AMBLYOPIA TREATMENT STUDY (ATS24)

A Randomized Trial of Dichoptic Treatment for Amblyopia in Children 8 to 12 Years of Age

Statistical Analysis Plan

Version 1.0

Version History

This SAP was written with reference to protocol version 1.0. If the protocol is subsequently updated, then this SAP will be reviewed to ensure consistency with the new protocol. The SAP will not be revised unless the protocol changes require modification of the analyses.

Version	Protocol Version	Author	Approver	Effective Date	Study Stage
1.0	1.0	Desirae Sutherland	Wesley Beaulieu	20 Jun 2024	First participant has not yet been enrolled.

Version	Revision Description
1.0	Original Version

Approvals

Role	Digital Signature or Handwritten Signature/Date
Author (Statistician)	Desirae Sutherland Digitally signed by Desirae Sutherland DN: cn=Desirae Sutherland ou=North Wing Reason: I am the author of this document Location: Date: 2024-07-23 14:19-04:00
Approver (Senior Statistician)	Wesley Beaulieu I agree to the terms defined by the placement of my signature in this document 2024-07-23 14:05-04:00

1. Study Overview

This document outlines the statistical analyses to be performed for the ATS24 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 18 weeks in participants randomized to treatment with the Luminopia One headset (1 hour per day of watching dichoptic movies, 6 days per week), Vivid Vision (25 minutes per day of playing dichoptic games, 6 days per week), or continued optical correction (full-time, if needed).

The aforementioned treatment regimens of Luminopia One, Vivid Vision, and continued optical correction shall subsequently be referred to as LUMINOPIA, VIVID VISION, and GLASSES, respectively. Note that although the name refers to glasses wear, it is acknowledged that participants randomized to GLASSES may be treated with either 1) glasses, 2) contact lenses, or 3) no optical correction; this will be in accordance with the participant's current prescribed treatment at the time of enrollment. Randomization will be performed using a permutated block design stratified by VA in the amblyopic eye as moderate (20/40 to 20/80 [72 to 53 letters]) versus severe (20/100 to 20/200 [52 to 33 letters]).

Approximately 252 participants will be enrolled and randomized in a 1:1:1 ratio to either LUMINOPIA, VIVID VISION, or GLASSES treatment (84 per group). This sample size was calculated assuming a mean difference of 3.75 letters between each active treatment group (LUMINOPIA and VIVID VISION) versus the control group (GLASSES), a common standard deviation of 7 letters, 5% alpha, 90% power, and 10% loss to follow-up. The study protocol provides further details on the sample size calculations.

At the 18-week primary outcome visit, participants who were randomly assigned to receive GLASSES treatment and have an interocular difference (IOD) of at least 1 line (≥ 5 letters) will be offered dichoptic treatment with either LUMINOPIA or VIVID VISION. Participants who agree to the extended follow-up will be randomized to one of the dichoptic treatments and will return for visits at 27 and 36 weeks post-randomization. Otherwise, for all other participants the study will end at 18 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 visual acuity (VA) (measured in letters) from randomization at 18 weeks. Change in letters will be calculated as [outcome VA] – [randomization VA] such that a positive change indicates improvement in VA letter scores, and a negative change indicates worsening.

The study is designed to test two, two-sided superiority hypotheses, each designed to evaluate whether the mean change in VA from baseline at 18 weeks with GLASSES is significantly different than either dichoptic treatment (with LUMINOPIA or with VIVID VISION):

Superiority Test 1:

H_0: $\mu_{\text{LUMINOPIA}}-\mu_{\text{GLASSES}} = 0$ letters

H_a: $\mu_{\text{LUMINOPIA}}-\mu_{\text{GLASSES}} \neq 0$ letters

Superiority Test 2:

H_0: $\mu_{\text{VIVID VISION}}-\mu_{\text{GLASSES}} = 0$ letters

H_a: $\mu_{\text{VIVID VISION}}-\mu_{\text{GLASSES}} \neq 0$ letters

For each hypothesis, the difference in mean VA change at 18 weeks between treatment groups (LUMINOPIA minus GLASSES and VIVID VISION minus GLASSES), and a two-sided 95% confidence interval (CI) for the difference will be constructed. Each hypothesis will be tested independently, utilizing GLASSES as a shared control group. As such, each will be conducted with an alpha level of 0.05 without formal adjustment to the familywise error rate (see section 16, "Multiple Comparisons/Multiplicity," for details).

