Institut Straumann AG

Immediate Placement of the Straumann® Bone Level Tapered Implant with Early Loading in Single Tooth Gaps in the Maxilla and Mandible Compared to Delayed Placement

CR 01/14
Protocol Version 4.0

Statistical Analysis Plan

Version: 1.0 – 25 Nov 2019

Institut Straumann AG
Peter Merian Weg 12
CH- 4002 Basel
Switzerland

CONFIDENTIAL **DO NOT COPY**

Primary Author:
Matt Partridge
Biostatistician II
Regulatory and Clinical Research Institute, Inc.
Minneapolis MN, 55416
Ph: (952) 746-8020

25 Nov 2019
Signature / Date

Primary Sponsor Contact:
Michel Mallaun
Head Clinical Operations
Institut Straumann AG
Peter Merian-Weg 12
CH-4052 Basel
Switzerland
Ph: +41 (0)61 965 14 23

26 Nov 2019
Signature / Date

Primary Sponsor Contact:
Robert Calhoun
Clinical Study Manager
Straumann USA, LLC
60 Minuteman Road
Andover, MA 01810
PH: (978) 747-2531

02 Dec 2019
Signature / Date

TABLE OF CONTENTS

1 Purpose of the Statistical Analysis Plan

The purpose of this Statistical Analysis Plan (SAP) is to summarize the statistical methodologies being used to confirm the primary effectiveness endpoint and assess the secondary effectiveness, additional effectiveness, and safety endpoints of the immediate implantation of the Straumann® Bone Level Tapered Implant (study device). The SAP will be use should there be any discrepancies with the protocol.

2 Study Design

2.1 Study Design/Overview

This is a post-market, prospective, randomized, controlled, multi-center study with five years of follow-up. The study is being performed at three centers within the United States. A total of 53 subjects were enrolled and randomized with 46 subjects having completed the 12 months post-implant loading (primary endpoint) follow-up. No more than 30 subjects were enrolled per center and subjects were randomized equally to have immediate (test treatment) or delayed (control treatment) implantation.

2.2 Device Description

The Straumann® Bone Level Tapered implant is used for both the treatment and control groups as the investigation is of the implant timing. The implant is available in three different diameters (3.3, 4.1, and 4.8 mm) and five different lengths (8, 10, 12, 14, and 16 mm). In addition to the implant device, clinicians may also use Straumann® restorative components that are appropriate for the implant.

2.3 Study Schedule

All subjects were screened (Visit 1) for study eligibility within 45 days of the tooth extraction and randomization visit (Visit 2a). Subject were randomized during the extraction procedure after the tooth had been extracted. Each subject was randomized to have the implantation surgery (Visit 2b) either immediately, on the same day as the extraction (test), or delayed, after 17 weeks from the extraction (control).

Subjects had a first post-operative assessment 10 days after implantation (Visit 3) to assess wound healing and remove sutures. The implant was then uncovered (Visit 4) and loaded (Visit 5) after 10 ± 3 weeks from implantation, which may have been performed at the same visit. An impression for final restoration was taken 5 weeks after implant loading (Visit 6) with the final restoration having been performed 10 weeks after implant loading (Visit 7).

Subjects have had follow-up visits after 6 and 12 (primary endpoint) months from implant loading (Visits 8 and 9 respectively) and will have additional follow-up visits after 24, 36, 48, and 60 months from implant loading (Visits 10, 11, 12, and 13 respectively).

2.4 Study Success

The study will be deemed a success if the change in mean peri-implant marginal bone levels (mesial and distal) from implant loading to 12 months-post loading for the test treatment is no worse than that the control treatment.

2.5 Randomization and Enrollment

Randomization occurred during the tooth extraction visit immediately after the tooth had been extracted and the inclusion and exclusion criteria had been confirmed. Subjects were enrolled upon randomization.

