A Prospective, Multicenter, Single Arm Clinical Study Evaluating the Use of the Renuvion Dermal System for Dermal Resurfacing

Statistical Analysis Plan

Study Protocol No: VP-1909

Revision No.: 2

Version Date: December 10, 2019

Table of Contents

List of Tables

LIST OF ABBREVIATIONS

AE	Adverse Event
CIP	Clinical Investigational Plan
CRF	Case Report Form
CRO	Clinical Research Organization
CSR	Clinical Study Report

DCF	Data Clarification Form
DRM	Data Review Meeting
DR-(S)AE	Device-Related (Serious) Adverse Event
FAS	Full Analysis Set
FDA	Food and Drug Administration
FSS	Fitzpatrick Skin Scale
FWS	Fitzpatrick Wrinkle and Elastosis Scale
GAIS	Global Aesthetic Improvement Scale
GCP	Good Clinical Practice
ICH	International Conference for Harmonization of Technical Requirements of Pharmaceuticals for Human Use
IPR	Independent Photographic Reviewer
ITT	Intent-to-Treat
MDD	Medical Device Directive
mITT	Modified Intent-to-Treat
PP	Per Protocol
PPS	Per Protocol Set
PR-(S)AE	Procedure-Related (Serious) Adverse Event
RF	Radiofrequency
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
UADE	Unanticipated Adverse Device Effect
VAS	Visual Analog Scale

1 PURPOSE

The purpose of this Statistical Analysis Plan (SAP) is to document the statistical design considerations for the study of a new medical device, the Renuvion Dermal System for Dermal Resurfacing. This SAP documents the sample size estimation, study objectives, endpoints, data collection tools, and analysis plan. The requirements of the United States FDA and European MDD have been considered in the development of this document.

Once the study is completed a 510(k) will be submitted to the agency to request FDA clearance.

2 REFERENCE DOCUMENTS

This SAP corresponds to the Apyx Clinical Investigation Protocol (CIP), "A Prospective, Multicenter, Single Arm Clinical Study Evaluating the Use of the Renuvion Dermal System for Dermal Resurfacing", Protocol # VP-1909, Rev 3 (the "Dermal Resurfacing Study"). Because detailed clinical information is included in the study protocol, the clinical aspects are *not* repeated in this SAP document. The information included in this SAP relates to the study design and data analysis.

3 OVERVIEW OF STATISTICAL DESIGN

This is a multi-center, single arm, single-blind (evaluator) prospective study of up to 55 study subjects who are seeking a procedure to reduce the appearance of wrinkles and rhytides at up to 5 investigational centers in the United States. The maximum enrollment per site is 20 subjects.

Study subjects that meet study eligibility criteria and have provided informed consent will be enrolled in the study. During the procedure, the investigators will use Renuvion® Dermal System on applicable facial zones to reduce wrinkles and rhytides.

Study subjects will be followed immediately following the procedure, at 1, 6, 10, 30, 90, and 180-days post-procedure for study assessments. Assessments of endpoints will use standardized evaluation tools and blinded evaluators as applicable. This includes the use of an independent 3-member evaluation team to eliminate the bias that is likely inherent in the investigators' assessments.

Study enrollment is expected to occur over 3-6 months. Imaging and study assessments will continue through 6 months post-procedure. Total study duration is expected to be approximately 9-12 months. It is expected that the 510(k) application for the device will be submitted based on 90-day post-procedure results. However, this clinical trial will continue until every enrolled subject has reached 180-days following their procedure and all adverse events are resolved. At that time, the trial will be considered complete, the final results will be determined, and a final report will be prepared.

Although every possible step will be taken to mitigate the procedure risks, it is possible that dermal treatments can cause pain, discoloration, and even scarring. Thus, adverse events are carefully collected and analyzed as part of this study. Table 1 below shows the required evaluations and the schedule on which they are to occur.