If mean 18-week change in VA with LUMINOPIA and VIVID VISION are both superior to GLASSES, then a hypothesis test will evaluate whether there is a difference between active treatments with no adjustment to alpha (per the fixed sequence method). The difference between treatment groups (LUMINOPIA minus VIVID VISION), and a two-sided 95% CI for the difference will be constructed, with *p*-value.

Superiority Test 3:

H_0: $\mu_{\text{LUMINOPIA}}-\mu_{\text{VIVID VISION}} = 0$ letters

H_a: $\mu_{\text{LUMINOPIA}}-\mu_{\text{VIVID VISION}} \neq 0$ letters

However, if the mean 18-week change in VA with either LUMINOPIA or VIVID VISION is not significantly different than GLASSES, then the difference between active treatment groups will be considered exploratory and a *p*-value will not be reported.

3. Outcome Measures

3.1. Primary Efficacy Endpoint

- Change in amblyopic eye distance VA from baseline at 18 weeks

3.2. Secondary Efficacy Endpoints

- Change in child and proxy PedEyeQ Functional Vision domain scores from baseline at 18 weeks
- Change in child and proxy PedEyeQ Social domain scores from baseline at 9 weeks
- Change in child and proxy PedEyeQ Frustration/Worry domain scores from baseline at 9 weeks

3.3. Exploratory Efficacy Endpoints

- Change in amblyopic eye distance VA from baseline at 9 weeks
- Change in amblyopic eye distance VA from baseline over 18 weeks (area under the curve)
- Improvement of amblyopic eye distance VA by 2 or more lines (≥ 10 letters) from baseline at 9 and 18 weeks
- Change in binocular function score from baseline at 9 weeks and 18 weeks
- Resolution of amblyopia from baseline at 9 and 18 weeks
- Child, proxy, and parent Treatment Impact Questionnaire scores at 9 and 18 weeks

Post-primary Outcome Follow Up

- Change in amblyopic eye distance VA from 18 weeks at 27 weeks and 36 weeks
- Change in amblyopic eye distance VA from 18 weeks to 36 weeks (area under the curve)
- Improvement of amblyopic eye distance VA by 2 or more lines (≥ 10 letters) from 18 weeks at 27 and 36 weeks
- Resolution of amblyopia from 18 weeks at 27 and 36 weeks
- Change in binocular function score from 18 weeks at 27 weeks and 36 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 distance VA letter score from baseline at 18 weeks, is a continuous variable that will be analyzed using an analysis of covariance (ANCOVA) model to estimate the adjusted mean difference between GLASSES and LUMINOPIA, as well as between GLASSES and VIVID VISION. The model will adjust for baseline amblyopic-eye distance VA, with multiple imputation for missing data. The adjusted between-group mean differences and two-sided 95% CIs and *p*-values will be reported.

Participants who do not complete the 18-week visit will have their 18-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 amblyopic-eye VA at baseline, 9, and 18 weeks. Imputation will be carried out separately for each treatment group.[1] Reasons for which a participant may not complete the 18-week visit are outlined in section 8, "Primary Estimand."

If both dichoptic treatments are declared superior to GLASSES (p-value for null hypothesis that mean difference is zero $\leq .05$), then a test of superiority between LUMINOPIA and VIVID VISION will be performed without further adjustment for multiplicity (see section 16). The same analysis approach will be used. If either of the dichoptic treatments are not declared superior to GLASSES, then LUMINOPIA and VIVID VISION will still be compared, however, the comparison will be considered exploratory, and a p-value will not be presented. Separate ANCOVA models will be used for each.

A boxplot showing changes in VA at 9 and 18 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.

5.2.1. Per protocol (Sensitivity Analysis #1)

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. Missing data will be imputed using multiple imputation.

5.2.2. Complete cases (Sensitivity Analysis #2)

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

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 18 weeks will be modeled with robust regression using the Huber M-estimator instead of ANCOVA. Missing data will 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. Missing data will be imputed using multiple imputation.

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 control for baseline amblyopic-eye VA to produce the adjusted between-group mean difference of the change in VA at 18 weeks and two-sided 95% CI. Missing data will 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)

The effect of amblyopia on quality of life will be evaluated using the PedEyeQ questionnaire. Scores on Functional Vision, Frustration/Worry, and Social domains will be assessed for both child and proxy (parent answering on behalf of the child) respondents at baseline as well as at the visit week indicated below (Table 1). The responses of child and proxy will be Rasch scored according to reference tables and standardized on a ratio scale ranging from 0 to 100.[2]

Table 1. Structure of the PedEyeQ Analysis: Domains and Respondents

Respondent Level	Domain			Outcomes
	Social (9 weeks)	Frustration/Worry (9 weeks)	Functional Vision (18 weeks)	
Child	1	1	1	3
Proxy	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 1.The PedEyeQ outcomes will be analyzed in separate ANCOVA models for each treatment group comparison (6 outcomes x 3 treatment comparisons) and will be adjusted for 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 9-week or 18-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 (child and proxy) at the enrollment and outcome visits. Imputation will be carried out separately for LUMINOPIA, VIVID VISION, and GLASSES.

7. Visit Windows

To be included in analyses, visits must be completed within the specified visit windows (Table 2). Values from visits outside of the indicated analysis window will be considered missing data.

Table 2. Analysis Windows for Primary and Post-primary Outcome Follow-up Visits

Primary Outcome Follow-up (Randomization to 18-Week Visit)		
Visit	**Target Day Post-Randomization**	**Allowable Window Post-Randomization**
9-Week Office Visit	63 days	42 days to 104 days
18-Week Primary Outcome	126 days	105 days to 168 days
Post-primary Outcome Follow-up (27-Week and 36-Week Visits)		
Visit	**Target Day Post-18-week***	**Allowable Window Post-18-week***
27-Week Office Visit	63 days	42 days to 104 days
36-Week Office Visit	126 days	105 days to 168 days

** The target day and allowable window for post-primary outcome follow-up will be calculated from the date the 18-week visit was completed.*

8. Primary Estimand

The primary outcome is derived from VA measurements at 18 weeks. The clinical question is whether the change in VA with LUMINOPIA or VIVID VISION is significantly different than with GLASSES. The population-level summary measure is the mean difference comparing the LUMINOPIA and VIVID VISION groups with the GLASSES group.

Table 3 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 3. 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 18-week Visit)

At 9 weeks and 18 weeks, the investigator will assess participant adherence to the assigned treatment. For each participant randomized to LUMINOPIA or VIVID VISION, the number of dichoptic treatment hours will be categorized according to percentage of prescribed treatment time as 75-100%, 50-75%, or <50%. Calendar data for the GLASSES group will not be analyzed other than a subjective assessment by the investigator of adherence at 9 and 18 weeks as Excellent, Good, Fair, or Poor after review of the calendar and interview with the 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 (27-week Visit and 36-week Visit)

Participants assigned to GLASSES who choose to continue follow-up after the primary outcome visit will be randomized to either LUMINOPIA or VIVID VISION. At 27 weeks and 36 weeks, the total amount of time utilizing the assigned dichoptic therapy 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[3] 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 18 weeks) and during the post-primary outcome phase (18 weeks to 36 weeks) for GLASSES participants who elect to receive a dichoptic therapy. During the initial study phase, the proportions will be compared statistically between all three groups using Fisher's Exact Test; if the p-value is $\leq .05$, then pairwise tests will be performed without further adjustment for multiplicity. 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 the study is not specifically powered to detect differences in safety outcomes and that the absence of a significant difference should not be viewed as evidence for the absence of a true difference. The proportion of adverse events occurring during the post-primary phase for original GLASSES participants will be tabulated within each dichoptic treatment group (LUMINOPIA or VIVID VISION) without formal statistical comparison.

