Spectacles for Patients with Down Syndrome

NCT03367793

Statistical Analysis Plan

November 22, 2019

Statistical Design and Power

Thirty individuals with Down syndrome will be enrolled and dispensed three different spectacle prescriptions (metric #1, metric #2, clinical) in random order for 2 months wear each. The primary outcome measure will be 2 month adapted visual acuity compared across spectacle prescription type. Differences in acuity will be compared across spectacle type for all subjects using dependent t-tests to identify the prescription type(s). For our sample size, we are powered to detect a difference as small as 1.9 letters. The anticipated effect size is 5 letters for the metric spectacle corrections versus the clinical.

Additional outcome measures include average daily wear time as determined by an objective temperature sensor data logger and score on a subjective survey of spectacle wear perceptions.

Statistical Analysis Plan:

We will report descriptive statistics by period and treatment for primary outcome visual acuity and additional objective compliance and subjective prescription preference/quality. A mixed-effects linear modeling statistical approach will be used to compare differences among the three spectacle prescriptions with study period as a fixed effect and participant as a random effect to account for within-subject and between-subject variability as well as to evaluate any period, sequence, and carryover effects. The overall effect of treatment via an overall F-test, which accounts for the covariance structure of the variance-covariance matrix, will be used in SAS (PROC GLIMMIX). Follow-up testing (via Tukey post-hoc) will be performed to further elucidate differences between experimental prescriptions. While carryover effects are not expected, we will implement Grizzle's approach to evaluating carryover. Should a carryover effect be detected, period 1 will be used for primary analysis. Assuming the objective measure of compliance, via temperature sensors, is continuous and normally distributed, we will adopt a similar statistical approach. Subjective prescription quality or satisfaction is based on a 5-point scale. We propose generalized estimating equations (GEE) to analyze the subjective quality outcome (ranging from 1-5) as a categorical response (e.g., as binary, we will define reported 4/5 as satisfactory; 0-otherwise).

Power Analysis:

Primary Outcome: A simulation based approach was used to carry out the power analysis and based on the generalized mixed effects model and associated f test corresponding to the treatment effect. For each of the 1,000 datasets generated based on the generalized mixed model under the alternative hypothesis, the overall F-test was conducted and the decision of rejection based on α=.05 was collected. Power, which by definition is the rejection of the null hypothesis assuming the alternative is true, was calculated as the proportion of tests rejected among the 1,000 simulated datasets. This power analysis was conducted under various specifications of standardized effect size (i.e., assumed average treatment difference between metric and clinic divided by the standard deviation), rho (intraclass correlation coefficient [ICC} associated with subject random effect), and sample size scenarios. {Note that when ICC or effect sizes increase, required sample size to achieve desired power decreases.} Results indicated that a sample size of 29 achieves 89% power to detect larger effect sizes of 0.25 and 0.3, respectively, with lower ICC (0.3). A sample size as low as n=11 would be required to achieve 91% power assuming larger effect sizes (0.3) and large ICC (0.7). Given assumed effect sizes of 0.25, moderate ICC of 0.5, a

sample size of n=26 yields 90% power. Although this population has been known to be highly compliant, assuming a 15% dropout rate over the course of the study, we aim to sample 30 participants. *Outcome Temperature Sensor:* Referring to results above, using the same statistical rationale, n=30 is justified to yield at least 80% power to detect larger effects, even with a smaller ICC for temperature sensor (compliance) assuming a continuous response. *Outcome Subjective Quality:* To justify the sample size based on distribution-free methods in a 3x3 crossover with potential fixed effect is not common. Thus, sample size justification is based on two approaches for this outcome: 1) use of the mixed model simulation approach realizing that parametric testing require less observations; 2) binary outcome approach. From (1), given that a sample size as low as n=9 is required to achieve 83% power to assuming effect sizes of 0.3 and ICC=0.7, n=30 (a tri-fold increase) is deemed sufficient to investigate the secondary subjective outcome. For (2), assume that a reported value of 4 or 5 is deemed 'desirable to wear' and termed a successful outcome, power can be calculated based on two separate McNemar's tests to compare metric #1 with the clinical prescription and metric #2 with the clinical spectacle, assuming a conservative Type I error of 0.025. The test is based on the number of discordant pairs from each occasion, where p12 is the probability of success at occasion 1 and failure at occasion 2, and p21 is the probability of failure at occasion 1 and success at occasion 2. It is expected that in the absence of a treatment effect, that the distribution of total discordance is equally likely for either intervention (i.e., p12-p21=0 under Ho). A sample size of n=28 yields 82% power to detect a difference of 0.4. This test assumes no period effects.