AMBLYOPIA TREATMENT STUDY (ATS23)

A Randomized Trial of Dichoptic Treatment for Amblyopia in Children 4 to 7 Years of Age

Statistical Analysis Plan

Version 1.0

Version History

Version	Protocol Version	Author	Approver	Effective Date	Study Stage
1.0	1.0	Desirae Sutherland	Wesley Beaulieu	4/17/2024	First participant has not yet been enrolled.

Version	Revision Description
1.0	Original Version

Approvals

Role	Digital Signature or Handwritten Signature/Date
Author: Desirae Sutherland	Desirae Sutherland Digitally signed by Desirae Sutherland DN: cn=Desirae Sutherland ou=North Wing Reason: I am the author of this document Location: Date: 2024-04-18 14:10-04:00
Senior Statistician: **Wesley Beaulieu**	Wesley Beaulieu I agree to the terms defined by the placement of my signature in this document 2024-04-18 09:55-04:00
Jaeb Principal Investigator: **Raymond Kraker**	Ray Kraker I agree to the terms defined by the placement of my signature in this document 2024-04-18 10:13-04:00
Study Principal Investigator: **Robert Henderson**	Robert Henderson Digitally signed by Robert Henderson DN: cn=Robert Henderson ou=South Wing Reason: I have reviewed this document Location: Date: 2024-04-18 10:18-04:00

1. Study Overview

This document outlines the statistical analyses to be performed for the ATS23 Trial and to be included in the primary manuscript data packet.

The protocol is a multicenter trial designed to compare the change in amblyopic eye distance VA from randomization to 26 weeks in participants randomized to treatment with the Luminopia One headset (1 hour per day of watching dichoptic movies, 6 days per week) or patching (2 hours of patching per day, 7 days per week).

The aforementioned treatment regimens of Luminopia One and patching shall subsequently be referred to as LUMINOPIA and PATCHING, respectively.

Approximately 238 participants will be enrolled and randomized in a 1:1 ratio to either LUMINOPIA or PATCHING treatment (119 per group).

At the 26-week primary outcome visit, participants who were randomly assigned to receive PATCHING treatment and have an interocular difference (IOD) of 1 line (0.10 logMAR) or more will be offered Luminopia dichoptic therapy. Participants who agree to the treatment will continue follow-up with visits at 39- and 52-weeks post-randomization. Otherwise, for all other participants the study will end at 26 weeks.

The study protocol specifies the eligibility criteria and schedule of study visits and procedures.

2. Statistical Hypotheses

The primary efficacy outcome will be the change in amblyopic eye distance VA (measured as logMAR) from randomization to 26 weeks. Change in logMAR will be calculated as [outcome VA] – [randomization VA] such that a negative change indicates improvement in visual acuity, and a positive change indicates worsening.

The study is designed to test a one-sided null hypothesis that the mean change in VA from baseline at 26 weeks with LUMINOPIA is inferior to PATCHING by 0.0625 logMAR (i.e., 5/8 of one line) or more in favor of the alternative hypothesis that LUMINOPIA is non-inferior to PATCHING.

$$\mathbf{H_0: \mu_{PATCHING} - \mu_{LUMINOPIA} \leq -0.0625\ logMAR}$$ (LUMINOPIA inferior to PATCHING)

$$\mathbf{H_a: \mu_{PATCHING} - \mu_{LUMINOPIA} > -0.0625\ logMAR}$$ (LUMINOPIA *not* inferior to PATCHING)