Table 1: Study Required Procedures

	Baseline/ Pre-Procedure Screening[1]	**Procedure**	**1 Day**	**6, 10 Days**	**30 Days**	**90 Days**	**180 Days**
			1 day	6+2 days and 10+4/-1 days	30±7 days	90±10 days	180±14 days
Informed Consent	√						
Assess Inclusion/Exclusion Criteria	√						
Urine Pregnancy Test[2]	√	√					
Medical History	√						
General Physical Exam	√						
Review Medications	√		√	√	√	√	√
Photographic Images[3]	√[9]		√	√	√	√[9]	√
Fitzpatrick Skin Type Scale (FST)	√						
Fitzpatrick Wrinkle and Elastosis Scale (FWS) [4]	√				√	√	√
Visual Analog Scale (11-point VAS) [5]		√	√	√	√	√	√
Study Procedure		√					
Debride Treatment Area (PRN)			√	√			
Subject Diary (11-point VAS)[6]		√	√	√			
Adverse Event Assessment		√	√	√	√	√	√
Re-epithelization and Down Time[7]			√	√	√	√	
Modified Global Aesthetic Improvement Scale (GAIS) [8]					√	√	√
Subject Satisfaction Survey						√	

[1] Pre-procedure Screening assessments to take place within 30 days prior to undergoing the procedure.
[2] Up to two urine pregnancy tests must be obtained prior to study procedure for females with child-bearing potential (one at pre-procedure screening and one on the day of the procedure prior to the procedure if screening and procedure are not performed on the same day).
[3]Digital photographs of the subject's face will be taken and labeled according to Photography Instructions.
[4]To be completed by Investigator.
[5] To be completed by the study subject on a day of the procedure (prior to the procedure and immediately following the procedure, e.g. within 60 minutes) and at specified follow-up visits.
[6] To be completed by the study subject daily starting from the day of procedure (after procedure, at home) until the 10 day follow-up visit.
[7] To be completed by Investigator to capture achievement of epidermal recovery status at follow-up visits; and date when study subject felt comfortable, willing and able to go in public following study procedure (assessed at the 10 day follow-up visit).
[8] To be completed by Investigator at all follow-up visits and study subject at 30, 90, and 180-day follow-up visits.
[9] Images taken on this visit as primary endpoint for effectiveness.

Table 1. Study Required Procedures

4 STUDY OBJECTIVES AND ENDPOINTS

4.1 Primary Effectiveness Objective and Endpoint

The primary effectiveness objective is to assess the proportion of subjects that show clinically significant improvement in the texture of the skin at 90 days post-procedure.

Three experienced, blinded Independent Photographic Reviewers (IPRS) will perform an analysis/review of the pre-treatment and post-treatment sets of images of each subject in a blinded and randomized order. The endpoint is the proportion of subjects with at least one-point improvement from baseline in the Fitzpatrick Wrinkle and Elastosis Scale (FWS) at 90 days as determined by 2 out of 3 blinded IPRS**.** On each subject, the appearance of facial wrinkles and rhytides on the baseline image and the 90-day image will be assessed by the IPRs using the FWS. Each subject will have three change scores; if 2 or more of the 3 change scores are one point or higher, the subject will be considered a success (Y or 1) and if 2 or more of the 3 change scores are less than one point, the subject will be considered a failure (N or 0).

4.2 Primary Safety Objective and Endpoint

The primary safety objective is to estimate the rates of the following classifications of adverse events (AE) through 90-day post-procedure:

- All AEs
- All serious AEs (SAEs)
- Device- and treatment-related AEs (DR-AEs)
- Procedure-related AEs (PR-AEs)
- DR-SAEs
- PR-SAEs

The primary safety endpoint is the enumeration of each subject's AEs up to the 90-day visit after treatment. Each AE will be categorized by cause, severity, seriousness, and relatedness to the procedure, and the device or treatment. The rating scales for these categories is provided later in this document; definitions of each are provided in the CIP.

4.3 Secondary Safety Objective and Endpoint

The secondary safety objective is to characterize the amount of pain that patients experience in the 10 days beginning with the procedure.

The secondary safety endpoint is the evaluation of the change in pain and discomfort after treatment (baseline, within 60 minutes following the procedure) experienced in the period up to the 10-day follow-up visit. Pain or discomfort will be recorded daily by each participant in a diary using an 11-point Visual Analogue Scale (VAS).

4.4 Additional Endpoints

Other endpoints to be evaluated include:

1. Whether or not (Yes/No) at least 2 out of 3 blinded Independent Photographic Reviewers (IPRs) correctly identify the 90-day image of a subject from the pair of baseline and 90-day images.
2. Magnitude of improvement measured by the mean change in FWS from baseline to 90-dayvisit as determined by Investigators.
3. Subject modified GAIS at 90-day FUV.
4. Investigator modified GAIS at 90-day FUV.
5. Subject satisfaction with procedure recorded at the 90-day visit.
6. Achievement of re-epithelialization by facial zone and across all facial zones at the 10 day, 30-day and 90-day follow-up visits as reported by the investigator.
7. Mean duration until study subject feels comfortable going in public after treatment as reported by the study subject.
8. Daily 11-point Visual Analog Scale (VAS) pain assessment following treatment through the 10-day follow-up visit by diary day with a change from the VAS pain score at baseline.

5 EVALUATION TOOLS

The following evaluation tools will be used in this study. More information about them, including how to administer them, is included in the protocol. The schedule on which they are administered is provided in Table 1.

5.1 Fitzpatrick Skin Type Scale (FST)

Assessment of subject's skin color will be determined prior to study procedure by the Investigator using the FST. The scale delineates skin color into the categories as shown in Table 2.

Table 2: Fitzpatrick Skin Type Scale (FST)

Table 2: Fitzpatrick Skin Type Scale Evaluation	
Skin Type	Description
Type I	White skin that never tans and always burns easily
Type II	White skin that tans slightly and always burns easily
Type III	Light brown skin that tans gradually and can burn moderately
Type IV	Moderately brown skin that tans well and burns slightly
Type V	Dark brown skin that tans profusely and burns rarely
Type VI	Black skin with deep pigmentation that never burns

5.2 Fitzpatrick Wrinkle and Elastosis Scale (FWS)

Assessment of each subject's wrinkles at baseline and at the 30, 90, and 180-day follow-up visits will be performed by the Investigator. As well, assessment of each subject's baseline and the 90-day follow-up visit images viewed simultaneously will be performed by the Independent Photographic Reviewers (IPRs).

Table 3: Fitzpatrick Wrinkle and Elastosis Scale

Table 3: Fitzpatrick Wrinkle and Elastosis Scale			
Class	Description	Score	Description
I	Fine wrinkles	1-3	**Mild:** Fine texture changes with subtly accentuated skin lines.
II	Fine to moderate depth wrinkles, Moderate number of lines	4-6	**Moderate:** Distinct papular elastosis (individual papules with yellow translucency under direct lighting) and dyschromia.
III	Fine to deep wrinkles, numerous lines, with or without redundant skin folds	7-9	**Severe:** Multipapular and confluent elastosis (thickened, yellow and pallid) approaching or consistent with cutis rhomboidalis.

5.3 Modified Global Aesthetic Improvement Scale (GAIS)

The Global Aesthetic Improvement Scale (GAIS) is a subjective rating of improvement in treatment results compared to pre-treatment. A modification of the GAIS to include "much worse" and "very much worse" as rating options will be used in this study. The Investigator will grade the overall improvement of treatment area as indicated in Table 4a by comparing the subject's appearance at follow-up visits against a photograph taken prior to procedure. Likewise, the subject will also rate their improvement compared to pre-treatment as shown in Table 4b.

The modified GAIS results will be collected at the 30, 90, and 180-day follow-up visits.

Table 4: Modified Global Aesthetic Improvement Scale Evaluation

Table 4a: Modified Global Aesthetic Improvement Scale Evaluation (GAIS): Investigator	
Rating	Description
Very much improved	Optimal cosmetic result from this procedure in this subject
Much improved	Marked improvement in appearance from the initial condition, but not completely optimal for this subject

Improved	Obvious improvement in appearance from the initial condition
No change	The appearance is essentially the same as the original condition
Worse	The appearance is worse than the original condition
Much worse	The appearance is much worse than the original condition
Very much worse	The appearance is very much worse than the original condition

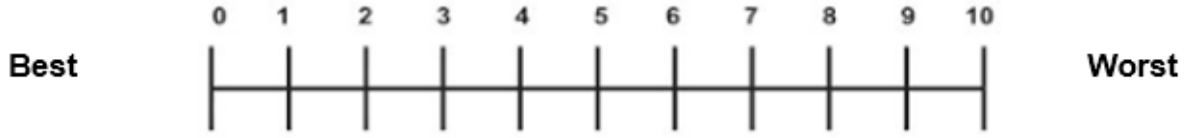
Table 4b: Modified Global Aesthetic Improvement Scale Evaluation (GAIS): Subject Rating	
Very much improved	□ Optimal cosmetic result.
Much improved	□ Marked improvement in appearance from the initial condition, but not completely optimal.
Improved	□ Obvious improvement in appearance from initial condition.
No change	□ The appearance is essentially the same as the original condition
Worse	□ The appearance is worse than the original condition.
Much worse	□ The appearance is much worse than the original condition.
Very much worse	□ The appearance is very much worse than the original condition.

5.4 Visual Analog Scale (VAS)

The study subjects will be asked to complete an 11-point Visual Analog Scale (VAS) for the level of pain and discomfort associated with study procedure (prior to the procedure and immediately following the procedure), and the follow-up assessments at the 10-day follow-up visit, and at all follow-up visits.

Scoring for the VAS will consist of making a mark on a 10-cm line demarcated at 1-cm intervals. Each end of the line will be awarded a score of 0 or 10 according to the extreme points of reference pertaining to an individual measure.

Figure 1: Visual Analog Scale



Subject satisfaction with treatment will be assessed at the 3-month visit using the Subject Satisfaction Survey. Additionally, the subjects will be asked if they would recommend the treatment to friends and acquaintances (yes, perhaps, or no) and improvements noted (if any).

Figure 2: Subject Satisfaction Survey

1. Are you happy with your results of the procedure performed on your face? ☐ Yes ☐ No

2. Would you recommend the procedure performed on your face to a friend? ☐ Yes ☐ No

3. Would you consider having the procedure that was performed on your face **performed again?** ☐ Yes ☐ No

4. Which, if any, changes do you see in the area treated?
   - **a. Skin Texture Improvement** ☐ Yes ☐ No
   - b. Skin Tone Improvement ☐ Yes ☐ No
   - c. Skin Pigmentation Improvement ☐ Yes ☐ No
   - d. Fine Lines & Wrinkles Improvement ☐ Yes ☐ No
   - e. Skin Pore Size Improvement ☐ Yes ☐ No
   - f. Skin Feels Better ☐ Yes ☐ No
   - g. Skin Feels Tighter ☐ Yes ☐ No
   - h. Skin Appears Tighter ☐ Yes ☐ No
   - i. Skin Looks more Radiant ☐ Yes ☐ No
   - j. Skin Appears Brighter ☐ Yes ☐ No
   - k. Skin Seems more Youthful ☐ Yes ☐ No
   - l. Other: _____

   ☐ **NONE**

5. Since your treatment have you felt…

a. An improvement in your social life?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

b. An improvement in your family life?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

c. More secure?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

d. An improvement in how people respect you?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

e. An improvement in your mood?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

f. An improvement in your daily quality of life?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

g. An improvement in your self-esteem?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

h. More confident?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

i. An improvement in your sensuality?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

j. An improvement in your vitality?
☐ **Completely** ☐ **A great deal** ☐ **Somewhat** ☐ **Not much** ☐ **Not at all** ☐ **It is worse**

6 ADVERSE EVENTS ASSESSMENT REPORTING

The definitions of adverse events and the subtypes are provided in the CIP. Adverse events will be classified by the investigator as to:

- Anticipated vs unanticipated;
- Serious vs not serious;
- Severity: mild, moderate, severe;
- Device causality: not related, related, undetermined;
- Procedure causality: not related, related, undetermined.

7 SAMPLE SIZE ESTIMATION

7.1 Determination of Sample Size

The primary efficacy objective of this study is to demonstrate response to treatment with the Renuvion Dermal System. The IPR will assess each subject image and assess using the FWS. The proportion of subjects achieving at least a one-point improvement in 2 out of 3 IPR assessors will be calculated; this is the proportion of treatment successes.

The sample sizes were estimated using PASS 2019[1] with the following inputs:

- A performance goal (PG) of 50% success;
- A one-sample, one-sided t-test against the PG;
- $\alpha = 0.05$;
- Power = 90%
- Renuvion success proportions (P) of 70%
- Test is Fisher's Exact Test.

Table 5: Numeric Results for Testing One Proportion using the Exact Test

Table 5: Numeric Results for Testing One Proportion using the Exact Test						
Alternative Hypothesis: One-Sided (H0: P ≤ P0 vs. H1: P > P0)						
Power	N	Performance Goal (PG)	Renuvion Proportion	Δ = Renuvion - PG	Alpha	Reject Ho if Successes ≥
0.90	50	0.5	0.70	0.20	0.05	32

The 50 subjects will be augmented by 10% to 55 subjects to accommodate dropouts and losses to follow-up.

[1] PASS 2019 Power Analysis and Sample Size Software (2019). NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/pass.

7.2 Performance Goal Rationale

A performance goal of 50% will be used for testing. A performance goal of 50% is clinically relevant to ensure that there are significant benefits of the procedure to outweigh its risk. The lower bound of the confidence interval of the proportion of subject achieving treatment successes will be compared against this Performance Goal. This primary effectiveness endpoint definition and performance goal have been used in prior clinical studies to support their 510(k) clearances (i.e., eTwo Skin Treatment System, cleared under K141507 on December 8, 2014[2]). Moreover, the expected Renuvion proportion used to determine the study sample size is within the observed results range of the clinical studies conducted to support this indication for other FDA cleared devices (e.g., Picosure Workstation, cleared under K140719 on May 23, 2016[3] and the eTwo Skin Treatment System mentioned above).

8 ANALYSIS POPULATIONS AND HANDLING MISSING DATA

8.1 Analysis Sets

The full analysis dataset will be used in our main analyses.

Full Analysis Dataset (Intent-to-Treat Sample, ITT)
Participants enrolled in the study who had baseline photographs taken will be included. Those with missing data items at 90-day visit will be imputed using multiple imputation, *unless the subject leaves the study due to lack of effectiveness*. In the case of lack of effect, the imputed data will be the same as the baseline data. The primary effectiveness endpoints will use this data set.

The primary safety endpoint will also use this data set, but without multiple imputation. The reason is that the analysis method takes into account missing data (see the Analysis section below) and can handle subjects who exit the study due to an adverse event without biasing the results.

Modified Analysis Datasets (modified Intent-to-Treat Samples, mITT)
Participants enrolled in the study who had baseline data and endpoint data will be included. No imputation will be done for missing endpoints unless there is evidence that the endpoint is missing for cause. The Secondary and Additional endpoints will use these datasets. The reason that there are more than one mITT data set is that the data set is dependent upon the endpoint in questions; only those subjects with baseline and endpoint data *for that endpoint* will be included in the analysis.

Per Protocol Dataset (Per-Protocol Sample, PP)
This will be a subset of the full analysis dataset comprising participants without major protocol deviations. Participants with major protocol deviations will be identified at the data review meeting that takes place before database lock. The effectiveness and safety analysis will also be performed in this per protocol population.

[2] https://www.accessdata.fda.gov/cdrh_docs/pdf14/K141507.pdf
[3] https://www.accessdata.fda.gov/cdrh_docs/pdf14/K140719.pdf

Subjects not Included in the mITT and PP Samples
Participants enrolled in the study who were excluded from the mITT and PP samples will have their reasons for missing data or protocol violations examined and reported.

8.2 Assessment of Poolability Data

The primary effectiveness endpoint will be summarized by randomized group and investigational site. Investigational sites with fewer than 10 subjects will be combined for this purpose. A 2-way contingency table of the primary endpoints of patients by center will be analyzed by the Fisher-Freeman-Halton Exact Test to see if the proportion of correctly identified photographs differs by center. If the test is not significant at α=0.05, the centers will be combined in a single analysis of the primary effectiveness objective. See the section on the analysis of the primary endpoint for a discussion of the analysis that will be used if the test is significant.

9 ANALYSIS OF STUDY ENDPOINTS AND OTHER DATA

9.1 Statistical Analysis of Demographic and Vital Signs Data

Demographic and vital signs data will be analyzed using descriptive statistics and reported.

- Continuous variables (e.g. age) will be reported as the mean, number of observations, minimum, maximum and 95% confidence interval (CI) of the mean.
- Categorical variables (e.g. sex) will be reported as the percentage and number of observations and the 95% CI of the percentage.

9.2 Statistical Analysis of the Effectiveness Endpoint: Success Rate

For the primary effectiveness endpoint, the proportion of successful patients (P) will be tested against the PG. If the lower bound of the 95% confidence interval of the proportion of subject achieving treatment successes is greater than the Performance Goal, the effectiveness endpoint would be met. The statistical test will be a one-sided Fisher's Exact Test at $\alpha = 0.05$. For the effectiveness endpoint to be met, Ho must be rejected.

$$\text{Ho: } P \leq 50\% \quad \text{vs.} \quad \text{Ha: } P > 50\%$$

If the test of poolability of study results is significant, the analysis will be conducted as a meta-analysis in which the centers serve as the individual studies in a random effects model. The global p-value will be reported.

9.3 Statistical Analysis of Safety Endpoints

9.3.1 Primary Safety Endpoint: AE rate

All of the various AE rates will be estimated using Kaplan-Meier time-to-event analyses. The monthly event rates and their 95% confidence intervals (CIs) will be reported by for each category of AE.

9.3.2 Secondary Safety Endpoint: Pain

The mean VAS score will be reported at each measurement along with it 95% CI, the number of observations, the minimum and maximum, and the 95% CI. The change from baseline will be similarly reported.

9.4 Statistical Analysis of Additional Endpoints

These analyses will be performed as follows. All analyses will be descriptive.

1. "After" image correctly identified. The analysis will be reported as the percentage and number of observations and the 95% CI of the percentage.
2. Change in FWS from baseline to 90-day follow-up visit will be reported as the mean, number of observations, minimum, maximum and 95% confidence interval (CI) of the mean.
3. Subject modified GAIS at 90-days will be reported as the mean, number of observations, minimum, maximum and 95% confidence interval (CI) of the mean.
4. Investigator modified GAIS at 90-days will be reported as the mean, number of observations, minimum, maximum and 95% confidence interval (CI) of the mean.
5. Subject satisfaction at the 90-days will be reported as the mean, number of observations, minimum, maximum and 95% confidence interval (CI) of the mean.
6. Achievement of re-epithelialization by facial zone and across all facial zones at the 10 day, 30 day and 90 day follow-up visits will be reported as the percentage and number of observations and the 95% CI of the percentage.
7. Mean duration until study subject feels comfortable going in public will be reported as the mean, number of observations, minimum, maximum and 95% confidence interval (CI) of the mean.
8. Daily 11-point Visual Analog Scale (VAS) pain assessment following treatment through 10 days will be reported as the mean, number of observations, minimum, maximum and 95% confidence interval (CI) of the mean.

9.5 Subgroup Analysis

Subgroup analyses will be performed to understand the effect of demographic and other patient-level variables on the magnitude of the improvement at 90-days. A multiple linear regression will be used to model the effect of on age, gender, race/ethnicity, baseline FWS and Fitzpatrick Skin Scale (FSS) on the change in FWS. (Age *et al* are called "independent variables" and the change in FWS is the dependent or response variable.)

If any of the independent variables are significant, the effect size and better responders will be reported. Also, the possibility of clinically relevant interaction terms will be tested.