LTF	Lost to follow-up
MBD	Metastatic Bone Disease
MedDRA	Medical Dictionary for Drug Regulatory Activities
MMRM	Mixed Model for Repeated Measures
MSTS	Muscoloskeletal Tumor Society
N, n	Number of subjects/subjects in a sample from a population or analysis group
PA	Psychosocial Aspects
PBSS	IlluminOss Photodynamic Bone Stabilization System
PC	Pain Characteristics
PET	polyethylene terephthalate
PI	Principal Investigator
PP	Per protocol
PS	Painful Sites
PT	Preferred Term
ROM	Range of Motion
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SD	Standard Deviation
SOC	System Organ Class
SP	Safety Population
TEAE	Treatment-emergent AE
UADE	Unanticipated Adverse Device Effect
VAS	Visual Analog Scale

LIST OF ABBREVIATIONS

Abbreviation	Term
WHO	World Health Organization

REVISION HISTORY

Version and Date	Changes
Version 1.0: 14 October 2015	Not applicable
Version 2.0: 25 March 2016	• Clarified analysis of primary endpoint • Clarified Enrolled subject to be consistent with protocol • Clarified DBL and CSR activities at all patients complete Day 90 visits, and another DBL and CSR extension by the end of study • Clarified analyses will be conducted at two DBL timepoints • Removed missing data convention for incomplete diagnosis or treatment date • Removed missing date imputation for AE and CM date • Removed prior medication from prior/concomitant medication • Added summary for % change of VAS values from baseline to Day 90 visit
Version 3.0: 18OCT 2016	• Clarify baseline definition, added baseline definition for ROM in Section 4.2. • Clarify that only subjects with non-missing baseline and any post baseline values will be included in the analysis in Section 6.6.1. • Clarify Safety success in 6.7.1.1. Radiographic success will be determined by examining device malfunction, device fracture, device removal recorded on CRF. Safety success will be determined by composite criteria on clinical and radiographic success. Statistical review meeting is removed from the text. Defined evaluable subjects in the safety summary. • Clarify ranges for ROM tests, added percent change from baseline for summary. • Added new analysis comparing mean VAS change from baseline at 90 Day F/U between subjects with actual and impending fracture in section 6.6.1.4

	• Summaries for disability status and return to work status have been removed due to sponsor request
Version 3.1: 15NOV 2016	• Removed "as the date of data extraction" from the definition of evaluable subject in safety success analysis in section 6.7.1.1.

1.0 INTRODUCTION

The purpose of this document is to provide further details about the statistical analysis methods, data derivations and data summaries to be employed in the study protocol 14-03-PATHOLHUM-02: *A Prospective, Multi-Center Study of the IlluminOss® Photodynamic Bone Stabilization System for the Treatment of Impending and Actual Pathological Fractures in the Humerus from Metastatic Bone Disease*. This statistical analysis plan (SAP) has been based on International Conference on Harmonisation (ICH) E3 and E9 guidelines and in reference to protocol version 5.0: date 6 January 2016 and Annotated Case Report Form (aCRF) version 7.0: date 27 January 2016. The statistical analysis plan covers statistical analysis, tabulations and listings of all data including effectiveness and safety data.

Any major deviations from the methods specified in this document and the protocol will be discussed and documented.

2.0 STUDY OBJECTIVES

The primary objective of this study is to evaluate safety and performance data of the IlluminOss Photodynamic Bone Stabilization System (PBSS) when used for the treatment of painful impending and actual fractures of the humerus secondary to metastatic malignancy.

3.0 STUDY DESIGN

3.1 Description of the Disease

Metastases from cancer are the most common malignancy involving the skeletal system.[1] In the United States alone, of the over 1.4M (2011) patients diagnosed with cancer annually, over 700,000 will be diagnosed with metastases to bone in addition to an underlying prevalence of 280,000 cases.[1, 2] Compared to the relatively few primary bone cancers per year, with an incidence of 3,010 (2013)[3], the economic burden of metastatic disease to bone is enormous: consuming $12.6B (2007) in healthcare spend and 17% of the $74B in total direct cost of oncologic care estimated by the National Institutes of Health.[4] Prevalence, incidence, and cost estimates for metastatic bone disease (MBD) are projected to continue growing as with improved medical treatment of many primary cancers, patients are living longer, becoming more likely to develop distant bone metastases.[5]

The skeleton is the third most common target of distant metastases[4], following lung and liver. The axial skeleton is predominantly affected, with MBD of the spine occurring 40 times more frequently than all primary bone tumors combined.[6] After the axial skeleton, MBD affects the femur, humerus, and tibia in decreasing incidence. Please refer protocol section 1.1 for more details.

3.2 Study Design and Population

Study 14-03-PATHOLHUM-02 is a prospective, multi-center, open label study to evaluate the PBSS in the treatment of impending and actual pathological fractures of the humerus for the purposes of Food and Drug Administration (FDA) marketing clearance.

The investigation population consists of skeletally mature adults, suffering from pain due to impending and actual pathological fractures of the humerus secondary to confirmed metastatic malignancy.

A subject is considered enrolled in the clinical investigation after the following two conditions are met:

- The patient has provided informed consent
- The sheath assembly used to guide the balloon in place within the intramedullary canal has entered the body.

Patients who fail one or more of the eligibility criteria prior to this point are considered screening/enrollment failures and should not be enrolled in the study. If they fail eligibility criteria prior to surgery then they will not be recorded in the study database.

After obtaining informed consent, the screening procedures will be performed at the screening visit as per the schedule of activities (Appendix 1). A total of up to 80 subjects will be recruited from up to 20 sites in the United States.

The enrollment period is expected to be approximately 12 months. Subjects will be followed for 90 days after treatment and then followed in an extended follow up phase for up to 360 days.

Once the subject has been enrolled in the study, the subject may withdraw consent to participate in the study at any time without prejudice. Participation in this clinical investigation is entirely voluntary. The Principal Investigator (PI) may prematurely discontinue any subject's participation in the study if the PI feels that the subject can no longer fully comply with the requirements of the study or if any of the study procedures are deemed potentially harmful to the subject. If a subject prematurely discontinues from the study, the reason for study termination will be recorded, if available. If termination was the result of an AE or death, completion of the appropriate AE forms and/or notifications will ensue. Whenever possible, an exit/final visit examination will be conducted prior to termination from the study. If possible, permission for contacting the subject for the assessment of long-term outcomes will be obtained. All reasonable efforts will be made to obtain complete data for all subjects.

3.3 Study Treatments and Assessments

PBSS is the only device used in this study. Neither placebo nor any comparator is used. The PBSS is comprised of an inflatable, thin walled polyethylene terephthalate (PET; Dacron™) balloon mounted on an insertion catheter. This balloon catheter system is designed to deliver the monomer cement to the fracture site via the medullary canal of the bone. The study treatment is a onetime procedure and surgery will be done on visit 2. The follow-up evaluations will be scheduled for 7-14 (visit 3), 30 (visit 4), 90 (visit 5) days post-index procedure and the extended follow up at 180 (visit 6) and 360 (visit 7) days post-index procedure.

Throughout the study, subjects will be assessed for the study objectives as outlined in the schedule of activities (Appendix 1). The visit window details are outlined in the Appendix 2.

3.4 Randomization and Blinding

This study is an open-label, prospective clinical trial. There is no randomization of subjects or blinding.

3.5 Sample Size Justification

Approximately 80 subjects will be enrolled into this study. The following text will give sample size justification details.

VAS pain improvement: The primary endpoint is change in VAS from baseline to Day 90. All patients will be included in the primary analysis of this endpoint through the use of a mixed model repeated measures (MMRM) model assuming an unstructured covariance matrix. To be conservative, statistical power is evaluated on the basis of a single-sample t-test with a 1-sided alpha=0.05 based on the change at Day 90. The null and alternative hypotheses of interest are:

Ho: mean improvement in VAS at Day 90≤ 53.8

H_a: mean improvement in VAS at Day 90> 53.8

Based on the historical control data in protocol Section 6.3, an estimate of the standard deviation of the primary endpoint is 13.3. Under the assumption this is the true standard deviation, under the assumption the true mean improvement from baseline VAS at Day 90 is 58, then 68 evaluable patients yields over 80% power to reject the null hypothesis in favor of the alternative. In order to account for 15% premature withdrawal rate, a total of 80 patients will be enrolled. Note that the assumption of 58 for the true mean improvement from baseline VAS at 90 days may be conservative given the historical control data in protocol Section 6.3 (where the overall mean change, or reference value, was estimated to be approximately 67, over a shorter period than 90 days).

MSTS improvement: The second primary endpoint, to be tested if the above null hypothesis for improvement in VAS is rejected, is change in MSTS from baseline at Day 90; this will be analyzed a similar MMRM manner as the primary endpoint above. The null and alternative hypotheses of interest for this endpoint are:

H_o: mean improvement in MSTS at Day 90 ≤ 23.7

H_a: mean improvement in MSTS at Day 90 > 23.7

The hypothesis regarding mean improvement in MSTS will be tested conditionally on rejection of the null hypothesis concerning VAS pain improvements (1-sided $p<0.05$). Based on a single-sample t-test (1-sided $\alpha=0.05$), assuming that the true standard deviation at each visit is equal to 20.8, assuming a sample size of 68 (accounting for potential LTF), then at least 80% power will be achieved to reject the null hypothesis that the mean improvement in MSTS <23.7 if the true mean change is at least 30.3 at Day 90. This value is similar to the mean improvement for the Kim historical control (mean = 29.6). Thus, as long as the true mean improvement across post-baseline visits is similar to that the mean improvement

reported in Kim, there will be good statistical power to reject the conditional hypothesis concerning the co-primary endpoint of improvement in upper arm function as reflected in change in MSTS at Day 90.

4.0 STATISTICAL CONSIDERATIONS

4.1 General Considerations

This study has primary analyses at Day 90 visit, followed by a long follow-up period from Day 180 to Day 360 visits. The efficacy and safety endpoints will be collected at baseline and at each post baseline visit, including Day 7-14, Day 30, Day 90, Day 180 and Day 360 visits. The database will be locked once all data collected through Day 90 have been entered, cleaned and declared complete and final. The Clinical Study Report (CSR) will be finalized after results from the primary endpoints up to Day 90 visits are complete; all available data post-Day 90 pertaining safety and efficacy endpoints will be provided. At the completion of the study, the CSR will be amended to include results from the long-term safety follow-up period (data collected at Day 180 and Day 360 visits). No statistical analyses will be performed with efficacy endpoints collected at Day 180 and Day 360 visits, they will be summarized descriptively instead.

Statistical analyses will be performed using SAS software, version 9.2 or higher. Formal hypothesis will be tested in this study and it will be explained in detail on following sections. The Intent-to-Treat (ITT) population will be used as the analysis population for the primary endpoints, secondary endpoints, and safety endpoints. Per-protocol (PP) population will be used for the secondary analysis of the primary endpoints.

Continuous data will be summarized using number (n), mean, standard deviation (SD), minimum value, median, maximum value and, if appropriate, number of missing values. Unless noted otherwise, summaries will be produced by visit (where applicable). If more than one value is reported at a scheduled visit, the value collected closest to the intended visit date will be used. In the case of a tie, the latest version will be summarized. If there are more than one value reported at the same date, take the average value of multiple values.

Categorical data will be summarized using number (n) and percentages. Percentages will be calculated based on the number of non-missing values. The number of missing values will be presented as a separate category with no percentage, if 1 or more subjects have missing data for the summary. Otherwise, all categories will be presented (even if no subjects are counted in the category). Counts of zero in any category will be presented without percentage.

Precision of summary statistics:

- Integer – Sample size (n, N) and number of missing responses (if displayed).
- One additional decimal place than reported/collected on the case report form (CRF) – mean and median
- Two additional decimal places than reported/collected on the CRF – standard deviation.
- Same number of decimal places as reported/collected on the CRF – minimum, maximum.
- Percentages – one decimal place.

- P-values will be presented for 3 decimal points (ie, 0.xxx).

Precision of output display:

- Report output will be generated using SAS (version 9.2 or higher) ODS RTF with no borders or framing around table elements.
- Font – Courier New font with minimum of 9 point font size
- Margins – For US, minimum of 3/4" bound edge margin and 3/8" other margins on 8.5"x11" paper.

All data will be presented on listings in order of subject, assessment date/time and assessment (in order collected on Electronic Case Report Form (eCRF), unless specified otherwise). Dates will be presented in the format DDMMMYYYY. Times will be displayed in 24-hour clock format. Numbering of tables, figures and listings will follow ICH E3 guidelines.

Unscheduled measurements will be excluded from the descriptive statistical analysis. All the unscheduled measurements will be included in the listing.

4.2 Specification of Baseline Values

If not stated otherwise, baseline values are defined as the last non-missing assessment result collected prior to the date of surgery. For subjects who have Screening and surgery visits on the same day, the values from the date of surgery will be considered for baseline. For range of motion tests, baseline values are defined as the measurements at 7-14 Day F/U visit since they were not collected at the Screening visit or on day of surgery.

4.3 Multiple Comparisons and Multiplicity

The primary effectiveness hypothesis concerns improvements in VAS pain at Day 90. Only if the null hypothesis is rejected relative to historical control with $p<0.05$ will testing for improvement in MSTS function occur. Pre-specification of the order of testing within this hierarchical framework eliminate type 1 error inflation due to multiplicity. Therefore, no multiplicity adjustment will be made across the co-primary endpoints.

5.0 ANALYSIS POPULATION

The populations used for analysis will include ITT population and per-protocol (PP) population. The success of the primary effectiveness endpoints will be evaluated based on the ITT population; these endpoints will also be analyzed for the PP population. The evaluation of secondary and safety endpoints analysis will be based on ITT population.

5.1 Intent-to-Treat Population (ITT)

The Intent-to-Treat (ITT) analysis population is defined as all subjects who provide informed consent, undergo surgery, and have an implantation of the IlluminOss Photodynamic Bone Stabilization System into the intramedullary canal; this includes subjects for whom implantation of the IlluminOss device is attempted but fails; such patients will be followed and included in the

ITT population. The ITT population is equivalent to the Full Analysis Set (FAS) and enrolled population.

5.2 Safety Population (SP)

Since safety population (SP) and ITT for this study are equivalent, this SAP addresses only analyses performed on the ITT population.

5.3 Per-protocol Population (PP)

The per-protocol (PP) analysis population is defined as all ITT subjects who achieved the 90-day visit.

6.0 METHODS OF ANALYSES AND PRESENTATIONS

The following sub-sections would be considered for logical presentation of study data. In all cases, listings of data will include the enrolled (ITT) population unless otherwise indicated.

6.1 Subject Disposition

The number of subjects who underwent surgery, who were intra-operative screen failures, enrolled, and with device implanted as well as study completion status will be summarized. Subjects who discontinued from the study will be summarized with reason and listed by their primary reason for discontinuation. Patients from ITT and PP populations will be summarized using descriptive statistics displaying number and percentage. Inclusion/exclusion criteria evaluation will be listed. These summaries will include all patients who underwent surgery.

6.2 Protocol Deviations

Protocol deviation will be collected on CRF page. A summary of protocol deviations will be provided by visit and type. All protocol deviations collected on CRF will be listed. All enrolled patients will be included in the summary.

6.3 Demographic and Baseline Characteristics

Demographic data and subject characteristics at baseline will be summarized using descriptive summary statistics for ITT population. The demographic and baseline characteristics of age, sex, ethnicity, race, work status, occupation, disability status, history of pain, smoking history, hand dominance, primary cancer type and fracture characteristics if evident will be summarized. All demographic and baseline characteristics will be listed in full.

6.4 Medical History

All medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The incidence of medical history abnormalities will be summarized using descriptive statistics by system organ class (SOC) and preferred term (PT) for the enrolled (ITT) population. Patients are counted only once in each PT and SOC category. Summary table will be sorted in alphabetical order by SOC and in decreasing frequency of PT based on total number of incidence within each SOC.

Additionally, the incidence of primary cancer diagnosis will be summarized using descriptive statistics.

Medical history and primary cancer diagnosis will also be displayed in patient listings.

6.5 Concomitant Medication/Procedures

Concomitant medication and concomitant procedures are collected at each visit. Medications taken for a surgical procedure should also be included as concomitant medication.

All concomitant medications and concomitant procedures will be listed as recorded on CRF.

6.6 Effectiveness Endpoints and Analyses

The primary effectiveness analysis will be based on both ITT and PP population. Subgroup analyses with primary effectiveness endpoints will be based on ITT population. The secondary effective analysis will be based on ITT population. Descriptive summary statistics mentioned in the section 4.1 will be performed for all data used within the context of the effectiveness analysis models. Summaries will be presented by visit. Comprehensive listings of all data will be presented.

6.6.1 Primary Effectiveness Endpoints and Analyses

Primary endpoints of this study are pain reduction and functional improvement. Pain reduction will be assessed by using VAS pain score and functional improvement will be assessed by using MSTS score. The pain VAS is a single-item scale. It is assessed on scale range from 0-100mm with 0 means “no pain” and 100 means “pain as bad as it could be” or “worst imaginable pain”. MSTS score is a six-item scale. It ranges from 0-30. Higher score is associated with better function.

6.6.1.1 Pain Reduction

The primary effectiveness hypothesis concerning VAS pain improvements is:

H_o: Mean improvement in VAS at Day 90 ≤ 53.8 [ie. 80% of reference (inferior)]
H_a: Mean improvement in VAS at Day 90> 53.8 [ie. 80% of reference (not inferior)]

Further information on the reference value can be found in protocol Section 6.3.

A mixed model repeated measures (MMRM) model will be executed for the dependent variable of “change from baseline at each visit”, with the categorical main effect of visit (Day 7-14, Day 30, Day 90) as the independent variable. From this model, the estimate of the mean change from baseline at Day 90 and its standard error, will be established, and from this a one-sample t-test will be generated to test the null hypothesis of interest at a one-sided 0.05 level of significance. The MMRM model will be estimated using a direct likelihood approach as implemented in the SAS procedure PROC MIXED. The model will include the baseline value of the score as an additional covariate. The model is designed not only to account for correlations among responses over time, but to use these correlations to implicitly impute missing values from the non-missing values. Inferences based on this approach are unbiased

under the assumption of MAR (missing at random) which is a more generally true assumption than MCAR (missing completely at random)[7].

Unless otherwise specified, the MMRM will have the following characteristics:

- All subjects in the ITT analysis set will be included in the primary effectiveness test as long as there are non-missing baseline and any post baseline (days 7-14, 30, and 90) values.
- The response variable will be the change VAS score from baseline to each post-baseline (days 7-14, 30, and 90) visit. Only post-baseline measurements collected up until the day 90 visit will be considered response values.
- Repeated post-baseline measurements from each subject will be identified by subject identifier.
- Within-subject correlations will be modeled using an unstructured covariance matrix structure. In the unlikely situation that this model does not converge (i.e., the study has too few observations for the number of parameters estimated), the model will use a compound symmetry structure which assumes equal correlation for a subject's measurements, regardless of how far apart in time they were taken.
- The following covariate will be used in the models:
  - Visit as defined by visit windows
  - Baseline VAS pain score
- The following SAS statements will be employed:

```
proc mixed method=ml;
        class visit subjid;
        model CHGvas = visit baseVAS/ ddfm=kr;
        repeated visit / type=un subject=subjid r rcorr;
        lsmeans visit / diff adjust=GT2 adjdfe=row;
        lsmestimate visit "mean improvement in VAS at Day 90" 0 0 1/ lower
testvalue=-53.8;
ods output lsmestimates=lsmestimates;
run;
```

where the variable CHGVAS contains the change from baseline VAS score for that patient-visit, VISIT is a categorical variable which identifies the Day 7-14, Day 30, and Day 90 assessments, BASEVAS is the baseline VAS, and SUBJID is the subject ID.

Least square mean estimate for change of VAS from baseline to Day 90 will be obtained from LSMESTIMATE statement, along with its corresponding 95% confidence intervals. In case unstructured covariance structure dose not converge, use type=cs to replace type=un in repeated statement.

Baseline VAS values, observed VAS values, change from baseline values and percent change from baseline values at post-baseline visits will be summarized.

6.6.1.2 Functional Improvement

The primary effectiveness hypothesis concerning MSTS functional improvements is:

H_o: mean improvement in MSTS at Day 90$\leq$ 23.7 [ie. 80% of reference (inferior)]

H_a: mean improvement in MSTS at Day 90> 23.7 [ie. 80% of reference (not inferior)]

The MMRM analysis and data summarization MSTS score will be similar to that described for the VAS pain score; for information on the reference value, please see protocol Section 6.3. Testing for pain reduction will be performed before the evaluation of the MSTS endpoint. Only if the null hypothesis of inferiority in mean VAS pain improvement is rejected at $p<0.05$, will the co-primary endpoint MSTS functional improvement be similarly tested.

6.6.1.3 Subgroup Analyses

Whether subject has actual fracture or impending fracture of the target humerus is recorded on CRF page. A subgroup analysis of subjects who had actual fracture and impending fracture will be performed for VAS and MSTS based on the statistical methods specified for these primary effectiveness endpoints, see sections 6.6.1.1 and 6.6.1.2. Subgroup analyses with MSTS will be performed only if the null hypothesis of inferiority in mean VAS improvement is rejected.

6.6.1.4 Other Analyses

A MMRM model comparing mean VAS change from baseline to Day 90 visit between subjects with actual and impending fracture will be performed. The following SAS statements will be employed:

```
proc mixed method=ml;
        class visit fracture_type subjid;
        model CHGvas = baseVAS visit fracture_type  visit*fracture_type
baseVAS*visit baseVAS*fracture_type/ ddfm=kr;
         repeated visit / type=un subject=subjid;
         lsmeans visit*fracture_type/ cl pdiff;
         ods output diffs=diffs lsmeans=lsmeans;
run;
```

where the variables CHGVAS, VISIT, BASEVAS, and SUBJID are the same as those in section 6.6.1.1. FRACTURE_TYPE is a categorical variable for pathological fracture types (Actual vs. Impending). Visit*fracture_type, baseVAS*visit, and baseVAS*fracture_type are the two-way interaction terms among VISIT, FRACTURE_TYPE and BASEVAS.

Least square mean estimates for change of VAS from baseline to each post-baseline visit for subjects with actual and impending fractures will be obtained from ODS table LSMEANS. The difference of mean change of VAS from baseline to each post-baseline between subjects with actual and impending fractures, the SE, 95% CI, and p value from t distribution will be obtained from ODS table DIFFS.

6.6.2 Secondary Effectiveness Endpoints and Analyses

6.6.2.1 Procedure- and Device-related Complication Rate

When there is device performance issue/Malfunction, the date and time period when issue occurred will be recorded on CRF. Time period can be "During device preparation", "During device implantation", or "Post-Implantation". Date/Time, time period, and specific reason on device malfunction will be listed. The number and percentage of subjects who experience a procedure- or device-related complication by the day 90 visit will be presented along with the corresponding 95% binomial confidence interval. Subjects with more than one device malfunction will be counted only once. The percentage of subjects who experience the complications by the day 180 and 360 visits will also be summarized.

6.6.2.2 Duration of Index Procedure and Length of Hospital Stay

Date of surgery, start time and end time of procedure, date of admission, date of discharge, time of discharge are recorded on CRF. Duration of index procedure will be calculated as (End time of procedure - Start time of procedure) and summarized in hours. Length of hospital stay will be calculated as (Date of discharge - Date of admission) +1 and summarized in days. Duration of procedure and length of hospital stay during reoperation will also be summarized.

6.6.2.3 Return to Pre-fracture Mobility

Subject's pre-fracture mobility ("independent ADLs" or "ADLs with assistance") will be captured at Visit 1 as baseline mobility. At each post-baseline visit (Visits 2-7), subjects are asked whether they have returned to pre-fracture mobility in the index arm. Possible replies are "Yes" or "No". Shift table from pre-fracture mobility to each post-baseline visit will be summarized using counts and percentages.

6.6.2.4 No Pain on Palpation

Pain on palpation (Yes/No/Not done) is collected at each visit along with Grade (Mild/Moderate/Severe) and whether the pain is clinically significant (Yes/No). This information will be summarized.

6.6.2.5 MSTS Upper Extremity Functional Outcome

The MSTS upper extremity functional score evaluates the functional status of patients with impending and actual fractures of the humerus. In the MSTS system, numeric values (0 to 5) are assigned to each of the six categories of pain, function, emotional, hand positioning, manual dexterity, and lifting ability. They give a total score of between 0 and 30 with 0 indicating poor and 30 good. The total MSTS score will be recorded on CRF. It will be divided by 30 and multiplied by 100 to facilitate interpretation.

Baseline values, observed values, change from baseline values, and percent change from baseline values at post-baseline visits, will be summarized.

6.6.2.6 Range of Motion

At each visit, five tests for range of motion are performed: Flexion, Extension, Abduction, Lateral Rotation, and Medial Rotation. For each test, active and passive range of motion values are collected for both affected arm and unaffected arm. The observed test values range from 0 to 180, 0-180, 0-180, 0-100, 0-120 degrees for flexion, extension, abduction, lateral rotation, and medial rotation tests, respectively. Subjects will be classified into different categories based on the change of range of motion values from baseline: <0, >=0 and <=10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, …. degrees. The last category for flexion, extension, abduction, lateral rotation, and medial rotation tests are 171-180, 171-180, 171-180, 91-100, and 111-120, respectively. Number and percentage of subjects in each category will be summarized by visit and test. Range of motion test values, change from baseline values, and percent change from baseline values will be summarized by visit and test. Within each test the values will be summarized by active or passive and affected or unaffected arm.

6.6.3 Optional Analyses

All analysis and summaries listed in this section will be performed in ITT population. Data identified in this section will be displayed in subject listings.

6.6.3.1 Fracture Location/Type of Primary Cancer

Analysis of primary effectiveness endpoints as described in Sections 6.6.1.1 and 6.6.1.2 will be performed on subgroups in addition to those described in Section 6.6.1.3. These subgroups will be defined by fracture location and by type of primary cancer. Fracture location is recorded on CRF page with three categories: proximal, diaphyseal, and distal fractures. CRF recorded primary cancer type will be grouped into different categories based on the MedDRA High Level Group Terms (HLGT).

6.6.3.2 Activities of Daily Living Score

The activities of daily living will be assessed by EORTC-QLQ-BM22, which is a 22-item questionnaire assessment composed of multi-item measures. It has been divided into the following categories: Painful Sites (PS), Pain Characteristics (PC), Functional Interference (FI), and Psychosocial Aspects (PA) (Table 1).

Table 1. EORTC QLQ-BM22 Scales

Scale Name	**Number of Items**	**Items Numbers**	**Item Range**
Symptom scales			
Painful Sites (PS)	5	1-5	3
Pain Characteristics (PC)	3	6-8	3

Functional scales			
Functional Interference (FI)	8	9-16	3
Psychosocial Aspects (PA)	6	17-22	3

Range is defined as the maximum and minimum scale values for a particular item. Since all items in BM22 questionnaire have maximum of 4 and minimum of 1, the range for all items are 3. All of the scales will be transformed to range in score from 0 to 100. A high score for the symptom scales represents a high level of symptomatology or problems, whilst a high score for the functional scales represents a high level of functioning.

With regards to the positive phrasing of questions 21 and 22, the scoring must be reversed prior to statistical analysis.

The detailed scoring procedure for the EORTC QLQ-BM22 is outlined below:

PS: Painful Sites Scale (items 1-5)

- Compute the raw score (mean scale score) if at least 3 of the items have a valid score. XPS = Mean of (Q1 – Q5)
- Carry out a linear transformation to convert to a 0-100 scale: PS = {(XPS - 1) / 3} * 100
- If fewer than 3 of the items have a valid score treat the scale as missing.

PC: Painful Characteristics Scale (items 6-8)

- Compute the raw score (mean scale score) if at least 2 of the items have a valid score. XPC = Mean of (Q6 – Q8)
- Carry out a linear transformation to convert to a 0-100 scale: PC = {(XPC - 1) / 3} * 100
- If fewer than 2 of the items have a valid score treat the scale as missing.

FI: Functional Interference Scale (items 9-16)

- Compute the raw score (mean scale score) if at least 4 of the items have a valid score. XFI = Mean of (Q9 – Q16)
- Carry out a linear transformation to convert to a 0-100 scale: FI = {1 - (XFI - 1) / 3} * 100
- If fewer than 4 of the items have a valid score treat the scale as missing.

PA: Psychosocial Aspects Scale (items 17-22)

- Reverse observed score for item 21-22 by 5 subtracted by Q21 or Q22
- Compute the raw score (mean scale score) if at least 3 of the items have a valid score. XPA = Mean of (Q17 – Q22)

- Carry out a linear transformation to convert to a 0-100 scale: PA = {1 - (XPA - 1) / 3} * 100
- If fewer than 3 of the items have a valid score treat the scale as missing.

Observed values and changes from baseline for the EORTC-QLQ-BM22 two symptom scales (PS, PC) and two function scales (FI, PA) will be summarized by visit.

6.6.3.3 Assessment of Post-Surgery Physical Therapy, Supportive Orthopedic Device and Analgesic Medication Use

At surgery visit and each post-surgery visit, subjects are asked whether physical therapy, supportive orthopedic device, or analgesic medication is prescribed. The possible answers for physical therapy and supportive orthopedic device are "Yes" or "No". The possible answers for analgesic medication are "None", "Over-the-counter", and "Prescription". At each post-surgery visit, subjects are asked whether they completed physical therapy, supportive orthopedic device, or take analgesic medication since last visit. The possible answers are the same as at surgery visit. The prescription status for physical therapy, supportive orthopedic device, and analgesic medication at surgery and each post-surgery visit will be summarized using counts and percentages. Subjects who are prescribed with "Over-the-counter" or "Prescription" for analgesic medication will be counted as "Yes" to receive analgesic prescription at each visit.

6.6.3.4 Survivability from Time of Index Procedure to Death

Overall survival is defined as (date of death due to any cause or date of last contact – date of surgery). A Kaplan-Meier plot for overall survival will be presented. Subjects will be censored based on their last date in the study if the event of death is not identified.

6.7 Safety Endpoints and Analyses

All safety endpoints will be summarized for ITT population. The safety endpoints will be evaluated through Day 90 and for the extended follow up portion of the trial at Days 180 and 360.

6.7.1.1 Safety Success (Clinical and Radiographic)

Safety Success is evaluated according to a composite endpoint by meeting all of the following criteria:

- Clinical
    - No Serious Device Related Complications
    - No additional surgical interventions:
        - Revisions, supplements, fixations, or removals

If subject does not have device related SAE or no reoperation CRF page filled, he/she can be considered as clinical success, meeting both criteria: "No Serious Device Related Complications" and "No additional surgical interventions: Revisions, supplements, fixations, or removals".

- Radiographic
    - No device fracture, migrations, mal-alignment or loss of reduction or fixation

If subject does not have device malfunction (fracture, migration, other) recorded on the CRF, subject can be considered as radiographic success, meeting radiographic criteria.

Subject achieves safety success if subject meets both clinical and radiographic success. The number and percentage of subjects achieving the safety success endpoint will be reported cumulatively at Day 7-14, Day 30, and Day 90, as well as at each of the extended follow up periods and with 95% 2-sided exact binomial confidence intervals. Only evaluable subjects will be included in the summary at each time point. Evaluable subject is defined as subject staying in the study long enough to reach the time point.

6.7.1.2 Adverse Events (AEs)

Adverse events (AEs) will be collected from the time of surgery until the last study visit or premature discontinuation from the study.

Treatment-emergent adverse events (TEAEs) are those AEs which worsen in severity on or after date of surgery or with onset date on or after date of surgery. AEs with missing start dates, but with stop dates either overlapping into the treatment period or missing, will be considered TEAEs. Treatment-related includes possibly or probably or definitely relationship of the event to the investigational device, procedure, anesthesia, or oncology medicine.

All AEs will be coded by System Organ Class (SOC) and Preferred Term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA). Incidences of TEAEs will be tabulated by SOC and PT, cumulatively per time point (Up to Day 90, Up to Day 180, and Up to Day 360). Incidences of TEAEs will also be presented by maximum severity, seriousness, unanticipated adverse device effect (UADE) and relatedness (device or procedure related) cumulatively per time point. This includes the reporting of device revisions, reoperations, removals, supplemental fixations, or other procedures. Subjects who discontinued due to TEAEs will be summarized by SOC and PT in a separate table. Patients are counted only once in each SOC, and only once in each PT category, cumulatively per time point.

The following tables will be developed for summarizing AEs cumulatively per time point.

- Overall summary of TEAEs, it includes number of treatment-emergent adverse events, subjects with any TEAE, subjects with any serious adverse events (SAE), subjects with any severe TEAEs, subjects with any device related TEAEs, subjects with any procedure related TEAEs, subjects with any anesthesia related TEAEs, subjects with any Oncology Medicine related TEAEs, subjects with any unanticipated adverse device effect (UADE), subjects with TEAEs leading to study discontinuation and death.
- Summary of TEAEs by system organ class and preferred terms
- Summary of serious TEAEs by system organ class and preferred terms
- Summary of TEAEs by system organ class, preferred terms and maximum severity

- Summary of treatment-related TEAEs by system organ class, preferred terms (device related, procedure related, anesthesia related and oncology medicine related).
- Summary of unanticipated adverse device effect (UADE) by system organ class and preferred terms
- Deaths or discontinuations due to TEAEs by system organ class and preferred terms

All AEs, SAEs and deaths will be listed.

6.7.1.3 Physical exam

Physical examination results will be displayed in listing.

7.0 HANDLING OF MISSING VALUES

In general, for all variables only the observed data from the patients will be used in the analysis.

All patients in the ITT analysis set will be included in the primary effectiveness test as long as there are non-missing baseline values and any post baseline (days 7-14, 30, and 90) values. The MMRM provides implicit imputation of missing post-baseline values through the covariance matrix reflecting the associations among responses within patient over time.

Missing data handling for the EORTC-QLQ-BM22 is described in Section 6.6.3.2.

No imputation of values for other missing data will be performed. Data that are potentially spurious or erroneous will be examined according to standard data management operating procedures. Adverse events with missing severity will be summarized as severe. Adverse events with missing causal relationship will be summarized as related. Missing safety data will not be replaced.

8.0 INTERIM ANALYSIS

An interim MMRM analysis on change from baseline VAS to Day 90 will be carried out. In an effort to have 30 evaluable patients with VAS data at 90 days, the first 37 patients to enroll in the study will be selected for inclusion in the interim analysis. The interim analysis will be conducted once those 37 enrolled patients have all either discontinued from the study prior to the 90 day visit or have reached the 90 day visit. The purpose of this interim analysis is not to stop the trial for overwhelming efficacy, but rather to (a) potentially stop for futility; and (b) potentially increase the sample size if the mean VAS improvement from baseline to Day 90 is large but not as large as anticipated in the original sample size calculations. The interim analysis will be reviewed and validated by an independent statistician not otherwise involved in the study, who will report to the Sponsor of whether or not a sample size increase is needed, but with no other details regarding the interim results.

The null and alternative hypotheses of interest are

$$H_o: \text{mean improvement in VAS at Day 90} \leq 53.8$$

$$H_a: \text{mean improvement in VAS at Day 90} > 53.8$$

At the interim stage, the conditional power for rejecting the above null VAS hypothesis by the time of the final analysis (a planned 68 evaluable patients) will be calculated, conditioned on the interim observed data. Patient is evaluable if he/she has VAS measurement at both baseline and Day 90 visit. The conditional power will be calculated using 1000 Monte Carlo simulations via a SAS program (see the appendix 3 for the SAS program, embedded with comments describing the program). The simulation of conditional power requires an assumption of (a) the true mean change in baseline VAS at each visit; (b) the true standard deviation of the change from baseline VAS at each visit; and (c) the true within-patient correlation of VAS. These values will be estimated from the interim observed sample, and these estimated values will be used as the assumption of the true values in the conditional power simulation. Once the interim observed sample is obtained, 1000 post-interim samples will be generated assuming that 38 (i.e. 68-30) evaluable patients will be enrolled post-interim (since, again, the total evaluable sample size is 68 patients, and assume that 30 evaluable patients among the first 37 patients are in the interim analysis, actual number of evaluable patients in interim analysis will be determined with actual interim data), and each of these simulated samples will be added to the one interim observed sample to create 1000 simulated complete-study datasets of 68 evaluable patients. On each simulated dataset, the following SAS statements will be used to test the above null hypothesis:

```
proc mixed method=ml;
        by simulation;
        class visit subjid;
        model CHGvas = visit baseVAS/ ddfm=kr;
         repeated visit / type=un subject=subjid r rcorr;
        lsmeans visit / ;
         lsmestimate visit 0 0 1/lower testvalue=-53.8;
run;
```

where the input dataset has 3 records per patient (one record per visit) where the variable CHGVAS contains the change from baseline VAS score for that patient-visit, VISIT is a categorical variable which identifies the Day 7-14, Day 30, and Day 90 assessments, baseVAS is the baseline VAS, and SUBJID is the subject ID. The interim estimated mean change from baseline VAS at Day 90, the test-statistic, and the corresponding one-sided p-value for the above-mentioned null hypothesis are obtained from the "LSMESTIMATE" statement. The proportion of simulated datasets for which the one-sided p-value is below 0.05 is the simulated conditional power. Note that the above PROC MIXED statements will be used for the final analysis once all data are collected.

If the above simulated conditional power is ≥ 80% when the final evaluable sample size for the simulated samples is equal to the pre-planned number of evaluable subjects (68), the study will continue as is. If the simulated conditional power is between 10% and 50%, the study will also continue as is. If the simulated conditional power is <10%, the study may be stopped for futility. If, however, the conditional power is between 50% and 80%, then the evaluable sample size

increase required to achieve 80% simulated conditional power will be calculated using the Chen-DeMets-Lan (CDL) approach[8]. Specifically, if the simulated conditional power is between 50% and 80% with a total sample of size 68 evaluable patients, then the evaluable sample size increase (from beyond the pre-planned 68) required to achieve a simulated conditional power of 80% will be determined by plugging in increasing values of post-interim sample sizes into the conditional power program, and determining the sample size at which the conditional power achieves 80%.

Under the above CDL approach, an "alpha-penalty" does not need to be made for any sample size increase in this setting as long as the maximum sample size increase required to maintain 80% power is below the bound 100*R% of the original sample size, where R is calculated as follows:

$$\sqrt{1+\mathrm{R}}\left(\sqrt{1+\mathrm{R}}-1\right)/\sqrt{1+\mathrm{R}-\mathrm{s}} = \mathrm{t}_{0.80,\mathrm{df}}/\mathrm{t}_{0.95,\mathrm{df}}$$

where s is the proportion of information at the interim analysis (s = 30/68 = 0.441), t0.80,df is the 80th percentile of the tdf distribution (to correspond to the desire to achieve 80% conditional power) and t0.95,df is the t critical value corresponding to a one-sided alpha of 0.05 (or i.e., the 95th percentile of the tdf distribution), where df = final degrees of freedom for the final t-statistic obtained from the PROC MIXED model. The df is unknown at this time as it depends on the interim observed sample and on the simulated data, but it will be between 50 and 70. To be conservative, the df for the numerator of the above equation will be set to 70 and the df for the denominator will be set to 50. The value of t0.80,df and t0.95,df are then 0.84 and 1.68, respectively. In that case, R will be 1.08, so as long as the final evaluable sample size required to achieve 80% simulated conditional power is <108% larger than the original pre-planned evaluable sample size of 68 patients (or i.e., as long as the final evaluable sample size is <142), the evaluable sample size may be increased to maintain 80% conditional power, without requiring a penalty on the one-sided alpha, at the final analysis.

The analyses described in section 6.6.1.1 will be generated for interim analysis.

9.0 CHANGES FROM ANALYSIS METHODS PLANNED IN THE PROTOCOL

The following changes from protocol were implemented in this SAP:

1. Secondary endpoints have been identified per Section 7 of the protocol. This prioritization may vary from the study synopsis description.
2. Health economic data in protocol section 7.9.2 has been summarized in SAP sections 6.6.3.3.
3. Protocol section 7.8.6 specified "Range of motion will be assessed using ordinal variables reflecting the forward flexion, lateral elevation, external rotation, and internal rotation of the arm." In order to reflect data collected on CRF, the range of motion measures in section 6.6.2.6 have been modified to "At each visit, five tests for range of motion are performed: Flexion, Extension, Abduction, Lateral Rotation, and Medial Rotation."

4. Protocol section 6.3 specified that "The values of the MSTS range from 0 to 30. The values have been divided by 30 and multiplied by 100% to facilitate interpretation." In SAP section 6.6.2.4, MSTS score is specified to be divided by 30 and multiplied by 100 to be consistent with hypothesis testing value.
5. Protocol section 7.10 listed z critical values for calculating R, and the proportion of information was 0.5. In SAP section 8, these have been updated with using t critical values and the proportion of information being 0.441.
6. Protocol section 7.10 stated that interim analysis will be performed when 30 evaluable subjects have been treated and followed for 90 days. In SAP section 8.0, this has been updated to "In an effort to have 30 evaluable patients with VAS data at 90 days, the first 37 patients to enroll in the study will be selected for inclusion in the interim analysis. The interim analysis will be conducted once those 37 enrolled patients have all either discontinued from the study prior to the 90 day visit or have reached the 90 day visit."
7. SAP section 6.6.1.4 added a MMRM model to compare mean change from baseline of VAS score for subjects with actual fracture type with those with impending fracture type. This analysis was not originally included in the protocol.
8. Protocol section 7.8.5 specified "The number and percentage of subjects who are considered disabled, per Investigator assessment, will be presented at each visit where disability status is collected." Summary of disability status is removed from the SAP due to sponsor request.
9. Protocol section 7.8.7 specified "The number and percentage of subjects who returned to work, and in what capacity (e.g., full-time, part-time, with limitations) will be presented at each follow-up visit". Summary of return to work status is removed from the SAP due to sponsor request.
10. Protocol section 8.1.2 specified that the baseline values are collected at Screening and baseline visit prior to index treatment. In SAP section 4.2, baseline values for ROM are defined as measurements collected at Day7-14 visit due to no data collected at Screening and surgery visits.

10.0 REFERENCES

1. Shuling Yi, Yi Peng, Arneson TJ. Estimated number of prevalent cases of metastatic disease in the US adult population. *Clin Epidemiol.* 2012;4:87-93.

2. Torbert JT, Lackman RD. (2011) Pathological Fractures. Pignolo RJ, Keenan MA, Hebala NM. *Fractures in the Elderly: A Guide to Practical Management* (12th Ed., pp. 43 – 52). Springer.

3. Surveillance, Epidemiology, and End Results (SEER) Program Database. National Institutes of Health. Accessed 7 December 2013: http://seer.cancer.gov/statfacts/html/bones.html

4. Schulman KL, Kohles J. Economic burden of metastatic bone disease in the U.S. *Cancer.* 2007;109(11):2334-2342.

5. National Program of Cancer Registries Database. Center for Disease Control and Prevention. Accessed: 7 December 2013: http://apps.nccd.cdc.gov/uscs/

6. Harrington KD. Metastatic disease of the spine. In: Harrington KD, ed. Orthopedic management of metastatic bone disease. St. Louis: Mosby, 1988:309-383.

7. Rubin DB (1976), "Inference and missing data." Biometrika, 63:581-592.

8. Chen YHJ, DeMets DL, Lan KKG (2004). "Increasing the sample size when the interim result is promising." Statistics in Medicine. 23: 1023-1038

11.0 APPENDIX

Appendix 1: Schedule of Activities

	Visit 1	**Visit 2**	**Visit 3**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**
	Screening and Baseline	Surgery and Discharge	**7-14 Day F/U (+3 days)**	**30 Day F/U (± 7 days)**	**90 Day F/U (± 14 days)**	**180 Day F/U (± 30 days)**	**360 Day F/U (± 30 days)**
Inclusion/Exclusion Review	X						
Informed Consent	X						
Patient Medical History	X						
Physical Exam	X						
Intraoperative Data[1]		X					
Clinical Assessments[2]			X	X	X	X	X
MSTS Score and EORTC QLQ BM22	X		X	X	X	X	X
Pain Visual Analog Scale (VAS)	X		X	X	X	X	X
Radiographs of Fracture Area	X		X	X	X	X	X
Adverse Events		X	X	X	X	X	X
Concomitant/Pain Medications	X	X	X	X	X	X	X
Health Economic Assessment	X	X	X	X	X	X	X

[1] Including date of surgery, duration of surgery, impending fracture treatment used.

[2] Clinical assessments: standard fracture evaluations and other assessments, as appropriate, including but not limited to assessments for pain, return to pre-treatment mobility, range of motion, return to work, disability status, use of pain medication, and physical therapy prescription status.

Appendix 2: Classification of Visits

Timing of assessment (days relative to surgery [a])	Visit	Visit Name to display for Analysis
Screening	Visit 1	Screening
Last assessment result collected prior to the surgery[b]	Baseline	Baseline
Surgery and Discharge (ie Day 1)	Visit 2	Surgery and Discharge
7-14 Day F/U (+3 days)	Visit 3	7-14 Day F/U
30 Day F/U (± 7 days)	Visit 4	30 Day F/U
90 Day F/U (± 14 days)	Visit 5	90 Day F/U
180 Day F/U (± 30 days)	Visit 6	180 Day F/U
360 Day F/U (± 30 days)	Visit 7	360 Day F/U

a. Day 1 is the day of surgery.
b. For subjects who have Screening and surgery visits on the same day, the values from the date of surgery will be considered for baseline. For range of motion tests, baseline values are defined as the measurements at 7-14 Day F/U visit.

Appendix 3: SAS Program for Conditional Power Calculation

```
*******************************************************************************
* CLIENT:    ILLUMINOSS
* PROJECT:   14-03-PATHOLHUM-02
*******************************************************************************
* PROGRAM NAME:      conditional_power_single_arm_three_timepoints
* PROGRAM LOCATION:  c:\proj\proj\icon\illuminoss\anllib
* PROGARM PURPOSE:   CALCULATES CONDITIONAL POWER GIVEN AN INTERIM OBSERVED DATASET.
*                                       DETAILS:  THE PRIMARY NULL AND ALTERNATIVE HYPOTHESES ARE:
                                        Ho: mean improvement in VAS at Day 90 <=  53.8
                                        Ha: mean improvement in VAS at Day 90 >  53.8
                                        WHERE VAS IS MEASURED AT BASELINE AND AT 3 POST-BASELINE VISITS
                                        (DAYS 7-4, 30 AND 90).  THE NULL HYPOTHESIS WILL BE TESTED USING
                                        THE SAMPLE CHANGE FROM BASELINE VAS to at Day 90,
                                        AS CALCULATED BY AN MMRM MODEL (SEE BELOW FOR THE PROC MIXED CALL).
                                        AFTER the 37th PATIENTS who underwent surgery had Day 90 visit
                                     , AN INTERIM MMRM ANALYSIS ON CHANGE FROM BASELINE VAS to Day 90 WILL
BE CARRIED OUT.  THE PURPOSE OF THIS INTERIM ANALYSIS
                                        IS NOT TO STOP THE TRIAL FOR OVERWHELMING EFFICACY, BUT RATHER
                                        TO (A) POTENTIALLY STOP FOR FUTILITY AND (B) POTENTIALLY
                                        INCREASE THE SAMPLE SIZE IF THE MEAN VAS IMPROVEMENT FROM BASELINE
                                        to at Day 90 IS LARGE BUT NOT AS LARGE AS ANTICIPATED IN THE
                                        ORIGINAL SAMPLE SIZE CALCULATIONS.  IN ORDER TO DO THIS,
                                        CONDITIONAL POWER (CP), CONDITIONED ON THE INTERIM OBSERVED
                                        DATA, NEEDS TO BE CALCULATED.  THIS PROGRAM CALCULATES CP USING
                                        COMPUTER SIMULATIONS.   1000 SIMULATED POST-INTERIM DATASETS
                                        ARE CALCULATED UNDER THE EFFECT SIZES AND WITHIN-SUBJECT
CORRELATIONS
                                        ESTIMATED FROM THE INTERIM OBSERVED DATA. EACH SIMULATED DATASET
                                        WILL BE ADDED TO THE INTERIM OBSERVED DATA, AND THE MMRM-BASED
                                        TESTING OF THE NULL HYPOTHESIS WILL BE CARRIED OUT.  THE PROPORTION
                                  OF SIMULATED DATASETS FOR WHICH THE NULL HYPOTHESIS IS REJECTED IS
                                        THE SIMULATED CONDITIONAL POWER.   NOTE: CONDITIONAL POWER IS
                                        CALCULATED UNDER THE ASSUMPTION THAT THE CHARACTERISTICS (MEAN
                                        AND STANDARD DEVIATION OF CHANGE FROM BASELINE VAS AT EACH VISIT,
                                        WTIHIN-PATIENT CORRELATION OF VAS BETWEEN VISITS) ARE THE TRUE
                                        POPULATION CHARACTERISTICS.
* SAS VERSION:       9.4
* AUTHOR:            Joe Massaro
* DATE:              06Oct2015
* revision history:  7Mar2016-modified to test VAS change from baseline to Day 90 visit, per protocol amendment
v5.0;
*-------------------------------------------------------------------------------;

/* FOR NOW, WE SIMULATE ONE "INTERIM" OBSERVED DATASET FROM THE POPULATION. */
/* NOTE THAT IN THE SIMULATIONS TO ESTIMATE CONDITIONAL POWER (CP) THAT     */
/* WILL BE CARRIED OUT AT THE TRUE INTERIM ANALYSIS, THIS SIMULATED INTERIM */
```

```
/* OBSERVED DATA WILL BE REPLACED WITH THE ACTUAL INTERIM OBSERVED DATA     */
/* OF POST-BASELINE CHANGE FROM BASELINE VAS FOR EVALUABLE SUBJECTS.        */

/* TO SIMULATE THIS INTERIM OBSERVE DATASET, WE FIRST NEED TO ENTER         */
/* THE CHARACTERISTICS OF THE ASSUMED POPULATION IN A DATASET. SPECIFICALLY:*/
/* THE FIRST ROW OF DATA IS THE ASSUMED MEAN OF BASELINE VAS AND OF THE     */
/* CHANGE FROM BASELINE VAS TO AT VISITS 1, 2 AND 3.  THE SECOND ROW IS THE */
/* ASSUMED S.D. OF VAS AT AND OF THE CHANGE FROM BASELINE VAS AT EACH OF    */
/* VISITS 1, 2 AND 3.   THE THIRD ROW IS THE EVALUABLE SAMPLE SIZES         */
/* ANTICIPATED AT EACH VISIT FOR THE INTERIM ANALYSIS.  ROWS 4-7 ARE        */
/* THE ASSUMED CORRELATION MATRIX (WITHIN-PATIENT CORRELATION OF CHANGE     */
/* FROM BASELINE VAS BETWEEN VISITS AND VS. BAESLINE VAS).                  */
data INTERIM_CHARACTERISTICS(type=corr);
      input _TYPE_ $ 1-4 _NAME_ $ 6-12 BASEVAS CHGVAS1 CHGVAS2 CHGVAS3;
cards;
MEAN          80    -38   -45   -58
STD           16    26    30    18
N             30    30    30    30
CORR BASEVAS 1.00 0.30 0.30 0.30
CORR CHGVAS1 0.30 1.00 0.40 0.16
CORR CHGVAS2 0.30 0.40 1.00 0.40
CORR CHGVAS3 0.30 0.16 0.40 1.00
run;

/* GENERATE A MULTIVARIATE NORMALLY DISTRIBUTED SAMPLE FROM THE POPULATION */
/* WITH THE ABOVE CHARACTERISTICS.  THIS WILL TAKE THE PLACE OF            */
/* THE INTERIM OBSERVED SAMPLE FOR NOW . THIS DATASET WILL HAVE ONE         */
/* OBSERVATION PER PATIENT, WITH THE VARIABLES BASEVAS AND CHGVAS1-CHGVAS3,*/
/* CONTAINING BASELINE VAS AND CHANGE FROM BASELINE VAS TO VISITS 1 - 3,    */
/* RESPECTIVELY.                                                           */

/* AGAIN, IN THE TRUE CP SIMULATIONS TO BE DONE AT THE INTERIM ANALYSIS,    */
/* THIS SIMULATED INTERIM OBSERVED DATA WILL BE REPLACED WITH THE ACTUAL    */
/* INTERIM OBSERVED DATA.                                                  */
proc simnormal data=INTERIM_CHARACTERISTICS numreal=30
         /* NUMREAL IS THE # OF ANTICIPATED OBSERVATIONS AT INTERIM, */
                                          seed=321248888 out=INTERIM_DATA;
         var baseVAS CHGvas1-CHGvas3;
run;

/* NOW RUN PROC MIXED ON THE INTERIM OBSERVED DATA, PRIMARILY TO CALCULATE  */
/* AN ESTIMATE OF THE CORRELATION MATRIX BETWEEN VISITS AND THE ESTIMATE    */
/* OF THE INTERIM OBSERVED MEAN AND STANDARD ERROR.  THESE CHARACTERSISTICS */
/* WILL BE LATER USED TO SIMULATE 1000 POST-INTERIM VAS DATASETS FROM A     */
/* POPULATION WITH THE SAME CHARACTERISTICS OF THE INTERIM OBSERVED DATA,   */
/* IN ORDER TO EVENTUALLY CALCULATED CP UNDER THE ASSUMPTION THAT THE       */
/* INTERIM DATASET CHARACTERISTICS ARE THE SAME CHARACTERISTICS AS THE      */
/* POPULATION.                                                             */
data INTERIM_DATA_STACKED;
```

```
        set INTERIM_DATA;
        subjid=rnum;
        visit=1; CHGvas=CHGvas1; output;
        visit=2; CHGvas=CHGvas2; output;
        visit=3; CHGvas=CHGvas3; output;
        keep subjid baseVAS visit CHGvas;
run;
/* THE RCORR OPTION BELOW YIELDS AN ESTIMATE OF THE CORRELATION MATRIX,    */
/* THE LSMEANS STATEMENT YIELDS AN ESTIMATE OF THE MEAN CHANGE FROM        */
/* BASELINE VAS AND ITS STANDARD ERROR AT EACH OF VISITS 1-3.              */
proc mixed data=INTERIM_DATA_STACKED method=ml;
    class visit subjid;
    model CHGvas = visit baseVAS/ ddfm=kr;
    repeated visit / type=un subject=subjid r rcorr;
    lsmeans visit / /*diff adjust=GT2 adjdfe=row*/;
run;
quit;
/* OBTAIN AN ESTIMATE OFF THE BASELINE MEAN VAS AND THE CORRELATION OF      */
/* BASELINE VAS WITH EACH CHANGE FROM BASELINE VAS.                         */
proc corr data=INTERIM_DATA;
        var baseVAS;
        with CHGvas1-CHGvas3;
run;

/* NOW TAKE THE OPERATIONAL CHARACTERISTICS ESTIMATED FROM ABOVE STATEMENTS,*/
/* AND ASSUME THAT THESE CHARACTERISTICS DEFINE THE POPULATION FROM WHICH   */
/* THE POST-INTERIM SAMPLE IS TAKEN.  WE WILL SIMULATE 1000 POST-INTERIM    */
/* SAMPLES FROM THIS POPULATION.  FOR EACH SIMULATED SAMPLE, WE WILL        */
/* COMBINE IT WITH THE ABOVE INTERIM OBSERVED SAMPLE. WE WILL THEN          */
/* CALCULATE THE PROPORTION OF THESE 1000 COMBINED DATASETS FOR WHICH THE   */
/* NULL HYPOTHESIS OF THE STUDY IS REJECTED.  THIS IS OUR CONDITIONAL       */
/* POWER (CP) UNDER THE ASSUMPTION THAT THE CHARACTERISTICS OF THE          */
/* INTERIM OBSERVED DATASET ARE THE TRUE CHARACTERISTICS OF THE POPULATION. */

/* TO START OUT, CREATE A DATASET CONTAINING THE CHARACTERISTICS OF THE     */
/* POPULATION.  FOR THE DATASET BELOW:
/* THE FIRST ROW OF DATA IS THE INTERIM OBSERVED MEAN OF BASELINE VAS        */
/* (ESTIMATED FROM THE ABOVE PROC CORR) AND OF THE CHANGE FROM BASELINE VAS  */
/* TO VISITS 1 - 3 (ESTIMATED FROM THE LSMEANS STATEMENT OF THE ABOVE PROC   */
/* MIXED STATEMENT. THE SECOND ROW OF DATA IS THE INTERIM OBSERVED           */
/* STANDARD DEVIATION OF BASELINE VAS ESTIMATED FROM THE ABOVE PROC CORR)    */
/* AND OF THE CHANGE IN BASELINE VAS TO VISITS 1, 2 AND 3 (ESTIMATED         */
/* BY TAKING THE ESTIMATED STANDARD ERRORS FROM THE LSMEANS STATEMENT        */
/* OF THE ABOVE PROC MIXED AND MULTIPLYING THEM BY THE SQUARE ROOT           */
/* OF THE SAMPLE SIZE AT EACH VISIT). THE THIRD ROW ARE THE PLANNED EVALUABLE*/
/* SAMPLE SIZES AT EACH VISIT POST-INTERIM.  ROWS 4-7 ARE THE ESTIMATED      */
/* CORRELATION MATRIX FROM THE INTERIM OBSERVED DATA.                                                  */
data POST_INTERIM_CHARACTERISTICS(type=corr);
```

```
      input _TYPE_ $ 1-4 _NAME_ $ 6-12 BASEVAS CHGVAS1 CHGVAS2 CHGVAS3;
cards;
MEAN          77.7 -40.9 -46.2 -60.4
STD           13.5 23.1 28.1 16.5
N             38   38   38    38
CORR BASEVAS  1.00 0.44 0.43  0.02
CORR CHGVAS1  0.44 1.00 0.20  0.26
CORR CHGVAS2  0.43 0.20 1.00  0.10
CORR CHGVAS3  0.02 0.26 0.10  1.00
run;

/* SIMULATE THE 1000 POST-INTERIM ANALYSIS DATASETS. EACH DATASET WILL      */
/* BE COMBINED WITH THE INTERIM OBSERVED DATASET IN ORDER TO CREATE 1000    */
/* COMPLETE SIMULATED CLINICAL TRIALS.  FOR EACH TRIAL, A PROC MIXED        */
/* ANALYSIS WILL BE CARRIED OUT TO TEST THE STUDY'S NULL HYPOTHESIS.        */
/* THE PROPORTION OF SIMULATED DATASETS FOR WHICH THE NULL HYPOTHESIS IS    */
/* REJECTED IS THE SIMULATED CONDITIONAL POWER.                             */
%MACRO DOIT2;
data POST_INTERIM_CHARACTERISTICS(type=corr);
        %DO I=1 %TO 1000;
                simulation=&I;
                set POST_INTERIM_CHARACTERISTICS;
                OUTPUT;
        %END;
run;
%MEND DOIT2;
%DOIT2;
proc sort data=POST_INTERIM_CHARACTERISTICS;
        by simulation;
run;
/* SIMULATE A MULTIVARIATE NORMAL DISTRIBUTION FOR BASELINE AND CHANGE FROM         */
/* BASELINE VAS TO VISITS 1 - 3 UNDER THE ABOVE CHARACTERISTICS FOR EACH SIMULATION.*/
proc simnormal data=POST_INTERIM_CHARACTERISTICS numreal=38 seed=315893282 out=POST_INTERIM_DATA;
        by simulation;
        var baseVAS CHGvas1-CHGvas3;
run;
data POST_INTERIM_DATA;
        set POST_INTERIM_DATA;
        subjid=rnum+100000; /* MAKE SURE SUBJIDSFOR POST-INTERIM DATA ARE NOT  */
                                                /* THE SAME AS SUBJECT IDS FROM INTERIM DATA.      */
        drop rnum;
run;
/* COMBINE EACH SIMULATED DATASET WITH THE INTERIM OBSERVED DATASET.       */
/* THIS IS DONE BY FIRST MAKING 1000 COPIES OF THE INTERIM OBSERVED DATA,  */
/* AND THEN ADDING ONE COPY TO EACH OF THE 1000 SIMULATED POST-INTERIM     */
/* DATASETS.                                                               */
%MACRO DOIT3;
data INTERIM_DATA;
        %DO I=1 %TO 1000;
```

```
                simulation=&I;
                set INTERIM_DATA;
                subjid=rnum;
                OUTPUT;
        %END;
        drop rnum;
run;
%MEND DOIT3;
%DOIT3;

data FINAL;
        set INTERIM_DATA POST_INTERIM_DATA;
run;
proc sort data=FINAL NODUPKEY;
        by simulation subjid;
run;

/* ARRANGE THE DATA SO THAT PROC MIXED CAN BE CARRIED OUT FOR EACH SIMULATED */
/* DATASET.                                                                  */
data FINAL_STACKED;
        set FINAL;
        visit=1; CHGvas=CHGvas1; output;
        visit=2; CHGvas=CHGvas2; output;
        visit=3; CHGvas=CHGvas3; output;
        keep simulation subjid baseVAS visit CHGvas;
run;

ods select none;
/* THE LSMESTIMATE STATEMENT BELOW CONDUCTS A TEST OF THE ABOVE NULL  */
/* HYPOTHESIS OF INTEREST.                                           */
proc mixed data=FINAL_STACKED method=ml;
        by simulation;
    class visit subjid;
    model CHGvas = visit baseVAS/ ddfm=kr;
    repeated visit / type=un subject=subjid r rcorr;
    lsmeans visit / /*diff adjust=GT2 adjdfe=row*/;
         lsmestimate visit 0 0 1/lower testvalue=-53.8;
         ods output lsmestimates=lsmestimates;
run;
quit;
ods select all;

/* DETERMINE THE PROPORTION OF SIMULATIONS FOR WHICH THE NULL HYPOTHESIS */
/* IS REJECTED.  THIS IS THE CONDITIONAL POWER.                          */
data FinalResult;
        set lsmestimates;if probt<=0.05 then reject=1;
        else reject=0;
run;
```

```
proc freq data=FinalResult;
        table reject;
run;
```