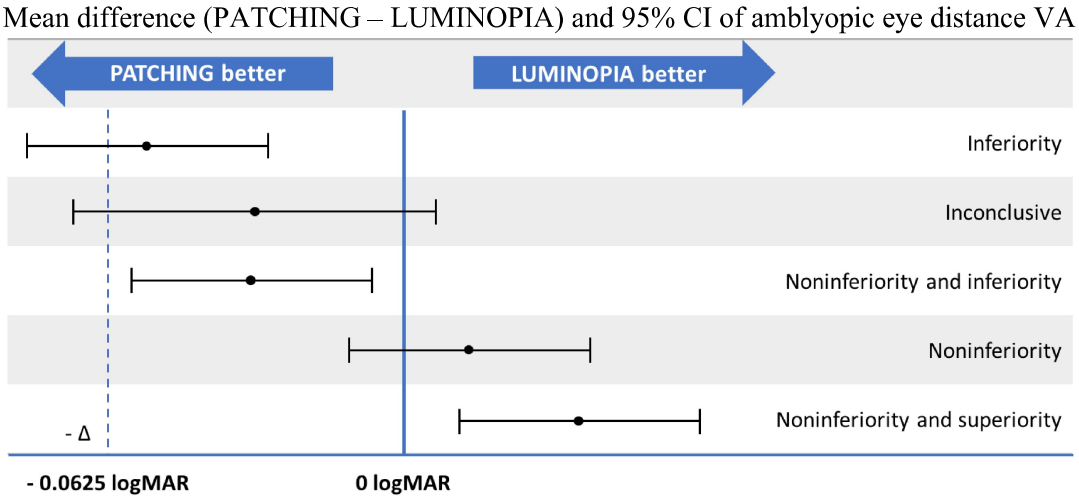
To represent the difference between treatment groups (PATCHING minus LUMINOPIA), a two-sided 95% confidence interval (CI) will be constructed. Since the LOWER limit of a two-sided

95% CI is equivalent to the LOWER limit of a one-sided 97.5% CI, this will allocate a significance level of 0.025 to be used in testing noninferiority.

Non-inferiority of LUMINOPIA to PATCHING will be declared if the LOWER limit of the two-sided 95% CI for the difference between treatment groups is greater than the non-inferiority limit of -0.0625 logMAR favoring PATCHING (Figure 1).

If non-inferiority is declared, a test of no difference (superiority test) in mean VA change at 26 weeks for LUMINOPIA compared to PATCHING will be conducted.

Figure 1. Depiction of Null and Alternative Hypotheses for Treatment Group Difference



3. Outcome Measures

3.1. Primary Efficacy Endpoint

- Change in amblyopic eye distance VA from baseline at 26 weeks

3.2. Secondary Efficacy Endpoints

- Change in child and proxy PedEyeQ Functional Vision domain scores from baseline at 26 weeks
- Change in child and proxy PedEyeQ Social domain scores from baseline at 13 weeks
- Change in child and proxy PedEyeQ Frustration/Worry domain scores from baseline at 13 weeks

3.3. Exploratory Efficacy Endpoints

- Change in binocular function score from baseline at 13 weeks and 26 weeks
- Change in amblyopic eye distance VA from baseline at 13 weeks

- Change in amblyopic eye distance VA from baseline over 26 weeks (area under the curve)
- Improvement of amblyopic eye distance VA by 2 or more lines (0.2 logMAR) from baseline at 13 and 26 weeks
- Resolution of amblyopia from baseline at 13 and 26 weeks
- Child, proxy, and parent Treatment Impact Questionnaire scores at 13 and 26 weeks

Post-primary Outcome Follow Up

- Change in binocular function score from 26 weeks at 39 weeks and 52 weeks
- Change in amblyopic eye distance VA from 26 weeks at 39 weeks and 52 weeks
- Improvement of amblyopic eye distance VA by 2 or more lines (0.2 logMAR) from 26 weeks at 39 and 52 weeks
- Resolution of amblyopia from 26 weeks at 39 and 52 weeks

4. Description of Statistical Methods

Analyses will follow the intent-to-treat principle (ITT); all participants will be analyzed according to their randomized treatment group, irrespective of adherence or compliance. However, a per protocol analysis will be performed for the primary outcome to check sensitivity of the results (section 5.2.1). The intent-to-treat analysis is considered primary and if the results of the per-protocol analysis and intent-to-treat give inconsistent results, exploratory analyses will be performed to evaluate possible factors contributing to the differences.

5. Primary Efficacy Outcome

5.1. Analysis of the Primary Endpoint

The primary endpoint, change in amblyopic eye logMAR distance VA from baseline at 26 weeks, is a continuous variable that will be analyzed using an analysis of covariance (ANCOVA) model to estimate the adjusted mean difference between PATCHING and LUMINOPIA. The model will adjust for baseline amblyopic-eye distance VA and prior treatment for amblyopia (glasses only vs other treatment in addition to glasses). The adjusted between-group mean difference, two-sided 95% CI and *p*-value for a test of no difference will be reported.

