STATISTICAL ANALYSIS PLAN

Study Title: Effectiveness of Intense Pulsed Light (IPL) for improving signs and symptoms in Dry Eye Disease (DED) due to Meibomian Gland Dysfunction (MGD)

Study Number: LUM-VBU-M22-IPL-17-01

Author:	Gerry Gray Principal Data-Fi, LLC
Date:	3/26/2020
Version:	1.0

Study Synopsis

Study Device	M22 intense pulsed light (IPL) module is a class II device with an FDA clearance (K142860) for benign cutaneous vascular lesions, including erythema of rosacea and facial telangiectasia.
Protocol number	LUM-VBU-M22-IPL-17-01
Study Title	Effectiveness of Intense Pulsed Light (IPL) for improving signs and symptoms in Dry Eye Disease (DED) due to Meibomian Gland Dysfunction (MGD)
Study design	Multi-center, prospective, randomized, sham-controlled, superiority, non-significant risk
Projected Study period	Initiation Date: January 15, 2018 — Completion Date: July 1, 2018
Study population	Up to 83 male or female subjects, aged 22-85 with signs and symptoms of DED caused by MGD Enrollment will continue until at least 10 subjects without skin rosacea are enrolled.
Investigational device	Lumenis® M22 system with the IPL handpiece
Study Duration	From screening to termination, each subject will participate in the study for 10-11 weeks (depending on whether screening & enrollment occurred on the same day or up to 1 week later) The total study duration from the screening of the first subject to termination of the last subject is estimated to be 75 weeks.
Main Inclusion Criteria	• Tear break-up time (TBUT) ≤ 7 seconds • MGS ≤ 12 for 15 glands in the lower eyelid • At least 5 non-atrophied meibomian glands in the lower eyelid • Symptoms self-assessed using the OSDI questionnaire ≥ 23
Detailed Inclusion Criteria:	• Subject is able to read, understand and sign an IC form • 22-85 years of age • Fitzpatrick skin type I-IV • Subject is able and willing to comply with the treatment/FU schedule and requirements • In the study eye, TBUT ≤ 7 seconds (PMID: 21450918) • In the study eye, MGS ≤ 12 (PMID: 2222996) • In the study eye, at least 5 non-atrophied meibomian glands in the lower eyelid • Symptoms self-assessed using the OSDI questionnaire ≥ 23 (PMID: 20065224)
Main Exclusion Criteria:	• Fitzpatrick skin type V or VI • Contact lens wear within the month prior to screening • Unwilling to discontinue use of contact lenses for the duration of the study • Ocular surgery or eyelid surgery, within 6 months prior to screening

- Neuro-paralysis in the planned treatment area, within 6 months prior to screening
- Other uncontrolled eye disorders affecting the ocular surface, for example active allergies
- Current use of punctal plugs
- Pre-cancerous lesions, skin cancer or pigmented lesions in the planned treatment area
- Uncontrolled infections or uncontrolled immunosuppressive diseases
- Subjects with ocular infections, within 6 months prior to screening
- Prior history of cold sores or rashes in the perioral area or in the planned treatment area that could be stimulated by light at a wavelength of 560 nm to 1200 nm, including: Herpes simplex 1 & 2, Systemic Lupus erythematosus, and porphyria
- Within 3 months prior to screening, use of photosensitive medication and/or herbs that may cause sensitivity to 560-1200 nm light exposure, including: Isotretinoin, Tetracycline, Doxycycline, and St. John's Wort
- Over exposure to sun, within 4 weeks prior to screening
- Use of prescription eye drops for dry eye, within 7 days prior to screening, excluding artificial tears and glaucoma drops
- Radiation therapy to the head or neck, within 12 months prior to screening
- Planned radiation therapy, within 8 weeks after the last treatment session
- Treatment with chemotherapeutic agent, within 8 weeks prior to screening
- Planned chemotherapy, within 8 weeks after the last treatment session
- New topical treatments within the area to be treated, or oral therapies, within 3 months prior to screening- except over-the-counter acetaminophen-based analgesics for pain management, new oral omega 3 fatty acid supplements and topical artificial tears
- Change in dosage of any systemic medication, within 3 months prior to screening
- Anticipated relocation or extensive travel outside of the local study area preventing compliance with follow-up over the study period
- Legally blind in either eye
- History of migraines, seizures or epilepsy
- Facial IPL treatment, within 12 months prior to screening
- Any thermal treatment of the eyelids, including Lipiflow, within 6 months prior to screening
- Expression of the meibomian glands, within 6 months prior to screening
- In either eye, moderate to severe (Grade 3-4 on the EFRON scale) inflammation of the conjunctiva, including: allergic, vernal or giant papillary conjunctivitis
- In either eye, severe (Grade 4 on the EFRON scale) inflammation of the eyelid, including: blepharochalasis, staphylococcal blepharitis or seborrheic blepharitis
- Ocular surface abnormality that may compromise corneal integrity in either eye (e.g., prior chemical burn, recurrent corneal erosion, corneal epithelial defect, Grade 3 corneal fluorescein staining, or map dot fingerprint dystrophy)
- Eyelid abnormalities that affect lid function in either eye, including: entropion, ectropion, tumor, edema, blepharospasm, lagophthalmos, severe trichiasis, and severe ptosis
- Any systemic condition that may cause dry eye disease, including: Stevens-Johnson syndrome, vitamin A deficiency, rheumatoid arthritis, Wegener's granulomatosis, sarcoidosis, leukemia, Riley-Day syndrome, systemic lupus erythematosus, and Sjögren's syndrome

	Unwilling or unable to abstain from the use of medications known to cause dryness (e.g., isotretinoin, antihistamines) throughout the study duration. Subjects must discontinue these medications for at least 1 month prior to the baseline visit. Any condition revealed whereby the investigator deems the subject inappropriate for this study
Objectives	**Primary objectives** TBUT estimated at the single FU (Between 4 weeks -3 days and 4 weeks + 7 days after the final treatment) Measurement of TBUT will be implemented using FUL-GLO® fluorescein ophthalmic strips. Three successive readings will be taken and averaged to a single value. **Secondary objectives** Self-evaluation of OSDI at the single FU, using the OSDI questionnaire Self-evaluation of Eye Dryness Score (EDS) at the single FU, using a VAS
Criteria for Evaluation	**Primary endpoint:** The difference in the change of TBUT from BL to FU, between eyes in the study arm and eyes in the control arm The improvement of TBUT in eyes of the study arm is larger than the improvement of TBUT in eyes of the control arm, where improvement is defined as a positive change of TBUT from BL to FU **Secondary endpoints** The difference in the change of OSDI from BL to FU, between subjects in the study arm and subjects in the control arm The difference in the change of EDS from BL to FU, between subjects in the study arm and subjects in the control arm The improvement of OSDI in the study arm is larger than the improvement of OSDI in the control arm, where improvement is defined as a negative change of OSDI from BL to FU The improvement of EDS in the study arm is larger than the improvement of EDS in the control arm , where improvement is defined as a negative change of EDS from BL to the FU
Exploratory effectiveness endpoints	- The difference in the proportion of eyes with normal TBUT (TBUT >10 sec) at FU, between eyes in the study arm and eyes in the control arm - The difference in the proportion of subjects with normal OSDI (OSDI < 23) at FU, between subjects in the study arm and subjects in the control arm - The difference in eyelids appearance, as qualitatively evaluated by the study investigator, between eyes in the study arm and eyes in the control arm - The difference in the percentage of area loss of meibomian glands, as evaluated using meibography, between eyes in the study arm and eyes in the control arm

	• The difference in the change of MGS from BL to FU, between eyes in the study arm and eyes in the control arm
Safety Endpoints	Throughout the study: • the difference in the incidence of ocular adverse events, between subjects in the study arm and subjects in the control arm • the difference in the incidence of non-ocular adverse events, between subjects in the study arm and subjects in the control arm • the difference in the incidence of unanticipated serious adverse events, between subjects in the study arm and subjects in the control arm At the treatment sessions- • the difference in the change of bio-microscopy examinations pre- and post- treatment, between subjects in the study arm and subjects in the control arm • the difference in the self-assessment of pain/discomfort during IPL administration, between subjects in the study arm and subjects in the control arm • the difference in the self-assessment of pain/discomfort during MGX, between subjects in the study arm and subjects in the control arm
Screening and baseline (BL) visit	Subjects will sign the IC form within one week of passing an informed consent process and having been provided the informed consent form (ICF). After signing the ICF, subjects will undergo the following baseline (BL) assessments, in the specified order: 1. Three consecutive measurements of tear break-up time (TBUT) in both eyes 2. Routine ophthalmology tests in both eyes, including: biomicroscopy, best-corrected visual acuity (BCVA) and intra-ocular pressure (IOP) 3. Ocular Surface Disease Index (OSDI), self-assessed using an OSDI questionnaire 4. Eye dryness score (EDS), self-assessed using a visual analog scale (VAS) 5. High resolution close-up photos of both eyes, including: the iris, lid margins and eyelashes 6. Meibography of the upper and lower eyelids in both eyes 7. Meibomian gland secretion (MGS) score in 15 glands of the lower eyelid in both eyes
Randomization	• Eligible subjects will be randomized 1:1 to a study arm or to a control arm • The randomization process will adopt a blocked randomization strategy, using random block sizes of 2 and 4
Treatment schedule	Four (4) treatment sessions, 2 weeks (-3 days, + 7 days) apart. Each treatment session will consist of the following procedures, in the specified order:

	1. Subject's report of the daily usage (frequency and dose) of eye drops, warm compresses and lid hygiene since the previous visit 2. Pre-treatment biomicroscopy with the slit lamp (observation of lid margins, eyelashes, conjunctiva) 3. **<u>In the study arm</u>**: Active IPL administered on the malar region (from tragus to tragus, below the lower eyelids, including the nose) **<u>In the control arm</u>**: Sham IPL administered on the same facial areas 4. Meibomian gland expression (MGX) of the upper and lower eyelids in both eyes 5. Post-treatment biomicroscopy with the slit lamp 6. Self-assessment of pain/discomfort during IPL administration, using a VAS 7. Self-assessment of pain/discomfort during MGX, using a VAS
Follow-up (FU) visit	There will be a single follow-up (FU) visit, which will occur 4 weeks (-3 days, + 7 days) after the final treatment session. Subjects will undergo the following assessments, in the specified order: 1. Subject's report of the daily usage (frequency and dose) of eye drop, warm compresses and lid hygiene since the previous visit 2. Three consecutive measurements of TBUT in both eyes 3. Routine ophthalmology tests in both eyes, including: biomicroscopy, BCVA and IOP 4. OSDI, self-assessed using an OSDI questionnaire 5. EDS, self-assessed using a VAS 6. High resolution close-up photos of both eyes, including: iris, lid margins and eyelashes 7. Meibography of the upper and lower eyelids in both eyes 8. Meibomian gland secretion (MGS) score in 15 glands of the lower eyelid in both eyes
Safety outcome measures	• incidence of non-ocular adverse events throughout the study • incidence of ocular adverse events throughout the study • incidence of unanticipated serious adverse events throughout the study • self-assessment of pain/discomfort of IPL in treatment sessions (using a VAS) • self-assessment of pain/discomfort of MGX in treatment sessions (using a VAS)
Covariates	• Baseline TBUT (continuous) • Baseline skin rosacea (4-point scale)
Sample Size	The following assumptions were used to calculate the original sample size: (1) In the study arm, the change of TBUT from BL to FU $\approx 5 \pm 5$ sec (2) In the control arm, the change of TBUT from BL to FU $\approx 1 \pm 5$ sec

	(3) A type I error of 0.05 (two-tailed test) (4) A type II error of 0.2 (power = 0.8) (5) A 1:1 ratio of Treatment to Control With these assumptions, the minimal sample size was calculated to be 50 evaluable subjects (100 evaluable eyes): 25 subjects (50 eyes) in the control arm and 25 subjects (50 eyes) in the study arm. Assuming a dropout rate of 15%, the anticipated number of enrolled subjects was 59. Assuming a screening failure rate of 15%, the anticipated total number of screened subjects was 69. After study initiation (but before any data were unmasked) it was found that 8 IPL patients at one site were not treated in accordance with the protocol. Accordingly, patients were added to the study to restore the original power and to allow comparison before/after corrective actions were taken. The final sample size is 83 evaluable subjects (166 evaluable eyes). Based on an updated dropout rate of 7%, the anticipated number of enrolled subjects is 89. Based on an updated screening rate of 23%, the anticipated number of screened subjects is 116.
Number of sites	Three (3)-Four (4)

1. Study Objectives

1.1. Primary objective

The primary objective of the study is to determine the efficiency of IPL therapy, in improving TBUT in eyes with moderate to severe DED due to MGD. This objective was chosen because a reduced TBUT is one of the most common signs of DED due to MGD. TBUT is considered abnormal when shorter than 10 seconds. To include subjects with moderate to severe DED, in this study the relevant inclusion criterion is TBUT ≤ 7 seconds.

1.2. Secondary Objectives

- To determine the efficiency of IPL in improving *symptoms* of DED, in subjects with DED due to MGD

- To qualitatively assess the effect of IPL on the appearance of the eyelids, , in subjects with DED due to MGD

- To determine the safety of IPL therapy

2. Study Endpoint

2.1. Primary efficiency endpoint

The difference in the *change* of TBUT, between eyes in the study arm and eyes in the control arm, where *change* is defined as TBUT at the follow-up minus TBUT at the baseline.

2.2. Success criterion

The study will be determined as successful if it satisfies two conditions:

1. The difference in the change of TBUT, between eyes in the study arm and eyes in the control arm, is statistically significant at the $\alpha = 0.05$ level.

2. The *change* of TBUT is *higher* in eyes of the study arm, compared to eyes of the control arm

where the *change* of TBUT is defined as TBUT at the follow-up minus TBUT at the baseline.

2.3. Secondary efficiency endpoints

- The difference in the *change* of OSDI, between subjects in the study arm and subjects in the control arm

- The difference in the *change* of EDS, between subjects in the study arm and subjects in the control arm

where the *change* of the variable (OSDI or EDS) is defined as the value of the variable at the follow-up minus the value of the variable at the baseline.

2.4. Exploratory efficiency endpoints

- The difference in the proportions of eyes with TBUT > 10 sec (i.e., not consistent with DED) at the follow-up, between eyes in the study arm and eyes in the control arm

- The difference in the proportions of subjects with OSDI < 23 (i.e., not consistent with DED) at the follow-up, between subjects in the study arm and subjects in the control arm

- The difference in eyelids appearance, as qualitatively evaluated by the study investigator, between eyes in the study arm and eyes in the control arm

- The difference in the *change* of MGS, between eyes in the study arm and eyes in the control arm

- The difference in the *change* of Meiboscore (Meibography), between eyes in the study arm and eyes in the control arm

2.5. Safety endpoints

- The difference in the incidence of non-ocular adverse events, between subjects in the study arm and subjects in the control arm

- The difference in the incidence of ocular adverse events, between subjects in the study arm and subjects in the control arm

- The difference in the incidence of unanticipated serious adverse events, between subjects in the study arm and subjects in the control arm

- The difference in the level of pain/discomfort during IPL treatment, between subjects in the study arm and subjects in the control arm

- The difference in the level of pain/discomfort during MGX, between subjects in the study arm and subjects in the control arm

- The difference in the change of biomicroscopy immediately before and immediately after IPL treatment, between subjects in the study arm and subjects in the control arm

2.6. Study duration

- For an individual participant, from the screening visit to the follow-up visit, the study duration is expected to be between 10-11 weeks (depending on whether screening and randomization/Tx1 were performed on the same day, or up to 1 week later).

- For the entire sample, from the screening of the first subject to the follow-up of the last subject, the study duration is expected to be 75 weeks. This is based on the following assumptions: a final sample size of 83 subjects (166 evaluable eyes); 3-4 sites; a screening rate of 1.5 subjects per week per site; a 23% rate of screening failures; and a lost to follow-up rate of 7% (89 enrolled subjects and 116 screened subjects)

- Enrollment will continue until at least 10 enrolled subjects have no signs of skin rosacea, even if 89 subjects were already enrolled to the study. Hence, in principle study duration can extend beyond 75 weeks.

2.7. Concurrent Control

Subjects will be randomized 1:1 to a study arm and to a control arm. Subjects in the control arm will undergo exactly the same procedures as subjects in the study arm, with the exception that the IPL pulses will not be actually delivered. This applies to both test spots and the full treatment.

Due to the nature of the IPL treatment, it is not possible to ensure masking of the subjects, as subjects are generally aware that IPL treatment should be felt and should normally leave a temporary redness of the skin. This said, the following steps will be taken:

In the informed consent process, subjects will be told that in some cases the IPL treatment might be felt as a mild pain/discomfort on the skin, and that the treatment can cause redness of the skin that should disappear within a few minutes. During IPL administration, all subjects will wear protective goggles that will completely occlude their vision.

In the control arm, subjects will undergo a sham treatment that will mimic the IPL (IPL) therapy. The sham treatment will be implemented using a filter that blocks all light emitted by the M22-IPL (the ∞ filter described in the protocol). The tip of the lightguide will still be cooled and the device will still produce clicking sounds whenever the trigger button on the IPL handpiece is triggered, but no light will be actually transmitted to the skin. Following the sham IPL procedure, subjects will undergo meibomian gland expression.

2.8. Blinding

Every effort will be made to ensure examiners measuring primary outcomes are masked to the treatment allocation. However, due to the nature of the IPL treatment, it is not possible to completely ensure masking of the examiner. The examiner that will assess the outcome measures at the baseline visit and at the follow-up visit will not be the study investigator that will administer the IPL pulses. This will increase the probability that the examiner is masked to the allocation.

2.9. Interim analysis

There will be no interim analysis in this study.

3. Study Population and Subject Selection

3.1. Source and Sample Size

The aim of this study is to analyze up to 83 subjects (166 eyes) who completed the full schedule, randomized 1:1 to a study arm and a control arm. It is estimated that 23% of the subjects will be screening failures, and that 7%% of the randomized subjects will drop out or will be lost to follow-up. Therefore, to reach up to 83 subjects (166 eyes) who will complete the study,116 subjects are expected to be screened and 89 subjects are expected to be enrolled.

3.2. Eligibility

Subject eligibility is based on diagnosis of symptoms and signs compatible with DED due to MGD. See the protocol for a complete listing of inclusion/exclusion criteria.

3.3. Analysis Populations

The per-protocol analysis population includes all subjects who have at least one follow-up visit and have no major protocol deviations. The primary and secondary endpoint analysis will be performed based on the per-protocol population.

The intent-to-treat (ITT) analysis population includes all subjects who are randomized to either arm, regardless of treatment received.

4. Study Evaluations

4.1. Effectiveness

4.1.1. Primary effectiveness endpoint

The primary efficiency endpoint will be estimated as the difference in the *change* of TBUT from baseline to follow-up (10 weeks) between the study arm and the control arm, where the change of TBUT is defined as TBUT at the follow-up minus TBUT at the baseline.

A linear mixed-effect (LME) model will be used to test the null hypothesis that the changes of TBUT are not different between the two arms. The success criterion will include the following two conditions:

1. Two-sided p-value < 0.05
2. The change of TBUT in eyes of the study arm is *larger* than the change of TBUT in eyes of the control arm

The linear mixed-effect model will include a random effect for subjects to account for correlation between eyes within subjects.

4.1.2. Secondary effectiveness endpoints

The following secondary efficiency endpoints will be estimated:

- The difference in the *change* of EDS, between subjects in the study arm and subjects in the control arm

- The difference in the *change* of OSDI, between subjects in the study arm and subjects in the control arm

where *change* is defined as the value of the variable (OSDI or EDS) at the follow-up minus the value of the variable at the baseline.

4.1.3. Exploratory effectiveness endpoints

The following exploratory endpoints will be evaluated:

- The difference in the proportions of subjects with TBUT > 10 sec at the follow-up, between eyes in the study arm and eyes in the control arm

- The difference in the proportions of subjects with OSDI < 23 at the follow-up, between subjects in the study arm and subjects in the control arm

- The difference in eyelids appearance at the follow-up, as qualitatively evaluated by the study investigator, between eyes in the study arm and eyes in the control arm

- The difference in the percentage of area loss of meibomian glands, as evaluated using meibography, between eyes in the study arm and eyes in the control arm

- The difference in the change of MGS from BL to FU, between eyes in the study arm and eyes in the control arm

where the *change* of a variable is defined as the value of this variable at the follow-up minus its value at the baseline.

4.2. Safety

Safety will be assessed by reporting:

- The difference in the incidence of related adverse events, between subjects in the treatment arm and subjects in the control arm

- The difference in the incidence of unrelated adverse events, between subjects in the treatment arm and subjects in the control arm

- The difference in the level of pain/discomfort during IPL treatment, between subjects in the treatment arm and subjects in the control arm

- The difference in the level of pain/discomfort during MGX, between subjects in the treatment arm and subjects in the control arm

5. Study Analysis Plan

5.1. Primary Study Hypothesis

The *change* of TBUT (ΔTBUT) in eyes of the study arm is *more positive* than ΔTBUT in eyes of the control arm, where ΔTBUT is defined as TBUT at the follow-up minus TBUT at the baseline, and subscripts S, C indicate the study (S) or control (C) arms:

H0: $\Delta TBUT_S \leq \Delta TBUT_C$

H1: $\Delta TBUT_S > \Delta TBUT_C$

5.2. *Secondary Study Hypotheses*

1. The change of OSDI (ΔOSDI) in subjects of the study arm is *more <u>negative</u>* than ΔOSDI in subjects of the control arm, where ΔOSDI is defined as OSDI at the follow-up minus OSDI at the baseline:

   > H0: ΔOSDI$_S$ $\geq$ ΔOSDI$_C$
   >
   > H1: ΔOSDI$_S$ < ΔOSDI$_C$

2. The change of EDS (ΔEDS) in subjects of the study arm is *more <u>negative</u>* than ΔEDS in subjects of the control arm, where ΔEDS is defined as EDS at the follow-up minus EDS at the baseline:

   > H0: ΔEDS$_S$ $\geq$ ΔEDS$_C$
   >
   > H1: ΔEDS$_S$ < ΔEDS$_C$

5.3. *Sample Size*

Approximately 4 months after initiation of the study, a round of site monitoring visits revealed that one of the three sites in the study was not following the instructions above with respect to where the M22 IPL lightguide should be applied; for IPL and sham subjects the lightguide was being applied too far from the eyelid margin in contrast to what is described in Figure 5 above. As a result, it is possible that the IPL treatments among the initial cohort of subjects at that site was will prove to be ineffective, or of reduced effectiveness compared to the sham control. The lightguide was applied correctly at the other two sites.

Although no data were unmasked, soon after this issue was recognized by study monitors, Lumenis initiated a corrective action (retraining and greater emphasis in the protocol), to better assure the protocol instructions would be followed.

To restore the original statistical power, the sample size was increased. Below are the original and adjusted calculations of the sample size:

5.3.1. *Original calculation of the sample size*

Assuming a 1:1 randomization ratio, a mean of 5$\pm$5 sec for ΔTBUT of the study arm, a mean of 1$\pm$5 sec for ΔTBUT of the control arm, an alpha level of 0.05 and a power of 80%, the total sample size is 50 completed study eyes (25 for the study arm and 25 for the control arm). Assuming a dropout rate of 15%, the anticipated number of enrolled subjects is 58. Assuming a screening failure rate of 15%, the anticipated total number of screened subjects is 66.

In addition, MGD is less common without concomitant skin rosacea. Therefore, enrollment will continue until at least 10 subjects without skin rosacea are enrolled, even if 58 subjects were already enrolled.

5.3.2. Adjusted calculation of the sample size.

After the corrective action to remedy the treatment errors, the sample size was recalculated using a simulation where we made the following conservative assumptions:

- The initial 8 IPL subjects at the affected site had response equal to the control arm (1±5 seconds)
- Subsequent to the corrective action, IPL subjects at the affected site had a response equal to the IPL arm (5±5 seconds)
- The final analysis was ITT, which included the inappropriately treated patients in the IPL arm.
- The correlation between eyes within a subject is 0.5.

Under these assumptions the overall treatment effect is "diluted" by the initial set of patients who were treated inappropriately. Under these conservative assumptions, the sample size needed to restore the original power was calculated to be 136 evaluable eyes (68 subjects, i.e. 18 subjects more than the original 50 subjects).

In addition, in order to allow for full comparison of the effect of the corrective action, 15 patients were added to the enrollment at the affected site so there would be sufficient before/after patients to allow for clinical evaluation of the difference.

Adding these 33 (18+15) subjects to the original 50 completed subjects, the number of completed subjects is up to 83. Therefore, with a loss to follow-up rate of 7% the expected number of enrolled subjects is 89, and with a screening failure rate of 23% the expected number of screened subjects is 116.

All sample size re-calculations were done prior to any unmasking of any results from the study.

5.4. Subject Accountability

A complete accounting of subjects by Treatment Arm and visit will be provided, including reasons for dropout, if known.

5.5. Demographics and Baseline Characteristics

Demographic variables gender, race, ethnicity, and age will be summarized for all enrolled subjects, along with medical history, by Treatment Arm and combined. Descriptive statistical summaries of pre-treatment parameters (min, max, median, mean, standard deviation) will also be provided for each treatment arm. This includes baseline measurements for each of the study endpoints: TBUT, Total Meibomian Gland Secretion Score, OSDI Score, and Eye Dryness VAS Score, as well as Fitzpatrick skin type and skin rosacea.

All baseline measurements will be evaluated for adequate balance between the two treatment arms.

5.6. Safety Endpoints and Analysis

5.6.1. Descriptive Analyses

All adverse events, measures of pain and discomfort, changes in ocular surface staining, changes in IOP, and chances in BCVA (EDTRS) will be tabulated by visit and treatment group.

5.6.2. Safety Endpoints

- The difference in the incidence of non-ocular adverse events, between subjects in the study arm and subjects in the control arm

- The difference in the incidence of ocular adverse events, between subjects in the study arm and subjects in the control arm

- The difference in the incidence of unanticipated serious adverse events, between subjects in the study arm and subjects in the control arm

- The difference in the level of pain/discomfort during IPL treatment, between subjects in the study arm and subjects in the control arm

- The difference in the level of pain/discomfort during MGX, between subjects in the study arm and subjects in the control arm

- The difference in the change of biomicroscopy immediately before and immediately after IPL treatment, between subjects in the study arm and subjects in the control arm

5.6.3. Primary Safety Analysis

The safety endpoints will be summarized for each treatment group.

We will compute the incidence of ocular adverse events as a simple proportion, counting the number of subjects with any event (that is, counting the first event per person).

5.7. Effectiveness analyses

The primary efficacy outcome measure (TBUT) consist of a pre-treatment measurement (the baseline), a treatment intervention, and a post-treatment measurement (the follow-up). Both eyes will be included in the analysis. Since the two eyes of a subject are not independent, to estimate whether TBUT has improved from the baseline to the follow-up, a linear mixed effects model with random intercept, using the subject identity as the random effect, is the appropriate test. Since there are two eyes per person then there is only one possible covariance structure (i.e. symmetric, with a single correlation between eyes).

The secondary efficacy outcome measures are OSDI and EDS which will be collected at the baseline, and at a follow-up visit. In the case of OSDI, the value is estimated per subject. Hence, a single value will be collected per subject at each of these two visits. In this case a paired t-test is the appropriate statistical test to estimate whether dry eye symptoms have improved, using the change of OSDI from baseline to the follow-up. In the case of EDS, the value is estimated per eye. Hence, two values will be collected per subject at each of these two visits. Since the two eyes of a subject are not independent, to estimate whether EDS has improved from the baseline to the follow-up, a linear mixed effects model with random intercept, using the subject identity as the random effect, is the appropriate test. The level of

statistical significance is α = 0.05. Non-parametric tests will be performed if the underlying distributions are not normal.

For all other eye-level analyses, where appropriate, sample means of continuous variables will be analyzed with linear mixed effects model with random intercept (whenever both eyes of a subject are included in the analysis), or two-sample t-tests. Proportions will be analyzed with $\chi 2$ tests.

All statistical tests will be two-sided because potentially the treatment procedure may worsen the outcome. The level of statistical significance (type I error) is α = 0.05. Non-parametric tests will be performed if the underlying distribution is not normal.

For descriptive statistics, the N, Mean, Median, Standard deviation, Standard error of the mean, 95% confidence interval, Minimum, and Maximum will be reported for continuous variables; the frequency and proportion for each category will be reported for nominal and categorical variables.

A previous study demonstrated that in subjects with a lower TBUT at the baseline, the change in TBUT at the follow-up was larger [21]. Although not shown in this study, the severity level of skin rosacea at the baseline may also be a confounder. Hence, in addition to the primary analysis, we will carry out an analysis of covariance (ANCOVA) to adjust for these covariates.

The following baseline covariates will be included in the ANCOVA:

> (1) TBUT at baseline. This covariate will be treated as continuous.
> (2) The 4-point severity level of skin rosacea at baseline (0= none; 1 = mild; 2 = moderate; 3 = severe). This covariate will be treated as either a numeric score (continuous) or using orthogonal polynomial contrasts.

5.7.1.   Stratified Analysis

To account for the potential loss of IPL treatment effect at one site (hereafter called "A") a stratified analysis will be performed that includes a separate stratum comprising subjects at site A who were treated prior to the corrective action. The strata will be labeled as A1, A2, B, and C for sites A, B, and C, with A1 and A2 the two separate groups of patients at site A before and after the corrective action. Since the treatment is applied over a sequence of four visits, we will conservatively assign to stratum A2 only those patients who received all four of their treatments subsequent to the corrective action. Any patient who received any one of their four scheduled treatments prior to the corrective action will be assigned to stratum A1.

Thus, the following four strata are obtained:

Stratum	Corrective Action	Description
A1	No	Patients from site A with one or more treatment visits prior to the corrective action

A2	Yes	Patients from site A with all 4 treatment visits subsequent to the corrective action
B	No	Patients from site B
C	No	Patients from site C

An ANOVA analysis will be performed for the change in TBUT primary endpoint, with fixed effects for treatment and stratum and a random subject effect.

To examine the effect of the corrective action, we will group the strata into two sets: {A1}, for the pre-corrective-action subjects at site A, and {A2+B+C} for all post-corrective-action subjects at site A, combined with all subjects at B or C. We will test for significance of the corrective action by creating a 1 degree of freedom contrast for "treatment by corrective action", using treatment (1 df) and the two corrective action sets (1 df), and test whether the contrast is significantly different from 0, using a two-sided alpha of 0.15.

- If the contrast is not significant, then the results will be analyzed as planned, pooling across all strata pre- and post- corrective action.
- If the contrast is significant then we will carry out the analyses separately for A1 and for the combined {A2+B+C}.

We will also directly compare outcomes between A1 and A2 by calculating the LSM for the difference between the two, along with the associated 95% CI.

5.7.2. Analysis Model

The linear mixed effects model for TBUT can be written as follows:

$$y_{ij} = \beta_0 + \beta_1 T_i + \beta_2 TBUT_{ij,0} + b_{0i} + \epsilon_{ij}$$

Where

- $TBUT_{ij,k}$ is the TBUT value for subject i, eye j, week k.
- y_{ij} is the 1 month change from baseline TBUT for subject i, eye j.
  $$y_{ij} = TBUT_{ij,4} - TBUT_{ij,0}$$

- T_i is the binary treatment indicator for subject i (0 = Control, 1 = IPL Laser)
- $\beta_0, \beta_1, and\ \beta_2$ are the fixed effect parameters:
  - β_0 is the mean change from baseline for the Control arm,
    β_1 is the difference between IPL and Control in mean change from baseline,
  - β_2 is the effect of each unit change in baseline TBUT.

- b_{0i} is the random effect (random intercept term) for subject i
  $$b_{0i} \sim N(0, \sigma_a^2)$$

- ϵ_{ij} is the normally-distributed error term
  $$\epsilon_{ij} \sim N(0, \sigma_e^2)$$

- $\{b_{0i}\}$ and $\{\epsilon_{ij}\}$ are independent

The linear mixed effects models can be fit in SAS or R as follows:

In SAS:

```
Proc Mixed data=IPLdat method=ml;
    class SUBJ TRT;
    model TBUT_DIFF =  TRT TBUT_BL  / solution;
    random intercept / subject=SUBJ;
```

In R:

```
Library(lme4)
fit.tbut <- lmer(TBUT.DIFF ~ TRT + TBUT.BL + (1|SUBJ) ,
REML=FALSE, data=IPLdat)
```

Note that since we are interested in testing the fixed effects, ML (maximum likelihood) versus REML (residual maximum likelihood) estimates are being used.

5.8. *Appendix: SAS sample size output*

5.8.1. *Alternative power levels*

```
                     The SAS System

T-test for non-inferiority (null = -2) and superiority (null = 0),
                    bigger is better

                   The POWER Procedure
            Two-Sample t Test for Mean Difference

                 Fixed Scenario Elements

            Distribution                 Normal
            Method                        Exact
            Number of Sides                   1
            Standard Deviation                5
            Group 1 Weight                    1
            Group 2 Weight                    1
            Alpha                          0.05
```

```
                          Computed N Total

              Null    Mean    Nominal   Actual       N
   Index      Diff    Diff    Power     Power     Total

     1        -2       5       0.80     0.845        16
     2        -2       5       0.85     0.884        18
     3        -2       5       0.90     0.914        20
     4        -2       4       0.80     0.825        20
     5        -2       4       0.85     0.858        22
     6        -2       4       0.90     0.908        26
     7        -2       3       0.80     0.824        28
     8        -2       3       0.85     0.868        32
     9        -2       3       0.90     0.902        36
    10        -2       2       0.80     0.817        42
    11        -2       2       0.85     0.861        48
    12        -2       2       0.90     0.905        56
    13        -2       1       0.80     0.809        72
    14        -2       1       0.85     0.853        82
    15        -2       1       0.90     0.904        98
    16        -2       0       0.80     0.800       156
    17        -2       0       0.85     0.852       182
    18        -2       0       0.90     0.901       216
    19         0       5       0.80     0.824        28
    20         0       5       0.85     0.868        32
    21         0       5       0.90     0.902        36
    22         0       4       0.80     0.817        42
    23         0       4       0.85     0.861        48
    24         0       4       0.90     0.905        56
    25         0       3       0.80     0.809        72
    26         0       3       0.85     0.853        82
    27         0       3       0.90     0.904        98
    28         0       2       0.80     0.800       156
    29         0       2       0.85     0.852       182
    30         0       2       0.90     0.901       216
    31         0       1       0.80     0.800       620
    32         0       1       0.85     0.851       722
    33         0       1       0.90     0.900       858
```

5.8.2. Power for a sample of size 42 (21 per arm)

```
                        The SAS System

T-test for non-inferiority (null = -2) and superiority (null = 0),
                      bigger is better

                    The POWER Procedure
            Two-Sample t Test for Mean Difference

                  Fixed Scenario Elements
```

```
           Distribution                    Normal
           Method                          Exact
           Number of Sides                     1
           Standard Deviation                  5
           Group 1 Weight                      1
           Group 2 Weight                      1
           Total Sample Size                  42
           Alpha                            0.05


                     Computed Power

                    Null     Mean
           Index    Diff     Diff     Power

             1       -2        5      0.998
             2       -2        4      0.985
             3       -2        3      0.938
             4       -2        2      0.817
             5       -2        1      0.605
             6       -2        0      0.355
             7        0        5      0.938
             8        0        4      0.817
             9        0        3      0.605
            10        0        2      0.355
            11        0        1      0.157
            12        0        0      0.050
```

5.8.3. Power for a sample of size 72 (36 per arm)

```
                   The SAS System

T-test for non-inferiority (null = -2) and superiority (null = 0),
                  bigger is better

                 The POWER Procedure
         Two-Sample t Test for Mean Difference

             Fixed Scenario Elements

           Distribution                    Normal
           Method                          Exact
           Number of Sides                     1
           Standard Deviation                  5
           Group 1 Weight                      1
           Group 2 Weight                      1
           Total Sample Size                  72
           Alpha                            0.05


                     Computed Power
```

| | Null | Mean | |
Index	Diff	Diff	Power
1	-2	5	>.999
2	-2	4	>.999
3	-2	3	0.995
4	-2	2	0.957
5	-2	1	0.809
6	-2	0	0.514
7	0	5	0.995
8	0	4	0.957
9	0	3	0.809
10	0	2	0.514
11	0	1	0.211
12	0	0	0.050