Photoreceptor Structure in a Phase 2 Study of Encapsulated Human NTC-201 Cell Implants Releasing Ciliary Neurotrophic Factor (CNTF) for Participants with Retinitis Pigmentosa Using Rates of Change in Cone Spacing and Density as the Primary Outcome

Statistical Analysis Plan

Confidential

Version 2.0 dated 20APR2022

Prepared by:
Travis C. Porco, PhD, MPH
Division of Preventive Medicine and Public Health
Department of Epidemiology and Biostatistics
490 Illinois St., Floor 2, Box 0944
San Francisco, CA 94107
University of California, San Francisco
San Francisco, California 94143-0412
Telephone: (415) 476-4101
Email: travis.porco@ucsf.edu

1 Introduction

This document (Statistical Analysis Plan, SAP) describes the planned analysis and reporting for the clinical trial entitled, "Photoreceptor Structure in a Phase 2 Study of Encapsulated Human NTC-201 Cell Implants Releasing Ciliary Neurotrophic Factor (CNTF) for Participants with Retinitis Pigmentosa Using Rates of Change in Cone Spacing and Density as the Primary Outcome", University of California, San Francisco (J. Duncan, PI). It includes specifications for the statistical analyses and tables to be prepared for the final Clinical Study Report.

The proposed trial is a Phase II clinical trial to estimate the effects of an implant which delivers sustained release CNTF (ciliary neurotrophic factor) to the retina (Neurotech USA) in treating retinitis pigmentosa and Usher Syndrome Types 2 and 3. The trial will also collect safety data.

The content of this Statistical Analysis Plan (see McCulloch, 2007) meets the requirements stated by the US Food and Drug Administration (Department of Health and Human Services, Food and Drug Administration, 1998) and conforms to the American Statistical Association's Ethical Guidelines (American Statistical Association, 1999).

The following documents were reviewed in preparation of this Statistical Analysis Plan:

- Proposal for "Photoreceptor Structure in a Phase 2 Study of Encapsulated Human NTC-201 Cell Implants Releasing Ciliary Neurotrophic Factor (CNTF) for Participants with Retinitis Pigmentosa Using Rates of Change in Cone Spacing and Density as the Primary Outcome"
- ICH Guidance on Statistical Principles for Clinical Trials (Department of Health and Human Services, Food and Drug Administration, 1998)

The planned analyses described in this Statistical Analysis Plan will be included in future manuscripts. Note, however, that exploratory analyses not necessarily identified in this Statistical Analysis Plan may be performed to support the analysis. All post-hoc or unplanned analyses which have not been delineated in this Statistical Analysis Plan will be clearly documented as such in the final Clinical Study Report, manuscripts, or any other document or submission.

2.1 Study Design

The trial, Photoreceptor Structure in a Phase 2 Study of Encapsulated Human NTC-201 Cell Implants Releasing Ciliary Neurotrophic Factor (CNTF) for Participants with Retinitis Pigmentosa Using Rates of Change in Cone Spacing and Density as the Primary Outcome ("CNTF Implant Trial"), is a Phase 2 clinical trial to determine whether treatment with a sustained-release CNTF implant (NT-501, Neurotech, USA) is safe and effective in reducing the rate of cone photoreceptor loss in eyes with RP and Usher syndrome types 2 and 3. The Specific Aims and rationale are indicated in the clinical trial Proposal.

2.2 Study Population

The study population is described in detail in Duncan (2010). Full details of the exclusion criteria are provided in that document.

2.3 Study Objectives and Endpoints

2.3.1 Primary Objective

The objectives of the study are to assess safety and efficacy of the sustained-release CNTF implant, based on cone spacing and density measures, supplemented with additional measurements.

The primary outcome measure for hypothesis testing is **cone photoreceptor spacing**, measured using AOSLO imaging in preselected regions of interest (ROI). Specifically, regions of interest are identified at baseline *prior to randomization* and followed longitudinally. Untreated, progression of disease exhibits continued longitudinal increase in cone spacing. Our objective is to compare the change in cone spacing in eyes treated with the CNTF implant with the change in cone spacing in contralateral eyes receiving sham surgery.

An additional primary outcome measure is cone spacing. Cone spacing increases progressively in untreated patients, and we wish to compare changes in cone density over time in eyes receiving CNTF implant with contralateral sham-treated eyes.

The study period will be 36 months.

Procedures for evaluating cone density and spacing using raters trained in grading images will be discussed below.

Additionally, we propose to evaluate the incidence of adverse events following application of the investigational product in all subjects.

The primary hypothesis test will be two-sided. If cone photoreceptor spacing increases (indicating a greater loss or thinning of cone cells) more in the treatment eyes than the control eyes, then this will provide evidence against the safety of the implant. However, safety will also be assessed using other laboratory measurements. (see §7.4 for further discussion). Similarly, if cone photoreceptor spacing decreases less in the treatment eyes than the control eyes, this will provide evidence of efficacy.

2.3.2 Secondary Objectives

Secondary outcome measures will compare rates of change in standardized measures of visual function between CNTF- and sham-treated eyes. Secondary outcomes will include functional and structural assessments:

- Mean and median change in best spectacle-corrected visual acuity (BCVA) using the ETDRS protocol,
- Changes in visual field sensitivity and foveal sensitivity in decibels (dB) using automated static perimetry using the Humphrey visual field 10-2 protocol and rod and cone-mediated fundus-guided microperimetry using a modified Nidek MP1 (see attached protocol, Appendix 5),
- Changes in ONL (outer nuclear layer) and OS+ (outer segment) layer thickness, and
- Changes in full-field ERG (electroretinogram) scotopic and photopic standard a- and b-wave amplitudes and timing from baseline to 36 months.

All parameters except full-field ERG will be measured every six months.

3 Study Design

3.1 Sample Size Planning

3.1.1 Alpha level

The study hypotheses are listed as bidirectional effects; we will never conduct one sided tests. All hypothesis tests proposed for the CNTF Implant Trial are two-sided.

Results will be judged significant using a two-tailed α of 0.05.

3.1.2 Power

We plan recruitment of sufficient subjects for an 80% power to detect an effect size of 0.084 arcminutes over two years.

3.1.3 Statistical Properties of the Primary Outcome

Based on our preliminary study, we expect the standard deviation in the difference between the regression slopes to be approximately 0.11 arcminutes.

3.1.4 Withdrawals and expected loss to follow-up

We conservatively allow for a 15% loss to follow-up based on prior studies.

3.1.5 Sample size calculation methodology

We used a simplified version of the analysis we plan in order to estimate the power for the more complex analysis. This reduces the analysis to a paired T-test. More specifically, we determined the rate of change over time of the cone spacing in our pilot patients, and computed a pooled standard deviation over all these regions (in untreated persons) from which we derived an expected standard deviation of the difference in the total change in cone spacing after two years of approximately 0.11. Using the standard formula for the sample size of a one-sample T-test

(e.g. Chow et al, 2007, formula 3.1.2), we find we would need a total of 17 individuals per group, or 20 adjusting for loss to follow-up.

3.2 Secondary outcomes

The sample size was determined with respect to the single prespecified primary outcome variable for hypothesis testing, cone spacing.

For the second primary outcome variable, **cone density**, we calculate the effect size for which our power is 80% as follows (although we will perform hypothesis tests here, this analysis is strictly secondary). We fitted a linear regression (ordinary least squares) to cone density over time in each region of interest in a small subset of patients (Talcott et al, 2011); we estimated the standard deviation of the regression slopes, and used this to estimate the standard deviation in the difference in the change at 2 years. Using this in the sample size formula for a one-sample T-test (Chow et al, 2007) yields an effect size of approximately 5% of the baseline value of cone density over two years. Expressed differently, if the true effect of the implant is to result in 5% more cones relative to baseline than the control eyes, we have 80% power to detect this effect. Assuming our standard deviation estimate is too optimistic, with the true value being 30% higher, we have power to detect a difference of 7% in cone density relative to baseline at 2 years.

For **visual acuity**, we have 80% power to detect a standardized effect size of 1.02 (using the standard formula for a one-sample T-test cited above), considering the outcome to be the change in the difference in acuity between the treatment and control (sham-implant) eyes. Using a standard deviation for the signed visual acuity difference between two eyes at baseline of 0.05 logMAR (e.g. Brown and Yap, 1995) and assuming independence of changes, we would expect greater than 80% power to detect an effect size of 0.1 logMAR in the change in the difference between two eyes. We do not have adequate information at this time regarding the standard deviation of the visual acuity difference between the eyes in our patient population or of the expected correlation between two time points assuming no treatment effect.

For changes in **visual field sensitivity**, **foveal sensitivity**, **outer nuclear layer**, **outer segment thickness**, and **changes in ERG measurements**, we will be able to obtain variance information which will be useful in planning future studies. We will report confidence intervals for estimated effects. Note that the trial will also provide information needed to plan further studies of retinal function and cone spacing/density.

Covariates, including age, gender, and diagnosis, will be used in selected exploratory or supplemental analyses.