Participants who do not complete the 26-week visit will have their 26-week amblyopic eye distance VA imputed. Markov chain Monte Carlo multiple imputation with 100 imputations will be used to impute missing data; variables in the imputation model will include prior treatment for amblyopia and amblyopic-eye VA at baseline, 13, and 26 weeks. Imputation will be carried out separately for PATCHING and LUMINOPIA. Reasons for which a participant may not complete the 26-week visit are outlined in section 8, “Primary Estimand.”

Non-inferiority of LUMINOPIA compared to PATCHING will be declared if the LOWER limit of the two-sided 95% CI for the difference between treatment groups (PATCHING minus LUMINOPIA) in mean change in logMAR distance VA from baseline to 26 weeks is greater than (i.e., to the right of) the non-inferiority limit of -0.0625 logMAR favoring PATCHING. Note that the LOWER limit of a two-sided 95% confidence interval is equivalent to the lower limit of a one-sided 97.5% confidence interval.

If non-inferiority is demonstrated, superiority of LUMINOPIA over PATCHING will be declared if the LOWER limit of the 95% CI is greater than zero. Conversely, superiority of PATCHING over LUMINOPIA will be declared if the UPPER limit of the 95% CI is less than zero; in this scenario, LUMINOPIA will be both inferior to PATCHING and non-inferior to PATCHING with a margin of 0.0625 logMAR distance VA (Figure 1).

A boxplot showing changes in VA at 13 and 26 weeks by treatment group will be presented to aid in interpretation.

5.2. Sensitivity Analyses of the Primary Endpoint

To explore the robustness of the primary analysis, sensitivity analyses will be conducted and are outlined below. All sensitivity analyses will control for baseline amblyopic-eye VA and prior treatment for amblyopia.

5.2.1. Complete cases (Sensitivity Analysis #1)

The primary outcome will be analyzed without imputation of missing data.

5.2.2. Per protocol (Sensitivity Analysis #2)

The primary outcome will be analyzed using the same methods but participants who discontinued their assigned treatment or initiated non-randomized treatment will be excluded from both the imputation and analysis models.

5.2.3. Outliers (Sensitivity Analysis #3)

To ensure that statistical outliers do not have undue impact on analyses, the change in distance VA from baseline at 26 weeks will be modeled with robust regression using the Huber M-estimator instead of ANCOVA. Missing data will still be imputed using multiple imputation.

5.2.4. Confounding (Sensitivity Analysis #4)

To ensure that confounding does not affect study results, if an imbalance of baseline factors between treatment groups is observed, the primary analysis will be repeated, controlling for these potential confounders. The determination of a meaningful baseline imbalance will be based on clinical judgement.

5.2.5. Heteroscedasticity (Sensitivity Analysis #5)

To ensure that heteroscedasticity does not affect study results, a linear model that applies the residual-based estimator HC3 (Firores) will be used to estimate the empirical standard error. Note that this will be used as an alternative to model-based standard error which may be incorrect in the case of severe heteroscedasticity. The model will produce the adjusted between-group mean difference of the change in VA at 26-weeks and two-sided 95% CI. Missing data will still be imputed using multiple imputation.

6. Secondary Efficacy Outcomes

Secondary analyses will test the null hypothesis of no difference between treatment groups.

6.1. Pediatric Eye Disease Questionnaire (PedEyeQ)

Quality of life will be evaluated for children respondents aged 5 to 7 years in each treatment group using the PedEyeQ questionnaire. Additionally, the parent will answer on behalf of his/her child as a proxy for children 4 to 7 years of age. Scores on Functional Vision, Frustration/Worry, and Social domains will be assessed for both child and proxy at baseline as well as at the visit week indicated below (Table 2). Responses will be Rasch scored according to reference tables and standardized on a ratio scale ranging from 0 to 100.[1]

Table 2. Structure of the PedEyeQ Analysis: Domains and Respondents

		Domain			
Participant Age	**Respondent Level**	**Social (13 weeks)**	**Frustration/Worry (13 weeks)**	**Functional Vision (26 weeks)**	**Outcomes**
4-7 years	Proxy	1	1	1	3
5-7 years	Child	1	1	1	3
					Total = 6

Univariate analysis of covariance (ANCOVA) will be used to assess the difference between treatment groups across all domains and respondents (3 domains × 2 respondents = 6 outcomes) as shown in Table 2. Models will be adjusted for prior treatment for amblyopia and enrollment scores. The treatment effect will be summarized as a mean difference and 95% CI.

PedEyeQ scores will be imputed for any participants who did not respond to the questionnaires at the 13-week or 26-week visits. Markov chain Monte Carlo multiple imputation with 100 imputations will be used to impute missing scores for each domain. The imputation model will include 12 variables to represent the three domains (Functional Vision, Frustration/Worry, and Social) and two levels of respondents (proxy and child) at the enrollment and outcome visits, and will include a variable to indicate prior treatment for amblyopia. Imputation will be carried out separately for PATCHING and LUMINOPIA.

7. Visit Windows

To be included in analyses, visits must be completed within the specified visit windows (Table 3). Values from visits outside of the indicated analysis window will be considered missing data.

Table 3. Analysis Windows for Primary and Post-primary Outcome Follow-up Visits

Primary Outcome Follow-up (Randomization to 26-Week Visit)		
Visit	**Target Day Post-Randomization**	**Analysis Window around Target Day**
13-Week Visit	91 days	**± 5 weeks (56 days to 125 days)** (8 weeks to <18 weeks)
26-Week Visit Primary Outcome	182 days	**± 8 weeks (126 days to 238 days)** (18 weeks to 34 weeks)
Post-primary Outcome Follow-up (39-Week and 52-Week Visits)		
Visit	**Target Day Post 26-week Visit**	**Analysis Window around Target Day**
39-Week Visit	91 days	**± 5 weeks (56 days to 125 days)** (8 weeks to <18 weeks)
52-Week Visit	182 days	**± 8 weeks (126 days to 238 days)** (18 weeks to 34 weeks)

8. Primary Estimand

The primary outcome is derived from VA measurements at 26 weeks. The clinical question is whether the change in VA with LUMINOPIA is non-inferior to PATCHING with a non-inferiority margin of 0.0625 logMAR. The population-level summary measure is the mean difference comparing the LUMINOPIA and PATCHING groups.

Table 4 specifies the foreseen intercurrent events, whether data will be imputed after the event, and the strategy as defined in *E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials*. Data that are missing due to death, loss to follow-up, or participant withdrawal will be imputed based on observed VA measurements. This is consistent with a hypothetical scenario in which the intercurrent events do not occur and assumes that outcomes in those dying, lost to follow-up, and withdrawn resemble outcomes of those without missing data due to these events. Treatment discontinuation, treatment crossover, receipt of treatment for a condition other than amblyopia, and receipt of alternative treatment for amblyopia allow for continued observation of the outcome but might affect the outcome itself. By using observed data from participants who experience these events, we are adopting a

treatment policy strategy in which the value for the variable of interest is used regardless of whether the intercurrent event occurs. This strategy aligns with the ITT Principle.

Table 4. Intercurrent Events, Censoring, and Treatment Effects for the Primary Outcome

Event	Data Imputed After Event?	Strategy
Death	Yes	Hypothetical
Loss to follow-up	Yes	Hypothetical
Withdrawal	Yes	Hypothetical
Treatment discontinuation	No*	Treatment policy
Treatment crossover	No*	Treatment policy
Receipt of treatment for a condition other than amblyopia	No*	Treatment policy
Receipt of alternative treatment for amblyopia (i.e., not allowed per protocol)	No*	Treatment policy

* Observed data will be used for analyses.

9. Missing Data

In general, the procedure for handling missing data is outlined in each section. Where not otherwise specified, missing data will be excluded, and only complete cases will be analyzed.

10. Intervention Adherence

The number of participants stopping study treatment along with reasons for stopping treatment will be tabulated for each group.

