INTERMITTENT EXOTROPIA STUDY 6 (IXT6)

A Pilot Randomized Clinical Trial of Base-in Prism Spectacles for Intermittent Exotropia

STATISTICAL ANALYSIS PLAN

Version 1.0
March 4, 2020

Based on Protocol version 1.0 (May 29, 2019)

Revision History

VERSION NUMBER		AUTHOR	APPROVER	EFFECTIVE DATE	REVISION DESCRIPTION (INCLUDING SECTIONS REVISED)
SAP	Protocol				
1.0	V1.0	A. Hercinovic D. Chandler	M. Melia	4-15-20	Initial version

1. Objective of IXT6 Pilot Study

The objective of this short-term, pilot randomized trial comparing spectacles with relieving prism to spectacles without prism is to determine whether to proceed to a full-scale, longer-term randomized trial. This decision will be based primarily on assessing the initial (8-week) response to prism and the preliminary estimates of treatment effect in both the prism and non-prism group.

2. Primary Analysis – Comparison of Control of Distance Exodeviation Between Treatment Groups

The primary analysis will be an intent-to-treat comparison of mean 8-week control of the distance exodeviation (average of 3 measurements) between treatment groups using an analysis of covariance (ANCOVA) model, which adjusts for baseline distance control [1].

Distance control at the 8-week visit (average of up to 3 measurements) will be tabulated by treatment group into the following categories: 0 to <1, 1 to <2, 2 to <3, 3 to <4, 4 to <5, and 5 points. The mean treatment group difference (prism – non-prism) and corresponding one-sided 95% confidence interval and p-value will be calculated. Although a two-sided test would be used in the longer-term, full-scale trial, a one-sided test is being used for the pilot study given that the decision whether to proceed to a full-scale trial is based only on whether overminus is better than the non-overminus control group. Model assumptions for the ANCOVA will be assessed, including linearity of the relationship with baseline distance control, normal distribution and equal variance across the treatment groups. If the linearity assumption is not met, a transformation for linearity will be performed, or baseline control will be categorized. Serious deviation of the mean 8-week control scores distribution from normality, conditional on treatment group and baseline control, is not expected given the outcome scale is bounded; but, if such deviation is observed, the van der Waerden normal scores transformation will be considered. An ANCOVA that allows for unequal variance in the 2 treatment groups will be used if the variance of the active treatment group differs from that of the control group by more than 20% [2].

When the protocol-specified three measures of control are not performed at the outcome exam, the mean of two tests will be used for analysis if only 2 distance control tests are completed. The single distance control score will be used for analysis if only 1 testing is completed. To account for missing data, a record for each randomized participant will be included in the analysis dataset containing all independent variables in the ANCOVA model along with outcome data when available, and maximum likelihood-based estimation will be used to obtain an estimate of treatment effect. The estimate from this model is unbiased when there are missing outcome data, as long as those data are missing at random.

2.1. Alternative Approaches to Primary Analysis

Results of a previous PEDIG pilot study of overminus lenses (IXT3) suggested a potential interaction between baseline control and treatment with overminus lenses. Hence, the interaction between baseline control and treatment will be tested for statistical significance in the current study by adding their interaction into the primary analysis model. It is recognized that power for this test is low, and lack of significance does not rule out interaction. For the interaction test, baseline control will be divided into two groups: 2 to less than 3.5 points, and 3.5 to 5 points. If there is significant interaction, an estimate of the treatment effect and 95% confidence interval will be obtained for each of the two baseline control subgroups.

The primary analysis will be repeated in sensitivity analyses as follows:

- The analysis will be limited to participants who received their study spectacles in time to allow for at least 4 weeks of spectacle wear relative to the outcome visit. The analysis will compare the date of spectacle receipt recorded on the 2-week phone call form and the date of the 8-week visit. For subjects who have not obtained the glasses by the time of the 2-week call, protocol monitors and site staff will follow up and have the glasses receipt date on the 2-week form edited once glasses are received.
- The analysis will be limited to subjects who were tested in their study spectacles at the 8-week visit (i.e. exclude subjects tested in trial frames).
- The analysis will be limited to subjects who have all 3 control scores obtained at the 8-week outcome visit (complete case analysis).

Note: records for participants with missing outcome data will be retained in these analyses if they meet the criteria for the analysis.

3. Secondary Outcomes

To aid in interpretation of the primary outcome comparison, 8-week control scores will be used to classify participants into those having no spontaneous tropia during all 3 control tests at distance and near, and additionally, those considered as treatment responders. The differences in treatment group proportions and 95% confidence intervals for each outcome will be calculated using the Farrington-Manning score test. If there is a significant difference ($p<0.05$) on the primary outcome comparison, p-values for these tests will be calculated using Barnard's test. It is recognized that power for these comparisons is low; hence, these outcomes will be considered exploratory and there will be no correction to p-values for multiplicity.

These proposed comparisons will not be adjusted for baseline distance control due to the high potential for small numbers of outcomes, making adjustment problematic. As baseline control is expected to be balanced between treatment groups due to randomization, the comparisons should be unbiased. Nonetheless, an analysis adjusted for baseline control using Poisson regression with the identity link and robust variance estimation will be performed, if possible, to confirm that adjustment does not affect conclusions from the unadjusted analyses.

3.1. Proportion with Treatment Response

"Treatment response" is defined as ≥ 1-point improvement in control of the distance exodeviation (average of 3 measurements) between baseline and the 8-week outcome exam.

3.2. No Spontaneous Tropia During Control Testing

No spontaneous tropia during control testing at the 8-week primary outcome exam is defined as a score of ≤2 (0, 1, or 2) on all three assessments of control at distance and at near.

It is acknowledged that participants who had a score of 2 on all three assessments of distance control at baseline and scores of 2 or better at near control could potentially meet this outcome criteria without any improvement in control. Given the randomization we expect to have similar numbers of such participants in each treatment group; however, if the proportion of participants with scores of 2 or better at distance and near control differs substantially between treatment groups, the proportion with no spontaneous tropia during control testing will be compared using logistic regression adjusting for baseline distance control.

4. Analyses of Additional Treatment Outcomes

Additional analyses will report the distribution of the outcome by treatment group, and an estimate of the difference between treatment groups and a 95% confidence interval (CI) if the outcome is continuous or quasi-continuous, i.e. distance control, near control, ocular alignment, near stereoacuity, distance visual acuity, and fusional convergence, will be estimated using ANCOVA, with adjustment for baseline level of the outcome where appropriate. If ANCOVA assumptions are not met, a non-parametric method such as the Wilcoxon rank sum test will be used instead, with no adjustment for baseline level of the outcome. For ordered categorical outcomes, i.e. symptoms and suppression, each level of the outcome will be assigned an integer score, starting with 0, and the bootstrap method will be used to obtain a median and 2.5^{th} and 97.5^{th} percentiles on the treatment group difference, instead of a mean and 95% confidence interval. P-values for treatment group comparisons will not be reported.

4.1 Change in Distance Control

The distribution of change in distance control from baseline to 8 weeks by treatment group, and treatment group means, standard deviation, and 95% CIs will be reported.

4.2 Near Control

Near control will be evaluated similarly to the distance control in the primary analysis, and as in *4.1*.

4.3 Ocular Alignment

The distribution of ocular alignment at distance and near fixation by PACT will be described for the enrollment exam and the outcome exam for each treatment group. Because participants in the prism group will have PACT measured at 8-weeks while

wearing study-prescribed relieving prism, the magnitude of the prescribed prism will be added to the deviation by PACT while wearing prism to obtain the total underlying deviation. The distribution of change in ocular alignment will also be described for each treatment group.

4.4 Distribution of Near Stereoacuity

The distribution of near stereoacuity by Preschool Randot Test will be described for the enrollment exam and the outcome exam for each treatment group. The distribution of change from baseline in near stereoacuity in log arc seconds will be described for each treatment group. Mean change in log arc seconds by treatment group, the treatment group difference, and a 95% confidence interval will be calculated.

4.5 Adverse Symptoms of Intermittent Exotropia and Prism Spectacle Wear

Adverse symptoms of IXT will be assessed at enrollment and 8-week outcome exam using a symptom survey that is administered to the child. Response options are based on frequency of observations: 0=never, 1=sometimes, and 2=always.

Similarly, adverse symptoms that may be associated with prism spectacle wear will be assessed at enrollment and at the 8-week outcome exam using a spectacle survey that is administered to the parent. Response options are based on frequency of observations: 0=never, 1=rarely, 2=sometimes, 3=often, 4=always, and not applicable. Percentage of not applicable responses will be calculated, but these responses will not be included in the treatment comparisons.

The distribution of scores on each survey item will be described for the enrollment exam and the outcome exam for each treatment group.

4.6 Distance Visual Acuity

Distance visual acuity will be assessed at enrollment and at the 8-week outcome exam. Any optotype method can be used for testing; however, the same method must be used at both the enrollment and 8-week outcome exam. Snellen visual acuities will be converted to logMAR for analysis.

The distribution of distance visual acuity in logMAR will be described for the enrollment exam and the outcome exam for each treatment group. The distribution of change in visual acuity in logMAR will also be described for each treatment group.

4.3 Suppression

The distribution of suppression level (0=none, 1=mild, 2=moderate, 3=severe) will be described for the enrollment exam and the outcome exam for each treatment group.

4.4 Fusional Convergence

The distribution of fusional convergence amplitude will be described for the enrollment exam and the outcome exam for each treatment group. Convergence amplitude was defined as first base-out prism that induced blur (blur point), and, if no blur, base-out prism that induced diplopia (break point). If testing was performed in prism glasses, the

prism value was analogously added or subtracted to yield a net convergence amplitude; credit was given for base-out prism, and base-in prism was subtracted.

5. Other Analyses

5.1 Mean Distance Control in Subgroups

As an exploratory analysis, the treatment group difference in mean distance control score within baseline mean distance control subgroups will be estimated. The hypothesis for a subgroup effect based on mean distance control severity at baseline (2 to <3, 3 to <4, 4 to 5) is that poorer control is associated with larger treatment effect as suggested in the PEDIG IXT3 pilot study of overminus spectacles, another form of non-surgical IXT treatment. Although the greater magnitude of response with poorer baseline control may have been at least partly attributable to regression to the mean and having more room for improvement, the same magnitude of response was not seen in the observation group, suggesting that the larger treatment effect with poorer baseline control observed when treating with overminus spectacles could be real.

This planned subgroup analysis will repeat the primary analysis, adding the baseline factor and the baseline factor by treatment interaction. It is acknowledged that this analysis has very low power and only very strong interactions will be able to be detected. Any observed subgroup effects will require confirmation in a full-scale trial to be considered true effects.

5.2 Prism Adaptation Test (PAT) Screening Study

In addition to the randomized trial, a separate analysis will include all participants undergoing prism adaptation testing as part of screening for the randomized trial, regardless of whether they are eligible for randomization based on the results of prism adaptation testing. The participants to be included will have met all randomized trial eligibility criteria other than that relating to the outcome of the prism adaptation test. The proportion and 95% confidence interval of prism adaptation test-screened participants who fully prism adapt will be estimated. Fully adapting to prism is defined as having magnitude of PACT at distance and/or near angle while wearing "trial" relieving prism for 30 minutes (measured through the prism) which is ≥ the original PACT measurement at the same testing distance. The PACT magnitudes with and without prism are directly compared with one another and do not take into account the amount of prism being worn for testing (in contrast to the 8-week outcome PACT described in section 4.3 in which the magnitude of the prescribed prism will be added to the deviation by PACT while wearing prism for analysis of the total underlying deviation).

For PACT at distance and near, scatterplots will be used to visualize the relationship between the original without-prism test and the subsequent with-prism test.

In addition, the proportion and 95% confidence interval of screened participants who are eligible for randomization following the PAT will be calculated. Participants who meet who meet all three of the following criteria while wearing relieving prism are eligible for randomization.

- Not fully prism adapted
- No NEW esotropia by cover test at near (while wearing relieving prism) – note that a participant with esotropia at near while wearing prism is eligible for randomization IF an esotropia at near (of any magnitude) was present during the original testing without prism.
- No esodeviation >6Δ on PACT at near (while wearing relieving prism)

Note that the analysis plan in the protocol erroneously refers to "repeat enrollment prism adaption testing;" repeat enrollments were not part of this protocol.

5.3 Compliance with Spectacle Wear

Parents were asked to complete a compliance calendar by recording the percentage of time their child wore the study-prescribed spectacle correction each day. Proportion of time worn each day will be described as excellent (76% to 100%), good (51% to 75%), fair (26% to 50%), poor (1 to 25%), or none (0%). Based on review of the calendars and discussion with parents at the 8-week outcome exam, the investigator recorded the total proportion of time worn as excellent (76% to 100%), good (51% to 75%), fair (26% to 50%), poor (1 to 25%), or none (0%: did not fill prescription or never picked up spectacles). The frequency distribution of compliance will be described for each treatment group at the outcome exam.

5.4 Masking Assessment

At the 8-week visit, the proportion of masked examiners who responded that the patient appears to be wearing prism glasses will be compared between treatment groups using a two-sided Barnard's test with alpha of 0.05, with calculation of a two-sided 95% confidence interval on the difference in proportions using the Farrington-Manning score method.

References

1. Raab GM, Day S, Sales J. How to Select Covariates to Include in the Analysis of a Clinical Trial. Control Clin Trials 2000;21:330-342.
2. http://support.sas.com/kb/22/526.html