

1 **Hyperopia Treatment Study (HTS1)**
 2 **Immediate versus Delayed Treatment In Children with Moderate**
 3 **Hyperopia**

4
 5 **Statistical Analysis Plan**
 6 **Version 3.0**

7
 8
 9
 10 Statistical Analysis Plan Version: 3.0, 25Jul18
 11 Protocol Version: 3.0, 28Jan15

12
 13
 14 **VERSION HISTORY**

The following table outlines changes for the analysis plan:

VERSION NUMBER	AUTHOR	APPROVER	EFFECTIVE DATE	REVISION DESCRIPTION*
1.0	T. Dean	M. Melia	12-19-17	Initial draft written after study was completed but before manuscript was written.
2.0*	T. Dean	M. Melia	05-14-18	Lead authors decided to redefine “best” stereoacuity and visual acuity at the 3-year outcome visit to be consistent with what was done for the IXT study. When preparing to unmask the study investigators to the results, the leads realized that it was not our intent to compare deterioration rates between the two groups. In the planning phase, the planning committee felt the deterioration rates in each group would aid in interpreting the results. The committee felt a formal statistical comparison was not needed because the rate of deterioration was expected to be higher in the observation group than in the glasses group.
3.0*	T. Dean	M. Melia	07-27-18	Lead authors decided to redefine <i>best</i> stereoacuity and <i>best</i> visual acuity at the 3-year outcome visit as the best of measurements reported with and without trial frames, during the initial assessment and retest, if required. During their final review of the manuscript, the leads realized that results reported for primary and secondary analyses were discordant. The leads determined that the definitions of <i>best</i> stereoacuity and <i>best</i> visual acuity needed updating to reflect the study design and to be in concordance with results reported in the primary analysis.

* indicate in the description if the changes were made after review of outcome data

15 **1.0 Primary Objective**

16 The primary study objective is to compare visual outcomes and development of strabismus after a
17 36-month follow-up period in two age cohorts (children age 12 to <36 months and children age
18 36 to <72 months) with moderate hyperopia (spherical equivalent +3.00D to +6.00D in either
19 eye) who are prescribed glasses either immediately or only after pre-specified deterioration
20 criteria are met.

21
22 **2.0 Primary Analysis**

23 The primary analysis will be a treatment group comparison of the proportion meeting failure
24 criteria (below) at 36 months post-randomization. The primary analysis will be performed
25 separately for each of the two primary cohorts.

26
27 The null hypothesis of no difference between the treatment groups will be rejected if the
28 proportion meeting failure criteria in immediate glasses is significantly lower or higher than the
29 proportion in observation, based on the Barnard's exact test with 2-sided type I error of
30 $\alpha=0.05$.

31
32 Although Fisher's exact test was initially proposed, this was changed to Barnard's exact test to
33 allow for the calculation of a confidence interval on the difference between proportions, which is
34 consistent with the p-value from the test.

35
36
37 **2.1 Primary outcome**

38 At the 36-month outcome visit, each subject's condition will be classified as either failure or not
39 failure as follows:

40
41 Failure = ANY of the following criteria are met at the masked 36-month visit both with and
42 without correction in trial frames (without prism or bifocal):

- 43
- 44 1. Any measurable heterotropia in primary gaze in the distance (3 meters) or at near (1/3
45 meter) not correctable with distance refractive correction alone
 - 46 2. Strabismus surgery prior to the 36-month outcome exam
 - 47 3. Visual acuity below age norms in either eye
48 a. To be below age-normal, visual acuity in the eye must fall below age-normal
49 values on both test and retest, both with and without correction.
50 b. Visual acuity will be obtained for each eye.
 - 51 4. ≥ 2 logMAR lines of IOD if visual acuity is 20/25 or worse in the better eye (applies to
52 IOD either with or without correction but not one eye with and one eye without)
 - 53 5. ≥ 3 logMAR lines of IOD if visual acuity is 20/20 or better in the better eye (applies to
54 IOD either with or without correction, but not one eye with and one eye without)
 - 55 6. Stereoacuity at near by Randot Preschool test below age normal values

56
57 Not Failure = NONE of the criteria for failure are met

58
59 If a subject meets any failure criteria (with the exception of strabismus surgery prior to the 36-
60 month outcome exam), the visual condition that is considered to be failed must be retested by the
61 masked examiner (*section 3.6.1*) using the same procedures outlined in *section 3.6.1* of the
62 protocol. The subject must be given a 10-minute break prior to retesting each failed criterion.
63 Failure criterion must be confirmed by a retest both with and without correction to be considered
64 a failure.

66 If failure is confirmed and the subject requires a change in glasses (*section 4.3* of the protocol),
67 the 36-month outcome exam is repeated on a different day 4 weeks \pm 1 week later (subsequently
68 referred to as the 37-month visit) by the masked examiner with the child wearing their new
69 glasses. Testing without correction does NOT need to be repeated. Failure will be assessed at this
70 post-36 month exam, and the subject will be classified as a failure if he/she meets any of the
71 failure criteria above. If the subject meets any failure criteria (with the exception of strabismus
72 surgery), the visual condition that is considered to be failed must be retested by the masked
73 examiner using the same procedures outlined above. Failure must be confirmed without
74 correction at the 36-month exam and with correction at this repeat outcome exam 4 weeks later to
75 be considered a failure.

77 **2.2 Principles to be Followed in Primary Analysis**

78 The following principles apply to the primary treatment comparison:

- 79 1. The primary treatment comparisons will be performed using a modified “intent-to-treat”
80 analysis principle. The data of all randomized subjects completing the 36-month exam
81 will be included in the analysis regardless of whether the assigned treatment was actually
82 received. Data from subjects who receive alternate treatment (i.e. glasses for subjects
83 assigned to observation who did not meet deterioration criteria and vice versa) will be
84 analyzed according to randomization group. There will be no imputation of data for
85 subjects who are lost to follow-up or withdrawn from study prior to the primary outcome,
86 defined as the failure status at the 36-month visit if the 37-month visit is not required and
87 failure status at the 37-month visit if required.
- 88 2. Subjects who meet failure criteria at the 36-month assessment and have significant
89 changes in their cycloplegic refraction are prescribed glasses and required to return
90 within 4 weeks to be retested in their updated correction. If this happens, the results of the
91 36-month examination without correction and the 37-month examination with correction
92 will be used to determine failure status in the primary analysis.
- 93 3. Subjects will be included in analysis as long as the 36-month visit was completed. Seeing
94 as failure must be confirmed, subjects who complete the 36-month visit will not be
95 considered to have met failure criteria for a condition if a required test or retest was not
96 performed.
- 97 4. Data will be considered missing if not completed within the analysis window. The 36-
98 month assessment must be completed no earlier than 33 months (1005 days) and no later
99 than 42 months (1282 days). Subjects will be excluded from analysis if the 36-month
100 visit is not completed within the analysis window.
- 101 5. The 37-month visit is an extension of the 36-month visit. No analysis window is defined
102 for the 37-month visit. If required and collected, data at the 37-month visit will be used in
103 analysis. Subjects will be included in analysis if the 36-month visit was completed within
104 the analysis window. See Table 1 and Figure 1 below for clarification.
- 105 6. Imbalances between groups in important covariates measured at baseline are not expected
106 to be of sufficient magnitude to produce confounding. However, as a complement to the
107 primary analysis, the presence of confounding will be evaluated by including in the
108 model factors that appear to be imbalanced between treatment groups and are associated
109 with visual acuity, stereoacuity, or strabismus.

110
111 Barnard’s exact test does not allow for the adjustment of covariates. Binomial regression was
112 used to compare treatment groups when the model was adjusted for possible confounders.

Table 1: Analysis of Data based on Visit Completion

36-Month Visit			37-Month Visit Required	37-Month Visit Completed	Primary and Secondary Analyses ^a	Additional Approaches to Primary Analysis (Imputation) ^b
In Window	Out of Window	Not Completed				
-	-	X	-	-	Excluded	Impute
X			No	-	Data included	Data included
X			Yes	Yes	Data included	Data included
X			Yes	No	Data included	Impute
	X		No	-	Excluded	Impute
	X		Yes	Yes	Excluded	Impute
	X		Yes	No	Excluded	Impute

X indicates when visit was completed

-Visits are considered completed if data was collected at the visit. Completion status is not dependent on the amount of data collected at the visit.

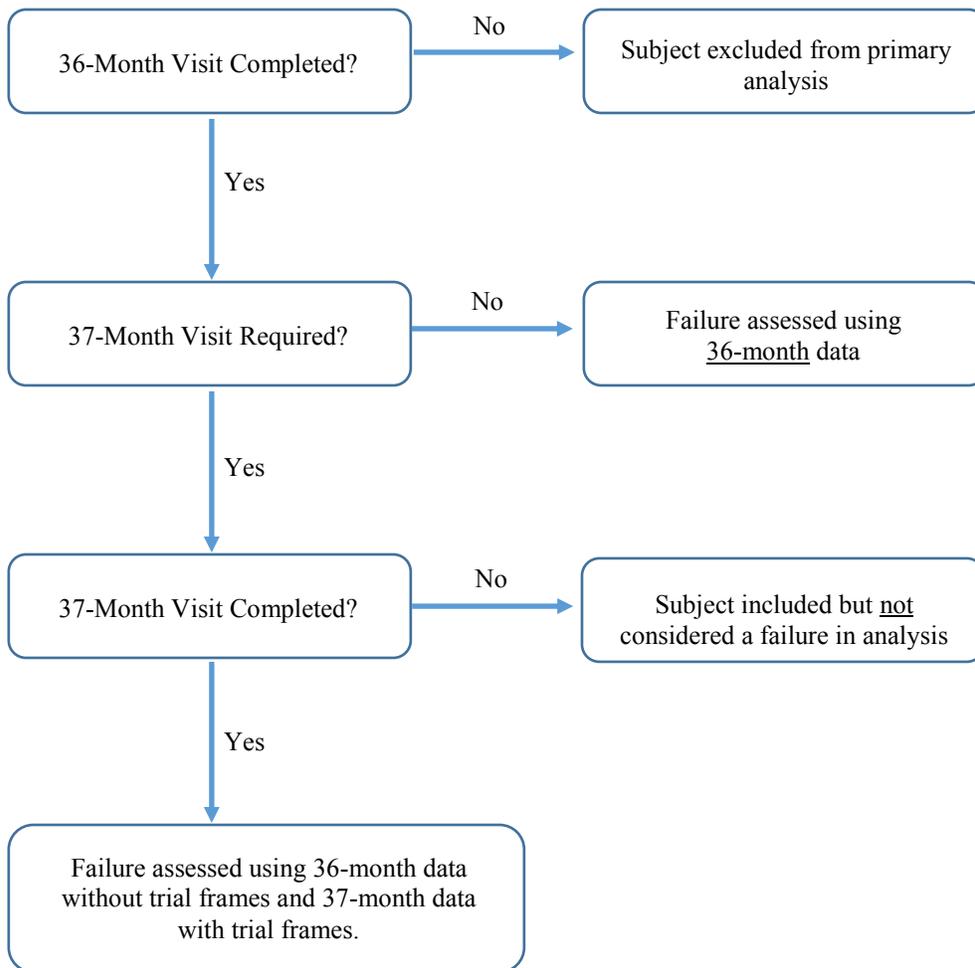
^aSubjects will be included in the primary and secondary analyses as long as the 36-month visit was completed. Seeing as failure must be confirmed, subjects who complete the 36-month visit will not be considered to have met failure criteria for a condition if a required test or retest was not performed.

^bData **will not** be imputed for subjects who have met failure criteria, even if all required testing (initial and retest, with and without correction) has not been completed. If failure criteria were met for any condition, the subjects is considered a failure at 36 months.

Table 2: Defining Color Code

Color	Meaning
	Data included
	Data excluded
	Data imputed

Figure 1: Assessment of Failure in Primary Analysis



*Note: Visits are considered to be completed if any data was collected for the subject at the visit. It does not mean that all required testing was performed at the visit.

2.3 Additional Approaches to Primary Analyses

The primary treatment comparisons will be repeated in the following ways:

1. Include data only from subjects who complete the 36-month visit within the protocol window (34 to 38 months after randomization).
2. The primary analysis will be repeated, using only the initial assessments to identify failure at 3 years. If a change in refractive correction was necessary after formal failure criteria were met and confirmed by a retest, measurements reported for initial assessments performed with trial frames at the 37-month visit will replace measurements reported with trial frames at the 36-month visit if the assessment was failed.
3. The primary analysis will be repeated, imputing data for subjects with incomplete or no data at the primary outcome visit. Data will be imputed for subjects classified as “not failure” at 3 years who meet any of the following criteria: 1) the 36-month visit was not completed or was completed out of window, 2) a required 37-month visit was not completed, 3) an initial assessment or required retest was not performed and the available data indicates that the subject is a failure, or 4) no data was reported for an assessment, or an eye in the case of visual acuity. The imputation procedure is described in section 2.5.
4. The primary analysis will be repeated, adjusting for age at 36-month visit. Barnard’s Exact test does not allow for adjustment. Instead, binomial regression will be used to obtain the adjusted difference in failure rate between the treatment groups. If the model does not converge, Poisson regression using robust variance estimation will be used instead. If the Poisson model does not converge, logistic regression will be used to test the difference in treatment groups, controlling for age.
5. The distribution of the following baseline factors will be reviewed by treatment group: age, visual acuity, and stereoacuity. If factors appear to be confounders, defined as imbalance between treatment groups, binomial regression will be used to evaluate the treatment group difference in failure rates, controlling for all confounding factors. The characteristics will be categorized as follows:
 - a. Age: 1 to <2, 2 to <3, 3 to <4, 4 to <5, 5 to <6 years
 - b. Visual acuity: 20/20 or better, 20/25, 20/32, 20/40, 20/50
 - c. Stereoacuity: 40, 60, 100, 200, 400

If the primary analysis and additional approaches produce similar results, the primary analysis will be considered the definitive analysis and the additional approaches 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.

2.4 Multiple Imputation

Multiple imputation using logistic regression will be used to impute 36-month failure status and will be performed separately for each primary cohort. Imputation in each cohort will be based on the presence of deterioration prior to 3 years and will be stratified by treatment group. The proportion of subjects who failed at 3 years will be tabulated for each imputed data set, and the results will be combined using the MiAnalyze procedure in SAS 9.4 to generate valid statistical inferences.

It is recognized that differential loss to follow up or a difference in the distribution of loss to follow-up times may introduce bias to the multiple imputation model. Prescribing or withholding treatment may influence a parent’s decision to keep their child enrolled in the study, therefore results may be subject to bias if subjects receiving one treatment regimen are more likely to withdraw from the study. Results may also be subject to bias if subjects in one treatment group tend to remain in the study longer than subjects in the other. Deterioration is identified during the

follow-up period, so the chance of meeting deterioration criteria increases with the number of completed visits.

The protocol states that the last-observation-carried-forward method would be used in addition to multiple imputation, but this is not possible as failure is only assessed at the 36-month visit.

3.0 Secondary Analyses

All secondary analyses will be performed separately for each of the two primary cohorts. The analyses below will follow the guidelines outlined in section 2.2. Secondary analyses will be considered significant if the p-value is less than 0.01.

3.1 Deterioration in Observation and Glasses Groups

For each treatment group, the Kaplan-Meier estimate of the cumulative proportion of subjects meeting deterioration criteria will be calculated along with the corresponding confidence interval. The deterioration rate in the observation group is expected to be higher than that of the glasses group, so deterioration rates will not be compared between treatment groups.

For the purposes of this analysis, a visit is defined by its corresponding analysis window. If deterioration criteria were met at a visit other than one specified by the protocol, the analysis window in which the subject met deterioration criteria will be defined as the visit at which deterioration criteria were met.

3.2 Best Distance Visual Acuity at 36-Months

A treatment group comparison of the best distance visual acuity (logMAR) in the better-seeing eyes at 36 months post-randomization will be performed using analysis of covariance (ANCOVA) models. A similar analysis will be performed to compare visual acuity in the worse-seeing eye between treatment groups. Both analyses will adjust for age at the 36-month exam and visual acuity in the eye at enrollment.

Best distance visual acuity in each eye is defined as the minimum logMAR visual acuity between measurements reported with and without trial frames, during initial assessment and retest, if required. If no retest is required at 36 months, best visual acuity is the minimum of logMAR visual acuities reported with trial frames and without trial frames during the initial assessment. If a retest is required at 36 months, best visual acuity is the minimum of logMAR visual acuities reported with and without trial frames, during the initial assessment and retest at 36 months. If a 37-month visit is required, best visual acuity is the minimum of logMAR visual acuities reported during initial assessment and retest without trial frames at the 36 months and during initial assessment with trial frames at the 37 months, if no retest is required at 37 months. If a retest is required at 37 months, best distance visual acuity is the minimum of logMAR visual acuities reported during initial assessment and retest without trial frames at 36 months and initial assessment and retest with trial frames at 37 months.

The better-seeing eye at 3 years will be defined as the eye with the smaller best logMAR visual acuity at 3 years.

For the purposes of this analysis, a value of 1.7 will be assigned to eyes that have a visual acuity worse than 20/800.

3.3 Proportion of Subjects with Age-Normal Distance Visual Acuity at 36 Months

A treatment group comparison of the proportion of subjects meeting age-normal distance visual acuity in **both** eyes at 36 months will be performed using Barnard's exact test.

The best reported visual acuity (*see section 3.2*) in each eye will be used to identify subjects who meet age-normal distance visual acuity. Subjects will be classified as meeting age-normal distance visual acuity if the best reported visual acuity in both eyes are within age-normal range.

3.4 Presence of Strabismus at 36 Months

A treatment group comparison of the proportion of subjects who meet failure criteria for strabismus at 3 years, defined as having measureable heterotropia at distance, having measureable heterotropia at near, or having strabismus surgery prior to 3 years, will be performed using Barnard's exact test.

Once deterioration criteria are met, masked exams were waived for the remainder of the study, except for the 36 month exam. If a subject were to deteriorate due to poor stereoacuity or visual acuity, they would not be retested if measurable heterotropia was present at a subsequent exam. Hence it is not possible to determine whether strabismus failure criteria, which require a retest, were met at any subsequent exam other than the 36 month exam

3.5 Best Near Stereoacuity at 36 Months

A treatment group comparison of best stereoacuity (log arcseconds) will be performed using an ANCOVA model, adjusting for age at the 36-month examination and the magnitude of anisometropia (using the most recent cycloplegic refraction).

Best near stereoacuity is defined as the minimum stereoacuity between measurements reported with and without trial frames, during initial assessment and retest, if required. If no retest is required at 36 months, best stereoacuity is the smaller of the measurements reported with and without trial frames during the initial assessment. If a retest is required at 36 months, best near stereoacuity is the minimum of measurements reported with and without trial frames, during initial and retest at 36 months. If a 37-month visit is required, best near stereoacuity is the minimum of measurements reported for initial assessment and retest without trial frames at 36 months and the initial assessment at 37 months, if no retest is required at 37 months. If a retest is required at 37 months, best near stereoacuity is the minimum of measurements reported during initial assessment and retest without trial frames at 36 months and initial assessment and retest with trial frames at 37 months.

For the purposes of this analysis, a value of 1600 will be assigned to participants with no detectable stereoacuity (nil).

3.6 Proportion of Subjects with Age-Normal Stereoacuity at 36-Months

A treatment group comparison of the proportion of subjects meeting age-normal near stereoacuity at 36 months will be performed using Barnard's exact test.

Subjects are considered to have age-normal stereoacuity if their best reported near stereoacuity (*see section 3.5*) is within age-normal range.

3.7 Subgroup Analysis

In accordance with NIH guidelines, a subgroup analysis of treatment effect according to gender, as well as race/ethnicity, will be conducted. Binomial regression will be used evaluate if the difference between treatment groups in failure rate differs by subgroup. A model will be fit for

each factor of interest and will include the factor, treatment group, and an interaction term between the treatment group and the factor of interest.

Treatment effects in subgroups based on baseline factors will also be assessed. The subgroups of interest include baseline spherical equivalent anisometropia and baseline spherical equivalent refractive error. Failure status at 36 months post-randomization will be tabulated by subgroup and reviewed for consistency. The subgroup definitions for the planned subgroup analyses are as follows:

- Spherical equivalent anisometropia: <0.50D, 0.50D - <1.00D, 1.00D to 1.50D
- Spherical equivalent refractive error: +3.00D to <+4.00D, +4.00D to <+5.00D, 5.00D to 6.00D, $\geq 6.00D$

A Poisson regression using robust variance estimation will be used if the binomial regression model does not converge. If convergence issues are encountered when using the Poisson model, logistic regression will be used to make statistical inferences. For consistency, the same model will be used for all analyses. If the Poisson model is required for one of the analyses, the Poisson model will be used in all analyses.

4.0 Exploratory Analyses

Exploratory analyses will be performed separately for each primary cohort. Analyses will be considered statistically significant if the p-value is less than 0.01.

4.1 Presence of Amblyopia at 36 Months

The number and proportion of subjects in each treatment group that have amblyopia at the 36-month outcome will be calculated separately in each primary cohort. The best reported visual acuity in each eye (*see section 3.2*) will be used to calculate interocular difference (IOD). Amblyopia will be defined as ≥ 2 logMAR lines of IOD if visual acuity is 20/25 or worse in the better-seeing eye, and ≥ 3 logMAR lines of IOD if visual acuity is 20/20 or better in the better-seeing eye. A treatment group comparison of the proportion of subjects who developed amblyopia will be performed using a Barnard's exact test.

4.2 Binocular Near Visual Acuity at 36 Months

A treatment group comparison of near visual acuity at the 36-month outcome exam will be performed using an ANCOVA model, adjusting for age at the 36-month visit.

4.3 Change in Refractive Error from Baseline to Outcome

An analysis of covariance model will be used to compare the change in spherical equivalent (SE) refractive error in the eye that was more hyperopic at baseline from baseline to 36 months between treatment groups. The model will adjust for SE refractive error at baseline. The analysis will be repeated for the eye classified as less hyperopic at baseline.

If the change in refractive error differs at 36 months, an analysis of covariance model will be used to evaluate whether the relationship between the relationship between age at baseline and change in refraction from baseline to 36 months in the more hyperopic eye differs by treatment group. The model will include age at baseline, treatment group, and an interaction between age at baseline and treatment group. Analyses will be repeated to evaluate whether the relationship between age at baseline and change in refractive error from baseline to 36 months in the less hyperopic eye differs by treatment group.

4.4 Factors Predictive of Failure and 36-Month Visual Outcomes

The following characteristics will be evaluated as potential predictors of failure rate, mean best reported distance visual acuity (logMAR), best reported stereoacuity (log arcsecond), and measureable heterotropia; all assessed at the 36-month examination, or 37-month exam if required:

- Accommodative response at baseline
- Family history of amblyopia or strabismus at baseline
- Use of ADHD medication at any time before the 36-month visit

For this analysis, the accommodative response will be defined as a composite of the lens power and the retinoscopy. Eyes for which the reflex is classified as “with motion” will maintain a positive lens power, and a negative lens will be converted to a negative value if the reflex is classified as “against motion”. If the reflex is “neutral”, a value of 0 will be imputed for the lens power.

Linear regression will be used to evaluate accommodative response as a predictor of mean best distance visual acuity and of best near stereoacuity. AN(C)OVA models will be used to evaluate family history of amblyopia or strabismus and use of ADHD medication as predictors of mean best distance visual acuity and of best near stereoacuity. Logistic regression will be used to evaluate each baseline characteristic as a predictor of failure rate and measureable heterotropia. Each model will evaluate the relationship between one potential predictor and one dependent variable.

Seeing as the study aims to obtain the best outcome for each subject, subjects will be classified as having measureable heterotropia if all reported results indicate that measureable heterotropia is present. If any of the required assessments indicate that measureable heterotropia is not present, the subject will not be classified as having measureable heterotropia.

5.0 General Principles for Analysis

5.1 Classifying Visits

The analysis visit window for the randomized clinical trial will be as follows:

1. 6-month – >3.5 to 9 months (106 to 274 days) after randomization
2. 12-month – >9 to 15 months (275 to 456 days) after randomization
3. 18-month – >15 to 21 months (457 to 639 days) after randomization
4. 24-month – >21 to 27 months (640 to 822 days) after randomization
5. 30-month – >27 to 33 months (823 to 1004 days) after randomization
6. 36-month – >33 to 42 months (1005 to 1282 days) after randomization

A visit will be considered missed if it is completed out of window or not completed, but a later visit is completed. There is no analysis window for the 37-month visit. Data from the 37-month visit will be used in analysis if the visit is completed.