Bilateral Lateral Rectus Recession Versus Unilateral Recess-Resect for Intermittent Exotropia

Statistical Analysis Plan / Technical Plan

March 1, 2017

Version 1.1

IXT1 Three-Year Analysis

1.1 Objective

To compare 3-year outcomes between patients treated with bilateral lateral rectus muscle recession (BLR) versus those treated with unilateral lateral rectus recession (R/R)

1.2 Cohort of Interest

197 patients with basic-type IXT with largest preoperative exodeviation between 15 and 40 PD by PACT at remote distance, distance, or near (101 in BLR group, 96 in R/R group).

1.3 Primary Outcome – Surgical Failure by 3 Years

The primary outcome of surgical failure by 3 years is defined as follows:

Failure = ANY of the following criteria are met at masked exam occurring between 6 months and 3 years after randomization:

1. **Exotropia** at distance OR near at any time during the exam (i.e., can be constant or intermittent; determined by a cover/uncover test) with a magnitude of atleast 10 PD by SPCT, confirmed by a retest
2. **Constant esotropia** at distance OR near (determined by at least 3 cover/uncover tests—one must be before any dissociation) with a magnitude of atleast 6 PD by SPCT, confirmed by a retest
3. Decrease in Preschool Randot near **stereoacuity** atleast 2 octaves (atleast 0.6 log arcsec) (*see Table 3*) from the enrollment measurement, or to nil, confirmed by a retest

Table 1: Preschool Randot Stereotest

Baseline stereoacuity at enrollment, in arcsec	Level needed at follow up visit to meet surgical failure criteria, in arcsec
40”	200” or worse
60”	400” or worse
100”	400” or worse
200”	800” or worse
400”	Nil

Patients will also be considered a surgical failure for analysis if they undergo **reoperation** or treatment with botulinum toxin at any time during the study.

1.4 Primary Analysis

The cumulative proportion of patients meeting criteria for failure by 3 years will be obtained using the Kaplan-Meier method and compared between treatment groups using the Z test. This will allow patients who drop out prior to 3 years to contribute to the estimation of the proportion of surgical failure at 3 years. In this analysis, all patients who meet surgical failure criteria prior to 3 years will be counted as failures at the first visit at which surgical failure criteria are met.

Patients who withdraw from the study or are lost to follow up without having met surgical failure criteria or being reoperated will be right-censored as non-failures at the last study visit completed.

1.5 Principles to be followed in Primary Analysis

- The primary analysis will follow the intention-to-treat principle in that all patients will be analyzed according to their randomized treatment group, regardless of whether/what treatment was received.
- The primary analysis will include all patients, including those who were enrolled but later found to be ineligible.
- The primary analysis will also include patients who did not receive surgery, so that each randomized patient can be accounted for. Inclusion of patients who did not receive surgery has no impact on the K-M cumulative probability of failure because these patients withdrew from the study without completing any follow up visits and are therefore considered censored at time 0, before the first failure occurs and the cumulative probability is calculated.
- For determining whether surgical failure criteria are met, the masked exams from all protocol-specified and unspecified visits will be evaluated. *It was acknowledged that inclusion of unspecified visits may bias the treatment group comparison if one treatment group is seen more frequently than the other, and thus has more opportunities for the event to be observed, and more opportunity for misclassification. It was agreed to discuss the issue further before deciding on how to handle this in the manuscript; however, unspecified visits are included in the abstract analyses. They have little impact given that all patients who met failure criteria at an unspecified visit were reoperated a short time afterward (and so would have been considered failure because of the reoperation).*
- All masked exams that were at least partly completed will be evaluated for whether surgical failure criteria are met. For example, a patient could meet surgical failure due to meeting constant esotropia criteria even if stereoacuity was not able to be obtained at the masked exam (e.g. stereo test was not at the location where the patient was seen). Patients who did not meet surgical failure on the basis of partial masked exam data were classified as not meeting failure criteria for that visit.
- Patients who appear to have met surgical failure criteria by initial testing but who did not complete all required retesting for that criteria are retained in the analysis and are considered not to have met surgical failure criteria.
- Patients who have not yet met surgical failure are considered to retain their non-failure status throughout any subsequent consecutive missed visit(s) until this status is potentially changed at a completed visit. For example, a patient who is a non-failure at a completed 1 year visit, misses the 18-month and 2-year visits, and is classified as a failure at a completed 30-month visit, the non-failure status from the 1 year visit is maintained until the 30-month visit.

1.6 Secondary Analysis -- Surgical Failure at 3 Year Time point

The binomial proportion of patients who meet surgical failure criteria *at* the 3 year visit (as opposed to *by* the 3 year visit) will be estimated for each treatment group and compared using Fisher's exact test.

Patients who do not return for the 3 year visit will not be included in the analysis, including patients who met surgical failure criteria at an intermediate visit or were reoperated. Patients

who complete the visit will be classified based on their status at 3 years, regardless of whether they met surgical failure criteria at an earlier time point, unless they have been re-operated (or treated with botulinum toxin), in which case they will be classified as a surgical failure.

The potential for bias in this treatment group comparison is recognized. Once a patient has met the clinical criteria for surgical failure criteria at an interim follow up visit, the decision to reoperate—and thus permanently classify the patient as a surgical failure for the analysis *at* 3 years—is at the discretion of an unmasked investigator and therefore could be related to treatment group. To assist in assessing for potential bias, the association between treatment group and reoperation in those meeting surgical failure criteria will be evaluated.

1.7 Secondary Analysis -- Reoperation by 3 Years

The cumulative proportion of patients undergoing reoperation or treatment with botulinum toxin by 3 years will be obtained using the Kaplan-Meier method and compared between treatment groups using the Z test. This outcome will include all cases of reoperation—cases where reoperation was completed after surgical failure was met in addition to cases where reoperation occurred without surgical failure having been met (i.e. against protocol).

The potential for bias in this treatment group comparison is recognized. Once a patient has met the clinical criteria for surgical failure criteria at an interim follow up visit, the decision to reoperate is at the discretion of an unmasked investigator and therefore could be related to treatment group. To assist in assessing for potential bias, the association between treatment group and reoperation in those meeting surgical failure criteria will be evaluated.

1.8 Secondary Analysis – 3-Year Exotropia Control and Angle Magnitude

Secondary outcomes of 3-year exotropia control (distance and near) and 3-year angle magnitude by the Prism and Alternate Cover Test (distance and near) will be assessed in all patients who complete the 3-year visit. All 3-year visit data will be analyzed regardless of what treatment(s) a patient has received and regardless of whether the patient has undergone reoperation. These 3-year control and PACT outcomes will be analyzed as continuous variables and compared between treatment groups using analysis of covariance (ANOVA) models that adjust for the corresponding baseline value (e.g. ANCOVA model of 3-year distance control will adjust for baseline distance control).

Objective #1: Define the cohort of interest

Objective #2: Compare the cumulative probability of surgical failure BY 3 years between BLR and R/R treatment groups (primary outcome)

Objective #3: Compare the binomial proportion of surgical failure AT 3 years between BLR and R/R treatment groups

Objective #4: Compare the cumulative probability reoperation by 3 years between BLR and R/R treatment groups

Objective #5: Compare 3-year control and PACT values between BLR and R/R treatment groups

Datasets Used

BASELINE - one-record per baseline exam for all patients enrolled into in IXT1 (regardless of whether randomized) N=277

MASKEDEXAMS - one-record per IXT1 masked exam (protocol-specified or unspecified) that was at least partially completed IXT1 N=1344

ROSTER – one-record per randomized patient analysis dataset N=265

Note that the above permanent datasets include all IXT1 patients, but the analysis was limited to the cohort of interest for this abstract.

Objective #1: Define the cohort of interest

1. 197 patients with basic-type IXT with largest preoperative exodeviation between 15 and 40 PD by PACT at remote distance, distance, or near (101 in BLR group, 96 in R/R group)

Technical plan

1. Limit the patient-level dataset ROSTER to patients where the STRATUM variable from tblStratum, the variable used to stratify the randomization, = 'Basic IXT with 15-40PD angle'

Dataset used: ROSTER

Objective #2: Compare the cumulative probability of surgical failure BY 3 years between BLR and R/R treatment groups (primary outcome)

1. Define the outcome
2. Obtain masked exam records
3. Determine whether exotropia failure criterion was met for each masked exam
4. Determine whether constant esotropia failure criterion was met for each masked exam
5. Determine whether stereoacuity failure was met for each masked exam
5. Determine whether patient was reoperated or underwent treatment with botulinum toxin
6. Calculate surgical failure at patient level and set timing variable for survival analysis (time to failure or censoring time)
7. Get cumulative probability of surgical failure by 3 years for each treatment group – from K-M
8. Compare cumulative probability of surgical failure by 3 years between treatment groups using a two-sided Z-test
9. Calculate the treatment group difference (and 95% CI) in the cumulative probability of surgical failure by 3 years

Technical plan

1. Define the outcome.

Failure = ANY of the following criteria are met at masked exam occurring between 6 months and 3 years after randomization:

1. **Exotropia** at distance OR near at any time during the exam (i.e., can be constant or intermittent; determined by a cover/uncover test) with a magnitude of atleast 10 PD by SPCT, confirmed by a retest
2. **Constant esotropia** at distance OR near (determined by at least 3 cover/uncover tests—one must be before any dissociation) with a magnitude of atleast 6 PD by SPCT, confirmed by a retest
3. Decrease in Preschool Randot near **stereoacuity** atleast 2 octaves (atleast 0.6 log arcsec) (*see Table 3*) from the enrollment measurement, or to nil, confirmed by a retest

Table 3. Preschool Randot Stereoacuity

Baseline Preschool Randot stereoacuity at enrollment, in arcsec	Level needed at follow up visit to meet surgical failure criteria, in arcsec
40”	200” or worse
60”	400” or worse
100”	400” or worse
200”	800” or worse
400”	Nil

Patients will also be considered a surgical failure for analysis if they undergo **reoperation** or treatment with botulinum toxin at any time during the study.

2. Obtain masked exam records from tblIXT1MaskedExam.

Include masked exams from all protocol-specified and unspecified visits.

- *It was acknowledged that inclusion of unspecified visits may bias the treatment group comparison if one treatment group is seen more frequently than the other, and thus has more opportunities for the event to be observed, and more opportunity for misclassification. It was agreed to discuss the issue further before deciding on how to handle this in the manuscript; however, unspecified visits are included in the abstract analyses but have little impact given that all patients who met failure criteria at an unspecified visit were reoperated a short time afterward (and so would have been considered failure because of the reoperation).*

The masked exam form was a required section of data entry on the web for all follow up visits, regardless of whether the masked exam was completed or was even required. Masked exams records where the field maskedexamnotdone (for protocol-specified visits) or the fields maskedexamnotreq or maskedexamreqnotdone (for unspecified visits) are set to 1 represent masked exams that were not completed either because they could not be completed or because they were not required. These records should be reviewed to confirm that they do not contain data and then excluded from the analysis.

3. Determine whether exotropia failure criteria were met for each masked exam

- Evaluate all masked exams. *Even though only the first masked exam where failure criteria is met is relevant to the primary outcome, save exotropia failure criteria flag in a masked-exam-level dataset because interested in whether this criteria is met at the 3 year visit also, and may also be interested in other visits as well.*
- Create a numeric variable for SPCT magnitude by setting '>50' equal to a nonsense value of 888. *Note that no means will be calculated on this variable.*
- Exotropia failure criteria is met if the masked exam shows the patient has an exotropia of 10 or greater at distance *or* near by SPCT, confirmed by a retest. If the worsening was not confirmed by the retest or the retest was not completed, the patient was considered not to have met exotropia failure criteria. Requires the following:
  - SPCT of >=10PD at distance or near on initial testing and retesting
  - Tropia type = 'Exo' at distance or near on initial testing and retesting
  - Note that corresponding size and type must meet above criteria for the same distance for initial and retest.
- *Unlike IXT2, the exotropia <u>does not</u> need to be constant to meet criteria and <u>does not</u> need to occur at both distance and near.*

4. Determine whether constant esotropia failure criteria were met for each masked exam

- Evaluate all masked exams. *Even though only the first masked exam where failure criteria is met is relevant to the primary outcome, save constant esotropia failure criteria flag in a masked-exam-level dataset because interested in whether this criteria is met at the 3 year visit also, and may also be interested in other visits as well.*
- For each masked exam, determine whether constant esotropia failure criterion was met.
- Use SPCT variables created above.

- Constant esotropia failure criterion is met if the masked exam shows the patient has an esotropia of 6 or greater at distance *or* near (throughout exam) by SPCT, confirmed by a retest. If the worsening was not confirmed by the retest or the retest was not completed, the patient was considered not to have met constant esotropia failure criteria. Requires the following:
    - SPCT of >=6PD at distance and at near on initial testing and retesting
    - Tropia type = 'Eso' at distance and at near on initial testing and retesting
    - Assessment of esodeviation throughout exam = 'Constant esotropia' at time of initial testing and at time of retesting

5. Determine whether stereoacuity failure was met for each masked exam
- Evaluate all masked exams. *Even though only the first masked exam where failure criteria is met is relevant to the primary outcome, save stereoacuity failure criteria flag in a masked-exam-level dataset because interested in whether this criteria is met at the 3 year visit also, and may also be interested in other visits as well.*
- Determine the best stereoacuity at the baseline visit. Note that stereo was to be retested unless the patient scored 40 arcsec on the initial test.
- Compare masked exam initial to best baseline stereo and determine whether meets criteria for 2 or more level worsening (use Table 1 under step 1).
- If the worsening was not confirmed by the retest or the retest was not completed, the patient was considered not to have met stereoacuity failure criteria.

6. Determine whether patient was reoperated or underwent treatment with botulinum toxin
- Get treatment used records from all visits, regardless of whether a masked exam was completed
- If REOPERATION = 1 for any treatment used, and set patient-level reoperation to 1 and capture reoperation date

7. Calculate surgical failure at patient level and set timing variable for survival analysis (time to failure or censoring time)
- Loop through masked exam records for each patient and determine the first masked exam at which any of the three objective failure criteria were met.
- In patient-level dataset:
    - If reoperation occurs and surgical failure has not been met (either not at all or not by the time of reoperation), failure = 1 and failure time = months between surgery and reoperation
    - If one of the surgical failure criteria were met, either before reoperation or in a patient who is not reoperated, failure = 1 and failure time is based on visit type (e.g. 6 months, 12 months, etc.) or months between failure and surgery if failure occurs at an unspecified visit
    - If patient does not meet surgical failure and is not reoperated, failure = 0 and failure time is based on type of last completed visit (e.g. 6 months, 12 months, etc.)

8. Get cumulative probability of surgical failure by 3 years for each treatment group – from Kaplan-Meier survival analysis

- Run Kaplan-Meier survival analysis using proc lifetest, and specifying the method as Kaplan-Meier.
- Output survival probabilities and confidence intervals to a dataset
- Create failure estimates and confidence intervals

```
/*****************************************************************
   PERFORM K-M ANALYSIS
   PRIMARY OUTCOME
   CUMULATIVE PROBABILITY OF SURGICAL FAILURE
   NOTES FOR K-M SURVIVAL ANALYSIS
   NUMBER AT RISK = NUMBER AT RISK GOING INTO THE VISIT
                    (E.G. LAST PERSON AT 3 MONTHS BEFORE 6 MONTHS)
   OUTSURV OPTION IN PROC LIFETEST STATEMENT CREATES AN OUTPUT DATASET
   THAT CONTAINS SURVIVAL ESTIMATES AND CONFIDENCE LIMITS.
   PARENTHETICAL IN TIME STATEMENT INDICATES WHAT CENSORING VALUE IS
*****************************************************************/

%sort (roster, trtgroup);
proc lifetest data = roster method=km  outsurv=failresults plots=none alpha=.05;
      time failtime*fail(0);
         by trtgroup;
         title2 'K-M Survival Analysis for Surgical Failure';
run;

proc print data = failresults;
     title2'Review Output Dataset from K-M Survival Analysis for Surgical Failure';
run;

data failresultCIs;
     set failresults;

        /* create failure estimates (rather than survival)*/
        failure = 1 - survival;
        failureLCL = 1 - SDF_UCL;
        failureUCL = 1 - SDF_LCL;

     /* limit to records where the survival estimate has changed
          (i.e. records where the CI is not null) */

     if SDF_LCL NE . then output;

        label failure = 'Cum. probability of surgical failure';
        label failureLCL = 'Lower limit of CI for surgical failure';
        label failureUCL = 'Upper limit of CI for surgical failure ';

run;

%sort (failresults, trtgroup);
proc print data = failresultCIs;
     by trtgroup;
     title2'Review Output Dataset from K-M Survival Analysis for Surgical Failure';
run;
```

9. Compare cumulative probability of surgical failure by 3 years for each treatment group using a Z-test code

See below—combined with step #10.

10. Calculate the treatment group difference (and 95% CI) in the cumulative probability of surgical failure by 3 years.

```
/* Manually enter cumulative probabilities and standard error from K-M into short
program to calculate Z-score, its corresponding P value, the treatment group
difference and 95% CI */

data check;
   input probblr seblr probrr serr;
   datalines;
   0.4594 0.0518 0.3731 0.0520
;
run;

data check;
     set check;
        diff = probblr - probrr;
         sumofsquaredse = sqrt (seblr**2+serr**2);
        cilower = diff - (1.96*sumofsquaredse);
        ciupper = diff + (1.96*sumofsquaredse);
        zscore = diff/sumofsquaredse;
     pvalue = 2*(1 - probnorm(zscore));
run;

proc print data = check noobs;
     var analysis probblr seblr probrr serr sumofsquaredse diff cilower ciupper zscore
pvalue;
        title'Calculate CIs and P values;
run;
```

Datasets used: MASKEDEXAMS, BASELINE, ROSTER, tblIXT1Treatused (SQL)

Objective #3: Compare the binomial proportion of surgical failure AT 3 years between BLR and R/R treatment groups

1. Define the outcome
2. Determine whether exotropia failure criterion was met at the 3-year masked exam
3. Determine whether constant esotropia failure criterion at the 3-year masked exam
4. Determine whether stereoacuity failure criterion was met at the 3-year masked exam
5. Determine whether patient was reoperated or underwent treatment with botulinum toxin, regardless of whether he/she first met one of the three surgical failure criteria
6. Create surgical failure at 3 years
7. Calculate binomial proportion, treatment group difference, and 95% exact confidence intervals

Technical plan

1. Define the outcome
   - The secondary outcome of surgical failure AT 3 years (not BY 3 years) is defined as follows:

Failure = ANY of the following criteria are met atthe3-yearmaskedexam:

1. **Exotropia** at distance OR near at any time during the exam (i.e., can be constant or intermittent; determined by a cover/uncover test) with a magnitude of atleast 10 PD by SPCT, confirmed by a retest
2. **Constant esotropia** at distance OR near (determined by at least 3 cover/uncover tests—one must be before any dissociation) with a magnitude of atleast 6 PD by SPCT, confirmed by a retest
3. Decrease in Preschool Randot near **stereoacuity** atleast 2 octaves (atleast 0.6 log arcsec) (*see Table 3*) from the enrollment measurement, or to nil, confirmed by a retest (see Table 1 from objective #2)

Patients will also be considered a surgical failure at 3 years if they undergo **reoperation** or treatment with botulinum toxin at any time during the study.

The outcome is assessed only in patients who complete the 3-year visit. To prevent biasing the estimates, patients who were reoperated before being lost to follow up will not contribute to the analysis (even though their surgical failure at 3 years status would have been permanently set when they were reoperated, if had they completed the 3-year visit).

2. Determine whether exotropia failure criterion was met at the 3-year masked exam
   - Created as part of objective #2 -- get data from 3-year visit record from MASKEDEXAMS dataset.

3. Determine whether constant esotropia failure criterion at the 3-year masked exam
   - Created as part of objective #2 -- get data from 3-year visit record from MASKEDEXAMS dataset.

4. Determine whether stereoacuity failure criterion was met at the 3-year masked exam
   - Created as part of objective #2 -- get data from 3-year visit record from MASKEDEXAMS dataset.

5. Determine whether patient was reoperated or underwent treatment with botulinum toxin, regardless of whether he/she first met one of the three surgical failure criteria
   - Use patient-level reoperation flag and reoperation date created for objective #2.

6. Create surgical failure at 3 years
   - For patients who completed the 3-year visit:
     - Code as 1 if any of the three surgical failure criteria are met at 3 years or if reoperation occurred at any time.
     - Otherwise code as 0

7. Calculate binomial proportion for each treatment and compare between treatment groups with a Fisher's exact test. Calculate treatment group difference, and 95% exact confidence intervals.

```
proc freq data = roster;
     tables trtgroup*failat36;
        exact fisher riskdiff;
        where vis_36 = 'Completed';
        title1 'Comparison of Crude % with Failure ***AT*** 3 Years';
run;
```

Datasets used: MASKEDEXAMS, ROSTER

Objective #4: Compare the binomial proportion with reoperation by 3 years between BLR and R/R treatment groups

1. Define the outcome
2. Determine whether patient was reoperated or underwent treatment with botulinum toxin, regardless of whether patient first met one of the three surgical failure criteria
3. Get cumulative probability of reoperation by 3 years for each treatment group – from K-M
4. Compare cumulative probability of reoperation by 3 years for each treatment group using two-sided Z-test
5. Calculate the treatment group difference (and 95% CI) in the cumulative probability of reoperation by 3 years.

Technical Plan

1. Define the outcome

Reoperation or botulinum toxin treatment at any time during the study, including cases where reoperation was completed after surgical failure was met and cases where reoperation occurred without surgical failure first having been met (i.e. against protocol).

2. Determine whether patient was reoperated or underwent treatment with botulinum toxin, regardless of whether he/she first met one of the three surgical failure criteria
   - Use reoperation outcome and time to reoperation variables created for objective #2.

3. Get cumulative probability of reoperation by 3 years for each treatment group – from K-M
4. Compare cumulative probability of reoperation by 3 years for each treatment group using Z-test
5. Calculate the treatment group difference (and 95% CI) in the cumulative probability of reoperation by 3 years.

Repeat steps #8 - #10 for objective #1 using the reoperation outcome and time to reoperation variables.

Objective #5: Compare 3-year control and PACT values between BLR and R/R treatment groups

1. Define the outcomes
2. Define the cohort
3. Compare mean 3-year outcomes between treatment groups

Technical Plan

1. Define outcomes.

a. PACT size (distance and near)

- Code according to size and type
  - Exodeviations will be coded as positive values (same as in database)
  - Esodeviations will be changed to negative values

b. Exotropia Control (distance and near)

- For both distance and near, create numeric values for exotropia control where 'not applicable' will be assigned a score of 0, the same as the score for a pure phoria. Note that 'not applicable' was entered on the form when no exodeviation was present.

Table #1: Intermittent Exotropia Control Scale Scoring

5	Constant Exotropia
4	Exotropia > 50% of the 30-second period before dissociation
3	Exotropia < 50% of the 30-second period before dissociation
2	No exotropia unless dissociated, recovers in >5 seconds
1	No exotropia unless dissociated, recovers in 1-5 seconds
0	No exotropia unless dissociated, recovers in <1 second (phoria)
Not applicable	No exodeviation present

For both PACT and control outcomes, use the 3-year visit data for all patients who completed the 3-year visit, regardless of what treatment(s) were received or if the patient had undergone reoperation.

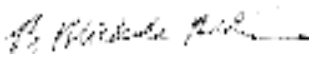
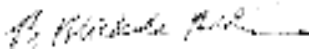
2. Limit the analysis to patients who completed the 3-year visit.

3. Compare mean 3-year outcome between treatment groups using ANCOVA model.

```
proc genmod data = roster;
   class trtgroup;
   model controlnumdi_36 = trtgroup controlnumdi_0;
   where comp_36 = 1;
   title1'Comparison of 3-year distance control between treatment groups';
run;
```

Datasets used: MASKEDEXAMS, ROSTER

511 **Version History**

Version Number	Author	Approver	Effective Date	Revision Description
1.0	Danielle Chandler	Michele Melia 	1-19-17	Original SAP for outcome data included in submitted AAPOS abstract (note that the analyses were specified in the protocol).
1.1	Danielle Chandler	Michele Melia 	3-1-17	For the purpose of the AAPOS presentation, added section 1.8 on secondary analyses of 3-year exotropia control and 3-year angle magnitude (note that these analyses were specified in the protocol).

512

IXT1 Primary Manuscript on Basic Primary Cohort

Statistical Analysis Plan
for Secondary Outcomes Not Covered in Previous Analysis Plans

May 30, 2018

Revision History

VERSION NUMBER		AUTHOR	APPROVER	EFFECTIVE DATE	REVISION DESCRIPTION (INCLUDING SECTIONS REVISED)
SAP	Protocol				
1.0	5.0 6-21-17	Danielle Chandler	Michele Melia	5/14/18	Initial version
1.1	5.0 6-21-17	Danielle Chandler	Michele Melia	6/11/18	In the SAS code on line 425 that creates a p value using a z-score, changed to using the absolute value of the z score (instead of the z score itself) to account for situations in which a two-sided test yields a negative z-score.

Important Notes:

The primary analysis and many of the secondary analyses were completed for several abstracts and presentations preceding completion of the primary manuscript. Below is a list of outcomes which are documented in previous analysis plans plus outcomes which are covered herein.

AAPOS Abstract and Presentation 2017

- **Surgical failure by 3 years (primary outcome) (relabeled "suboptimal surgical outcome" at 3 years)**
- **Surgical failure at 3 years (relabeled "suboptimal surgical outcome" at 3 years)**
- **Reoperation by 3 years**
- **3-year PACT magnitude (distance and near)**
- **3-year control (distance and near)**

Notethattheabovevariableswereoriginallydocumentedandcreatedatthetimeoftheaapos presentationusingpre-closeoutdatainJanuary2017;however,aspartoftheverificationoftheESA2017 presentationinAugust2017,thesevariableswererunusingpost-closeoutdataandincorporatedintothe ROSTERdatasetusedforESA. Themanuscriptdatasetwillstartwithesadatasetandaddadditional dataneededforthemanuscript.

The current document contains the following outcomes:

- **Exotropia failure by 3 years (using exotropia criteria of surgical failure)**
- **Constant esotropia failure outcome (using constant esotropia criteria of surgical failure)**
- **Stereoacuity failure outcome (using stereoacuity loss criteria of surgical failure)**
- **3-year stereoacuity (distance and near)**
- **Complete or near-complete resolution at 3 years without regard to previous surgical failure (post hoc)**
- **Complete or near-complete resolution at 3 years with no previous surgical failure (pre-specified)**
- **IXT Questionnaire (IXTQ) scores for parent, proxy and child components**
- **Additional tabulations**

1.1 Objective

To compare 3-year outcomes between patients treated with bilateral lateral rectus muscle recession (BLR) versus those treated with unilateral lateral rectus recession (R/R)

1.2 Cohort of Interest

Briefly, the IXT1 study enrolled patients who were 3 to <11 years of age, had stereoacuity of 400 arcsec or better, and were undergoing surgery for intermittent exotropia (N = 265 overall). The cohort of interest is the primary cohort of 197 patients with basic-type IXT and with largest preoperative exodeviation between 15 and 40 PD by PACT at remote distance, distance, or near (101 in BLR group, 96 in R/R group).

1.3 Cause-specific surgical failure outcomes

As secondary analyses, three cause-specific suboptimal outcomes by 3 years were specified post hoc, for the exotropia, constant esotropia, and stereo loss criteria defined in the primary outcome (Table 1). These cause-specific outcomes differ from the primary outcome in two ways: 1) the primary outcome refers to the first occurrence of *any* suboptimal outcome criterion (or re-operation) being met, whereas the cause-specific outcomes refer to the first occurrence of the *particular* suboptimal outcome criterion being met, and 2) reoperation prior to meeting a particular suboptimal outcome criteria was considered an *suboptimal outcome* for the primary analysis but was censored as a *non-outcome* in the analysis that evaluated cause-specific outcomes. For each of the three cause-specific outcomes, participants who met criteria *other than the particular criteria being assessed* remained "at risk" for the criterion of interest unless they underwent reoperation. For example, participants who met the stereo loss outcome remained "at risk" for the exotropia and constant esotropia outcomes until they either met them or underwent reoperation. The cumulative probability of each cause-specific outcome by 3 years and a 95% CI were obtained using the K-M method. It is acknowledged that the three cause-specific outcomes are not independent because reoperation is a competing risk for each (e.g., participants who met the exotropia outcome and underwent reoperation were no longer at risk for meeting stereo loss or constant esotropia outcomes).

The cumulative probability (and 95% confidence interval) of each cause-specific outcome will be determined using Kaplan-Meier method.

Table 1: Three Objective Criteria for Surgical Failure

Based on a masked exam occurring between 6 months and 3 years after randomization:

1. **Exotropia** at distance OR near at any time during the exam (i.e., can be constant or intermittent; determined by a cover/uncover test) with a magnitude of atleast 10 PD by SPCT, confirmed by a retest
2. **Constant esotropia** at distance OR near (determined by at least 3 cover/uncover tests—one must be before any dissociation) with a magnitude of atleast 6 PD by SPCT, confirmed by a retest
3. Decrease in Preschool Randot near **stereoacuity** atleast 2 octaves (atleast 0.6 log arcsec) (*see Table 3*) from the enrollment measurement, or to nil, confirmed by a retest

Baseline stereoacuity at enrollment, in arcsec	Level needed at follow up visit to meet surgical failure criteria, in arcsec
40"	200" or worse
60"	400" or worse
100"	400" or worse
200"	800" or worse
400"	Nil

Note that reoperation without having met any of the three objective criteria was also considered a surgical failure in the primary analysis.

1.4 Secondary Analysis – 3-Year Stereoacuity

Secondary outcomes of 3-year stereoacuity (distance and near) will be assessed in all patients who complete the 3-year visit. The 3-year visit stereoacuity will be analyzed regardless of what treatment(s) a patient has received and regardless of whether the patient has undergone reoperation. These 3-year stereoacuity outcomes will be analyzed as continuous variables and compared between treatment groups using analysis of covariance (ANOVA) models that adjust for the corresponding baseline stereoacuity.

1.5 Complete or Near-complete Resolution at 3 Years

Complete or near-complete resolution at 3 years was defined post hoc as meeting all of the following at the 3-year visit:

1. Exodeviation <10 PD (tropia, phoria, or no deviation) by both SPCT and PACT at distance and near and ≥10 PD reduction in PACT magnitude from distance and near angles at enrollment provided the corresponding angle was at ≥10PD at baseline.
    - This criterion was originally written as ≥10 PD reduction in PACT magnitude from *largest of* distance and near angles at enrollment. After asking the protocol chairs to clarify which angle needed to be reduced by ≥10 PD at 3 years in cases where the distance and near angles at enrollment are the same magnitude (i.e. neither is the 'largest'), it was decided that both angles should be reduced by ≥10 PD in order to meet the criterion, if the angle was at ≥10PD at baseline, OR should be reduced to orthodeviation (0 PD) if the angle was at <10PD at baseline. *Note that because all patients in the basic primary cohort have distance and near PACT ≥10PD, this caveat is not cited in the manuscript as being part of the criteria. Only one patient in one of the secondary cohorts was <10PD at baseline—a 3PD near angle which was ortho at 3 years.* The revised criterion was extended to all patients, not only those whose enrollment PACT was the same magnitude at distance and near, for consistency. For example, if a patient with 20 PD at distance and near at baseline is required to have a reduction ≥10PD at both distance and near at 3 years in order to be eligible to meet complete or near-complete resolution, then a patient with 25 PD at distance and 15 PD at near should also be required to have both reduced by ≥10 PD at 3 years, otherwise the first patient is being penalized simply for having the same measurement at both distances.
2. Esotropia <6 PD (tropia or no tropia, *note that phoria does not apply to SPCT*) at distance and near by SPCT
3. No decrease in Preschool Randot stereoacuity of ≥2 octaves from the enrollment stereoacuity or to nil
4. No reoperation or treatment with botulinum toxin
5. No non-surgical treatment for a recurrent or residual exodeviation

Although complete or near-complete resolution was specified post hoc, a three-level failure/indeterminate/success outcome was prespecified in the protocol; the only difference between the "success" level and "complete or near-complete resolution" is that patients who met suboptimal outcome criteria at a previous visit but not at the 3-year visits were considered failures in the pre-specified outcome but could potentially be considered complete or near-complete resolutions if all other criteria were met.

The treatment group difference in the proportion of participants with complete or near complete resolution at 3 years was compared using Barnard's exact test and calculating a 95% CI using Farrington-Manning scores. Originally the Fisher's exact test was used, but the corresponding exact CI is calculated in a way in SAS that would permit potential disagreements on statistical significance between p values of <0.05 and the 95% CI, whereas Barnard's test and Farrington-Manning scores for the 95% CIs will agree on statistical significance.

Note that complete or near-complete resolution *without meeting suboptimal surgical outcome at any time* was also reported in the manuscript (consistent with the prespecified "success" criteria in the protocol) and compared between treatment groups using similar methods.

1.6 IXT Questionnaire (IXTQ)

For the IXTQ proxy questionnaires and for each of the three parent questionnaire subscales (psychosocial, functional, and surgical), mean Rasch-based QOL scores at 3 years were compared between treatment groups using linear regression models adjusting for the baseline score.

For the IXTQ child questionnaire, because some participants were too young for the IXTQ at baseline (those enrolled 3-<5 years of age) and because some children completed the 5 to 7-year old IXTQ version at baseline and the 8 years and older version at 3 years, the two age versions were evaluated separately and mean QOL scores at 3 years were compared between treatment groups using linear regression models that did not adjust for baseline. Because the Rasch scores can be difficult to interpret, the mean of the 0 to 100 scores based on the Rasch scores will be cited as the mean in each treatment group.

The assumptions of ANOVA/ANCOVA will be tested and if violated, a non-parametric test such as Wilcoxon rank sum test will be used.

1.8 Nonsurgical Treatment

Postoperative nonsurgical treatment for XT, ET, and/or diplopia was tabulated for each treatment group. Because the reason for this nonsurgical treatment was not specified other than being prescribed for IXT, ET, or diplopia, the type of deviation that was present when the nonsurgical treatment was prescribed was reported. This data was used to report the proportions of participants with non-surgical treatment prescribed when exodeviation was present, when esodeviation was present and when exodeviation and esodeviations were present at different times during the study. Among participants who met the constant esotropia suboptimal surgical outcome during the study, the proportion who had nonsurgical treatment prescribed was reported.

1.7 Additional Tabulations and Analyses

The following will be tabulated for each treatment group:

- Baseline demographic and clinical characteristics
- Baseline demographic and clinical characteristics for study completers vs. not
    - Study completion refers to patients who completed the 3-year visit or were withdrawn from the study after they met the primary outcome of suboptimal surgical outcome (for meeting objective criteria or for reoperation). It does not include patients who were withdrawn from the study without ever having met primary outcome of suboptimal surgical outcome.
- Percentage with completion of each follow up visit

- Listing of each complications that occurred either during surgery, or were reported at the 1-week or the 8-week postoperative visits
- Outcomes for participants meeting exotropia or esotropia suboptimal outcome criteria
    - Separately for participants who met the exotropia and esotropia components of suboptimal outcome criteria, the proportion out of those meeting the criterion who received nonsurgical treatment, and the proportion of non-reoperated cases in which the criterion of interest was not present at 3 years was reported.
- In subjects who completed the 3-year visit, 3-year status was evaluated according to whether suboptimal surgical outcome had occurred before 3 years
    - 3-year status was defined as: reoperation before 3 years, suboptimal surgical outcome at 3 years, complete or near-complete resolution at 3 years, or non-reoperated and meeting neither suboptimal surgical outcome or complete or near-complete resolution at 3 years.
    - Timing of suboptimal surgical outcome was categorized as: never met suboptimal surgical outcome, suboptimal surgical outcome before 3 years, suboptimal surgical outcome met only at 3 years

Datasets/Databases Used

SOURCE DATASETS/DATABASES USED TO CREATE FINAL IXT1 DATASETS

PREVIOUSLY VERIFIED SAS DATASETS CREATED FOR ESA 2017 PRESENTATION

BASELINE - one-record per baseline exam for all patients enrolled into in IXT1 (regardless of whether randomized) N=277
Location: F:\user\PEDIG\Manuscripts-Presentations\Manuscripts\IXT\IXT1\Manuscripts\Primary MS\Datasets\8-23-17 (verified)

MASKEDEXAMS - one-record per IXT1 masked exam (protocol-specified or unspecified) that was at least partially completed IXT1 N=1344
Location: F:\user\PEDIG\Manuscripts-Presentations\Manuscripts\IXT\IXT1\Manuscripts\Primary MS\Datasets\8-23-17 (verified)

ROSTER – one-record per randomized patient analysis dataset N=265
Location: F:\user\PEDIG\Manuscripts-Presentations\Abstracts and Presentations\ESA\ESA 2017\IXT1\Dataset\8-21-17

Note that the above permanent datasets include all IXT1 patients, but the analysis was limited to the primary cohort of basic-type IXT participants with angles ranging from 15 to 40Δ (N=197).

IXTQ.IXTQALL_19OCT2017 – one-record per IXTQ completed at any visit in IXT1 and IXT2 studies (N = 4790)
Location: F:\user\PEDIG\Manuscripts-Presentations\Manuscripts\IXT\IXT2\IXTQ\Manuscript Analysis\Datasets

- Note that the above permanent dataset was created by taking a previously-verified program using IXT2 data and re-running to read in data from both IXT1 and IXT2 (discussed at 2/13/18 manuscript meeting and confirmed that no additional verification is required). Note that the dataset includes all IXTQs completed at any visit in IXT1 and IXT2 studies; however, only data from enrollment and 3-year visits for IXT1 patients was added to the IXT1 final dataset.

POST CLOSEOUT SQL DATABASE: PEDIG_IXT1_3yrCloseout_20jun2017

FINAL DATASETS CREATED

ROSTER – one-record per randomized patient in IXT1 N=265
Location: F:\user\PEDIG\Manuscripts-Presentations\Manuscripts\IXT\IXT1\Manuscripts\Primary MS\Datasets\#-#-18

MASKEDEXAMS - one-record per IXT1 masked exam (protocol-specified or unspecified) that was at least partially completed IXT1 N=1344
Location: F:\user\PEDIG\Manuscripts-Presentations\Manuscripts\IXT\IXT1\Manuscripts\Primary MS\Datasets\#-#-18

BASELINE - one-record per baseline exam for all patients enrolled into in IXT1 (regardless of whether randomized) N=277
Location: F:\user\PEDIG\Manuscripts-Presentations\Manuscripts\IXT\IXT1\Manuscripts\Primary MS\Datasets\#-#-18

Note that #-#-18 is placeholder for date of final run for final datasets.

List of Objectives

Objective #1: Define the cohort of interest

Objective #2: Determine the cumulative probability of meeting each of the three cause-specific suboptimal surgical outcome criteria by 3 years (post hoc outcome)

Objective #3: Compare 3-year stereoacuity between BLR and R/R treatment groups

Objective #4: Compare complete or near-complete resolution at 3 years between BLR and R/R treatment groups

Objective #5: Compare 3-year IXT Questionnaire (IXTQ) scores between treatment groups

Objective #6: Tabulate non-surgical treatment prescribed for BLR and R/R treatment groups

Objective #1: Define the cohort of interest

1. 197 patients with basic-type IXT with largest preoperative exodeviation between 15 and 40 PD by PACT at remote distance, distance, or near (101 in BLR group, 96 in R/R group)

Technical plan

1. Limit the patient-level dataset ROSTER to patients where the STRATUM variable from tblStratum, the variable used to stratify the randomization, = 'Basic IXT with 15-40PD angle'

Dataset used: ROSTER

Objective #2: Determine the cumulative probability of meeting each of the three cause-specific suboptimal surgical outcome criteria by 3 years (post hoc outcome)

1. Define three causes of meeting suboptimal surgical outcome:
   - Exotropia suboptimal outcome criterion
   - Constant esotropia suboptimal outcome criterion
   - Stereoacuity suboptimal outcome criterion
2. Calculate patient-level cause-specific suboptimal surgical outcomes.
3. Get cumulative probability of each cause-specific suboptimal surgical outcome by 3 years for each treatment group from Kaplan-Meier (K-M) survival analysis.
4. Compare cumulative probability of each cause-specific suboptimal surgical outcome by 3 years between treatment groups using a two-sided Z-test.
5. Calculate the treatment group difference (and 95% CI) in the cumulative probability of each cause-specific suboptimal surgical outcome 3 years.

Technical plan

1. Define three causes of meeting suboptimal outcome criteria:

   - **Exotropia** at distance OR near at any time during the exam (i.e., can be constant or intermittent; determined by a cover/uncover test) with a magnitude of atleast 10 PD by SPCT, confirmed by a retest
   - **Constant esotropia** at distance OR near (determined by at least 3 cover/uncover tests—one must be before any dissociation) with a magnitude of atleast 6 PD by SPCT, confirmed by a retest
   - Decrease in Preschool Randot near **stereoacuity** atleast 2 octaves (atleast 0.6 log arcsec) (*see Table 3*) from the enrollment measurement, or to nil, confirmed by a retest

Table 1: Preschool Randot Stereotest

Baseline stereoacuity at enrollment, in arcsec	Level needed at follow up visit to meet surgical failure criteria, in arcsec
40”	200” or worse
60”	400” or worse
100”	400” or worse
200”	800” or worse
400”	Nil

311
312

310

evel time-to-event outcome for meeting the criterion at any time by 3 years.

2. For each of the three causes, create a separate patient-l

Rules for Classification of Each Cause Specific Outcome

a. Patients who meet the specified criterion at any time without having first undergone reoperation are counted as having met the outcome the first time the specific outcome occurs, regardless of whether any other suboptimal surgical outcome criterion was met at any time.

b. Patientswhoundergoreoperationwithoutfirstmeetingthespecifiedcriterion will be considered not to have met the outcome and will be censored at the reoperation date month (i.e. the end of their 'uncontaminated' time).

c. Patientswhodonotmeetthespecifiedcriterionby3yearsandhavenotundergonereoperation are right-censored for the specified outcome at the last visit date.

d. Patientswhomeetcriteriaotherthanthespecifiedcriterion(e.g.exotropiaorconstant esotropiawhen consideringthestereoasthespecificoutcome) continue to be 'at risk' for the specified criterion provided they have not been reoperated, and are eventually classified as either a, b, or c above.

Because the primary outcome of suboptimal surgical outcome relates to the first occurrence of deterioration by any method (exotropia, constant esotropia, or stereoacuity loss), need to create the following for each patient

- Whether/when they meet stereo deterioration regardless of whether exotropia and/or constant esotropia deterioration were met first.
- Whether/when they meet exotropia deterioration regardless of whether constant esotropia and/or stereo deterioration were met first.
- Whether/when they meet constant esotropia deterioration regardless of whether exotropia and/or stereo deterioration were met first.

- Use the existing verified MASKEDEXAMS dataset created for the ESA 2017 presentation, which has one-record per maskedexam and defines whether stereo SSO criteria has been met at the specified masked exam.

```
/* VERIFIED MASKEDEXAMS DATASETS CREATED AT TIME OF ESA 2017 PRESENTATION */
libname ixt1prev 'F:\user\PEDIG\Manuscripts-
Presentations\Manuscripts\IXT\IXT1\Manuscripts\Primary MS\Datasets\8-23-17
(verified)';
```

Using STEREO suboptimal surgical outcome as an example:

- the verified MASKEDEXAMS dataset (one record per masked exam) contains a flag variable (STEREODETER) to indicate whether stereo deterioration was met at that visit. If stereo deterioration is met at the visit, the visit type (6, 12, …36 months) (STEREODETERVISIT) and the date of the visit (STEREODETERDT) are also defined.

- For each patient, obtain from the MASKEDEXAMS dataset the first record at which stereo deterioration was met.

- Merge stereo deterioration flag, date, and visit from this record into the one record per patient dataset ROSTER.

3. Run Kaplan-Meier survival analysis on cause-specific outcome
   a. Run Kaplan-Meier survival analysis using proc lifetest, specifying the method as Kaplan-Meier. Use same K-M macro as was used for primary outcome.
   b. Output survival probabilities and confidence intervals to a dataset.
   c. Create failure estimates and confidence intervals.

```
/*****************************************************************************************
  CREATE MACRO TO GET K-M CUMULATIVE PROBABILITY OF A GIVEN OUTCOME BY 3 YEARS
*****************************************************************************************/

%macro kmestimates (outcome, outcometime);

title1 "K-M Survival Analysis for '&outcome' Outcome";

proc lifetest data = roster method=km outsurv=failresults stderr plots=none alpha=.05;
      time &outcometime*&outcome(0);
         strata trtgroup;
run;

data failresultCIs;
     set failresults;
        /* create failure estimates (rather than survival)*/
        failure = round ((1-survival), 0.01);
         failureLCL = round ((1-SDF_UCL), 0.01);
        failureUCL = round ((1-SDF_LCL), 0.01);
     /* limit to records where the survival estimate has changed
          (i.e. records where the CI is not null) */
     if SDF_LCL NE . then output;
        label failure = "Cum. probability of '@outcome' Outcome";
        label failureLCL = "Lower limit of CI for Cum. probability of '@outcome' Outcome";
        label failureUCL = "Upper limit of CI for Cum. probability of '@outcome' Outcome";
run;

%sort (failresultCIs, trtgroup);
proc print data = failresultCIs;
     by trtgroup;
     title2"Review Output Dataset from K-M Survival Analysis for '&outcome' Outcome";
run;

%mend;
```

3. Compare cumulative probability of each cause-specific outcome by 3 years for each treatment group using code to perform a Z-test. Manually input probabilities and standard error for each outcome (get from K-M output).

```
/* Calculate Z-Test */

/* Need to manually input probabilities and standard error for each outcome
  (get from K-M output) */

data check;
   length outcome $20;
   input outcome probblr seblr probrr serr;
   datalines;
   exofail 0.#### 0.####  0.#### 0.####
   conesofail 0.#### 0.####  0.#### 0.####
   stereofail 0.#### 0.####  0.#### 0.####
;
run;

/* note that 0.#### is a placeholder for use in SAP.   Actual values taken from K-M output */

data check;
     set check;
        diff = probblr - probrr;
     sumofsquaredse = sqrt (seblr**2+serr**2);
        cilower = diff - (1.96*sumofsquaredse);
        ciupper = diff + (1.96*sumofsquaredse);
```

```
        zscore = diff/sumofsquaredse;
     pvalue = 2*(1 - probnorm(abs(zscore)));
run;


proc print data = check noobs;
     var outcome probblr seblr probrr serr sumofsquaredse diff cilower ciupper zscore pvalue;
        title'Calculate CIs and P values';
   run;
```

Datasets used: MASKEDEXAMS, ROSTER

Objective #3: Compare 3-year stereoacuity between BLR and R/R treatment groups

1. Limit the analysis to patients who completed the 3-year visit.
2. Define the outcomes as distance stereoacuity and near stereoacuity.
3. Create change in stereoacuity between baseline and 3 years.
4. Obtain distribution of 3-year stereo and change in 3-year stereo.
5. Compare mean 3-year outcomes between treatment groups adjusting for corresponding baseline stereoacuity.

Technical Plan

1. Limit the analysis to patients who completed the 3-year visit. Use the 3-year visit data for all patients who completed the 3-year visit, regardless of what treatment(s) were received or if the patient had undergone reoperation.

2. Define outcomes as distance and near stereoacuity, using existing log scale stereoacuity.

3. Create change in stereoacuity as the baseline value minus the 3-year value, so positive values = improvement.

4. Run proc means to obtain distribution of 3-year stereo and change in 3-year stereo.

5. Compare mean 3-year outcome between treatment groups using ANCOVA model adjusting for corresponding baseline stereoacuity.

```
proc genmod data = roster;
   class trtgroup;
   model stereodi_36 = trtgroup stereodi_0;
   where comp_36 = 1;
   title1'Comparison of 3-year distance stereoacuity between treatment groups';
run;
```

Datasets used: ROSTER

Objective #4: Compare complete or near-complete resolution at 3 years between BLR and R/R treatment groups

1. Define complete or near-complete resolution outcomes:
   - Complete or near-complete resolution at 3 years without regard to previous failure (post hoc)
   - Complete or near-complete resolution at 3 years with no previous failure (consistent with pre-specified "success" criteria in the protocol).

2. Calculate complete or near-complete resolution outcomes using 3-year alignment, 3-year stereoacuity, nonsurgical treatment for IXT during the study, reoperation data. In addition, suboptimal surgical outcome **by** 3 years will also be used to calculate complete or near-complete resolution AT 3 years with no previous failure.

3. For each definition, compare proportion of participants with complete or near-complete resolution at 3 years between treatment groups and calculate 95% CI.

Technical Plan

1. Complete or near-complete resolution AT 3 years without regard to previous failure (post hoc) was defined as meeting all of the following at the 3-year visit:

    1. Exodeviation <10 PD (tropia or phoria) by both SPCT and PACT at distance and near and either ≥10 PD reduction in 3-year PACT magnitude from both the distance and near preoperative* angles ≥10PD if the corresponding preoperative angle was ≥10 PD, or reduction to orthodeviation by PACT if corresponding preoperative angle was <10PD. *Because all patients in the basic primary cohort had distance and near PACT ≥10 PD, the last phrase was not applicable and therefore not included in the definition in the manuscript.*
    2. Esotropia <6 PD at distance and near by SPCT
    3. No decrease in Preschool Randot of ≥2 octaves from enrollment stereoacuity or to nil
    4. No reoperation or treatment with botulinum toxin
    5. No non-surgical treatment for a recurrent or residual exodeviation (nonsurgical treatment for esodeviation was allowed)

    **Preoperative angle represents the largest deviation by PACT at distance, near, and remote distance at the enrollment visit if no additional preoperative (pre-randomization), PACT measurements were taken closer to the surgery date; OR the PACT deviation entered on the randomization form if any additional preoperative PACT measurements were taken closer to the surgery date.*

    Complete or near-complete resolution AT 3 years with no previous failure was defined as follows:
    - Patients who had suboptimal surgical outcome at any time will be considered not resolved
    - All other patients will have same values as for complete or near-complete resolution AT 3 years without regard to previous failure.

2. Calculate complete or near-complete resolution outcomes using baseline alignment (see note), 3-year alignment, 3-year stereoacuity, nonsurgical treatment for IXT during the study, reoperation

data. In addition, suboptimal surgical outcome **by** 3 years will also be used to calculate complete or near-complete resolution AT 3 years with no previous failure.

- Note for baseline alignment for calculating whether distance and near angles have had the requisite reduction in PACT: In addition to distance and near PACT which were measured at enrollment, one or more PACT angles may have been re-measured closer to (but before) randomization. The single PACT measurement that was entered on the randomization form was the largest recorded at near, distance, or remote distance fixation, and was the angle upon which surgical dose would be based--and may have been one of the enrollment angles or an angle measured closer to (but before) randomization. If the angle entered on the randomization form was measured at remote distance, the enrollment distance and near were used for the baseline alignment; if the angle entered was measured at distance or near, then this angle was used for the corresponding baseline measurement and the enrollment data used for the remaining measurement.

3. For each outcome, compare the proportion of participants with complete or near-complete resolution using Barnard's exact test and specifying the FMSCORE option to obtain 95% confidence intervals from Farrington-Manning score.

```
proc freq data = roster;
     tables trtgroup*compresolve / nocol nopercent;
     exact barnard riskdiff (method=fmscore);
     where comp_36 = 1;
     title1 "Comparison of Complete or Near-complete Resolution Without Regard to
Previous Failure";
run;
```

Datasets used: ROSTER

Objective #5: Compare 3-Year IXT Questionnaire (IXTQ) Scores Between Treatment Groups

1. Obtain IXTQ data from existing verified SAS dataset **IXTQ.IXTQALL_19OCT2017** provided by Trevano. Include the following:
    a. **Rasch score for each component IXTQ (or parent subscale)**
    b. **0 to 100 score for each component IXTQ (or parent subscale)**
2. Test ANOVA assumptions for normality, homogeneity of variance.
3. Test ANCOVA assumption of homogeneity of slopes for parent and proxy only (child versions using ANOVA).
4. For older and younger child IXTQ, parent proxy IXTQ, and for each of parent subscales between compare distribution Rasch proxy score between treatment groups.
5. List median 0 to 100 score (based on Rasch score) for each treatment group.

Technical Plan

1. Use SAS IXTQ dataset **IXTQ.IXTQALL_19OCT2017** provided by Trevano

```
*IXT1 IXTQ;
libname IXTQ 'F:\user\PEDIG\Manuscripts-
Presentations\Manuscripts\IXT\IXT2\IXTQ\Manuscript Analysis\Datasets';
```

- Contains one record per IXTQ completed at enrollment and 3 years for both IXT1 and IXT2 studies.
- Take enrollment and 3-year questionnaire scores for IXT1 patients from **IXTQ.IXTQALL_19OCT2017** and add to ROSTER dataset.
- Scores are also included as follows:
    a. For the child IXTQ component, 5 to 7 year version scores are include for children aged 5 to 7 at the time of testing; 8 year and older version scores are include for children aged 8 or older at the time of testing; no child component scores are included for children younger than 5 years at the time of the visit. The variable AGE indicates the child's age at the time of testing.
    b. For the parent proxy IXTQ component
    c. For the parent IXTQ component: each of the three parent IXTQ subscales of psychosocial, function, and surgery.
- For each IXTQ component or subscale, a Rasch-based score and a 0 to 100 score (based on Rasch score) are included.
- Note that some participants were too young for the IXTQ at baseline (those enrolled 3-<5 years of age) but completed the 5 to 7 year old IXTQ version at interim visits and at the 3-year visit once the child turned 5 years old. Also note that some children completed the 5 to 7 year old IXTQ version at baseline and the 8 years and older version at 3 years.

2. Test ANOVA/ANCOVA assumptions for normality, homogeneity of variance, and homogeneity of slopes.
    - Test for normality using Shapiro-Wilkes test

```
577                     %macro normalitytest (outcome);
578                     %sort (roster, trtgroup);
579                     proc univariate data = roster normal;
```

```
                    var &outcome;
                    by trtgroup;
          run;
          %mend;


          %normalitytest (ParentPsychRaschMean_36);

/* Note that all look non-normal(Shapiro-Wilkes test P values all < 0.05 */
```

- Test for homogeneity of variance using Levene's test

```
%macro variancetest (outcome);
proc glm data = roster;
     class trtgroup;
     model &outcome = trtgroup;
     means trtgroup / hovtest=levene(type=abs) hovtest=bf;
    title1"ASSESSING ASSUMPTION OF EQUALITY OF VARIANCES FOR '&OUTCOME'";
run;
quit;
%mend;

%variancetest (ParentPsychRaschMean_36);


/* Note that all look OK except for child younger -- not equal */
```

- Test for homogeneity of regression slopes using regression model with interaction term

```
%macro slopetest (outcome, predictor);
proc glm data = roster;
     class trtgroup;
     model &outcome = trtgroup &predictor trtgroup*&predictor;
     title1"ASSESSING ASSUMPTION OF EQUALITY OF SLOPES FOR '&OUTCOME' BY INCLUDING
INTERACTION TERM IN THE ANCOVA MODEL";
run;
%mend;

%slopetest (ParentPsychRaschMean_36, ParentPsychRaschMean_0);

/* all look OK */
```

3. For each of the two age-specific child IXTQs, the IXTQ proxy questionnaire, and for each of the three parent questionnaire subscales (psychosocial, functional, and surgical), compare distribution of Rasch-based QOL scores at 3 years using Wilcoxon Rank Sum test.

```
%macro wilcoxon (outcome);
proc npar1way data = roster plots=none wilcoxon;
     class trtgroup;
     var &outcome;
     title1"NONPARAMETRIC WILCOXON RANK SUM COMPARING '&OUTCOME' BETWEEN TREATMENT
GROUPS";
run;
%mend;

%wilcoxon (ProxyRaschMean_36);
```

4. Since all analyses have at least one assumption violated, switch to using non-parametric Wilcoxon rank sum test to compare distributions between treatment group, instead of means using ANOVA or ANCOVA.

5. Calculate median of median 0 to 100 score at 3 years using proc means.

References

[1]*Leske DA, Holmes JM, Melia M, on behalf of Pediatric Eye Disease Investigator Group. Evaluation of the Intermittent Exotropia Questionnaire using Rasch analysis. JAMA Ophthalmol 2015;133:461-5.*

[2]*Leske DA, Hatt SR, Liebermann L, Holmes JM. Evaluation of the Adult Strabismus-20 (AS-20) Questionnaire Using Rasch Analysis. Invest Ophthalmol Vis Sci 2012.*

[3]*Leske DA, Hatt SR, Liebermann L, Holmes JM. Lookup Tables Versus Stacked Rasch Analysis in Comparing Pre- and Postintervention Adult Strabismus-20 Data. Transl Vis Sci Technol 2016;5:11.*

Datasets used: ROSTER, IXTQALL_19OCT2017

Objective #6: Tabulate non-surgical treatment prescribed for BLR and R/R treatment groups separately

1. **Define nonsurgical treatment of interest as that which was prescribed for XT, ET, or diplopia.**

   Because the reason for this nonsurgical treatment was not specified other than being prescribed for XT, ET, or diplopia, the type of deviation that was present when the nonsurgical treatment was prescribed was reported. This data was used to report the proportions of participants with non-surgical treatment prescribed when exodeviation was present, when esodeviation was present, and when exodeviation and esodeviations were present at different times during the study. Among participants who met the constant esotropia suboptimal surgical outcome during the study, the proportion who had nonsurgical treatment prescribed was reported.

   Do not include nonsurgical treatment prescribed for amblyopia, which was recorded in a different section of the data form.

2. **Obtain alignment data**

3. **Create patient level flag variables for the following:**
   - Nonsurgical treatment when XT is present anytime during study
   - Nonsurgical treatment when ET is present anytime during study
   - Non-surgical treatment when XT and ET are present (different times during study)

Technical Plan

1. **Get treatment prescribed records from the SQL data table TBLIXT1TREATRX. Obtain visit date from the login table.**

   Use: NonSurgTmtRxNone, NonSurgTmtRxPrism, NonSurgTmtRxPatch, NonSurgTmtRxOverMinus, NonSurgTmtRxVT, NonSurgTmtRxOther, NonSurgTmtRxOtherDs

   Do not use: AmbTrtRxNone, AmbTrtRxPatch, AmbTrtRxAtrp, AmbTrtRxVT, AmbTrtRxOverPlus, AmbTrtRxBang, AmbTrtRxOther, AmbTrtRxOtherDs

2. **Obtain SPCT and PACT alignment data for maskedexams from verified dataset MASKEDEXAMS. Obtain similar data from the 1 week and 8 week visits from the SQL data table TBLIXT1OCUALIGN.**

   Consider XT present if exotropia is present by SPCT or exodeviation is present by PACT
   Consider ET present if esotropia is present by SPCT or esodeviation is present by PACT

   Merge alignment data into treatment prescribed records.

3. **Create patient level flag variables for the following:**

687 • Nonsurgical treatment when XT is present anytime during study

- Nonsurgical treatment when ET is present anytime during study
- Non-surgical treatment when XT and ET are present (different times during study)

1. Loop through the treatment prescribed records, retaining flags for non-surgical treatment when XT is present and for non-surgical treatment when ET is present.
2. Take last record for each patient.
3. Add flags to roster dataset.
4. Create flag for non-surgical treatment when XT and ET using XT and ET treatment flags .