10.1. Primary Outcome Follow-up (Randomization to 26-week Visit)

At 13 weeks and 26 weeks, the investigator will assess participant adherence to the assigned treatment. For each participant randomized to LUMINOPIA, the number of movie-watching hours will be categorized according to percentage of prescribed treatment time as 75-100%, 50-75%, or <50%. PATCHING calendar data will not be analyzed other than a subjective assessment by the investigator of adherence at 13, and 26-weeks as Excellent, Good, Fair, or Poor after review of calendar and interview with parent. The tabulation of data related to treatment adherence is intended for exploratory purposes only, and therefore formal comparisons between treatment groups will not be performed.

10.2. Post-primary Outcome Follow-up (39-week Visit and 52-week Visit)

At 39-weeks and 52-weeks for PATCHING randomized participants that choose to receive Luminopia, the total amount of time watching movies for each treatment period will be

categorized by percentage and tabulated as described in section 10.1. However, no formal analyses will be conducted using adherence data.

11. Protocol Adherence and Retention

Protocol deviations and visit completion rates (excluding participants who die before the end of the visit window) will be tabulated for each treatment group. A CONSORT[2] flow diagram will be constructed showing the following for each group:

- Numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome
- Losses and exclusions after randomization, together with reasons

The number of participants who were consented but not randomized will also be provided. Reasons for not randomizing will not be systematically collected.

12. Safety Analyses

The cumulative proportions of each of the following adverse events by treatment group will be assessed at the initial study phase (enrollment to 26 weeks) and during the LUMINOPIA treatment phase for those originally randomized to PATCHING (26 weeks to 52 weeks). During the initial study phase, the proportions will be compared statistically with Barnard's Unconditional Exact Test considering the number of participants per group as fixed. As type II error (false negative) is more of a concern than type I error (false positive) in safety analyses, we will use $p \leq 0.05$, without adjustment for multiplicity, to define statistical significance in all safety analyses. It is noted that this study is not powered for safety analyses and that absence of a significant effect cannot be taken as evidence that a difference does not exist. The proportion of adverse events occurring during the LUMINOPIA treatment phase for original PATCHING participants will be tabulated.

- Worsening of best-corrected fellow-eye distance VA of 0.2 logMAR or more
- New onset strabismus >5 Δ by SPCT in participants with no strabismus at baseline
- Strabismus >10 Δ by SPCT in participants with strabismus at baseline
- Parental report of diplopia occurring more than once per week
- Skin irritation
- Headache
- Eyestrain
- Dizziness
- Night terrors
- Eye twitching
- Facial redness

The PEDIG DSMC will review safety data tabulated by treatment group at each of its semi-annual meetings and can request formal statistical comparison of any safety outcome at any time if they have cause for concern.

13. Baseline Descriptive Statistics

Demographic and clinical characteristics at enrollment will be tabulated by randomized treatment group, and summary statistics appropriate to their distributions will be reported. Variables will include participant age, sex, race, ethnicity, prior treatment for amblyopia, ocular alignment, refractive error, binocular function, amblyopic-eye VA, fellow-eye VA, and IOD.

14. Planned Interim Analyses

There is no plan for formal interim analyses. The Data and Safety Monitoring Committee (DSMC) will review tabulated and graphical displays of interim safety data at approximately 6-month intervals and will have the option to recommend stopping the study.

15. Subgroup Analyses

Subgroup analyses will be used to evaluate potential effect modification (interaction) between the randomized treatment and each pre-specified baseline variable. These analyses will be considered exploratory. Missing data will be imputed like the primary analyses except that the subgroup factors of interest, specified below, will be included in the imputation model, which will be stratified by treatment group. Within-subgroup mean differences for the treatment effects with 95% CIs will be estimated for each subgroup by adding an interaction term to the primary analysis models. Results will be presented as forest plots; *p*-values will not be presented.

The baseline factors to be evaluated in pre-planned exploratory subgroup analyses include:

- Amblyopic-eye distance VA
  - Moderate impairment (20/40 to 20/80, 72 to 53 letters)
  - Severe impairment (20/100 to 20/200, 52 to 33 letters)
- Type of amblyopia
  - Strabismic only
  - Anisometropic only
  - Both strabismic and anisometropic
- Ocular alignment at near
  - 0 < Heterotropia ≤ 5 Δ by SPCT
  - None by SPCT
- Prior treatment for amblyopia
  - Yes (prior amblyopia treatment and glasses)
    - ≥ 1 year of treatment
    - < 1 year of treatment
  - No (glasses only)
- Age
  - 4 to 5 years
  - >5 to 7 years

- Sex
  - Male
  - Female
- Race and Ethnicity
  - White and non-Hispanic
  - Non-white and/or Hispanic
- Binocular function at near
  - Randot Preschool Stereoacuity (1.6 to 2.9 log seconds of arc)
  - Randot Butterfly Stereoacuity (3.3 log seconds of arc)
  - Worth 4-Shape Stereoacuity (Fusion)
  - Nil stereoacuity

There are no data to suggest that the treatment effect will vary by sex, race, or ethnicity. However, each of these factors will be evaluated in exploratory subgroup analyses as mandated by National Institutes of Health (NIH) guidelines.

If there is insufficient sample size in a given subgroup (N < 20), the cut points for continuous measures may be reconfigured to correspond to the observed distribution of values, possibly using the median to determine the cut point.

16. Multiple Comparison/Multiplicity

For the primary outcome, a 95% confidence interval for the treatment group difference will be constructed and used to evaluate the primary hypothesis of non-inferiority and also the possibility of superiority or inferiority (Figure 1).[3]

For the PedEyeQ questionnaire secondary outcomes, the adaptive false discovery rate (FDR) method with two-stage testing will control the FDR at 5% to adjust *p*-values and confidence intervals for multiplicity.[4]

17. Exploratory analyses

Exploratory analyses will test the null hypothesis of no difference between treatment groups. *p*-values and CIs will not be adjusted for multiplicity.

17.1. Mean Change in Distance VA at 13 Weeks

Change in amblyopic eye VA from baseline to 13 weeks is a continuous outcome. Analysis, including imputation of missing data, will mirror the primary outcome; the model will be adjusted for baseline amblyopic-eye distance VA and prior treatment for amblyopia.

17.2. Mean Change in Distance VA over 26 weeks (area under the curve)

The change in amblyopic eye distance VA from baseline over 26 weeks (area under the curve) will be calculated for each participant with the trapezoidal rule using the following formula:

$$AUC = \sum_{i=1}^{n} \left(\frac{V_i + V_{i+1}}{2} \times d \right)$$

Where V_i is the VA measured at the i^{th} visit, d is the number of days between visits i and $i+1$ (based on the target day, not the actual date of completion), and n is the number of outcome visits included in the analysis. This analysis has $n = 3$ as it will include visits at baseline, 13 and 26 weeks; note that change in VA is 0 at baseline for all participants. For presentation, the AUC will be divided by the number of days between baseline and the 26-week visit based on the target day (i.e., 182 days) so that the value shown will have units of letters rather than letter·days. The area under the curve can be interpreted as a weighted average of change in VA over 26 weeks with weights proportional to the time between visits.

The area under the curve will be calculated after imputation of missing data. The analysis, including imputation of missing data, will mirror the primary outcome; the model will be adjusted for baseline amblyopic-eye distance VA and prior treatment for amblyopia. A boxplot showing AUC for each treatment group over 26 weeks will be constructed.

17.3. Improvement of Amblyopic-eye Distance VA by 2 or More Lines at 13 and 26 weeks

Improvement of amblyopic-eye distance VA of 2 or more lines (reduction of 0.2 logMAR) at 13 and 26 weeks, respectively, will be analyzed as binary outcomes. For each time point, the proportions with improvement ≥ 2 lines and likelihood-ratio 95% CIs for each treatment group will be calculated with logistic regression, adjusting for baseline amblyopic-eye VA and prior treatment for amblyopia.

The adjusted risk difference will be calculated using logistic regression with conditional standardization. The delta method will be implemented to construct a 95% CI on the risk difference[5] and the model-based two-sided p-value will be reported. Missing data will be imputed as described for the primary outcome.

17.4. Binocular Function Change at 13 Weeks and 26 Weeks

The change in binocular function score from enrollment to the 13 and 26-week visit is an ordinal outcome (Table 5). Components of binocularity include results from the following 3 tests: Randot Preschool Stereoacuity (RPS), Random Dot Butterfly, and Preschool Worth 4-Shape (W4S) at near. These tests will create a composite ordinal score of binocular function with 9 levels.[6] .

The difference between LUMINOPIA and PATCHING for the change in binocular function score from baseline to 13 and 26 weeks will be evaluated with the nonparametric Wilcoxon Rank-Sum test. The difference between groups will be estimated using the Hodges-Lehmann estimator with 95% CI. Analyses for binocular function score will be limited to complete case data at each respective outcome visit (13 weeks or 26 weeks).

Table 5. Levels of Binocular Function as Seconds of Arc on Near Stereoacuity Tests

Seconds of arc	Log_{10} seconds of arc	Stereoacuity Test (Measured at Near)	Binocularity Score (Ordinal)
40	1.60	Randot Preschool	1
60	1.78		2
100	2.00		3
200	2.30		4
400	2.60		5
800	2.90		6
2000	3.30	Randot Butterfly	7
n/a (Fusion)	n/a	Worth 4-Shape	8
n/a (Nil)	n/a		9
Missing	Missing (not included in analyses)		Missing

17.4.1. Binocular Function Sensitivity Analysis

In a sensitivity analysis, the difference between LUMINOPIA and PATCHING on the change in binocular function score at the 13- and 26-week visits, respectively, will be evaluated with parametric methods to allow adjustment for baseline binocular function score and prior treatment for amblyopia; such methods will also permit imputation of missing data. For this analysis, fusion and nil will be arbitrarily assigned values of 4000 and 8000 arcsec (each double the previous level), respectively. Using ANCOVA, the adjusted mean difference and 95% CI in log_{10} arcsec between the treatment groups will be presented. Missing binocular function data will be imputed using fully conditional specification (FCS) with logistic regression (cumulative logit) in 100 imputations.[7,8] Imputation will be carried out separately for PATCHING and LUMINOPIA. Variables in the imputation models will include prior treatment for amblyopia and binocular function scores at baseline, 13, and 26 weeks. This method of imputation is being used instead of Markov chain Monte Carlo so that the imputed values are consistent with the possible values of log_{10} arcsec from the binocular function score.

17.5. Resolution of Amblyopia at 13 weeks and 26 weeks

Resolution of amblyopia is defined as $\leq$0 lines IOD and fellow eye VA no worse than 1 line (0.10 logMAR) below baseline. The cumulative probability of amblyopia resolution at 13 and 26 weeks will be calculated using Cox proportional hazards regression with direct adjustment. Event times will be grouped based on the target day of the visit; all 13-week visits will have time set to 91 days and all 26-week visits will have time set to 182 days. Ties will be modeled using the exact method. The IOD in VA at baseline will be included as a covariate because the outcome is a function of the IOD; the model will also control for prior treatment of amblyopia.

Participants who are lost to follow up will be censored on the day of the last completed visit. For each visit, the rate of resolution (estimated using the survivor function) and 95% CI will be presented for each group along with the difference in rates and 95% CI, and *p*-value (based on a Z test). To aid in interpretation, Kaplan-Meier curves will be plotted and the number of participants at risk will be shown by visit.

17.6. Treatment Impact Questionnaire

The Treatment Impact Questionnaire (TIQ) will be used as a quantitative measure to evaluate child, proxy, and parent opinions regarding the burdens and impact of the randomized treatment at 13 weeks and 26 weeks (as questions for the child – the Child TIQ, for the parent about the child – the Proxy TIQ, and the parent themselves – the Parent TIQ). The Child-TIQ, Proxy-TIQ, and Parent-TIQ will undergo separate factor analysis to determine the number of domains for each TIQ. Each domain will be refined through the evaluation of misfitting items and will then be Rasch scored. Note that because the TIQ is not administered at baseline (because treatment has not been started), there will be no adjustment for baseline score in any analysis.

Additional methods to score and analyze the Treatment Impact Questionnaire will be detailed in a separate SAP.

17.7. Post Primary Outcome Follow-up

Participants that were randomized to PATCHING who have 1 line or more (0.10 logMAR) IOD residual amblyopia will be offered dichoptic treatment with LUMINOPIA after 26 weeks. These participants will be followed at 39 weeks and 52 weeks to evaluate safety and efficacy. The same safety, binocular function, and VA outcomes evaluated at 13 and 26 weeks will be evaluated at 39 and 52 weeks with 26 weeks considered the baseline visit for the extended follow-up. Analyses will be limited to PATCHING-randomized participants who agree to be treated with LUMINOPIA after completing the 26-week primary outcome visit.

17.7.1. Binocular Function Change at 39 and 52 Weeks

The change in binocular function score at 39 and 52 weeks is an ordinal outcome that will be analyzed as described in section 17.4. The change in binocular function score from 26 weeks to 39 and 52 weeks will be summarized using the median and interquartile range and the one-sample Hodges-Lehmann estimator with 95% confidence interval. Analyses for binocular function score will be limited to complete case data at each respective outcome visit (39 weeks or 52 weeks).

As a sensitivity analysis, mean change in binocular function score ($\log_{10}$ arcsec) will be estimated using ANCOVA with adjustment for 26-week score. Missing binocular function data will be imputed using fully conditional specification (FCS) with logistic regression (cumulative logit) in 100 imputations. Variables in the imputation models will include binocular function scores at 26, 39, and 52 weeks.

17.7.2. Mean Change in Distance VA at 39 and 52 Weeks

The mean difference and 95% CI of the change in VA from 26 weeks to 39 weeks, and from 26 weeks to 52 weeks, will be calculated for these participants using ANCOVA to adjust for 26-week primary outcome VA. Missing data will be imputed using Markov chain Monte Carlo multiple imputation with 100 imputations. Variables in the imputation model will include VA measured at 26, 39, and 52 weeks.

17.7.3. Improvement of Amblyopic-eye Distance VA by 2 or More Lines at 39 and 52 weeks

Improvement of amblyopic-eye distance VA of 2 or more lines (reduction of 0.2 logMAR) at 39 and 52 weeks, respectively, will be analyzed as binary outcomes. For each time point, the proportions with improvement ≥ 2 lines and likelihood-ratio 95% confidence intervals will be calculated in a logistic regression with intercept only. Missing data will be imputed as described in section 17.7.2.

17.7.4. Resolution of Amblyopia at 39 weeks and 52 weeks

For PATCHING-randomized participants treated with LUMINOPIA during the extended follow-up, the cumulative probability of amblyopia resolution ($\leq$0 lines IOD and fellow eye VA no worse than 1 line below baseline) at 39 weeks and 52 weeks will be calculated using Cox proportional hazards regression with intercept only. Event times will be grouped based on the target day of the visit; all 39-week visits will have time set to 273 days and all 52-week visits will have time set to 364 days. Ties will be modeled using the exact method. Participants who are lost to follow up will be censored on the day of the last completed visit. For each visit, rate of resolution (estimated using the survivor function) and 95% CI will be presented.

18. Assumptions

All model assumptions including linearity, normality, and homoscedasticity will be verified using graphical methods. If seriously violated, then transformations, robust methods, or nonparametric methods may be used instead.

19. References

1. Hatt SR, Leske DA, Castañeda YS, et al. Development of pediatric eye questionnaires for children with eye disease. *Am J Ophthalmol*. 2019;200:201-217.
2. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18.
3. US Food and Drug Administration. *Non-inferiority clinical trials to establish effectiveness: Guidance for industry* 2016. FDA-2010-D-0075. Accessed June 28, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/non-inferiority-clinical-trials
4. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*. 2006;93(3):491-507.

5. Austin PC. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. *Journal of clinical epidemiology*. 2010;63(1):2-6.
6. Webber AL, Wood JM, Thompson B, E. BE. From suppression to stereoacuity: a composite binocular function score for clinical research. *Ophthalmic Physiol Opt*. 2019;39:53-62.
7. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-99.
8. Berglund PA. Multiple Imputation Using the Fully Conditional Specification Method : A Comparison of SAS ® , Stata , IVEware , and. 2015: