ATS20: A Randomized Trial of Binocular Dig Rush Game for Treatment of Amblyopia

Statistical Analysis Plan Version 4.0 (09/16/2020)

Based on Protocol Version 4.0 (04/30/18)

Revision History

Author Date	Version Number	Description of Changes	Stage of Analysis	Reviewer Date
E. Lazar 4/18/17	1.0	• N/A – Initial version	After enrollment initiation but prior to any data tabulation	
E. Lazar 8/27/18	2.0	• Updated protocol version number/date • Minor edits to add clarification • Updated section 4.2 to define adjustment for multiple testing for secondary outcomes • Section 6.0: Added post hoc analysis for ocular alignment (proportion of participants with a microtropia at baseline who are classified as orthotropic at follow-up) of a group comparison. Therefore, only the raw values will be reported for both the level and change in amblyopic-eye visual acuity from the 8-week to the 16-week visit. • Diplopia analyses: Replaced the Cochrane-Armitage trend test with the Wilcoxon rank sum test in reference to treatment group comparison in levels of diplopia. Unlike diplopia level (dependent variable), treatment group (independent variable) is measured without error. Therefore, it is not appropriate to use the Cochrane-Armitage trend test, which switches the relationship of these two factors. Instead, the exact Wilcoxon rank sum test will be used for the treatment group comparison of ranked diplopia scores. • Post 8-week Phase: Change in amblyopic-eye visual acuity at 16-week visit will not be adjusted for the 8-week acuity. There is no rationale for this adjustment as no treatment group comparison is being performed.	Post hoc – Performed during manuscript review process but prior to journal submission	M. Melia 8/27/18
E. Lazar 9/13/18	3.0	• Added the rationale for performing interim monitoring for the younger cohort and provided the link to the folder with the saved plan (section 2.1) • Safety analyses: Provided a rationale for		M. Melia 9/19/18

		using a type I error rate of 1% for statistical significance for each formal comparison (section 4.4).		
E. Lazar 10/24/18	3.1	• Added post hoc subgroup analysis for younger cohort based on whether or not enrollment occurred prior to modification to the Dig Rush algorithm	After enrollment initiation but prior to any data tabulation	
Z. Li 09/16/2020	4.0	The following changes have been applied to the younger cohort: • Sections 4.1 & 4.2: Changed the overall type I error rate from 5% to 4.9% for the primary VA outcome and for the set of secondary VA outcomes per the Interim Monitoring Plan. • Section 4.1: Added a sensitivity analysis using winsorized VA data to examine if the primary analysis results are robust to outliers. • Section 4.2.3: Added detailed description of the analysis of binocular treatment effect on VA outcomes at 4 and 8 weeks by Dig Rush game algorithm. • Section 5.3: Analysis of dose-response relationship after 8 weeks of binocular treatment pooled across the original binocular group and the control group opted for binocular treatment will not be performed due to concerns such as non-mandatory masking for 16-week VA exam. • Section 5.4 (Fellow-eye VA): Mean change in fellow-eye VA from 8 to 16 weeks will be estimated without adjustment for the 8-week VA since no treatment group comparison will be performed (same rational as for the analysis of change in amblyopic-eye VA from 8 to 16 weeks). • Section 6.0: Added post hoc analyses to report separately the proportions of new heterotropia, worsening heterotropia, and baseline heterotropia no longer present at each follow-up visit.	Updated when drafting the younger cohort manuscript	M. Melia 10/7/20

8
9

1.0 Study Overview

Participants aged 4 to <13 years old with amblyopia due to anisometropia and/or strabismus are enrolled into the multi-center trial which consists of two phases:

1) an 8-week randomized trial phase comparing the Dig Rush binocular game play on an iPad® device (1 hour/day 5 days/week) and spectacle wear (if needed) versus continued spectacle wear only (if needed). After a 1-week phone call, all participants are seen at 4 weeks post-randomization (primary outcome) and again at 8 weeks post-randomization (secondary outcome).

The randomized trial phase is followed by:

2) an 8-week post-randomization phase limited to participants in the continued spectacle group who opt for 8 weeks of treatment with the binocular game (1 hour/day and 5 days/week).

The trial includes two sub-studies with identical protocols:

- Younger cohort: Participants aged 4 to <7 years old at enrollment
- Older cohort: Participants aged 7 to <13 years old at enrollment

This document describes the analyses that will be performed for both sub-studies. For both age cohorts, the primary objective is to compare mean change in amblyopic-eye visual acuity (VA) between prescribed binocular game play with spectacle wear (if needed) and continued spectacle wear (if needed) alone (subsequently referred to as the "binocular group" and "control group") after 4 weeks of treatment. A secondary analysis will compare mean change in amblyopic-eye VA after 8 weeks of treatment.

2.0 Sample Size / Re-estimation

A minimum sample size was calculated to be 116 participants in the younger cohort and 84 participants in the older cohort based on the primary treatment group comparison of mean VA change from baseline to 4 weeks. The sample size was estimated to provide 90% power to detect a treatment group difference in mean VA for each sub-study assuming a 2-sided type I error rate of 5%, a group difference of 0.75 logMAR lines (pooled standard deviation (SD) = 1.2 logMAR lines) in the younger age cohort, a group difference of 3.75 letters (pooled SD = 5 letters) for the older age cohort, including a 5% adjustment for loss to follow-up. Details of the sample size estimation are described in a separate document (F:\user\PEDIG\Studies\ATS\Protocols\Current Protocols\Binocular Game Play ATS20\Sample Size\Verification\Binocular Games Sample Size 10-05-16.docx).

Although we believe our estimates of variance are reasonable, a sample size re-estimation will be performed for each sub-study once approximately 50% of the pre-planned sample has completed the 4-week outcome visit. A pooled estimate of variance without respect to treatment group will be calculated and used to re-estimate sample size using a procedure that maintains masking and has a negligible effect on the Type I error rate.[1] Within each sub-study, if the observed standard deviation of change is larger than the anticipated estimate, the sample size will be increased up to a maximum limit of 182 participants (SD of change = 1.5 logMAR lines) and 206 participants (SD of change = 8 letters) for the younger and older cohorts, respectively, which includes a 5% loss to follow-up. Due to the short duration of the primary outcome (4 weeks) and expected rapid recruitment, no interim monitoring will be conducted for either sub-study. This decision will be re-evaluated if the sample size is increased or the recruitment rates are slower.

2.1 Interim Monitoring (Younger Cohort)

In April 2018, the sample size re-estimation was performed in the younger cohort (as described above) and the pooled SD of the 4-week change in amblyopic-eye VA was estimated to be 1.6 logMAR lines. Based on these results, the Data Safety Monitoring Committee recommended that the total sample size be increased to the pre-specified maximum of 182 participants and that an interim monitoring plan be developed. Details of the approved interim monitoring plan are described in a separate document located in the following folder:
F:\user\PEDIG\Studies\ATS\Protocols\Current Protocols\Binocular Game Play ATS20\Monitoring\Statistical Interim Monitoring Plan

3.0 General Principles for Analysis

3.1 Visual Acuity Outcomes

Two examination procedures will be used for measuring VA. For participants aged <7 years at enrollment (younger cohort), visual acuity will be measured using the ATS single-surround HOTV method. This procedure provides Snellen equivalent scores that will be converted to the logMAR scale for analyses. For participants aged 7 to <13 years at enrollment (older cohort), visual acuity will be measured using the E-ETDRS testing protocol on the Electronic Visual Acuity Tester and analyses will be performed using letter scores. The same testing protocol is to be used throughout the study regardless of the participant's age during follow-up.

3.2 Stereoacuity

Stereoacuity will be measured using the Randot Preschool stereoacuity test and the Randot Butterfly test at each visit. For participants who fail the 800 seconds of arc level of the Randot Preschool test or the pretest, stereoacuity will be analyzed as 2000 seconds of arc (correct response on the Randot Butterfly test) or as nil (incorrect response or not attempted for the Randot Butterfly test).

The number of participants classified as having nil stereoacuity in absence of a butterfly test will be reported by treatment group for each visit and will be flagged for further review. Analyzing stereoacuity as nil in absence of the butterfly test could introduce misclassification bias because it's possible that some participants may have had 2000 seconds of arc of stereoacuity had they attempted the test. In ATS18, it was rare that the Randot butterfly test was not attempted,[2] which is expected to also be true in this study. If these cases account for ≥ 10% of the data, a sensitivity analysis will be performed in which analyses are repeated after substituting missing values for these cases.

A logarithm (base 10) transformation will be applied to stereoacuity scores for analyses. Participants classified as having nil stereoacuity (worse than 2000 seconds of arc) will be assigned a log score of 3.6 (the next largest disparity level) to calculate the difference between log converted scores (baseline – follow-up).

3.3 Analysis Window

The analysis window for visits will be as follows:

- 4-week: 3 to <7 weeks (21 to <49 days) after randomization
- 8-week: 7 to <15 weeks (49 to <105 days) after randomization
- 16-week (visit occurring 8 weeks after initiating binocular treatment for participants initially assigned to spectacles only group): 15 to <23 weeks (105 to <161 days) after randomization

A visit will be considered missed if it is completed outside of the analysis window or not completed at all. Analyses of the amblyopic-eye VA and stereoacuity outcomes will be limited to visits completed within the analysis windows.

4.0 Analysis Plan for 8-week Randomized Trial Phase

Analyses outlined in this chapter are limited to data collected at the time of randomization (subsequently referred to as 'baseline') through the 8-week post-randomization follow-up visit.

4.1 Primary Analysis

Two treatment approaches will be evaluated within each of the sub-studies:

- Binocular treatment: Binocular 'Dig Rush' game played on an iPad device 1 hour per day 5 days per week and spectacle wear (if required)
- Continued spectacle wear (if required) only

The primary analysis will follow a modified intent-to-treat principle, limited to data from participants who complete the 4-week exam within the pre-specified analysis windows as defined in section 3.3. Data from participants with treatment crossover, those who received alternative treatment for ≥ 1 week and participants found to be ineligible after subsequent review of enrollment data will also be included in the primary analysis. There will be no imputation of data for participants who are lost to follow-up or withdraw from the study prior to the 4-week exam.

An analysis of covariance (ANCOVA) adjusting for baseline VA will be performed to compute the 4-week mean change in amblyopic-eye VA for each treatment group and the 95% confidence intervals (CIs), as well as the difference in mean VA change between the treatment groups and the 95% CI.

For the younger cohort, the type I error rate for the primary analysis is pre-specified as 4.9% in the Interim Monitoring Plan (section 2.1) because 0.1% was allocated for the review of VA outcomes by the Data and Safety Monitoring Committee. Therefore, the significance level of the confidence intervals will be adjusted to 95.1%.

Model Assumptions: Model assumptions for the ANCOVA will be assessed, including linearity of adjustment covariates, normality and equal variance of the outcome across the treatment groups. The linearity assumption of covariates will be evaluated using descriptive scatterplots and by categorizing each of the baseline factors in the model to check for approximate linearity of the coefficients across ordered categories. A covariate will be included as a continuous variable in the model if assumptions for linearity are met for that covariate; otherwise, it will be categorized.

Although the ANCOVA is relatively robust to departures from normality, potential outliers will be identified and a sensitivity analysis will be performed to evaluate the effect of these outliers on the primary outcome results. Residual values will be examined for an approximate normal distribution. If values are highly skewed, then either a transformation will be applied or alternative analysis strategies (robust regression, non-parametric methods) will be considered instead.

Confounding: Imbalances between groups in important baseline covariates are not expected to be of sufficient magnitude to produce confounding. However, as a complement to the primary analysis, the presence of confounding will be evaluated. If there is evidence of confounding based on one or more

factors, the primary analysis will be repeated with the factors included in the ANCOVA model as adjustment covariates. Results of the model will be compared with that of the primary analysis results to evaluate the effect of confounding on the treatment group comparison.

Sensitivity Analyses:

The primary analysis will be repeated in the following ways:

1. Perform multiple imputation using the Monte Carlo Markov Chain (MCMC) method that includes data from baseline and follow-up visits to impute 4-week VA data for participants who missed the exam
2. Exclude 4-week VA data from participants who completed the 4-week exam outside of the protocol window (4 ± 1 week after randomization)
3. Exclude 4-week VA data from participants found to be ineligible upon subsequent review of enrollment data, those with treatment crossover, or those who received alternative treatment for ≥ 1 week
4. Include cause of amblyopia as an adjustment covariate in the ANCOVA model
5. Winsorize baseline and 4-week VA data at the 10^{th} and 90^{th} percentiles by treatment group (younger cohort only)

If the primary analysis and sensitivity analyses produce similar results, the primary analysis will be considered the definitive analysis and the sensitivity analyses will be used to provide supportive evidence of the magnitude of treatment effect. However, if the results differ, exploratory analyses will be performed to evaluate the factors that have contributed to the differences.

4.2 Secondary Analyses

Secondary analyses will be conducted separately for each sub-study and all treatment group comparisons will consist of a 2-sided test of the null hypothesis of no difference between groups. Unless otherwise specified, analyses will only include participants with visits completed within the pre-specified analysis windows (section 3.3) and will follow the principles outlined in the primary analysis.

Due to the number of secondary outcomes, the 2-sided type I error rate for each secondary analysis (including confidence intervals calculated on the estimate) will be adjusted to account for multiple testing as follows:

- The Bonferroni method will be used to preserve the overall type 1 error rate at 5% (4.9% for the younger cohort; see section 4.1) for all secondary analyses of visual acuity (n=3 tests, sections 4.2.1 - 4.2.2) and at 5% across all stereoacuity analyses (n=4 tests, section 4.2.4)
- For each subgroup analysis (section 4.2.3), statistical significance of the interaction term will be tested using a type I error rate of 1%.

4.2.1 Visual Acuity Improvement at 8 Weeks

A treatment group comparison of mean VA change from baseline to 8 weeks will parallel the 4-week primary analysis. This analysis will only include data from participants who complete the 8-week exam within the pre-specified analysis window (section 3.3); there will be no imputation of data for participants with a missed 8-week exam.

4.2.2 Visual Acuity Improvement Defined as a Binary Outcome

A secondary analysis will estimate and compare the proportion of participants with amblyopic-eye VA improvement of ≥ 2 logMAR lines (≥ 10 letters if E-ETDRS) from baseline to 4 weeks by treatment

group. The proportion of participants who achieve this outcome will be tabulated by treatment group. For the treatment group comparison, a p-value, an estimate (proportion) of the group difference, and the corresponding 98.4% CI on the estimate will be computed using binomial regression with adjustment for the baseline VA. If the binomial regression model does not converge, Poisson regression with robust variance estimation or an exact method (without adjustment for baseline VA) will be used to estimate the treatment group difference.

The aforementioned secondary analysis will be repeated to estimate and compare the proportion of participants with amblyopic-eye VA improvement of ≥ 2 logMAR lines from baseline to 8 weeks by treatment group.

4.2.3 Subgroup Analyses

The 4-week treatment effect will be assessed in subgroups of participants based on baseline factors. Subgroup analyses will be considered exploratory and used to suggest hypotheses for further investigation and future studies. The baseline subgroups of interest include age, amblyopic-eye VA, stereoacuity, the presence of a heterotropia at near, and prior amblyopia treatment. In accordance with NIH guidelines, a subgroup analysis of treatment effect according to gender, as well as race/ethnicity, will be conducted. However, based on results from previous ATS studies, there are no data to support a differential treatment effect by these variables.

The subgroup definitions for the pre-planned subgroup analyses are as follows:

- Amblyopic-eye VA at baseline: 20/40 (68 to 72 letters), 20/50 (63 to 67 letters), 20/63 (58 to 62 letters) and 20/80 or worse (<58 letters)
- Stereoacuity at baseline (nil versus better than nil)
- Presence of a near heterotropia (deviation 1 to 4Δ) at baseline measured by SPCT (yes/no)
- Age (years) at baseline (Younger cohort: 4 to <5, 5 to <7; Older cohort: 7 to <10, 10 to <13)
- Prior amblyopia treatment at baseline (yes/no)
- Prior amblyopia treatment with binocular therapy (yes/no)
- Sex (male/female)
- Race/ethnicity (non-Hispanic white versus other)

It is hypothesized that these subgroup factors will not influence the treatment effect for either sub-study. The purpose of the subgroup analyses is to provide evidence to support this hypothesis and the combining of the subgroups for the primary outcome analysis. For each subgroup factor, a formal analysis will only be performed if there is a minimum of 20 participants in every subgroup category for both treatment groups.

The general approach for the subgroup analyses is to conduct an ANCOVA similar to the primary analysis, adding a term for the main effect of the baseline subgroup factor and an interaction term between the treatment group and the baseline subgroup factor. Interpretation of the subgroup analyses will depend on whether the overall analysis demonstrates a significant treatment group difference. In the absence of an overall difference, these subgroup analyses will be interpreted with caution.

A significant interaction term ($p < 0.01$) will be taken as an indication that subgroup effects need to be explored for full interpretation of the study results. A non-statistically significant F-test for interaction ($p \geq 0.01$) will not be interpreted as conclusive evidence of no subgroup effect given that power for the tests of interaction is low. The estimated treatment group difference and a 2-sided 95.1% CI will be computed

from the interaction model for each of the subgroups. Baseline age and amblyopic-eye VA will be treated as continuous variables to compute the p-value for interaction.

For the younger cohort, additional exploratory analyses will be conducted to assess the binocular treatment effect at 4 and 8 weeks according to the game algorithm. An ANCOVA adjusting for baseline VA and age will be performed to compute the mean VA change by the algorithm for fellow-eye contrast increment (≥ 30 versus ≥ 15 minutes of game play), as well as the difference in mean VA change between the two algorithms and the 95% CI. The same approach will be used for the 4-week and 8-week analyses.

4.2.4 Stereoacuity

The distribution of stereoacuity scores will be tabulated by treatment group at baseline and at each follow-up visit. Medians and ranges of stereoacuity scores for all visits will be computed by treatment group. Change in ranked stereoacuity scores (≥ 2 levels worse, within 1 level, ≥ 2 levels better) from baseline to the 4- and 8-week visits will be tabulated for each group and compared between treatment groups using the exact Wilcoxon rank sum test.

The above analyses for stereoacuity and change in stereoacuity will be repeated in participants with no history of strabismus.

4.3 Treatment Compliance, Dose & Game Performance with Binocular Therapy (Binocular Treatment Group)

In addition to subjective compliance with prescribed treatment based on parent-reported calendars, an objective measure of compliance will be obtained from the automated iPad log files for those assigned to binocular treatment. The iPad log files record total time playing the game and the contrast level presented to the fellow eye. The following sections describe exploratory analyses for compliance measures, treatment dose, and game performance in the binocular group. The analyses will be limited to participants who completed the follow-up visits within the pre-specified analysis windows (section 3.3). No adjustment for multiplicity will be made to these exploratory analyses.

4.3.1 Binocular Treatment Dose, Compliance and Fellow-eye Contrast

Participants will be prescribed 1 hour per day of binocular game play for 5 days per week as per protocol. The cumulative amount of binocular treatment received since baseline (dose) and the percentage of prescribed treatment completed (compliance) will be calculated from the log file data for the 4- and 8-week visits. Compliance will be calculated as the total amount of binocular treatment received divided by the total number of prescribed hours at that time point since baseline. For each follow-up visit, the distribution of the cumulative hours of treatment received since baseline (0 to <10, 10 to <20, etc.) and the percentage of prescribed treatment completed (0% to 25%, >25% to 50%, >50 to 75%, >75%) will be tabulated with computation of descriptive statistics (median and range).

The change in contrast level presented to the fellow eye provides a measure of game performance because the contrast level in the fellow eye is incremented based on ≥ 15 minutes (or ≥ 30 minutes for participants enrolled on or after 8/24/2018) of successful game play on the previous day. The distribution of contrast level presented to the fellow eye (20%, >20% to 40%, >40% to 60%, >60% to 80%, >80% to <100%, 100%) at each follow-up visit will be tabulated with computation of descriptive statistics (median and range). Change in contrast will be defined based on the log file data as the current contrast level on the date of the visit minus the initial contrast level, which was set to 20% for all participants.

4.3.2 Relationship between Binocular Treatment Dose & Change in Fellow-eye Contrast

The relationship between cumulative binocular treatment dose and change in contrast presented to the fellow eye at the 4- and 8-week visits will be explored using scatterplots to examine whether there is evidence of association and the form of association. Correlation between treatment dose and change in contrast at 4 and 8 weeks will be computed using Pearson correlation coefficients if there is evidence of a linear trend based on the scatterplots.

4.3.3 Visual Acuity Change according to Binocular Treatment Dose & Change in Fellow-eye Contrast

The relationship between change in amblyopic-eye VA from baseline to the 4- and 8-week visits with respect to 1) cumulative binocular treatment dose and 2) change in fellow-eye contrast will be examined using scatterplots for evidence of association and form of association. Pairwise correlation between VA change and each of the two factors will be computed using Pearson correlation coefficients if there is evidence of a linear trend based on the scatterplots.

If there is evidence of a linear relationship between amblyopic-eye VA change from baseline to 4 weeks with either cumulative treatment dose or change in fellow-eye contrast at 4 weeks, a multivariable regression model that adjusts for baseline VA will be fit to describe this relationship. Collinearity diagnostics will be output from the multivariable regression model to assess whether it is appropriate to retain both factors in the model. If the two factors are highly collinear, separate regression models will be fit to evaluate the relationship between change in amblyopic-eye VA with each individual factor, adjusted for baseline VA. The analyses described above will also be repeated for the 8-week visit if appropriate.

Model assumptions for the multivariate regression models will be assessed as described in section 4.1. If the assumption of linearity between change in amblyopic-eye VA and any of the aforementioned factors is not met, then this factor will be categorized into quartiles based on the distribution of the data.

Given that all participants will be prescribed the same dose of binocular therapy, any differences in cumulative treatment dose between participants would only be due to treatment compliance, which could be affected by the efficacy of the binocular treatment, differences in motivation to comply with prescribed treatment and potentially other unmeasured factors. Therefore, results of these analyses will be interpreted with caution.

4.3.4 Stereoacuity Change according to Binocular Treatment Dose & Change in Fellow-eye Contrast

The relationship between change in stereoacuity from baseline to the 4- and 8-week visits with respect to 1) cumulative binocular treatment dose and 2) change in fellow-eye contrast from baseline will be examined using scatterplots and Pearson correlation coefficients as described in section 4.3.3. For this analysis, a logarithm (base 10) transformation will be applied to the raw stereoacuity scores and the difference between the log converted baseline and follow-up scores will be computed as described in section 3.2.

If there is evidence of a linear relationship between change in stereoacuity from baseline to 4 weeks with either cumulative treatment dose or change in fellow-eye contrast at 4 weeks, a multivariable regression model that adjusts for baseline stereoacuity will be fit to describe this relationship. The same modeling approach as described in section 4.3.3 will be used. If appropriate the analyses will also be applied to the 8-week visit. As noted in section 4.3.3, results of these analyses will be interpreted with caution.

4.4 Safety Analyses

Statistical significance for each formal group comparison of safety outcomes/adverse events was tested at a type I error rate of 1% to adjust for multiple testing. Given that spectacle wear and game play on an iPad device are not invasive treatments and pose minimal risk, if any, to participants, there are greater concerns about the possibility of falsely finding a group difference in safety outcomes/adverse events than in missing a difference. Safety analyses will include all participants who completed the follow-up visits regardless of the analysis windows.

4.4.1 Visual Acuity in the Fellow Eye

The distribution of change in fellow-eye VA from baseline to 4 weeks will be tabulated. The mean change in fellow-eye VA from baseline to 4 weeks will be calculated and compared between treatment groups using ANCOVA with adjustment for the baseline fellow-eye VA. The proportion of participants with loss of ≥ 2 logMAR lines (≥ 10 letters) of VA in the fellow-eye from baseline to the 4-week exam will be reported by treatment group and compared using Barnard's exact test. The analyses will be repeated for the 8-week visit.

4.4.2 Ocular Alignment

The proportion of participants with 1) no baseline heterotropia at distance and/or near who developed a new heterotropia (measured by SPCT) at 4 weeks or 2) a baseline heterotropia at distance and/or near (measured by SPCT) who had an increase of ≥ 10Δ in the pre-existing heterotropia at 4 weeks will be reported by treatment group and compared using Barnard's exact test. The analyses will be repeated for the 8-week visit.

4.4.3 Diplopia

The frequency of diplopia was reported as "Never", "Less than once a week", "Once a week", "Once a day", "Up to 10 times a day", "More than 10 times a day", or "All the time". The distribution of diplopia frequency at baseline, 4 weeks, and 8 weeks will be tabulated by treatment group. Each level of diplopia frequency will be assigned an ordered, numeric score. The change in frequency scores (increased by ≥ 2 levels, within 1 level, decreased by ≥ 2 levels) from baseline to 4 and 8 weeks will be tabulated by treatment group. The frequency scores and the change in scores at 4 and 8 weeks will be compared between treatment groups using the exact Wilcoxon rank sum test. The number of participants with monocular diplopia (diplopia that does not go away when the eye is closed or covered) will be reported by treatment group, but these cases will not be excluded from the analyses.

The above analyses will be performed separately for assessments completed by the participant and by the parent.

4.4.4 Adverse Symptoms

Parents were asked to complete a 5-item symptoms survey regarding the frequency of 1) headaches, 2) eyestrain, 3) blurry vision, and if wearing spectacles, how often they 4) look over them or 5) take them off. The frequency of each survey item was graded as "Never", "Almost never", "Sometimes", "Often", or "Almost Always". For each item, the distribution of symptom frequency at baseline, 4 weeks, and 8 weeks will be tabulated by treatment group. Each level of symptom frequency will be assigned an ordered, numerical score. The change in frequency scores (increased by ≥ 2 levels, within 1 level, decreased by ≥ 2 levels) from baseline to 4 and 8 weeks will be tabulated by treatment group. The frequency scores and the change in scores at 4 and 8 weeks will be compared between treatment groups

using the exact Wilcoxon rank sum test. Participants who were not wearing spectacles during the randomized trial will not be included in analyses for items regarding spectacle wear.

4.5 Protocol Adherence and Additional Descriptive Analyses

The following descriptive analyses will be performed:

1. Provide a flow chart accounting for all participants for all visits and the 1-week phone call according to treatment group
2. Calculate completion rates for each follow-up visit by treatment group
3. Calculate completion rate for the 1-week phone call by treatment group
4. Tabulate baseline characteristics according to treatment group
5. Report subjective compliance with prescribed treatment by treatment group based on parent-reported calendars
6. Report protocol deviations by treatment group
7. Report deviations to prescribed treatment (treatment crossover or alternative treatment received for ≥ 1 week during the study) by treatment group

5.0 Analysis Plan for 8-week Post-randomization Phase

Unless otherwise specified, analyses outlined in this section will be limited to data collected between the 8-week visit of the randomized trial and the 16-week visit for participants assigned to the spectacle only group who opt to receive 8 weeks of binocular treatment after completing the randomized trial phase. The 8-week visit will serve as the baseline for the 16-week visit.

5.1 Amblyopic-eye Visual Acuity after 8 Weeks of Binocular Treatment

The mean amblyopic-eye VA at 8 and 16 weeks will be estimated along with the 95% CIs using data completed within each corresponding analysis window. The mean change in amblyopic-eye VA from 8 to 16 weeks will be estimated along with the 95% CI using data completed within both analysis windows for the 8- and 16-week visits.

The proportion of participants with amblyopic-eye VA improvement of ≥ 2 logMAR lines (≥ 10 letters if E-ETDRS) after 8 weeks of binocular treatment will be calculated along with the exact 95% CI.

5.2 Stereoacuity after 8 Weeks of Binocular Treatment

The following analyses will include data completed within the analysis windows for the 8- and 16-week visits. The distribution of stereoacuity scores at the 8- and 16-week visits will be tabulated and compared using the exact Wilcoxon signed rank test. The change in ranked stereoacuity scores (≥ 2 levels worse, within 1 level, ≥ 2 levels better) between the two visits will also be tabulated. Medians will be computed for stereoacuity scores and change in stereoacuity scores.

The above analyses for stereoacuity and change in stereoacuity will be repeated in participants with no history of strabismus.

5.3 Treatment Dose, Compliance and Fellow-eye Contrast after 8 Weeks of Binocular Treatment (Older Cohort Only)

For the analyses described below, log file data recorded during an 8-week interval of binocular treatment will be pooled across participants assigned to the binocular group (randomization to 8-week visit) and those receiving binocular treatment in the post-randomization phase (8- to 16-week visits).

The cumulative amount of treatment (hours) received, the percentage of prescribed treatment completed, and the change in fellow-eye contrast after 8 weeks of binocular treatment will be calculated from the log file data and descriptive analyses will be performed as described in section 4.3.1. The relationship between cumulative binocular treatment dose and change in fellow-eye contrast after 8 weeks of binocular treatment will also be explored as described in section 4.3.2.

5.4 Safety Analyses after 8 Weeks of Binocular Treatment

The following safety assessments will be performed on all participants who completed the exams regardless of the analysis windows:

1. Fellow-eye VA: The mean change in fellow-eye VA from 8 to 16 weeks will be estimated along with the 95% CI. The proportion of participants with fellow-eye VA loss of ≥ 2 logMAR lines (≥ 10 letters) after 8 weeks of binocular treatment will be calculated along with the exact 95% CI.

2. Ocular alignment: The proportion of participants with 1) no baseline heterotropia at distance and/or near who developed a new heterotropia (measured by SPCT) at 16 weeks or 2) a baseline heterotropia at distance and/or near (measured by SPCT) who had an increase of ≥ 10Δ in the pre-existing heterotropia at 16 weeks will be reported.

3. Diplopia: The distribution of diplopia frequency at the 8- and 16-week visits will be tabulated and compared between the two visits using the Wilcoxon signed rank test. The change in diplopia frequency (increased by ≥ 2 levels, within 1 level, decreased by ≥ 2 levels) from 8 to 16 weeks will also be tabulated. The above analyses will be performed separately for participant- and parent-reported assessments.

4. Adverse Symptoms: For each item, the distribution of symptom frequency at the 8- and 16-week visits will be tabulated and compared between the two visits using the Wilcoxon signed rank test. The change in symptom frequency (increased by ≥ 2 levels, within 1 level, decreased by ≥ 2 levels) from 8 to 16 weeks will also be tabulated.

6.0 Post Hoc Analyses

The following post hoc analyses will be performed for the older cohort:

- Calculate the proportion of participants with a microtropia (maximum ocular deviation of <5Δ by SPCT) who are classified as orthotropic (no manifest tropia at near or distance by SPCT) at the 4- and 8-week visits.
  Rationale: Upon review of participants who developed a new ocular deviation, J. Holmes noted that "the new tropias were essentially all microtropias" and that some had only been identified at one distance (near or distance) at enrollment and now detected in the other direction. A microtropia could have been missed as its presence may be variable based on the condition and the attention of the participant and/or examiner. There was a suggestion to calculate the proportion of participants who were microtropic at baseline (maximum deviation size of <5Δ) with no deviation present (orthotropic) at the 4- and 8-week visits (separately) to determine whether this estimate is similar to the proportion who "developed" a new microtropia.

The following post hoc analyses will be performed for the younger cohort:

- Report separately the proportion of participants with 1) no baseline heterotropia at distance and/or near who developed a new heterotropia (measured by SPCT) and 2) a baseline heterotropia at distance and/or near (measured by SPCT) who had an increase of ≥ 10Δ in the pre-existing heterotropia at each follow-up visit.
- Calculate the proportion of participants with a baseline heterotropia at distance and/or near (measured by SPCT) which was no longer present at 4 and 8 weeks.

References

1. Gould AL, Shih WJ. Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. Comm Stat Theory Meth 1992; 21: 2833-53.
2. Pediatric Eye Disease Investigator Group. Effect of a binocular iPad game versus part-time patching in children aged 5 to 12 with amblyopia: a randomized clinical trial. JAMA Ophthalmol. 2016;134(12):1402-1408.