3.3 Randomization

Each patient will receive a unique identifier. Each patient will be randomized to receive the implant in the right or left eye (with the sham in the other eye). T. Porco will generate the randomization list using the statistical package R, creating an Excel® spreadsheet in which each unique identifier is listed on a row next to a treatment assignment. The name of the file will clearly identify it as a randomization list. TP will maintain a hard copy of the randomization list in a locked cabinet in a locked room, and will communicate the randomization lists to the Study Surgeon in person (or using secure encrypted email if absolutely necessary). To ensure against

information loss due to unforeseen circumstances, a backup copy of the randomization list will be maintained by the Safety Monitoring Committee Chair in a locked cabinet stored away from the main UCSF site, and also a backup copy of the randomization list will be maintained on an encrypted partition on the REDCap server; these copies would not be accessible to the PI or to any study personnel. Our procedures are designed to maintain absolute confidentiality of the list while ensuring integrity.

The randomization protocol is as follows. First, a seed integer for the random number generator will be chosen prior to any collection of data; the choice of the seed is arbitrary, but it determines the pseudorandom sequence generated by the algorithm, and it will be recorded in a sealed envelope and kept confidential. The algorithm will be the default algorithm used by the R statistics package. Ten assignments for the right eye and ten for the left eye will then be produced by randomly shuffling a sequence of ten "R" and ten "L" characters using the `sample` command from R. These assignments will be presented in a spreadsheet to the surgeon using a hand-delivered hard copy. Each line of the spreadsheet will contain a randomization code for each patient and an assignment

3.4 Masking

Masking procedures and procedures designed to maintain masking are discussed in Appendix 2 and the grant Proposal.

4 Analysis Populations

4.1 Summary

The following analysis populations are planned for this study:

- The **screening population**, which is to include all subjects who provide (a) baseline screening (including any demographic, visual acuity, or photographic data), and (b) informed consent.
- The **safety population**, which is to include all patients who receive the investigational intervention.
- The **intent-to-treat efficacy population**, which is to include all patients who are randomized. This is the primary population for the efficacy analyses.
- The **per-protocol efficacy population**, which is to include all patients in the intent-to-treat efficacy population, excluding patients with major protocol deviations.

4.2 Major protocol deviations

The incidence of deviations from the inclusion and exclusion criteria will be summarized using counts and percentages, and the treatment groups compared for the overall frequency of deviations using a $2\times N$ Fisher's exact test. Similar deviations will be grouped into general categories of deviations for a more condensed summary. A listing of deviations by participant will also be produced. Any major deviations from the protocol will be listed and/or summarized, including, but not limited to, participants who:

- never received the intervention
- were subsequently found to be ineligible for the study
- never returned for a follow-up visit
- have follow-up visits outside any prescribed visit window

The number and percentage of randomized subjects actually receiving study medication will be summarized. The treatment groups will be analyzed for the proportion of and reason for study discontinuation using the chi-square test. A summary of volunteer status at the end of the study period will also be generated with categories including lost to follow-up.

5 Data Collection and Quality Assurance

Data collection and quality assurance is described in Appendix 2 and the grant Proposal.

6 Human Subjects

6.1 Summary of final dispositions

All subjects who provide informed consent will be accounted for in this study. The frequency of subjects in each population will be presented. We will also present the frequency of subjects in each subgroup, the frequency of withdrawal and loss to follow-up, and any major protocol violations.

6.2 Safety Monitoring Committee

6.2.1 Scope

A Safety Monitoring Committee (SMC) will be empaneled. The committee will meet by teleconference at study initiation and will convene biannual teleconferences for progress reports. *Ad hoc* meetings as needed may also be convened. All study protocols will be subject to review and approval by Institutional Review Boards at the UCSF and by the SMC.

No planned interim efficacy analysis is planned.

6.2.2 Meetings

All teleconference meetings of the SMC and study personnel will consist of (a) "open" sessions, which may be attended as needed by masked study personnel, and (b) "closed" sessions, which may only be attended by unmasked study personnel (TP). Care will be taken so that *no* treatment assignments, data which would allow treatment assignments to be determined, or outcome data based on treatment assignments will be revealed during the open sessions.

Interim reports for the SMC will be prepared by study biostatistician (TP). These reports will include (a) recruitment overall, and by study site, (b) compliance, and (c) retention. The reports will also list study outcomes, and all adverse outcomes, including mortality and perforations. The SMC will determine the database closure dates for each report in advance; archival copies of the (a) main SQL database, and (b) study analysis file as they exist at the time of each report will be maintained. All reports will be sent using secure email to the members of the SMC two weeks prior

to each meeting. Each interim report will be labeled clearly as confidential, printed in binding so that the contents are not visible from the outside, and labeled with the name of each person authorized to receive it. In addition, redacted versions of the interim reports will be prepared which contain no masked study information, and which are suitable for restricted distribution to other personnel on an as-needed basis. All hard copies will be destroyed at the end of each meeting, except for a copy to be kept in a locked file cabinet accessible only to TP.

6.2.3 Decisions

The SMC will make nonbinding recommendations. Guidelines will be agreed upon at the initial meeting, and are expected to include (a) safety, (b) effect of baseline covariates, or (c) validity.

Benefits. In this relatively small Phase II trial, we do not propose unmasked interim analyses to determine whether or not sufficient evidence has accumulated to justify stopping the trial because one treatment is clearly superior (and therefore should be extended to all future cases). Such interim analyses may, however, be requested by the SMC, in which case we will utilize available procedures. If needed, a flexible alpha spending function will be prespecified at the beginning of the trial and included in version 2.0 of this SAP.

Harm. Stopping for harm may be recommended. Several endpoints will be examined, including (a) visual acuity, (b) cone density and spacing as measured by an AOSLO, (c) visual sensitivity measurements, (d) adverse events, and especially (e) serious adverse events, including mortality. While the analysis would consider maldistribution of predictive factors such as baseline visual function in the treatment eye and underlying medical history, it is recognized that ethical considerations require careful considerations of statistical tests as well as qualitative judgments in the light of experience. Any additional analyses required by the SMC will be conducted by TP as needed.

Note that serious adverse events (SAE) are reported directly to the SMC Chair with 24 hours of the time the study site learns of them. The SMC Chair receives notification of the event, the timing of the event, a medical narrative from the medical monitor, the site, the patient identification number. The statistician reports the study treatment assignment to the SMC Chair.

We do not anticipate any difficulties with respect to safety, because the implant device has been studied in previous clinical trials and no safety issues have been observed. Nevertheless, if the CNTF implant use results in an unacceptable increase in the risk of adverse outcomes, then the study will be stopped. It is difficult to fully prescribe boundaries for monitoring safety because there need not be strong evidence to discontinue the study if it appears that the treatment is harmful.

Futility. For our Phase II trial, we are not proposing decision rules for early discontinuation due to the unlikeliness of significant findings conditional on interim results. Should the SMC recommend such procedures, we would propose to use available procedures for conducting unmasked interim analysis together with sample size re-estimation.

7 Statistical Analysis

7.1 Summary and Descriptive Statistics

7.1.1 Demographics and Patient History

All demographic and history variables determined at presentation or enrollment will be summarized by counts and percentages tabulated by treatment assignment. In particular, we propose to collect age, sex, eye involved, previous surgery, visual acuity, cone spacing, cone density, all current medications, and all current medical problems. Laboratory measurements are reviewed in §2.3.2 and in detail in the study Proposal.

7.1.2 Prior and concurrent medication

We will present the percentage taking any ophthalmic medications at baseline.

7.1.3 Baseline comorbidities and history

Clinical variables at baseline will be presented by gender and age. We will also tabulate the baseline visual function measures and laboratory measurements. Measurements of physiological function (i.e., Humphrey visual field measurements, and all other measurements outlined in §2.3.2) will be tabulated by treatment assignment (note that treatment assignment is random, independent of these measurements). In addition, the mean distance from the fovea of all prospectively identified regions of interest in each eye will be calculated (i.e., if D_{ijk} is the mean distance to the fovea of region k in eye j of person i, we compute $\overline{D}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} D_{ijk}$, where n_{ij} is the number of regions of interest in eye j of person i) and tabulated by subsequent treatment assignment of the eye. Note that the *mean foveal distance of the regions of interest is independent of treatment assignment*; this tabulation is designed to provide insight into whether the randomization process yielded any substantial imbalance in this factor.

7.2 Accrual

We anticipate accruing sufficient patients within 15 months following the beginning of the study. Details are given in the Proposal.

7.3 Interim and Final Analysis

As indicated earlier, no formal prespecified interim analysis of efficacy is planned. We expect, however, to report results to the SMC during the progress of the trial.

Stopping guidelines for harm are at the discretion of the medical monitor and/or the SMC. In this trial (a small phase II trial), we propose no formal stopping rules for benefit or futility (see Section 7.2). Any rules or guidelines would be determined at the first meeting of the SMC (see Section 7.2).

All final, planned analyses identified in this Statistical Analysis Plan will be performed only when the last patient has completed the final visit. The final analysis will be conducted by statistician TP.

7.4 Safety and Efficacy Outcomes

The primary outcome variable (cone spacing) will be analyzed using a two-sided test with symmetric rejection regions on each side. An relative increase in cone spacing in the treatment group will provide evidence of lack of safety; a relative decrease in cone spacing in the treatment group (the treatment group loses fewer cones than the control group) will provide evidence for efficacy.

While (as discussed in the proposal) the CNTF implant has been found safe in previous clinical trials, it is understood that failure to find evidence of harm does not demonstrate safety. Our purpose has been to refine these earlier safety assessments using more refined measurements. We recognize that multiple measures may be needed, together with clinical assessment by experts, to form a complete judgment; we note that we are not powered to detect unlikely events. *Our reporting on safety will always reflect these considerations.* Adjustments, such as an asymmetric rejection region for the primary outcome variable or the use of noninferiority procedures, may be requested by the SMC, and would be prespecified in this SAP prior to enrolling patients.

7.5 Hypothesis Tests

7.5.1 Analysis of primary outcome

We describe an analysis template for continuous outcomes which will be repeated as indicated.

7.5.1.1 Primary analysis

While we consider both cone spacing and density to be primary outcome measures, for the purpose of formal hypothesis testing for the efficacy aims of the trial, we chose the outcome variable to be cone spacing. We will analyze cone spacing as the primary prespecified outcome, but also conduct an analogous analysis of cone density. Cone spacing as well as cone density will be measured longitudinally in regions of interest *selected prior to randomization and surgery* and followed over time. Specifically, as detailed in the flow sheet of examination procedures in the Protocol, we will consider acquiring measurements at the following times:

Baseline 1, within 7 weeks of implant surgery

Baseline 2, within 4 weeks of baseline 1

Six months (plus or minus 7 days)

Twelve months (plus or minus 14 days)

Eighteen months (plus or minus 14 days)

Twenty-four months (plus or minus 14 days)

Thirty months (plus or minus 14 days)

Thirty-six months (plus or minus 14 days)—this measurement is the primary outcome and is required

A minimum of seven regions of interest per eye will ideally be established from the Baseline 1 images. These specific regions will be followed longitudinally. Note that if outcome data were not available at a specific visit due to inability to acquire images from patients, the study could

optionally be continued at the discretion of the DSMC. Such decisions would be made independent (masked to) all outcome data.

The primary hypothesis test uses the 36-month observations.

We use the following notation. The data are hierarchical, with region nested within eye and eye nested within person. Let each region within each eye be labeled k, k=1, ..., n_{ij}, where n_{ij} is the number of regions of interest identified within eye j of person i. We denote the control eye by j=0, and the treatment eye by j=1. We denote the study time by ℓ, with $\ell=0$ representing baseline, and $\ell=1$ representing the final followup. We let $Y_{ijk\ell mg}$ denote the outcome variable for person i, eye j, region k, time ℓ, repetition m, and grader g. In the case of the baseline visit, m=1, 2; for the followup, m=1; grader g=1,2 identifies the two graders. We denote the foveal distance of each region of interest as D_{ijk}; each region of interest is prospectively identified and followed over time, and these distances are therefore constant over time.

Z scores will be used as the prespecified outcome variable, based on analysis of normal subjects. Spacing is determined by analysis of images, by two independent graders.

The baseline score for a region of interest for each grader is $(Y_{ijk01g}+ Y_{ijk02g})/2$. The change score for each region of interest for grader g is $\Delta_{ijkg} = Y_{ijk\ell 1g}-(Y_{ijk01g}+ Y_{ijk02g})/2$. The mean change score (averaging over graders) is $\Delta_{ijk}=(\Delta_{ijk1}+ \Delta_{ijk2})/2$. The average change score for each eye is $\Delta_{ij}=(\Sigma_k \Delta_{ijk})/n_{ij}$. *The difference between the treat*ment and control eyes for person i is $U_i= \Delta_{i1}- \Delta_{i0}$.

*Under the null hypothesis, the differences U in change between treat*ment and control should have mean 0. We test this hypothesis using the Wilcoxon rank sum test.

Additional analyses may be reported. In each case, we will distinguish between such supplementary analyses and the primary analysis.

- The grader-specific difference in change scores, analyzed using Rosner's clustered version of the Wilcoxon rank sum test
- Analyses of covariance, in which the baseline is used as a regressor and the outcome variable is the final value. These models will be at the level of the individual ROI, clustering on eye and person.
- Subset analyses in which the analysis is restricted to regions of interest in which both graders' results are available at all visits. Note that this does not eliminate bias, since missingness of follow-up data occur post-randomization.

7.5.1.2 Cone spacing

The analytic template in §7.5.1.1 (and in §7.5.1.3 below) will be applied to the cone spacing measurements. Inferentially, we consider cone spacing as a secondary outcome.

7.5.1.3 Longitudinal analysis

As a complement to the above analysis, we will also model the clinical trial data using linear mixed models (e.g. Yasui et al, 2004), a modeling approach we have used in analysis of cone spacing and density data in the pilot study (Talcott et al., 2011)—provided interim observations are available.

7.5.1.4 Image grading

The outcome variables above will be derived by averaging cone spacing and cone density measurements for independent graders. These graders will be masked to the treatment during the entire trial.

The linear mixed model framework may be expanded to model the ratings individually by including a rater-specific random effect.

7.5.2 Secondary efficacy variables

The following methods will be used to assess secondary endpoints, as indicated. While we will report Holm-corrected p-values (correcting for six hypothesis tests), we note that such procedures are conservative and reduce our power to detect lack of safety. We emphasize that safety cannot be assessed by means of these hypothesis tests, that alternative lower bounds may be of interest to the SMC, and that a full assessment must take into account multiple factors and clinical judgment. As with the primary outcome variable, the analytic results in this section are relevant both to safety (when results are worse in the treated eye), and to efficacy (when the results are better).

7.5.2.1 Visual acuity

The template of §7.5.1.1 will be used here, except that we will simply model the difference in visual acuity between the eyes (regions of interest play no role, and we do not have multiple graders). We will test the hypothesis that the difference in visual acuity between the treatment and control eyes is constant in time, using a two-sided alpha of 0.05 (following Holm correction). Models including the covariates (such as age and diagnosis) will be explored.

7.5.2.2 Visual field sensitivity

As above, we will follow the template in §7.5.1.1, using visual field sensitivity in each quadrant to produce an overall measure for each eye. A two-sided alpha of 0.05 will be used (following Holm correction).

7.5.2.3 Foveal sensitivity

Foveal sensitivity will be examined in the same way as visual acuity (one measurement per eye), using the same strategy as in §7.5.2.1. A two-sided alpha of 0.05 will be used (following Holm correction).

7.5.2.4 Outer nuclear layer

Outer nuclear layer measurements will also be examined in the same way as visual acuity (one measurement per eye), using the same strategy as in §7.5.2.1. A two-sided alpha of 0.05 will be used (following Holm correction).

7.5.2.5 Outer segment thickness

Outer segment thickness will also be examined in the same way as visual acuity (one measurement per eye), using the same strategy as in §7.5.2.1. A two-sided alpha of 0.05 will be used (following Holm correction).

7.5.2.6 ERG measurements

We will model photopic b-wave measurements at the baseline, 12, 24, and 36 month times using the template of §7.5.1.1. This will be conducted at a two sided alpha of 0.05 (following Holm correction for multiple comparisons). Scotopic b-wave, as well as photopic and scotopic a-wave

measurements, will be reported descriptively and will provide additional information which can guide assessments of safety.

7.5.2.7 Other outcomes

For the incidence of adverse events following application of the investigational product in all subjects, we propose to use Poisson or negative binomial random-effects regression (extending to a more general hidden Markov model in the event of poor fit of the simpler models). Specifically, we will test the hypothesis that the underlying event rate per unit time is the same in the treatment and sham eyes. As a secondary analysis, we will analyze the time to first adverse event in each group using Cox proportional hazards regression (using baseline medical history as well as time since last treatment as covariates). We emphasize that while the statistical fit may be informative, decisions about drug safety (or trial continuation) will not be based solely on statistical criteria.

7.5.2.8 Other analyses

We will also explore multivariate models (using many variables as outcomes), and explore the correlation structure between these variables cross-sectionally and longitudinally.

Plots of the average cone spacing in the treatment eye as a function of average cone spacing in the treatment eye will also be examined, adjusting for mean distance to the fovea (and likewise for density, and all other ocular measurements). Exploratory models in which the spacing in the control eye is used as a predictor of the spacing in the treatment eye may be examined to provide additional statistical insight, but not as formal hypothesis tests.

7.6 Additional statistical considerations

7.6.1 Software

The standard software package R version 2.11 or later (`http://www.r-project.org`) for the MacIntosh OS X will be used for all descriptive and inferential analyses. SAS `PROC MIXED` (SAS Institute, Cary, North Carolina, USA) will be used in addition to R for fitting linear mixed models.

7.6.2 Missing data and loss to follow-up

Missing values of cone spacing or density will be handled by regression-based multiple imputation in several steps.

- Missing values for one grader at a baseline visit, when observations at the same visit are available from the other grader
- Missing values for both graders at a baseline visit, when observations at the other baseline visit are available
- Missing values for one grader at a follow-up visit, when observations are available at the same follow-up visit from the other grader
- Missing values for both graders at a follow-up visit, when values other regions of interest from that eye are available at that visit
- Missing values for all regions of interest for a given eye at the follow-up visit

Complete case analyses will be included in reports. We will report multiple imputation results when we impute baseline visits based on other baseline visits. We will then report multiple imputation results when we extend to impute follow-up data for regions of interest when the second grader is imputed from the first. We will finally report multiple imputation results when we impute outcome data for regions of interest for which no follow-up data are available.

The latter is the primary outcome for the trial, and will be sharply distinguished from other analyses. We note that complete case analyses are biased.

Subsidiary analyses will be conducted to assess the role of image quality on missingness. Results may be biased because of treatment-related loss of image quality.

7.6.3 Transformations and model adequacy

Because the legitimacy of the hypothesis test being conducted depends on the assumptions (i.e. normality and homoskedasticity for linear models, proportional hazards for Cox regression), the adequacy of the statistical model must be checked. Methods which will be employed will include (a) residual plots (vs. baseline value, vs. predicted values, and Q-Q plots), (b) jackknife influence estimates, and (c) tests for normality (including the Anderson-Darling and Shapiro-Wilk procedures).

7.6.4 Multiple comparisons

An alpha of 0.05 will be used for the primary efficacy analysis. For pre-specified secondary analyses, we will report both the P-value and the number of pre-specified analyses performed. Note that the results of secondary analyses are not independent events, making a Bonferroni correction very conservative.

7.7 Safety and tolerability

The analysis of safety in this study will include summaries of the following:

- Exposure
- Adverse events
- Adverse events and serious adverse events
- Adverse events leading to withdrawal
- Any deaths

The incidence of adverse events will be compared between treatment and sham eyes using the McNemar test.

Additional considerations are given in §7.4.

7.7.1 Exposure

Individual eyes are assumed to have exposure to the intervention assigned to the arm they were randomized to.

7.7.2 Adverse Events

7.7.2.1 Individual events

The proportion of subjects with at least one of the following safety-related events will be compared using Fisher's Exact Test. Non-serious adverse events (not requiring narrative form) are list here: irritation or infection from dilation drops, bruising from blood work.

Serious ocular adverse events (which must be reported within 24 hours and which require a narrative form) are list here: acute angle closure glaucoma precipitated by dilation, corneal abrasion after tonometry or electroretinography, risks associated with implantation surgery (cataract, retinal detachment, infection, bleeding).

In addition, we will compare the rate of each of the adverse events during the follow-up period using Poisson regression, which can take into account multiple instances of adverse events within a single subject.

7.7.2.2 Pooled adverse events

Adverse events will be analyzed according to four main categories:

- Proportion of subjects with *any ocular adverse event*
- Proportion of subjects with *any serious ocular adverse event*
- Proportion of subjects with *any non-ocular serious adverse event*
- Proportion of subjects with *any non-ocular adverse event*

The proportion of subjects with these events will be compared between the arms using Fisher's Exact Test. Poisson regression will be applied to compare the rates of overall adverse events, including recurrent events.

8 Reporting conventions

- All tables and data listings will be presented in landscape orientation, unless presented as part of the text of the final report.
- Figures will be presented in landscape orientation, unless the information is substantially easier to interpret in portrait orientation.
- Direct annotation of figures will be preferred to legends. All figures with more than one variable or item will contain either direct annotation or legends. All annotation will be unambiguously identifiable as such.
- Color will be used in figures only when needed to enhance clarity of communication. All color schemes will be evaluated for visual clarity for individuals with diminished color vision. All color encodings will be identified. Redundant encodings (such as the use of different plot symbols or line dash patterns) will be used in addition to color, so that all figures are interpretable after monochrome reproduction at 100 dots per inch. All dash patterns and line widths will be adequate to be distinguishable after monochrome reproduction at 100 dots per inch. Any distinction between plot symbols (circles, filled circles, diamonds, etc.) will remain clear after monochrome reproduction at 100 dots per inch.

- Sans serif fonts will be used for all labeling (Helvetica, Arial, Futura, or Computer Modern Sans Serif (CMSS)).
- Boldface and italics will not be used unless substantial value is added.
- Decorative fonts and enhancements, including borders and shading, will not be used. Decorative presentation methods, such as ribbon graphs, will never be used.
- All information given in figures will also be presented in summary tables (perhaps only included in an Appendix or in supplementary materials).
- Only standard characters will be used in tables and data listings.
- All titles will be centered. The first title line will be the number of the table, figure, or listing. The second and possibly third lines will be the description of the table, figure, or data listing. The ICH numbering convention will be used for all.
- All footnotes will be left justified and at the page bottom. Footnotes will be used sparingly. Reference footnotes will be complete enough to locate any reference based on the information provided (Author, Journal, Pages, Date, or PubMed accession number).
- Missing values for numeric or character variables will be unambiguously identified as such using the special string `NA` (not available) in all settings; `NA` is the standard missing value code for our software. Each figure or table caption in which `NA` is used will indicate the meaning of `NA` in that figure or table. The abbreviation `NA` will never be used for any other purpose.
- All date values will presented in the form DDmmmYYYY format (e.g. 01jan2008), using four digit years. June will be encoded as `jne` (otherwise `jan` and `jun` would differ by only a single character), and July as `jly` (so that the lowercase letter `l`, easily confused with the digit `1`, will not be adjacent to any numerals).
- All tables, figures, and data listings will have the name of the program and a date/time stamp on the bottom of the output.

9 Abbreviations and acronyms

AES Advanced Encryption Standard

AOSLO Adaptive Optics Scanning Laser Ophthalmoscope

BCVA, BSCVA Best spectacle-corrected visual acuity

CAR(1) continuous autoregression process of order 1 (Ornstein-Uhlenbeck process)

CNTF Ciliary Neurotrophic Factor

CNTF Implant Trial The trial discussed in this SAP; official title: *Photoreceptor Structure in a Phase 2 Study of Encapsulated Human NTC-201 Cell Implants Releasing Ciliary Neurotrophic Factor (CNTF) for Participants with Retinitis Pigmentosa Using Rates of Change in Cone Spacing and Density as the Primary Outcome*

DSMC Data and Safety Monitoring Committee

ERG Electroretinography

ETDRS Early Treatment of Diabetic Retinopathy Study

FDA U.S. Food and Drug Administration

FIPS Federal Information Processing Standard

JD Jacque Duncan

NIST National Institute of Standards and Technology

PI Principal Investigator

ROI region of interest

SAP Statistical Analysis Plan

SMC Safety Monitoring Committee

TP T. Porco

UCSF University of California, San Francisco

References

American Statistical Association. 1999. *Ethical Guidelines for Statistical Practice*. Alexandria, Virginia: ASA.

Brown B. & Yap M. K. H. 1995. Differences in visual acuity between the eyes: determination of normal limits in a clinical population. *Opthal. Physiol. Opt.* 15:163-169.

Chow, S. C., Shao, J., & Wang, H. 2007. *Sample size calculations in clinical research, 2d ed*. Boca Raton, Florida: Taylor and Francis.

Department of Health and Human Services, Food and Drug Administration. 1998. International Conference on Harmonization; Guidance on Statistical Principles for Clinical Trials. *U. S. Federal Register*, **63**(179), 49583–49598. Docket 97D-0174.

Duncan, J. 2010. IND BB-IND 10,931, Serial #0060: Protocol Amendment: "Photoreceptor Structure in a Phase 2 Study of Encapsulated Human NTC-201 Cell Implants Releasing Ciliary Neurotrophic Factor (CNTF) for participants with Retinitis Pigmentosa Using Rates of Change in Cone Spacing and Density as the Primary Outcome", submitted to the FDA on 22 December 2010. Appendix 2.

McCulloch, C. 2007. Statistical Analysis Plan, Steroids for Corneal Ulcers Trial (T. Lietman, PI), unpublished.

Nobre, J. S. & Singer, J. 2007. Residual analysis for linear mixed models. *Biometrical Journal* 49:863-875.

Talcott, K. E., Ratnam, K., Sundquist, S. M., Lucero, A. S., Lujan, B. J., Tao, W., Porco, T. C., Roorda, A., Duncan, J. L. 2010. Longitudinal study of cone photoreceptors during retinal degeneration and in response to Ciliary Neurotrophic Factor treatment. *Invest Ophthalmol Vis Sci* 2011;52:2219-2226.

Yasui Y, Feng Z, Diehr P, McLerran D, Beresford SA, McCulloch CE. 2004. Evaluation of community-intervention trials via generalized linear mixed models. *Biometrics* 60:1043-1052.