3 Study Objectives

This study is designed to assess the clinical and radiographic outcomes of using a Straumann® Bone Level Tapered implant for immediate implantation following extraction of a tooth in the pre-molar and anterior region of the maxilla and mandible (test) compared to the outcomes of placing this implant in healed sites (control). The outcomes that will be compared include the following:

- Change in mean peri-implant marginal bone levels (mesial and distal)
- Implant success and survival
- Buccal bone dimensional change
- Implant stability
- Soft tissue changes
- Subject satisfaction
- Adverse events

4 Effectiveness and Safety Endpoints

4.1 Primary Effectiveness Endpoint

The primary effectiveness endpoint is defined as the change in the mean of the mesial and distal aspect crestal bone levels from implant loading to 12 months post-implant loading.

4.2 Secondary Effectiveness Endpoints

The five secondary effectiveness endpoints are defined as follows:

- Implant success and survival at 12 months post-implant loading
- Buccal bone dimensional change, measured on cone beam computed tomography (CBCT) images, from implant loading to 12 months post-implant loading
- Implant stability measured by radio frequency analysis (RFA) at implant placement, implant loading, and final restoration
- Gingival and papilla soft tissue change from final restoration to 12 months post-implant loading
- Subject satisfaction of esthetics, function and level of pain at 12 months post-implant loading

4.3 Additional Effectiveness Endpoints

The five additional effectiveness endpoints are each evaluated at 24, 36, 48, and 60 months post-implant loading follow-ups and are defined as follows:

- Change in mean crestal bone levels from implant loading to follow-up
- Implant success and survival at follow-up
- Buccal bone dimensional change from implant loading to follow-up
- Gingival and papilla soft tissue change from final restoration to follow-up
- Subject satisfaction of esthetics, function and level of pain at follow-up

4.4 Safety Endpoints

The safety endpoints are defined the as the number of adverse events or adverse device effects experienced and the number and proportion of subjects experiencing an adverse event or adverse device effect throughout the study.

5 Analysis of Primary Effectiveness Endpoint

The primary effectiveness endpoint evaluates the difference in the change in mean crestal bone levels from implant loading to 12 months post-implant loading between the treatment and control groups. The change is defined as the mean crestal bone level of the mesial and distal aspects of the implant at 12

months post-implant loading minus at implant loading and the difference is defined as the change for the treatment group minus the control group.

The mean crestal bone levels at each visit, and the change in mean crestal bone levels will be summarized for each group with the count, average, standard deviation, median, minimum, and maximum. The mean change in crestal bone levels will also be categorized as Gain (>0), No change (0), Loss <1 mm, Loss <2 mm, Loss <3 mm, or loss ≥3 mm and summarized with the number and proportion of subjects in each category for each treatment group.

5.1 Sample Size

This is a post-market, prospective, randomized, controlled, multi-center study. The sample size and study duration were determined to allow for adequate healing and the opportunity for comparable changes. In addition, a statistically powered sample size of 28 subjects (14 subjects per treatment group) is required for the primary endpoint. The sample size estimation is based on the following assumptions:

- Subjects are assigned to either the test or control treatment group
- The expected difference between treatment groups is 0 mm
- The standard deviation for both the test and control treatments is 0.5 mm
- A clinically significant difference between treatment groups is ≥ 0.5 mm (non-inferiority margin, δ)
- Testing will use a one-sided t-test having at least 80% power and a 5% significance level

5.2 Primary Analysis

The null hypothesis for the primary analysis is that the difference between the mean change in crestal bone levels for the treatment and control groups is at least the non-inferiority margin (indicating inferior bone gain). Rejection of the null hypothesis indicates the observed data supports the alternative hypothesis that the difference between the mean change in crestal bone levels for the treatment and control groups is less than the non-inferiority margin (indicating non-inferior bone gain). The hypotheses associated with the primary analysis are defined as:

- H_0: $\mu_c - \mu_s \geq \delta$
- H_1: $\mu_c - \mu_s < \delta$

where μ_c is the mean change in crestal bone levels from implant loading to 12 months post-implant loading for the treatment group, μ_s is the mean change for the control group, and δ is the 0.5 mm non-inferiority margin. The hypotheses will be tested using a one-sided t-test with a 5% significance level.

5.3 Secondary Analysis

If the null hypothesis for the primary analysis is rejected with a one-sided 5% significance level, an additional secondary analysis will be performed testing superiority, where the superiority margin is $-\delta$. The null hypothesis for the secondary analysis is that the difference between the mean change in crestal bone levels for the treatment and control groups is at most the superiority margin (indicating non-superior bone gain). Rejection of the null hypothesis indicates the observed data supports the alternative hypothesis that the difference between the mean change in crestal bone levels for the treatment and control groups is greater than the superiority margin (indicating superior bone gain). The hypotheses associated with the secondary analysis are defined as:

- H_0: $\mu_c - \mu_s \leq -\delta$
- H_1: $\mu_c - \mu_s > -\delta$

where μ_c, μ_s, and δ are defined as above. The hypotheses will be tested using a one-sided t-test with a 5% significance level.

6 Analysis of Secondary Effectiveness Endpoints

All secondary effectiveness endpoints will be summarized descriptively without hypothesis testing.

6.1 Implant Success and Survival

A surviving implant is defined as the implant being in place at 12 months post-loading. An implant success is defined as a surviving implant that is absent of persistent subjective complaints, recurrent peri-implant infection with suppuration, mobility, and continuous radiolucency around the implant at 12 months post-loading. The implant survival/success at 12 months post-loading will be summarized with the number and percentage of subjects with a surviving/successful implant.

6.2 Buccal Bone Dimensional Change

Buccal bone dimensional change is defined as the buccal bone dimensional thickness at 12 months post-loading minus the dimension thickness at implant loading. The buccal bone dimensional thickness at implant loading and 12 months post-loading will be summarized in addition to the change at 12 months post-loading with the count, mean, standard deviation, median, minimum, and maximum.

6.3 Implant Stability

Implant stability is defined as the ISQ category. The facial-lingual and mesial-distal implant stability at implant placement, implant loading, and final restoration will be summarized with the number and percentage of subjects in each category.

6.4 Soft Tissue Change

Soft tissue change is defined by the gingival (CLT_m, CLT_d, and CLI) and papilla (IP_m and IP_d) margin measurements. The soft tissue measurements at final restoration and 12 months post-loading will be summarized in addition to the change at 12 months post-loading with the count, mean, standard deviation, median, minimum, and maximum.

6.5 Subject Satisfaction

Subject satisfaction is defined as the VAS value. The satisfaction with esthetics, function, and level of pain will be summarized at 12 months post-loading with the count, mean, standard deviation, median, minimum, and maximum.

7 Analysis of Additional Effectiveness Endpoints

All additional effectiveness endpoints will be summarized descriptively without hypothesis testing. The endpoints will be summarized as defined above evaluated at 24, 36, 48, and 60 months post-loading.

8 Analysis of Safety Endpoints

All safety endpoints over the course of the study will be summarized descriptively without hypothesis testing. Events meeting at least one of the criteria for seriousness will be considered serious. The adverse events and adverse device effects will be summarized and tabulated by type, seriousness, relationship to the study device or the procedure, severity, outcome, and expectedness with the number of events/effects and the number and percentage of subjects experiencing the event/effect. Except where indicated, a subject experiencing the same event/effect more than once will still be counted as one subject experiencing the event/effect.

8.1 Relationship to Study Device or Procedure

The causal relationship between the adverse event and the study device or procedure will be defined as follows:

- Definitely related: Reasonable causal and temporal relationship
- Possibly related: Causal and temporal relationship is less likely; however, the determination that there is no relationship cannot be made
- Not related: Causal relationship can be definitely excluded

For purposes of dichotomizing the causal relationship in safety summaries, events that are "Definitely related" or "Possibly related" will be considered related.

8.2 Severity

The severity, or intensity of an event experienced by a subject, will be defined as follows:

- Mild: Discomfort noticed, but no disruption in daily activities; the event is easily tolerated by the subject
- Moderate: Discomfort sufficiently enough to reduce or affect normal daily activity
- Severe: Inability to work or perform normal daily activity and/or the subject's life is at risk from the adverse event

8.3 Outcome

The outcome, or status of the adverse event at the time of recording, will be defined as follows:

- Resolved: Full recovered from the event without any sequelae or if it is unknown whether there are sequelae
- Resolved with sequelae: Condition stabilized despite the persistence of sequelae (e.g., lesion or medical condition which is a consequence of the event)
- Ongoing: Not yet recovered
- Worsened: Severity of the AE/ADE increased
- Fatal: Related to a death; whether it caused death or contributed to it
- Unknown: Knowledge of the current status of the AE/ADE is truly not available to the investigator (i.e. event was ongoing at last observation, but no further contact with the subject could be established)

8.4 Expectedness

The expectedness of adverse events determined to be related to the device will be defined as follows:

- Expected: Reaction is consistent with the effects of the device listed in the IFU and protocol.
- Unexpected: Reaction is not consistent with the effects listed in the IFU and protocol.

9 Measurement Scales

9.1 Implant Stability Quotient (ISQ)

The Implant Stability Quotient (ISQ) is a value from 1-100 indicating the stability of the implant where a higher ISQ means a more stable implant. The ISQ is then categorized as ISQ < 60, ISQ 60 to 69, and ISQ ≥ 70. The ISQ is provided by an Osstell device using the RFA method and both the facial-lingual and mesial-distal implant stability will be obtained.

9.2 Gingival Margins (CLT_m, CLT_d, and CLI)

The gingival margins are the lengths (mm) of the crown from the highest point of soft tissue to the incisal edge of the adjacent mesial (CLT_m) and distal (CLT_d) teeth and to the incisal edge (CLI).

9.3 Papilla Margins (IP_m and IP_d)

The papilla margins are the distances (mm) from the top of the papilla to the incisal edge mesial (IPm) and distal (IPd) of the implant crown.

9.4 Visual Analogs Scale (VAS)

The Visual Analogs Scale (VAS) value is the distance (mm) from the beginning of a 100 mm line to the point on the line indicating the subject's satisfaction with the factor being assessed, where a value that is larger (closer to end of the line) means greater satisfaction. Possible values are from 0-100 mm.

10 General Statistical Considerations

10.1 Analysis Sets

There are three analysis sets planned for this study.

10.1.1 Intention to Treat (ITT) Analysis Set

The ITT analysis set consists of all subjects who are consented in the study and randomized (i.e., enrolled). This population includes all subjects, regardless of any protocol deviations and/or premature termination. Subjects are classified based on their randomization assignment rather than the treatment received. The ITT analysis set will be used for presenting baseline data, procedural information, adverse events and protocol deviations. The ITT analysis set will also be used to provide a sensitivity analysis of the primary effectiveness endpoint assessing the robustness of the findings based on the mITT analysis set.

10.1.2 Modified Intention to Treat (mITT) Analysis Set

The mITT analysis set is a subset of the ITT analysis set. This population only includes subjects that received the study treatment, regardless of any protocol deviations and/or premature termination. Subjects are classified based on their randomization assignment rather than the treatment received. The mITT analysis set will be used as the primary analysis set for all primary and secondary effectiveness endpoints.

10.1.3 Per Protocol (PP) Analysis Set

The PP analysis set is a subset of the mITT analysis set.

This population only includes subjects who received the study treatment according to the protocol and randomization schedule, had no significant protocol deviations, and reached primary stability. The list of subjects to be excluded from the PP analysis set will be made while blinded to subject randomization and primary endpoint result and prior to final analysis. The PP analysis set will be used to provide a sensitivity analysis of the primary effectiveness endpoint assessing the robustness of the findings based on the mITT analysis set.

10.2 Control of Systemic Bias

Minimization of study bias has been designed into the study in the following, and other, ways:

- A multi-center trial helps minimize investigator, site, or subject enrollment bias
- This document specifies appropriate statistical methodology to ensure the minimization of bias
- The primary effectiveness measures will be based on the readings from a centralized radiologist, ensuring consistent and objective reporting

10.3 Pooling Data

10.3.1 Pooling Across Centers

The three centers each enrolled no more than 30 subjects in the study. For the purpose of performing statistical analyses, data from all study centers will be pooled.

10.3.2 Pooling Across Protocol Versions at Enrollment

During the study, some subjects were enrolled on version 2 of the protocol (cohort 1), which does not state requirements for bone grafting of the extraction socket, and some on version 3 of the protocol (cohort 2), which states all subjects will have bone grafting of the extraction socket performed, regardless of randomization assignment. Due to the procedural difference caused by the protocol revision, the validity of pooling the cohorts will be statistically analyzed based on the primary effectiveness endpoint.

The treatment effect difference on the change in mean crestal bone levels from implant loading to 12 months post-loading between the two cohorts will be evaluated with a linear regression model. The model will be defined as:

- $\Delta = \beta_0 + \beta_1 T + \beta_2 C + \beta_3 TC + \epsilon$

where Δ is the change in mean crestal bone levels, T is an indicator of the treatment group (0=control, 1=treatment), C is an indicator of the cohort (0=Cohort 1, 1=Cohort 2), TC is the interaction between the treatment group and the cohort, β_0, β_1, β_2, and β_3 are the parameter estimates, and ϵ is the residual error.

The null hypothesis is that treatment effect on the change in mean crestal bone levels for cohort 1 is equivalent to cohort 2 (indicating the cohorts can be pooled). Rejection of the null hypothesis indicates the observed data supports the alternative hypothesis that the treatment effect on the change in mean crestal bone levels for cohort 1 is different than cohort 2 (indicating the cohorts remain separated). The hypotheses associated with this analysis are defined as:

- H_0: $\beta_3 = 0$
- H_1: $\beta_3 \neq 0$

where β_3 is the parameter estimate for the interaction between the treatment group and cohort from the linear regression model. The hypotheses will be tested using a two-sided t test with a 5% significance level.

Rejection of the null hypothesis will require additional presentation of separated cohorts for all summaries and endpoints described in this document. The linear regression model will be summarized with the parameter estimates, 95% confidence intervals for the parameter estimates, standard errors, t statistics, and p-values. In addition, the change in mean crestal bone levels for each cohort will be summarized as previously defined. Population characteristics for each cohort will also be summarized with the count, mean, standard deviation, median, minimum and maximum for continuous characteristics or with the number and percentage of subjects for categorical characteristics.

10.4 Handling of Missing Data

All practical monitoring and follow-up steps will be taken to ensure complete and accurate data collection. All analyses will be based on available data only; no imputation for missing data is planned.

10.5 Other Data Summaries

Subject disposition and population characteristics (such as age, gender, medical history, etc.) will be summarized descriptively. Unless otherwise stated, characteristics will be summarized with the count, mean, standard deviation, median, minimum, and maximum for continuous characteristics and with frequency tables (frequencies, percentages, confidence intervals) for categorical characteristics. Hypothesis tests will use a two-sided 5% significance level and will be based on two sample t-tests for continuous characteristics and on Chi-square (or Fisher exact) tests for categorical characteristics. Non-parametric Wilcoxon signed rank test may be used for continuous characteristics when underlying data are found to be non-normal.

10.6 Changes to Planned Analyses

Changes made prior to the analysis of the primary effectiveness endpoint will be documented in an amended SAP approved prior to said analysis. All other changes will be documented in the clinical study report with the deviation reason.

10.7 Statistical Software

The statistical software package SAS® 9.4 or later will be used for all the data derivations, summarization, data listings and statistical analyses. Additional statistical software may be used for graphics or validation purposes as appropriate.