FERRING
PHARMACEUTICALS

Statistical Analysis Plan

Title of trial: A multi-center, randomized, pivotal study evaluating AMPLEX compared to autogenous bone graft in subjects indicated for arthrodesis surgery involving the hindfoot or ankle
NCT number: NCT03028415
Sponsor trial code: 000226
Date: 27 March 2017

STATISTICAL ANALYSIS PLAN

A Multi-Center, Randomized, Pivotal Study Evaluating AMPLEX® Compared To Autogenous Bone Graft in Subjects Indicated for Arthrodesis Surgery Involving the Hindfoot or Ankle

000226

Investigational Product:	AMPLEX (B2A Enhanced Ceramic Granules)
Indication:	AMPLEX is indicated for use as an alternative to autograft in hindfoot and ankle fusion procedures that require supplemental graft material, including tibiotalar, talocalcaneal, talonavicular and calcaneocuboid fusions
Phase:	Pivotal IDE
Author:	[redacted], PhD
Date of issue:	27-Mar-2017
Version:	1.0

Change log

Version No.	Effective Date	Reason for the Change / Revision	Supersedes

Signed agreement on Statistical Analysis Plan

Author:

- [redacted], PhD, Director, Biostatistics, Ferring International PharmaScience Center U.S., Inc.

Reviewer:

This statistical analysis plan was reviewed and signed electronically in the REAL system by:

- [redacted], MD, Vice President, Clinical Development, Ferring International PharmaScience Center U.S., Inc.
- [redacted], MD, Associate Vice President, US Pharmacovigilance, Ferring Pharmaceuticals, Inc.
- [redacted], MsC, PhD, Senior Health Economics & Outcomes Research Manager, Global Health Economics & Outcomes Research, Ferring Pharmaceuticals A/S
- [redacted], Statistical Programmer, Global Biometrics, Ferring Pharmaceuticals A/S
- [redacted], Manager, Clinical Operations, Ferring International PharmaScience Center U.S., Inc.

[redacted] Date (dd/mm/yyyy)

Approval:

This statistical analysis plan was reviewed, approved and signed electronically in the REAL system by:

- [redacted], PhD, Director of Biostatistics, Ferring International PharmaScience Center China, Ltd.

TABLE OF CONTENTS

1 Introduction 6
1.1 Definitions/ Abbreviations 6
1.1.1 Definition of Terms 6
1.1.2 Abbreviations 6

2 Trial Objectives and Endpoints 8
2.1 Objectives 8
2.2 Endpoints 8
2.2.1 Primary Endpoint 8
2.2.2 Key Secondary Endpoints 8
2.2.3 Other Secondary Endpoints 8
2.2.4 Safety Endpoints 9
2.2.5 Pharmacokinetic Endpoints 9

3 Trial Design 10
3.1 General Design Considerations 11
3.2 Determination of Sample Size 11

4 Subject Disposition 12

5 Protocol Deviations 13

6 Analysis Sets 14
6.1 Intention-To-Treat Analysis Set 14
6.2 Modified Intention-To-Treat Analysis Set 14
6.3 Per Protocol Analysis Set 14
6.4 Safety Analysis Set 14

7 Trial population 15
7.1 Demographics and Other Baseline Characteristics 15
7.1.1 Demographics 15
7.1.2 Body Measurements 15
7.1.3 Comorbidity 15
7.1.4 History and Physical Examination of Ankle/Hindfoot 16
7.1.5 Surgical Site and Non-union Risk Factors 16
7.1.6 Other Baseline Characteristics 16
7.2 Medical History 16
7.3 Prior and Concomitant Medication 17
7.4 Perioperative Characteristics 17
7.4.1 Procedure 17
7.4.2 Hardware 17
7.4.3 Medications 17

8 Exposure and Treatment Compliance 19
8.1.1 Extent of Exposure 19
8.1.2 Treatment Compliance 19

9 Efficacy 20
9.1 General Considerations 20
9.2 Primary Endpoint 21
9.2.1 Primary Variable Analysis 21

9.2.2 Sensitivity Analysis 22
9.2.3 Pooling of Site Data 23
9.3 Secondary Endpoint(s) 24
9.3.1 Key Secondary Endpoint 24
9.3.2 Other Secondary Endpoints 24

10 Safety 26
10.1 General Considerations 26
10.2 Adverse Events 26
10.2.1 Overview of Treatment-Emergent Adverse Events 26
10.2.2 Incidence of Adverse Events 27
10.3 Secondary Surgical Intervention 28
10.4 Radiographic Assessment 28
10.4.1 Image Quality 28
10.4.2 Post-operative Fusion Site 28
10.4.3 Safety 28
10.5 Anti-B2A Antibody 28
10.6 Safety Laboratory Variables 28
10.6.1 Summary Statistics 28
10.6.2 Laboratory Variable Changes Relative to Normal Range 29
10.6.3 Markedly Abnormal Changes 29
10.6.4 Data Listings 29
10.7 Vital Signs 29
10.7.1 Summary Statistics 29
10.7.2 Vital Signs Abnormal Values 29
10.7.3 Data Listings 29

11 Interim analyses 31

12 Deviations from protocol analysis 32

13 References 33

Appendices 34
Appendix 1 Markedly Abnormal Laboratory Values 34
Appendix 2 FAAM-ADL 35
Appendix 3 Schedule of Assessment 36

List of Tables

Table 1 Scheduled trial day and window for analysis visits 20

1 Introduction

This document describes the planned statistical analyses for the AMPLEX pivotal study 000226 based on the Protocol Ver. 6.0 dated December 14, 2016.

1.1 Definitions/ Abbreviations

1.1.1 Definition of Terms

Terms	Definitions
Screened	Signed the informed consent form
Screen failed	Discontinued from the trial prior to randomization
Device	AMPLEX or ABG
Procedure	Index surgical fusion procedure or bone graft harvest procedure

1.1.2 Abbreviations

Abbreviations	Meaning of abbreviations in document
ABG	Autogenous bone graft
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine transaminase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
AUC	Area under the curve
BMI	Body mass index
BVF	Bone void filler
CEC	Clinical events committee
CI	Confidence interval
CT	Computerized tomography
DMC	Data monitoring committee
eCRF	Electronic case report form
FAAM-ADL	Foot and ankle ability measure; activities of daily living
FDA	Food and drug administration
IDE	Investigational device exemption
LOCF	Last observation carried forward
ITT	Intention to treat
MCS	Mental component summary
MedDRA	Medical dictionary for regulatory activities
mITT	Modified intention to treat
PCS	Physical component summary
PK	Pharmacokinetics
PP	Per protocol
PT	Preferred term
SAE	Serious adverse event

Abbreviations	Meaning of abbreviations in document
SAP	Statistical analysis plan
SF-12	Short form-12
SOC	System organ class
SPC	Subject performance composite
TEAE	Treatment-emergent AE
US	United States
VAS	Visual analog scale

2 Trial Objectives and Endpoints

2.1 Objectives

The primary objective of this pivotal study is to demonstrate that AMPLEX is non-inferior to ABG for bone fusion in a population indicated for single, double, or triple hindfoot arthrodesis or ankle arthrodesis surgery with supplemental graft material.

2.2 Endpoints

2.2.1 Primary Endpoint

The primary efficacy endpoint is the proportion of subjects who meet all the following criteria for the Subject Performance Composite (SPC) Endpoint at 52 weeks:

- Improvement in pain on weight-bearing at fusion site (≥ 20 mm reduction from baseline on 100 mm VAS)
- Absence of significant graft harvest site pain (< 20 mm on 100 mm VAS)
- Improvement in Foot and Ankle Ability Measure Activities of Daily Living subscale (FAAM-ADL) (≥ 8 points improvement from baseline)
- Absence of device related or procedure related SAEs (up to Week 52)
- Absence of secondary surgical or nonsurgical interventions intended to promote fusion (up to Week 52)

2.2.2 Key Secondary Endpoints

The key secondary endpoint is the proportion of subjects achieving CT radiographic fusion success at 52 weeks:

- Radiographic evidence of fusion by CT scan (≥ 50% bone bridging across the joint space for the full complement of joints in the absence of secondary surgical or nonsurgical interventions intended to promote fusion)

2.2.3 Other Secondary Endpoints

- Proportion of subjects achieving CT radiographic fusion success at 12 and 24 weeks (in the absence of secondary surgical or nonsurgical interventions intended to promote fusion)
- Change from baseline in pain on weight-bearing at fusion site at 12, 24, and 52 weeks (> 20 mm reduction from baseline on 100 mm VAS)
- ABG harvest site pain at 2, 6, 12, 24, and 52 weeks (< 20 mm on 100 mm VAS)
- Change from baseline in FAAM-ADL at 12, 24 and 52 weeks
- SPC at 12 and 24 weeks
- Change from baseline in Short Form-12 (SF-12) at 24 and 52 weeks

2.2.4 Safety Endpoints

- Frequency, severity and seriousness of Adverse Events (AEs)
- Device or Procedure Related SAEs
- Secondary surgical intervention including revision, removal, reoperation or supplemental fixation
- Subsidence, device migration, nonunion, osteolysis and/or heterotopic ossification in the area surrounding the implant site by radiographic assessment
- Presence of anti-B2A antibodies (Ab) in serum

2.2.5 Pharmacokinetic Endpoints

- Plasma B2A levels at Days 1, 2, 4, 7, and 15
- AUC, AUCt, %AUC extrap, Cmax, Tmax, t½, lamda z, CL/F, Vz/F for B2A

Note that the analysis of the pharmacokinetic (PK) endpoints are not in the scope of this SAP. The PK analysis will be performed by Translational Medicine Department, Ferring Pharmaceuticals A/S.

3 Trial Design

This is a prospective, multi-center clinical trial in subjects (≥18 and <75 years of age) undergoing tibiotalar (ankle) arthrodesis or single, double, or triple hindfoot (talocalcaneal-, calcaneocuboid- or talonavicular-joint) arthrodesis with supplemental graft material and internal fixation with up to 3 screws per joint.

AMPLEX, the investigational device, is a synthetic bone graft substitute comprised of two parts; 1) a synthetic bone void filler (BVF) that functions as an osteoconductive scaffold and 2) a synthetic peptide (B2A) that augments osteodifferentiation. The control material, ABG, is harvested at the discretion of the surgeon from the calcaneus, proximal tibia, distal tibia or iliac crest and administered by surgical implant.

Prior to surgery, each subject will have baseline information collected which will include medical history, injury/disease etiology, diagnosis, clinical laboratory assessments (including serum for anti-B2A antibody analysis), pain and disability assessments, in addition to radiographic imaging.

As part of eligibility determination, the investigator will review any relevant diagnostic imaging results and comorbidities to ensure that subjects meet the appropriate eligibility criteria and that ankle or hindfoot arthrodesis surgery with supplemental graft material is indicated. Eligible subjects will be randomized as close to the procedure as possible, preferably on the day of the surgery.

Subjects will be randomly assigned on a two to one (2:1) basis to either the treatment group (AMPLEX) or the control group (ABG). Randomization will be stratified by clinical center, surgical site (hindfoot vs. ankle), and nonunion risk factors (none vs. any of: obesity, diabetes, previous hindfoot/ankle surgery, or smoking history). Permuted block randomization will be performed within strata. To minimize the opportunity for the sequence to be predicted, the block size will be variable.

Surgeries will be performed by board-certified/board eligible orthopedic surgeons. Perioperative prophylactic antibiotics will be administered before surgical incision in accordance with investigational site guidelines. Standard anesthesia procedures will be used. Standard surgical technique will be employed to gain access to each fusion site and to ensure rigid fixation of the fusion site. The joint will be exposed and prepared as per standard practice. The graft material should be placed in the joint space to minimize the chance of extrusion while maximizing bone-on-bone contact. Once the graft is placed, internal fixation screws will be placed across the joint. Up to three (3) screws can be used per joint. Supplemental pins and staples may be used, as well as supplemental screws and plates external to the fusion site(s). The surgical site will be closed in layers using standard surgical technique.

The trial is designed as a randomized, controlled trial comparing AMPLEX to ABG for bone fusion at 52 weeks. Although the primary efficacy endpoint will be assessed at 52 weeks, all subjects will be followed up for 78 weeks for the safety assessment. Post-discharge clinical visits will include

visits at 2, 6, 12, 24, 52 and 78 weeks. Adverse event data will be collected from randomization until the final follow up. The schedule of assessments is described in Appendix 3.

3.1 General Design Considerations

Three independent parties will be established for endpoint assessment and safety monitoring.

An independent clinical events committee (CEC) will adjudicate device, procedural, or ancillary device related AEs, all events associated with secondary surgical interventions, and any other events specified in the CEC Charter.

An independent imaging core laboratory will centrally review all images. Osseous bridging will be assessed on a per-joint basis. The extent of osseous bridging will be semi-quantitatively scored (i.e., absent, <25%, ≥25 to <50%, ≥50 to <75%, and ≥75 to 100%) by assessing the area of bridged joint surface relative to the total area of joint surface available to be bridged (i.e., Bridged Surface Area/Total Surface Area = % Osseous Bridging).

An independent data monitoring committee (DMC) will periodically review clinical study data to ensure participants' safety, trial integrity, and scientific rigor. The DMC will develop and ratify a Charter that provides guidelines for oversight of trial safety.

3.2 Determination of Sample Size

Assuming true success rates of the primary endpoint to be 60% for ABG and 62% for AMPLEX groups, respectively, 396 subjects are needed to maintain 85% power to demonstrate non-inferiority of AMPLEX to ABG with a non-inferiority margin of -12% (AMPLEX minus ABG) at a type I error rate of 1-sided 5% in a 2:1 randomization (264 for AMPLEX and 132 for ABG). To maintain this sample size for the 78-week completers (per the FDA's recommendation) after taking account for approximately 15% discontinuation (including lost to follow-up) rate, a total of 480 subjects (320 for AMPLEX and 160 for ABG) will be randomized.

In this calculation, the true success rate for AMPLEX is assumed to be slightly higher than that for ABG because of the potential benefit for AMPLEX on the absence of significant graft harvest site pain component. The assumption on the true success rate for ABG is made based on the clinical endpoint results in DiGiovanni et al. (2013)1 (1) and the Augment SSED (2015) (4). These data indicate relatively high success rates for each component at Week 52. The estimated success rate as a composite endpoint is approximately 60%, when it is assumed that the success of each component is independent of each other.

4 Subject Disposition

All subjects who signed the informed consent form will be considered screened for the trial. All screened subjects who discontinued from the trial prior to randomization will be classified as screening failures. Screening failures and their primary reason for screening failure will be tabulated. The number of screened subjects will be used as the denominator when the percentage of each primary reason for screening failure is calculated.

Subjects who were randomized, received the procedure (surgery), were treated with the graft material, prematurely discontinued from the study, and completed the study will be tabulated with the reason of premature discontinuations for all randomized subjects. Similar tabulations will be made by investigational site and by region (US or Canada).

The number and percentage of subjects who attended Visits 2, 3, 4, 5, 6, 7, and 8 will be tabulated for all randomized subjects based on the nominal visit data recorded on the eCRF.

The time from randomization to discontinuation will be summarized by the Kaplan-Meier method.

Subjects who discontinued from the trial after randomization will be listed by including information on the last visit attended. The list will be sorted by investigational site and subject ID.

5 Protocol Deviations

Subjects with important protocol deviations from the study conduct perspective will be tabulated and listed for all randomized subjects. The list will be sorted by investigational site and subject ID. The classification of important or non-important deviations will be completed by the sponsor and documented prior to the database lock.

The following protocol deviations will result in the exclusion of subjects for the per protocol (PP) analysis set:

- Deviations from the required pre-existing conditions that could affect the bone fusion
- Deviations from the procedural requirements that could affect the bone fusion or negatively impact the assessment of the bone fusion

Other unforeseeable types of protocol deviations may also result in the exclusion of subjects for the PP analysis set, and the final determination of those protocol deviations will be completed and documented prior to the database lock.

6 Analysis Sets

Subjects who are included and not included in each analysis set will be tabulated with the reason of the exclusion. The percentages will be calculated based on all randomized subjects. Subjects who are not included in each analysis set will be listed with the reason. The list will be sorted by investigational site and subject ID.

6.1 Intention-To-Treat Analysis Set

The intention-to-treat (ITT) analysis set will consist of all randomized subjects. Analyses for the ITT analysis set will be conducted according to the randomized treatment regardless of the actual treatment received.

6.2 Modified Intention-To-Treat Analysis Set

The modified ITT (mITT) analysis set will consist of all randomized subjects who have an attempted fusion of the index joints. If a subject becomes ineligible or withdraws prior to the initiation of surgery, the subject will not be included in the mITT analysis set. Once the surgery is initiated, all subjects will be included in the mITT analysis. Analyses for the mITT analysis set will be conducted according to the randomized treatment regardless of the actual treatment received.

6.3 Per Protocol Analysis Set

The per protocol (PP) analysis set will consist of all randomized subjects that: [1] complete the fusion procedure and receive either AMPLEX or ABG as assigned, [2] meet critical study eligibility criteria, and [3] have no significant protocol deviations (see Section 5).

6.4 Safety Analysis Set

The safety analysis set will consist of all treated subjects (i.e., received AMPLEX or ABG). Analyses for the safety analysis set will be conducted according to the actual treatment received.

7 Trial population

7.1 Demographics and Other Baseline Characteristics

Descriptive statistics of demographics and other relevant baseline characteristics will be presented for all subjects in the mITT and PP analysis sets by treatment group unless specified otherwise. In addition, they will be summarized by investigational site and by region (US or non-US) for the mITT analysis set.

Categorical data will be summarized using numbers and percentages. The percentages are based on the total number of subjects with a corresponding assessment. Continuous data will be presented using the number of subjects, mean, standard deviation, median, inter-quartile range, minimum and maximum.

All demographics and baseline data will be listed and sorted by investigational site and subject ID.

No missing data will be imputed unless specified otherwise.

7.1.1 Demographics

The following demographic information will be tabulated: age at randomization (years), age at randomization (<65 years, or ≥65 years), sex, ethnicity (Hispanic or Latino, Not Hispanic or Latino), race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White) and country (US, non-US).

7.1.2 Body Measurements

The following baseline body measurements made at screening will be summarized: body weight (kg), height (m), and BMI (kg/m^2). In addition, frequency tables for BMI (<20, 20 - <25, 25 - <30, 30 - <35, 35 - <40, 40 - <45, or ≥45 kg/m^2) will be produced.

7.1.3 Comorbidity

The following comorbidities that warrant the use of supplemental graft material will be recorded at the screening visit:

- Radiographic evidence of bone defect, deficit, subsidence or subchondral cyst
- More than one joint to be fused
- Involvement of other adjacent or nonadjacent joints
- Large surface area
- Intra-articular or extra-articular deformity
- Post-traumatic arthritis
- Diagnosis of osteoporosis

The prevalence of each comorbidity (yes, no) will be tabulated. A separate table will summarize the number of comorbidities per subject both as the actual number (i.e., 1 to 7) and in a cumulative manner (i.e., at least 1, at least 2, …, and at least 7).

7.1.4 History and Physical Examination of Ankle/Hindfoot

The affected side of ankle/hindfoot will be determined by the planned fusion side. If a subject indicates that the contralateral foot or ankle is also affected by the same condition, the affected side will be considered bilateral. The affected side (right, left, or bilateral), the primary etiology of the condition that required fusion (primary osteoarthritis, post-traumatic arthritis, or other), and the alignment of ankle/hindhoot (normal, pes plano valgus, cavovarus/equinovarus, hindfoot varus, or hindfoot valgus) will be summarized.

7.1.5 Surgical Site and Non-union Risk Factors

The following definitions of planned surgical site and non-union risk factors will be used for the stratified randomization:

- Planned surgical site: tibiotalar (ankle) or any of talocalcaneal, calcaneocuboid, or talonavicular joints (hindfoot)
- Obesity: BMI $\geq$ 30 kg/m^2
- Diabetes: History of Type 1 or Type 2 diabetes mellitus
- Previous hindfoot/ankle surgery: Previous surgery on the target hindfoot/ankle
- Smoking history: Any smoking history prior to screening

The randomization strata, that is, planned surgical site and presence or absence of any of obesity, diabetes, previous hindfoot/ankle surgery, or smoking history will be tabulated. The prevalence of each non-union risk factor and planned surgical site will also be tabulated.

7.1.6 Other Baseline Characteristics

Baseline weight-bearing pain, FAAM-ADL score, SF-12 physical component summary (PCS), and SF-12 mental component summary (MCS) will be summarized.

7.2 Medical History

Medical history will be coded for the system organ class (SOC) and preferred tem (PT) using the Medical Dictionary for Regulatory Activities (MedDRA) version effective at trial start or later. The version of the MedDRA used for the final analyses will be documented. Medical history will be summarized by the SOC (in alphabetical order) and PT (in decreasing order of frequency) for the mITT analysis set.

Medical history will be listed by investigational site and subject ID.

7.3 Prior and Concomitant Medication

All medications will be coded using the World Health Organization Drug Reference List and categorized as either prior or concomitant medications.

1) Prior medication; i.e. medication taken exclusively prior to treatment (i.e. with stop date before date of the surgery)
2) Concomitant medication, i.e. medication taken during the treatment period (i.e. medication that was not stopped before date of the surgery and not started after the end-of-trial visit)

If the timing of the dose of a medication cannot be established in relation to the date of the surgery, it will be considered as concomitant medication.

Prior and concomitant medications will be summarized by the Anatomical Therapeutic Chemical (ATC) classification 1st level (alphabetically) and ATC classification 2nd level (in decreasing order of frequency) for the mITT analysis set.

Prior and concomitant medications will be listed by investigational site and subject ID.

7.4 Perioperative Characteristics

Perioperative characteristics of the index procedure will be summarized by treatment group for the mITT analysis set. In addition, they will also be summarized by investigational site and by region (US or non-US) for the mITT analysis set. Perioperative characteristics will be listed by investigational site and subject ID.

7.4.1 Procedure

The length of the procedure, length of the hospital admission, treated side, treated joint(s), type of surgical incision, estimated blood loss at fusion site, estimated blood loss at harvest site, and blood transfusion (yes/no) will be summarized.

The length of the procedure will be calculated as the time between the first incision and skin closure. Treated joint(s) will be tabulated for ankle or any of the hindfoot joints with breakdown of all combinations of treated hindfoot joints.

7.4.2 Hardware

Number of screws used as the primary fixation per joint will be summarized for each joint as well as each type of fusion (i.e., joint(s) treated).

7.4.3 Medications

Perioperative medications will be summarized by the ATC classification 1st level (alphabetically), ATC classification 2nd level (in decreasing order of frequency). In addition, the type and length of anesthesia will be summarized.

The lengths of anesthesia will be calculated from the earliest start time and the latest end time of any type of anesthesia.

8 Exposure and Treatment Compliance

8.1.1 Extent of Exposure

The amount of graft material used will be summarized as a continuous variable in cm^3 as well as a categorical variable ($0<-\leq 5\ cm^3$, $5<-\leq 10\ cm^3$, or $>10\ cm^3$) by treatment group for the mITT analysis set in following manners:

- Total amount used
- Total amount used for each joint
- Total amount used for each combination of joint(s) treated
- Total amount used for procedures involving single, double, and triple hindfoot joints, respectively

For the ABG group, the harvest site, harvest side, type of incision at the harvest site, and volume harvested will be summarized.

The length of the follow-up from the day of the surgery will be summarized together with the following categorical summaries:

- < 2 weeks, 2 - <4 weeks, 4 – <12 weeks, 12 – <24 weeks, 24 – <36 weeks, 36 – <52 weeks, 52 – <64 weeks, 64 – <78 weeks, and 78 weeks or more
- At least 1 day, at least 2 weeks, at least 4 weeks, at least 12 weeks, at least 24 weeks, at least 36 weeks, at least 52 weeks, at least 64 weeks, and at least 78 weeks

8.1.2 Treatment Compliance

The investigator-reported response for the compliance to the protocol-defined procedure requirements and the instructions for use (if applicable) will be summarized by treatment group for the mITT analysis set.

9 Efficacy

9.1 General Considerations

Non-inferiority tests will be conducted at a one-sided 5% significance level whereas superiority tests will be conducted at a two-sided 5% significance level per an agreement with the FDA in the AMPLEX IDE.

All efficacy analyses will be conducted for the mITT analysis set unless specified otherwise. The primary and key secondary efficacy endpoints will be analyzed by a fixed-sequence procedure according to the following order to maintain the overall Type 1 error rate to a one-sided 5% for non-inferiority and a two-sided 5% for superiority testing.

1) Non-inferiority of AMPLEX to ABG in the primary efficacy endpoint (primary analysis)
2) Non-inferiority of AMPLEX to ABG in the key secondary efficacy endpoint
3) Superiority of AMPLEX to ABG in the key secondary efficacy endpoint
4) Superiority of AMPLEX to ABG in the primary efficacy endpoint

Testing procedure will begin with the non-inferiority test for the primary efficacy endpoint, and each test will be conducted without adjustment of multiplicity as long as all preceding tests are significant. All other secondary endpoints will be presented with the point estimate of the treatment group difference and the corresponding 2-sided 90% and 95% confidence intervals.

For visit-based data, the data collected on the day that is closest to the scheduled trial day within a window specified in Table 1 will be assigned to the corresponding analysis visit. If the closest day cannot be uniquely identified for an analysis visit, the data collected on an earlier trial day will be assigned to the analysis visit. Visit windows for the analysis are specified in a way to utilize all collected data as much as possible and are different from the protocol specified visit windows.

Table 1 Scheduled trial day and window for analysis visits

Analysis visit	Scheduled trial day	Analysis window
Baseline	Day -84 to Day 1	Up to prior to the initiation of surgery
Post-Operative CT and X-ray	Day 1	Day 1 to Day 29
Week 2	Day 15	Day 2 to Day 29
Week 6	Day 43	Day 30 to Day 64
Week 12	Day 85	Day 65 to Day 127
Week 24	Day 169	Day 128 to Day 267
Week 52	Day 365	Day 268 to Day 456
Week 78	Day 547	Day 457 to Day 637

9.2 Primary Endpoint

9.2.1 Primary Variable Analysis

The primary efficacy endpoint will be the proportion of subjects who meet all of the following criteria for the SPC at 52 weeks:

- Improvement in pain on weight-bearing at fusion site (≥ 20 mm reduction from baseline on 100 mm VAS)
- Absence of significant graft harvest site pain (< 20 mm on 100 mm VAS)
- Improvement in Foot and Ankle Ability Measure Activities of Daily Living subscale (FAAM-ADL) (≥ 8 points improvement from baseline)
- Absence of device related or procedure related SAEs (up to Week 52)
- Absence of secondary surgical or nonsurgical interventions intended to promote fusion (up to Week 52)

The primary analysis will assess a non-inferiority of AMPLEX to ABG, and it will be claimed if the lower bound of the 2-sided 90% CI for the difference of the proportions (AMPLEX minus ABG) is greater than -12.0%. The superiority will be claimed if the lower limit of the 2-sided 95% CI is greater than zero. The CI will be constructed with the normal approximation to the binomial. In addition, the response status for each component will be summarized.

A subject will be considered an SPC success if he/she meets all of the SPC criteria, otherwise, the subject will be considered an SPC failure. The SPC response status will be determined for 12, 24, and 52 weeks. The response status for the visit-based components (i.e., weight-bearing pain, graft harvest site pain, and FAAM-ADL) will be assessed for each time point based on the windowed analysis visits. The response status for the non-visit-based components (i.e., absence of device or procedure related SAEs and secondary interventions to promote fusion) will be assessed in a period up to the actual visit date that the corresponding visit-based variables for a particular analysis visit are collected. If the actual visit dates among the visit-based variables are different at the analysis visit, the latest visit date among them will be used to determine the period for the assessment of the non-visit-based variables. If any of the components of the SPC at an analysis visit is missing, the SPC response status will be considered missing for the analysis visit. Note that the graft harvest site pain will be collected only for the ABG group, and therefore, no data for this component will be recorded for the AMPLEX group.

For the primary analysis, the last available SPC response status will be used (i.e., the last observation carried forward; LOCF) if the SPC response status at 52 weeks is missing. If no SPC response status data are available at previous visits, the SPC response status at 52 weeks will be considered failure.

9.2.2 Sensitivity Analysis

The following sensitivity analyses will be conducted to assess the robustness of the primary analysis:

- As-treated analysis based on actually received treatment on the mITT analysis set
- Analysis on the ITT analysis set
- Analysis on the PP analysis set

The same analysis conducted for the primary analysis will be repeated for these sensitivity analyses.

In addition, sensitivity analyses using the following approaches for the handling of missing data will be conducted:

- Missing-equals-failure approach
- Multiple imputation
- Tipping point analysis

Missing-equals-failure approach:

All subjects with missing SPC status at 52 weeks will be considered failures.

Multiple imputation methods:

The missing data for each component of the SPC for 12, 24, and 52 weeks will be imputed sequentially by regression-based imputation models using treatment group, baseline characteristics, and outcomes from preceding visits as predictors. The missing SPC response status will be calculated based on the imputed data. This imputation procedure will be repeated 100 times, and the primary analysis method will be applied to each of the imputed datasets. A combined estimate of the difference of the proportions (AMPLEX minus ABG) and the corresponding two-sided 90% CI will be obtained by the methods described in Rubin (1987) (3).

Tipping point analyses:

Let n_{1m} and n_{2m} be the number of subjects with missing SPC response status at 52 weeks in AMPLEX and ABG group, respectively. The tipping point analysis will assess the non-inferiority in all possible combinations of missing data imputation (i.e., $(n_{1m}+1)*(n_{2m}+1)$ imputations), and the result will be presented graphically. Specifically, the horizontal and vertical axes of the graph will be the number of subjects imputed as success in AMPLEX group (i.e., ranges between 0 and n_{1m}) and the number of subjects imputed as success in ABG group (i.e., ranges between 0 and n_{2m}), respectively. The non-inferiority outcome (i.e., supportive or not supportive) will be plotted for each combination of the number of successes from each group in the graph by different symbols (e.g., a solid circle for supportive and an open circle for not supportive outcomes). Note that each result with the LOCF method (i.e., primary analysis) and with the missing-equals-failure approach will be represented as one of the combinations in the graph.

The primary endpoint will also be analyzed for the following subgroups for the mITT analysis set, and the treatment group difference of the proportions of subjects with SPC success for each subgroup will be presented in forest plots.

- Demographic characteristics: age, gender, and race
- Baseline comorbidities: obesity, diabetes, previous ankle/hindfoot surgery, smoking history, primary etiology
- Surgical site: affected side, fusion site (single, double, triple, or ankle)
- Investigational site
- Geographical region (US or non-US)

The treatment group comparisons after controlling each factor listed above will be conducted by stratified analyses using the Mantel Haenszel method as sensitivity analyses.

9.2.3 Pooling of Site Data

The primary analysis will be based on the pooled results. However, the homogeneity of study outcomes across investigational sites will be examined.

The primary justification for pooling is that investigational sites will be following the same protocol, using the same device system, and following the same Instructions for Use/AMPLEX Surgical Manual. Additionally, frequent contact and monitoring of the sites will be performed to ensure that all investigational sites are evaluating participants and recording study results in a reliable and reproducible manner. It is not anticipated that any individual investigational site will dominate the study results. Therefore, it is considered that these procedures will ensure that the data from these investigational sites can be combined and analyzed.

To evaluate differences among sites and regions (US or non-US) in the trial, summaries for important baseline variables, such as demographics, foot and ankle related medical history and baseline clinical variables, and variables related to procedures will be presented by site and by region as described in Section 7.

Poolability across investigational sites will be tested for the primary endpoint using the homogeneity test of risk difference proposed by Zhang, Yang, & Cho (2009) (5) at a two-sided 5% significance level. If all the subjects in a site are randomized to one treatment group, the site will be excluded from this assessment. The same test will also be used to assess the homogeneity of risk difference among regions. If there is statistically significant heterogeneity across investigational sites, the implications for the interpretation of the primary analysis will be discussed.

9.3 Secondary Endpoint(s)

9.3.1 Key Secondary Endpoint

The key secondary efficacy endpoint will be the proportion of subjects who meet the following criteria for the CT radiographic fusion success at 52 weeks:

- Radiographic evidence of fusion by CT scan (≥ 50% bone bridging across the joint space for the full complement of joints in the absence of secondary surgical or nonsurgical interventions intended to promote fusion)

Non-inferiority of AMPLEX to ABG will be claimed if the lower bound of the 2-sided 90% CI for the difference of the proportions (AMPLEX minus ABG) is greater than -12.0%. The superiority will be claimed if the lower limit of the 2-sided 95% CI is greater than zero. The CI will be constructed with the normal approximation to the binomial.

A subject will be considered a CT radiographic fusion success if he/she has ≥ 50% bone bridging across the joint space by the CT scan for all treated joints without secondary surgical or nonsurgical interventions intended to promote fusion, otherwise, the subject will be considered a CT radiographic fusion failure. The CT radiographic fusion response status will be determined for 12, 24, and 52 weeks. The response status will be assessed for each time point based on the windowed analysis visits. The response status for the absence of secondary interventions to promote fusion will be assessed in a period up to the actual visit date that the corresponding CT images for a particular analysis visit were collected. If any of the components of the CT radiographic fusion success criteria at an analysis visit is missing, the response status will be considered missing for the analysis visit.

For the key secondary analysis, the last available CT radiographic fusion response status will be used (i.e., the last observation carried forward; LOCF) if the response status at 52 weeks is missing. If no CT radiographic response status data are available at previous visits, the response status at 52 weeks will be considered failure.

The same set of sensitivity analyses as for the primary endpoint will be conducted for the key secondary endpoint. These sensitivity analyses will be conducted regardless of the testing results of the primary efficacy endpoint. Therefore, no multiplicity among sensitivity analyses will be adjusted.

9.3.2 Other Secondary Endpoints

The following other secondary efficacy endpoints will be assessed without multiplicity adjustments.

- Proportion of subjects achieving CT radiographic fusion success at 12 and 24 weeks (in the absence of secondary surgical or nonsurgical interventions intended to promote fusion)
- Change from baseline in pain on weight-bearing at fusion site at 12, 24, and 52 weeks (>20 mm reduction from baseline on 100 mm VAS)

- ABG harvest site pain at 2, 6, 12, 24 and 52 weeks (<20 mm on 100 mm VAS)
- Change from baseline in FAAM-ADL at 12, 24 and 52 weeks
- SPC at 12 and 24 weeks
- Change from baseline in Short Form-12 (SF-12) at 24 and 52 weeks

The difference of the proportions of subjects achieving CT radiographic fusion success between treatment groups will be estimated with the 2-sided 90% and 95% confidence intervals at 12 and 24 weeks. Similarly, the difference of the proportions of subjects meeting all the criteria for SPC between treatment groups will also be estimated at 12 and 24 weeks. In addition, the response status for each component of SPC will also be summarized. The missing values will be left as missing.

The change from baseline in FAAM-ADL and weight-bearing pain at fusion site at 12, 24 and 52 weeks will be analyzed by repeated-measures analysis of covariance (ANCOVA) models that includes treatment, time, treatment-by-time interaction, baseline value, presence of non-union risk factors (none vs. any of obesity, diabetes, and previous ankle/hindfoot surgery, or smoking history) at baseline, and fusion sites (single, double, or triple hindfoot, or ankle). The adjusted changes from baseline and their difference between treatment groups will be estimated for each time point. The overall difference over 52 weeks will be estimated as the main effect for treatment in the model. In addition, the proportion of subjects achieving ≥ 8 points improvement from baseline in FAAM-ADL and the proportion of subjects achieving ≥20 mm reduction from baseline in weight-bearing pain at fusion site will be summarized by treatment group at each visit. The missing values will be left as missing.

The ABG harvest site pain at 2, 6, 12, 24 and 52 weeks will be summarized for the ABG group. The proportion of subjects with < 20 mm in ABG harvest site pain will also be summarized at each visit.

The change from baseline in SF-12 scores (PCS and MCS) at 24 and 52 weeks will be analyzed by the repeated measures ANCOVA using the similar ANCOVA models described for FAAM-ADL. Each score will be calculated based on the algorithm to be provided by the licensee.

The FAAM-ADL score will be calculated based on the algorithm described in Appendix 2.

10 Safety

10.1 General Considerations

Safety parameters will be evaluated for the safety analysis set, and the data will be reported according to actual graft material received.

10.2 Adverse Events

Adverse events (AEs) will be classified according to the MedDRA version effective at trial start or later. The version of the MedDRA used for the final analyses will be documented.

A pre-treatment AE will be defined as an AE occurring before the initiation of the surgery. All AEs occurring after the initiation of the surgery will be considered treatment-emergent AEs (TEAEs).

If the timing of an AE cannot be established in relation to the initiation of the surgery, it will be considered as a treatment emergent AE.

Pre-treatment AEs will be reported in a listing only. This listing will be sorted by investigational site, subject ID and date of onset.

Any potential device, procedural, or ancillary device related AEs and all events associated with secondary surgical intervention will be adjudicated for the severity and relatedness by the Clinical Events Committee (CEC). The type of the secondary surgical intervention will also be adjudicated. In the adverse event summaries, the adjudicated data will be used for those events.

The relationship to device (i.e., graft material), ancillary device, index surgical fusion procedure, or bone graft harvest procedure will be evaluated as definitely, probably, possibly, unlikely, unrelated, or unknown. Any events evaluated as definitely, probably, possibly, or unknown will be classified as related.

Device related AEs will include all AEs classified as related to the graft material. Procedure related AEs will include all AEs classified as related to the index surgical fusion procedure or the bone graft harvest procedure.

10.2.1 Overview of Treatment-Emergent Adverse Events

An AE overview summary table will be prepared including the number of subjects reporting an AE, the percentage of subjects with an AE, and the number of events reported, for the following categories:

- Unanticipated adverse device effect
- Any TEAEs
- Device related AEs
- Procedure related AEs

- Deaths
- Serious adverse events (SAEs)
- Device related SAEs
- Procedure related SAEs
- AEs leading to secondary surgical interventions

The relationship to device or procedure will be

10.2.2 Incidence of Adverse Events

AEs will be summarized by SOC sorted alphabetically and PT sorted in decreasing frequency of occurrence. Tables will display the total number of subjects reporting an AE, the percentage of subjects with an AE, and the number of events reported.

Summary tables will be prepared for:

- Unanticipated adverse device effect
- Any TEAEs
- Common TEAEs with an incidence of >5% in any treatment group
- TEAEs by severity
- Device related AEs
- Procedure related AEs
- SAEs
- Device related SAEs
- Procedure related SAEs
- AEs leading to secondary surgical interventions

Supporting data listings will be provided for:

- Unanticipated adverse device effect
- Any TEAEs
- Device related AEs
- Procedure related AEs
- Deaths
- Serious adverse events (SAEs)
- AEs leading to secondary surgical interventions

10.3 Secondary Surgical Intervention

Numbers and percentages of subjects who required any secondary surgical interventions as well as each type of the secondary surgical intervention (revision, removal, reoperation, or supplemental fixation) will be summarized.

10.4 Radiographic Assessment

All radiographic images will be assessed by independent image core laboratory. It will include radiographic assessment of subsidence, device migration, nonunion, osteolysis and/or heterotopic ossification in the area surrounding the implant site. Further details of these assessments will be specified in the Radiographic Evaluation Protocol.

10.4.1 Image Quality

Overall assessments of fusion site visualization on CT and X-ray will be summarized at each visit.

10.4.2 Post-operative Fusion Site

Pre-existing bone abnormality and adequacy of graft fill within the fusion site from the first post-operative (baseline) CT will be summarized for each treated joint.

10.4.3 Safety

Overall hardware condition, abnormal bone changes, graft integrity, graft migration or misplacement, and non-union from CT and X-ray will be summarized for each treated joint (except for overall hardware condition) at each visit.

10.5 Anti-B2A Antibody

Numbers and percentages of subjects with anti-B2A antibodies at screening, 6 weeks, and 12 weeks will be summarized. If a subject had a positive test at 12 weeks, subsequent test results will be presented in a listing.

10.6 Safety Laboratory Variables

Baseline for all laboratory analyses will be the values obtained at the last assessment prior to the surgery. Treatment-emergent laboratory data will include tests completed after the initiation of the surgery.

Safety laboratory variables will be grouped under "Hematology" and "Chemistry".

10.6.1 Summary Statistics

Observed values and changes from baseline for the last post-baseline values will be presented for each laboratory variable. In addition, descriptive statistics will be presented for observed values and change from baseline at each time-point for each laboratory variable.

10.6.2 Laboratory Variable Changes Relative to Normal Range

Shift tables will be prepared to compare baseline values to the last post-baseline values using a categorization of low, normal, and high values for each laboratory variable.

Low, normal, and high will be defined according to the reference ranges provided by the central laboratory in the following manner:

- Low: values which are below the lower reference range limit
- Normal: values which are within the lower and upper reference ranges
- High: values which are above the upper reference range limit

10.6.3 Markedly Abnormal Changes

For livery function tests (AST, ALT, ALP, and total bilirubin levels), a summary table will be prepared displaying the proportion of subjects who have at least one post baseline markedly abnormal value (see Appendix 1). The denominator will be the number of subjects with baseline and at least one post baseline value. The table will also include a break-down by classification of the baseline value (low/normal/high) according to the reference ranges provided by the central laboratory.

10.6.4 Data Listings

All laboratory values will be listed by subject ID and time point. Values outside the reference range and markedly abnormal values will be flagged.

10.7 Vital Signs

Baseline for all vital signs will be the values obtained at the last assessment prior to the surgery. Treatment-emergent vital signs data will include vital signs obtained after the initiation of the surgery. End of trial will include the last post-baseline observation during the trial.

10.7.1 Summary Statistics

Observed values and changes from baseline at the end of trial visit will be presented for each vital sign variable. In addition, descriptive statistics will be presented for observed values and change from baseline at each time-point for each vital sign variable.

10.7.2 Vital Signs Abnormal Values

Shift tables will be prepared to compare baseline values to the end of trial visit using a categorization of normal or abnormal determined by the investigator for each vital sign variable.

10.7.3 Data Listings

Data listings will be prepared by treatment group for all subjects with any markedly abnormal vital sign values at any time-point (including baseline).

All vital sign values will be listed by subject ID and time point. Values determined abnormal by the investigator and markedly abnormal values will be flagged.

11 Interim analyses

No interim analysis is planned.

12 Deviations from protocol analysis

There is no deviation from the planned analysis described in the protocol.

13 References

1 DiGiovanni, C. W., Lin, S. S., Baumhauer, J. F., Daniels, T., Younger, Al., Glazebrook, M., Lynch, S. E., and the North American Orthopedic Foot and Ankle Study Group. (2013). Recombinant human platelet-derived growth factor-BB and beta-tricalcium phosphate (rhPDGF-BB/β-TCP): an alternative to autogenous bone graft. *Journal of Bone and Joint Surgery, 95*, 1184-1192.

2 Martin, R. L., Irrgang, J. J., Burdett, R. G., Conti, S. F., & Van Swearingen, J. M. (2005). Evidence of validity for the foot and ankle ability measure (FAAM). *Foot & Ankle International, 26,* 968-983.

3 Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

4 U.S. Food and Drug Administration. (2015). Proposed summary of safety and effectiveness data (SSED). Retrieved from http://www.accessdata.fda.gov/cdrh_docs/pdf10/P100006b.pdf.

5 Zhang, L., Yang, H., & Cho, I. (2009). Test homogeneity of risk difference across subgroups in clinical trials. *Journal of Biopharmaceutical Statistics, 19*, 67-76.

Appendices

Appendix 1 Markedly Abnormal Laboratory Values

Table 1: Markedly abnormal Criteria for Laboratory Tests

Variable	Markedly abnormal Criteria	
	Low	High
AST	Not applicable	> 3 times of upper normal limit
ALT	Not applicable	> 3 times of upper normal limit
ALP	Not applicable	> 1.5 times of upper normal limit
Total bilirubin	Not applicable	> 2 times of upper normal limit

Appendix 2 FAAM-ADL

The FAAM-ADL score will be calculated based on the following algorithm described in Martin et al. (2005) (2):

1) The response to each item on the FAAM-ADL will be scored from 4 to 0, with 4 being 'no difficulty' and 0 being 'unable to do'

2) Responses marked as not applicable will not be counted

3) Add together the scores on each of the items to get the item score total

4) Calculate the total number of items with a response

5) Multiply the value in 4) by 4 to obtain the highest potential score (e.g., 84 if all 21 items are responded)

6) The total item score calculated at 3) will be divided by the highest potential score calculated at 5) and then multiplied by 100 to produce the FAAM-ADL score that ranged from 0 to 100

Appendix 3 Schedule of Assessment

	Screening	Treatment	Follow-up					
Visit number	**V1**	**V2 Procedure**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**
Weeks	**-12**	**--**	**2**	**6**	**12**	**24**	**52**	**78**
Days	**Days -84 to -1**	**Day 1**	**Day 15 (+/- 3)**	**Day 43 (+/- 14)**	**Day 85 (+/- 14)**	**Day 169 (+/- 30)**	**Day 365 (+/- 30)**	**Day 547 (+/- 30)**
Informed consent	X							
Medical History/ comorbidities	X	X						
Supplemental graft requirement confirmation	X							
Eligibility criteria verification	X	X						
Physical assessment of foot/hindfoot, vital signs	X	X	X	X	X	X	X	X
Pregnancy test (if applicable)	X	X						
Standard Unrinalysis [1]	X							
Clin. labs: chemistry and hematology	X		X	X	X			
Immunological Testing for HBV, HCV and HIV	X							
Serum for B2A Ab testing [2]	X			X	X			
Identification of target arthrodesis site	X	X						
Subject randomization [3]		X						
Arthrodesis procedure		X						
Specific intraoperative data collection [4]		X						
Graft material volumes		X						
A/P, lateral and oblique X-rays of hindfoot/ankle	X†	X [5]		X	X	X	X	X
Foot and ankle CT scan		X [5]			X	X	X	
FAAM-ADL	X				X	X	X	
Weight bearing pain at fusion site VAS	X				X	X	X	
Graft Harvest site pain VAS			X	X	X	X	X	
SF-12	X					X	X	
Adverse event evaluation [6]		X	X	X	X	X	X	X
Concomitant Medications	X	X	X	X	X	X	X	X

† At the screening/pre-operative visit, A/P, oblique and lateral X-rays will be collected (historical X-rays taken up to 24 weeks prior to the planned date of surgery will also be accepted).
‡ Collect at "discharge", e.g., after implant but before hospital discharge.
[1] If positive for blood, leukocytes, or nitrite, microscopic unrinalysis will be performed.
[2] B2A Antibody testing may continue past the 12 week follow-up visit if there is a positive test at 12 weeks.
[3] The subject will be randomized as close to the time of surgery as possible, with the recommendation that the randomization be performed on the day of the surgical procedure.
[4] See Section **Error! Reference source not found.** referencing the Surgical Manual.
[5] Perform immediately following surgery or within 1 week following the procedure.
[6] Collected from time of randomization and will include pain at the fusion site.