- Worsening of best-corrected fellow-eye distance VA of 2 lines (10 letters) 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, ocular alignment, refractive error, binocular function, amblyopic-eye VA, fellow-eye VA, and IOD.

14. Planned Interim Analyses

The study may be discontinued by the Steering Committee (with approval of DSMC) prior to the preplanned completion of follow-up for all study participants. No formal analyses and/or guidelines for stopping the study for futility or efficacy are pre-specified.

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 $\leq 5\ \Delta$ by SPCT
  - No heterotropia by SPCT
- Prior treatment for amblyopia
  - Yes (prior amblyopia treatment and glasses)
    - ≥ 1 year of treatment
    - < 1 year of treatment
  - No (glasses only)

- Age
  - 8 to <10 years
  - 10 to <13 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 Fusion
  - Worth 4-Shape No Fusion (Suppression or Diplopia)

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, two tests of superiority for 18-week mean change in amblyopic eye distance VA will be conducted: LUMINOPIA vs GLASSES and VIVID VISION vs GLASSES. The tests will be performed independently in separate ANCOVA models, and each will be conducted with an alpha level of 0.05.

Although two pairwise comparisons are being evaluated, there will be no formal adjustment to the familywise error rate; because the main objective of this trial is to compare two dichoptic treatments with different mechanisms of action with a shared control group, and not one another, an adjustment (e.g., Bonferroni) is not needed.[4-6] The risk of a false positive finding with this approach is lower than if each hypothesis were evaluated in two separate studies with different control groups. The same logic applies to secondary, exploratory, safety, and subgroup analyses.

For the comparison of LUMINOPIA vs VIVID VISION, the familywise error rate will be controlled with a hierarchical (i.e., fixed sequence) approach. If the null hypotheses for LUMINOPIA vs GLASSES and VIVID VISION vs GLASSES are rejected, then LUMINOPIA and VIVID VISION will be compared in a separate ANCOVA model without further adjustment to the type 1 error rate.[7] If, however, both null hypotheses are not rejected, then the comparison of LUMINOPIA vs VIVID VISION will be considered exploratory and a *p*-value will not be presented. It is noted that for the comparison of LUMINOPIA vs VIVID VISION, the absence of

a statistically significant difference cannot rule out the presence of a clinically meaningful difference between active treatment groups. The study is powered assuming a difference in VA between treatments as small as 3.75 letters with a standard deviation of 7.0 letters. This hierarchical approach for the comparison of LUMINOPIA vs VIVID VISION will be employed in all primary, secondary, and exploratory analyses.

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 CIs for multiplicity.[8] Each treatment comparison (LUMINOPIA vs GLASSES, VIVID VISION vs GLASSES, and LUMINOPIA vs VIVID VISION) will be conducted separately and considered a separate family of tests. As such, the PedEyeQ outcomes will be modeled separately with ANCOVA for each treatment group comparison (6 outcomes x 3 treatment comparisons). The hierarchical testing approach will be employed for each domain and level of respondent. For example, the *p*-value for the comparison of LUMINOPIA vs VIVID VISION on the Child Social domain will only be reported if both null hypotheses are rejected when testing LUMINOPIA vs GLASSES and VIVID VISION vs GLASSES on the Child Social domain.

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 9 Weeks

Change in amblyopic-eye VA from baseline to 9 weeks is a continuous outcome. Analyses, including imputation of missing data, will mirror the primary outcome.

17.2. Mean Change in Distance VA over 18 weeks (area under the curve)

The change in amblyopic-eye distance VA from baseline over 18 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, 9 and 18 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 18-week visit based on the target day (i.e., 126 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 18 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 and adjustment for baseline amblyopic-eye distance VA, will mirror the primary outcome. A boxplot showing AUC for each treatment group over 18 weeks will be constructed.

17.3. Improvement of Amblyopic-eye Distance VA by ≥ 2 Lines at 9 and 18 weeks

Improvement of amblyopic-eye distance VA of 2 or more lines (≥ 10 letters) at 9 and 18 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.

The risk difference will be calculated using logistic regression with conditional standardization, centering on the mean amblyopic-eye VA at baseline. The delta method will be implemented to construct a 95% CI on the risk difference[9] 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 9 Weeks and 18 Weeks

The change in binocular function score from enrollment to the 9- and 18-week visits is an ordinal outcome (Table 4). Components of binocularity include results from the following 3 tests: Randot Preschool Stereoacuity (RPS), Random Dot Butterfly, and Preschool Worth 4-Shape at near. These tests will create a composite ordinal score of binocular function with 9 levels.[10]

The differences between treatment groups for the change in binocularity from baseline to 9 and 18 weeks will be evaluated with the nonparametric Wilcoxon Rank-Sum test. Differences 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 (9 weeks or 18 weeks).

Table 4. Levels of Binocular Function as Seconds of Arc on Near Stereoacuity Tests

Stereoacuity Test (Measured at Near)	Seconds of Arc	Log_{10} Seconds of Arc	Binocularity Score (Ordinal)
Randot Preschool	40	1.60	1
	60	1.78	2
	100	2.00	3
	200	2.30	4
	400	2.60	5
	800	2.90	6
Randot Butterfly	2000	3.30	7
Worth 4-Shape	Fusion	N/A	8
	No Fusion (Suppression or Diplopia)	N/A	9

17.4.1. Binocular Function Sensitivity Analysis

In a sensitivity analysis, the difference between each treatment group on the change in binocular function score at the 9- and 18-week visits, respectively, will be evaluated with parametric methods to allow adjustment for baseline binocular function score and imputation of missing data. For this analysis, values on Worth 4-Shape will be arbitrarily assigned as 4000 arcsec for Fusion and 8000 arcsec for No Fusion (i.e., Suppression or Diplopia). Using ANCOVA, the baseline-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.[11,12] Imputation will be carried out separately for each treatment group. Variables in the imputation models will include binocular function scores at baseline, 9, and 18 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 9 weeks and 18 weeks

Resolution of amblyopia is defined as $\leq$0 lines IOD and fellow-eye VA no worse than 1 line (5 letters) below baseline. The cumulative probability of amblyopia resolution at 9 and 18 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 9-week visits will have time set to 63 days and all 18-week visits will have time set to 126 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. 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 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 9 weeks and 18 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 independently of treatment assignment.

For each domain, mean treatment group scores will be compared using a t-test, and a 95% CI on the difference between groups will be calculated. However, if the data are severely non-normal in distribution, the treatment groups will be compared using Wilcoxon Rank-Sum and the difference estimated using the Hodges-Lehmann estimator with 95% CI. 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 regarding factor analysis and Rasch scoring for the Treatment Impact Questionnaire will be detailed in a separate analysis plan.

17.7. Post Primary Outcome Follow-up

Participants who were randomized to GLASSES who have 1 line or more (≥ 5 letters) IOD residual amblyopia will be offered dichoptic treatment with either LUMINOPIA or VIVID VISION after 18 weeks. These participants will be randomized to one of the dichoptic treatments and will have visits at 27 weeks and 36 weeks to evaluate safety and efficacy. The same safety, binocular function, and VA outcomes evaluated at 9 and 18 weeks will be evaluated at 27 and 36 weeks with 18 weeks considered the baseline visit for the extended follow-up. Within-group outcomes with 95% CIs will be presented. Because the study is not powered for this phase, between-group comparisons will not be conducted.

17.7.1. Mean Change in Distance VA at 27 and 36 Weeks

The mean difference and 95% CI of the change in VA from 18 weeks to 27 weeks, and from 18 weeks to 36 weeks, will be calculated for participants in each group using ANCOVA to adjust for 18-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 18, 27, and 36 weeks. Imputation will be carried out separately for participants randomized to LUMINOPIA and VIVID VISION.

17.7.2. Mean Change in Distance VA from 18 Weeks to 36 Weeks (area under the curve)

The change in amblyopic-eye distance VA from 18 weeks to 36 weeks (area under the curve) will be calculated for each participant with the trapezoidal rule and analyzed as described in section 17.2. The area under the curve will be calculated after imputation of missing data. Missing data will be imputed as described in section 17.7.1.

17.7.3. Improvement of Amblyopic-eye Distance VA by ≥ 2 Lines at 27 and 36 weeks

Improvement of amblyopic-eye distance VA of 2 or more lines (≥ 10 letters) from 18 weeks at 27 and 36 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 in a logistic regression, adjusting for 18-week amblyopic-eye VA. Missing data will be imputed as described in section 17.7.1.

17.7.4. Resolution of Amblyopia at 27 weeks and 36 weeks

For each treatment group, the cumulative probability of amblyopia resolution (≤0 lines IOD and fellow eye VA no worse than 1 line [5 letters] below baseline) at 27 weeks and 36 weeks will be calculated using Cox proportional hazards regression with direct adjustment. Baseline IOD will

be included as a covariate. Event times will be grouped based on the target day of the visit; all 27-week visits will have time set to 189 days and all 36-week visits will have time set to 252 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, the rate of resolution (estimated using the survivor function) and 95% CI will be presented by treatment group.

17.7.5. Binocular Function Change at 27 and 36 Weeks

The change in binocular function score at 27 and 36 weeks is an ordinal outcome that will be analyzed as described in section 17.4. The change in binocular function score from 18 weeks to 27 and 36 weeks will be summarized for each treatment group using the median and interquartile range and the one-sample Hodges-Lehmann estimator with 95% CI. Analyses for binocular function score will be limited to complete case data at each respective outcome visit (27 weeks or 36 weeks).

As a sensitivity analysis, mean change in binocular function score ($\log_{10}$ arcsec) will be estimated for each group using ANCOVA with adjustment for 18-week score . Missing binocular function data will be imputed using fully conditional specification (FCS) with logistic regression (cumulative logit) in 100 imputations. Imputation will be carried out separately for each treatment group. Variables in the imputation models will include binocular function scores at 18, 27, and 36 weeks.

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. Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res*. 2018;27(9):2610-2626.
2. 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.
3. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18.
4. Howard DR, Brown JM, Todd S, Gregory WM. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. *Stat Methods Med Res*. 2018;27(5):1513-1530.
5. Juszczak E, Altman DG, Hopewell S, Schulz K. Reporting of Multi-Arm Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA*. 2019;321(16):1610-1620.
6. Parker RA, Weir CJ. Non-adjustment for multiple testing in multi-arm trials of distinct treatments: Rationale and justification. *Clin Trials*. 2020;17(5):562-566.

7. Dmitrienko A, D'Agostino RB, Sr. Multiplicity Considerations in Clinical Trials. *N Engl J Med*. 2018;378(22):2115-2122.
8. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*. 2006;93(3):491-507.
9. Kleinman LC, Norton EC. What's the Risk? A simple approach for estimating adjusted risk measures from nonlinear models including logistic regression. *Health Serv Res*. 2009;44(1):288-302.
10. 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.
11. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-99.
12. Berglund PA. Multiple Imputation Using the Fully Conditional Specification Method : A Comparison of SAS ® , Stata , IVEware , and. 2015: