Official Title: An Exploratory Phase II Study to Determine the Tolerability, Safety, and Activity of a Novel Vasopressin 1a Receptor Antagonist (SRX246) in Irritable Subjects with Huntington's Disease (HD) - SAP

NCT02507284

Document Date: November 9, 2018

Statistical Analysis Plan

An Exploratory Phase II Study to Determine the Tolerability, Safety, and Activity of a Novel Vasopressin 1a Receptor Antagonist (SRX246) in Irritable Subjects with Huntington's Disease (STAIR)

Michael Brownstein, MD, PhD
Senior VP of Drug Development, Azevan Pharmaceuticals
Protocol Principal Investigator

Steve Hersch, MD, PhD
Professor of Neurology, Harvard Medical School
Principal Investigator

Jeffrey D. Long, PhD
Professor, Departments of Psychiatry & Biostatistics, University of Iowa
External Biostatistician

Christopher S. Coffey, PhD
University of Iowa Data Coordinating Center
Principal Investigator

Merit E. Cudkowicz, MD
Massachusetts General Hospital Clinical Coordinating Center
Principal Investigator

Neal G. Simon, PhD
Chief Executive Officer, Azevan Pharmaceuticals, Inc.
Co-Investigator

VERSION 1.0
November 9, 2018

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

An Exploratory Phase II Study to Determine the Tolerability, Safety, and Activity of a Novel Vasopressin 1a Receptor Antagonist (SRX246) in Irritable Subjects with Huntington's Disease

Protocol Principal Investigator Approval

Signature: Michael Brownstein **Date:** 11/12/18
Name: Michael Brownstein, MD, PhD
Affiliation: Azevan Pharmaceuticals, Inc.

U44 Principal Investigator Approval

Signature: Steven Hersch **Date:** 11/11/2018
Name: Steven Hersch, MD, PhD
Affiliation: Harvard Medical School

External Biostatistician Approval

Signature: Jeffrey Long **Date:** 11/12/2018
Name: Jeffrey Long, PhD
Affiliation: University of Iowa

NeuroNEXT Clinical Coordinating Center Approval

Signature: Merit Cudkowicz **Date:** 11/12/2018
Name: Merit Cudkowicz, MD
Affiliation: Harvard Medical School

NeuroNEXT Data Coordinating Center Approval

Signature: Chris S Coffey **Date:** 11/12/2018
Name: Christopher S. Coffey, PhD
Affiliation: University of Iowa

Study Sponsor Approval

Signature: Neal G Simon **Date:** 11/11/2018
Name: Neal G Simon, PhD
Affiliation: Azevan Pharmaceuticals, Inc.

Table of Contents

PREFACE

This Statistical Analysis Plan (SAP) describes the planned analyses for the STAIR study [NeuroNEXT NN105 funded by the National Institute of Neurological Disorders and Stroke (NINDS) grant #U44NS090616 to Azevan Pharmaceuticals, Inc., the Study Sponsor; ClinicalTrials.gov ID: NCT02507284; Azevan ID: AVN011]. The planned analyses identified in this SAP are intended to support the completion of the Final Study Report (FSR) and will be included in regulatory submissions and/or future manuscripts. All final, planned analyses identified in this SAP will be performed only after the last randomized study participant has completed the study, and all data have been cleaned and verified in accord with applicable National Institute of Health NeuroNEXT Network SOPs (https://neuronext.org/neuronext-standard-operating-procedures). Once all data have been cleaned and verified, a "locked" version of the data will be used for reporting the final study results. Key statistics and study results will be made available to the PPIs and CCC following database lock and prior to completion of the final FSR.

1. STUDY DESIGN

This is a 3-arm, multicenter, randomized, placebo-controlled, double-blind, 12-week, dose escalation study of SRX246 in irritable study participants with early symptomatic HD. We initially planned to randomize *N* = 108 participants to the study arms (36 per arm), with one arm being placebo and the other arms being different final doses of SRX246. The treatment for each study participant was assigned by a randomized code.

Following an initial screening visit, study participants fulfilling the study inclusion and exclusion criteria entered a pre-treatment screening phase, of no longer than 30 days, to permit evaluations to confirm eligibility for inclusion into the study. At the completion of the screening period, eligible participants were randomized at the baseline visit to receive either placebo or final doses of SRX246 of 120 mg twice daily or 160 mg twice daily during the double-blind treatment phase. At baseline, study participants in the active groups were to receive 80 mg twice daily for 2 weeks, then escalate to 120 mg twice daily for 4 weeks. Thereafter, one group of active study participants were to continue to take 120 mg of SRX246 twice daily for an additional 6 weeks, and the second group of participants was to increase their dose to 160 mg of SRX246 twice daily for 6 weeks. Total dosing duration was 12 weeks. Study participants in the placebo group were to receive the same number of capsules that are identical in appearance to the capsules that contain SRX246 during the trial, to preserve the blind. In all groups, step-up dose escalation occurred if patients did not experience dose-limiting adverse events. Patients who could not tolerate their final dose of drug (or placebo) could have their dose reduced without compromising the blinding.

Study participants had periodic scheduled visits either "in-person" or by "telephone". An electronic (smart phone or table) diary known as the eDiary prompted participants to take their capsules twice a day and asked whether they did so. The study participants and their study partners were asked to answer daily questions having to do with irritability and behavior (also recorded in the eDiary). Visits (with varying windows) were to occur at week 0 (baseline), 2, 4, 6, 8, 10, and 12. Data from all study participants who were given placebo will be combined, and two pairwise tests will be performed (120 mg vs placebo, 160 mg vs placebo).

1.1 Primary Objective

The primary objective is to evaluate the tolerability of SRX246 versus placebo.

1.2 Secondary Objective

The secondary objective is to evaluate the safety of SRX246 versus placebo.

2. PRIMARY ENDPOINT

The primary endpoint for tolerability is defined as the proportion (percentage) of study participants who completed the study while active on their assigned intervention arm. Study completion can occur despite temporary drug interruptions, protocol deviations, etc. Conversely, a non-completer is a study participant who withdraws from study drug (either participant-initiated or researcher-initiated) prior to the normal study termination as described in the protocol or who completes the study on a dose below their assigned dose.

Each treatment group will be compared to the placebo group, and the statistical test of equivalence of tolerability will involve two pairwise tests (120 mg vs. placebo, 160 mg vs. placebo).

3. SECONDARY ENDPOINT

The secondary endpoint for safety is defined as the proportion of study participants who experience any treatment-emergent (starting at or after the first dose of study medication) adverse event (AE) or serious adverse event (SAE). For the purposes of this study, a treatment emergent adverse event (AE) is any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related. FDA, Office of Human Research Protection (OHRP) and NeuroNEXT CIRB requirements for reporting AEs will be followed. AEs are generally detected in two ways:

- Clinical → Symptoms reported by the study participant or signs detected on examination
- Ancillary Tests → Abnormalities of vital signs, laboratory tests, and other diagnostic procedures

If discernible at the time of completing the AE source documentation, a specific disease or syndrome rather than individual associated signs and symptoms should be identified by the Site Investigator and recorded in the AE documentation. However, if an observed or reported sign, symptom, or clinically significant laboratory anomaly is not considered by the Site Investigator to be a component of a specific disease or syndrome, then it should be recorded as a separate AE in the source documentation. Clinically significant laboratory abnormalities, such as those that require intervention, are identified as such by the Site Investigator.

Each clinical study site's Principal Investigator and research team (co-investigators, research nurse, and clinical trial coordinator) are responsible for identifying and reporting AEs through the NeuroNEXT Online Adverse Event Reporting System. The AE definitions and reporting procedures used for this study comply with all applicable U. S. Food and Drug Administration (FDA) regulations and International Conference on Harmonization (ICH) guidelines. The Site Investigator will carefully monitor each study participant throughout the study for possible AEs. All AEs will be documented on CRFs designed specifically for this purpose. It is important to report all AEs, especially those that result in permanent discontinuation of the investigational product, whether serious or non-serious. Local site investigators are also required to fulfill all reporting requirements of their local institutions.

Study participants who sign consent and receive investigational treatment will be monitored for AEs from the time they sign consent through week 12 or the end of their study participation. This study utilized the CTCAE version 4.03 coding system for AE recording. AEs reported using CTCAE will be recoded into MedDRA terms by the DCC.

At each visit, the participant will be asked "Have you had any problems or symptoms since your last visit?" in order to determine the occurrence of AEs. If the study participant reports an AE, the Investigator will determine:

- Type of event
- Date of onset and resolution (duration)
- Severity (mild, moderate, severe, life-threatening, results in death)
- Seriousness (does the AE meet the definition of an SAE)
- Causality, relation to investigational product and disease
- Action taken regarding investigational product
- Outcome

All clinical AEs are recorded in the AE data entry template in the study participant's study binder. The sites should fill out the AE data entry template and enter the information into the Online Adverse Event Reporting System (AERS) within 5 working days / 7 calendar days of the site learning of a new AE or receiving an update on an existing AE. Entries in the AE data entry template will include the following:

- Name and severity of the AE
- Date of onset
- Date of resolution

- Relationship to study drug
- Action taken
- Primary outcome of event

The severity of all AEs will be graded according to CTCAE, version 4.03. Any AE not listed in the CTCAE will be graded as follows:

- Grade 1 – Mild: Transient or mild discomfort; No limitation in activity; No medical intervention/therapy required
- Grade 2 – Moderate: Mild to moderate limitation in activity; Some assistance may be needed; No or minimal medical intervention/therapy required
- Grade 3 – Severe: Marked limitation in activity, some assistance usually required; Medical intervention/therapy required; Hospitalizations possible
- Grade 4 – Potentially Life-Threatening: Extreme limitation in activity; Significant assistance required; Significant medical intervention/therapy required; Hospitalization or hospice care probable
- Grade 5 – Death: The AE results in death

The Site Investigator is initially responsible for classifying AEs as serious or non-serious. An AE is considered serious if it meets one or more of the following criteria:

- Results in death
- Is life-threatening (i.e., a study participant is at immediate risk of death at the time the AE occurs, not an event where occurrence in a more serious form might have caused death)
- Requires inpatient hospitalization or prolongation of existing hospitalization
    - Hospitalization for an elective procedure or a routinely scheduled treatment is not an SAE by this criterion because an elective or scheduled “procedure” or “treatment” is not an untoward medical occurrence
- Results in persistent or significant disability or incapacity
    - This serious criterion applies if the “disability” caused by the reported AE results in a substantial disruption of the study participant’s ability to carry out normal life function
- Results in a congenital anomaly / birth defect
- Necessitates medical or surgical intervention to preclude permanent impairment of a body function of permanent damage to the body structure
- Another important medical event
    - Important medical events that may not result in death, are not life-threatening, or do not require hospitalization may also be considered SAEs when, based on appropriate medical judgment, they may jeopardize the participant and may require medical or surgical intervention to prevent one of the outcomes listed above. Examples of such medical events include blood dyscracias or convulsions that do not result in inpatient hospitalization, or the development of drug dependency or drug abuse.

An inpatient hospital admission in the absence of a precipitating, treatment-emergent, clinical AE may meet criteria for “seriousness” but is not an *adverse* experience, and will therefore not be considered an SAE. An example of this would include a social admission (study participant admitted for reasons other than medical, e.g., lives far from the hospital, has no place to sleep).

Drs. Amy Lee Bredlau (through July 2017) and Andrew McGarry (since July 2017) served as the Medical Safety Monitors (MSMs) for this trial. They worked closely with the DCC, and used an online AE reporting system to review all SAEs in near real time in order to evaluate them to identify the need for timely intervention. For any reported SAEs, an automatic email was sent to the MSM to prompt a review of the event for determination of whether the event meets the criteria for an SAE and, if so, whether the SAE is unanticipated and/or related to study drug.

For the purposes of this study, a treatment-related AE (also referred to as an Adverse Drug Reaction) is defined as any noxious or unintended response to a medicinal product related to any dose. The phrase "responses to a medicinal product" means that a causal relationship between a medicinal product and an AE is at least a reasonable possibility, i.e., the relationship cannot be ruled out. Therefore, a subset of AEs can be classified as treatment related if there is thought to be a causal relationship to study drug. At the time of reporting, the relationship of the AE to the investigational product should be specified by the Site Investigator using the following definitions:

- Definitely Related: The reaction follows a reasonable temporal sequence from administration of investigational product; that follows a known or expected response pattern to the investigational product; and that is confirmed by improvement on stopping or reducing the dosage of the investigational product, and reappearance of the reaction on repeated exposure (suspected treatment related AE or ADR)
- Probably Related: The reaction follows a reasonably temporal sequence from administration of investigational product; is confirmed by discontinuation of the investigational product or by re-challenge; and cannot be reasonably explained by the known characteristics of the study participant's clinical state (suspected treatment-related AE or ADR)
- Possibly Related: The reaction follows a reasonably temporal sequence from administration of the investigational product and follows a known response pattern to the suspected investigational product; the reaction could have been produced by the investigational product or could have been produced by the study participant's clinical state or by other modes of therapy administered to the participant (suspected treatment-related AE or ADR)
- Unlikely to be Related: The reaction has little or no temporal sequence from administration of the investigational product, and/or a more likely alternative etiology exists
- Not Related: Concomitant illness, accident, or event with no reasonable association with treatment

For the purposes of this study, an SAE is considered to be treatment-related if the attribution is possible, probable, or definite. All AEs should be followed until resolution or a new baseline is established, but no longer than study week 13. As this is a double-blind study, the causality assessment should be made under the assumption that the study participant is receiving active study medication. If considering unblinding, this assessment should be made prior to unblinding to avoid bias.

An unexpected SAE is any SAE for which the specificity or severity is not consistent with the current Investigators Brochure or package insert described in the protocol. An unexpected and treatment-related SAE is an unexpected SAE that, in the opinion of the MSM, has a reasonable possibility that the investigational product caused the event. With the assistance of the coordinators at the DCC, the MSM had the option of requesting additional information about any SAE. They completed a form for each review, and this information was entered into the online data entry system.

4. ENROLLMENT & RANDOMIZATION

Following the screening phase, study participants who continue to meet entry criteria will be enrolled and randomly assigned in a 1:1 manner to one of three treatment groups: SRX246 120 mg, SRX246 160 mg, or matching placebo. A total of approximately 108 study participants will be randomized into the study. The treatment assignment for each participant will be assigned by a randomized code using randomization tables generated by the NeuroNEXT DCC using a block randomization scheme.

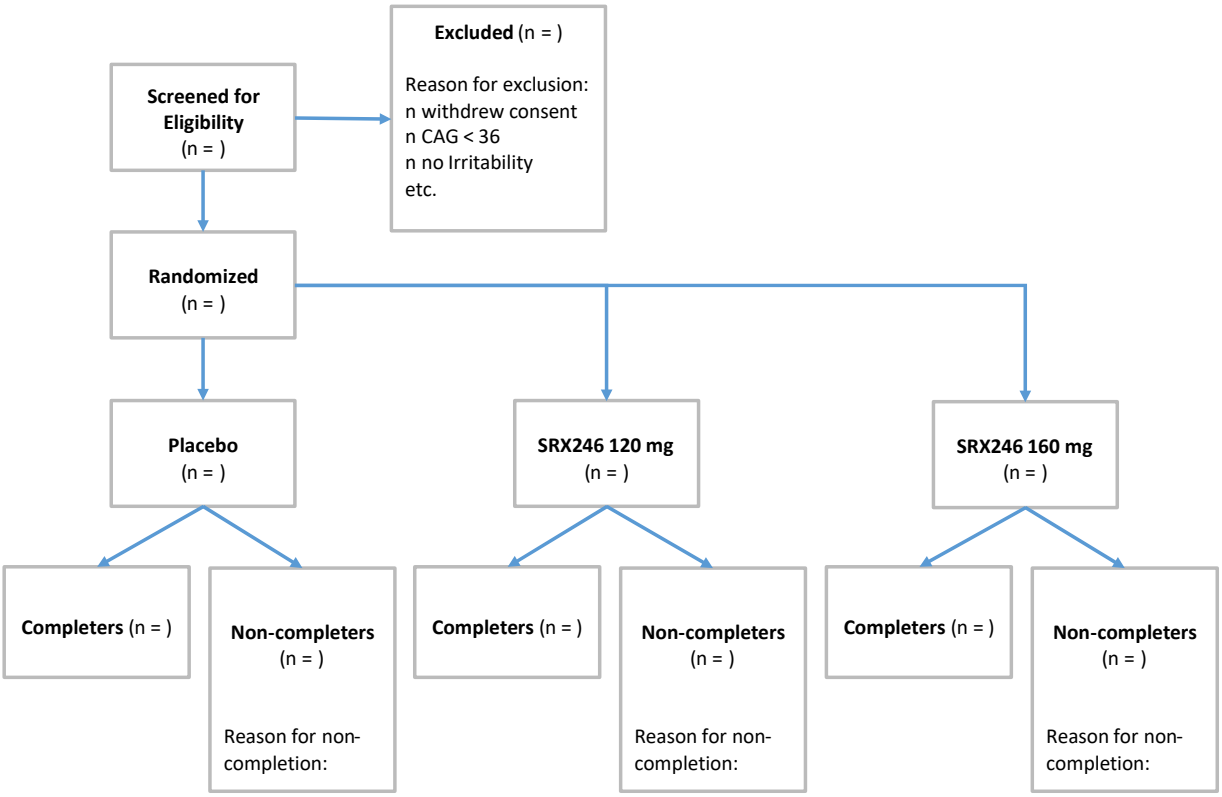
5. TABULATIONS

All study participants who provide informed consent will be accounted for in this study. The number of randomized participants and their study disposition will be reported overall, and by treatment group. The number of study participants who prematurely discontinued from the trial and the reason for discontinuation will be presented based on the categories on the Case Report Form (CRF). A CONSORT diagram summarizing the final status of all study participants will be provided in a flowchart similar to Figure 1 below.

Additional summary reports will describe:

- Number of study participants consented, eligible, and randomized by site
- Reasons for ineligibility
    - A listing of all study participants with one or more inclusion/exclusion exceptions
- Completeness of study visits and case report forms
- Concomitant medications at beginning and end of study
    - Any medication recorded during the trial (from screening to end of study) is considered a concomitant medication. Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHODrug)
    - Records of concomitant medications will be summarized separately for pre-dose (any medications with start date before baseline date) and end of study (any medication participant was on at the date of termination) using frequency counts for all medications at the Anatomical Therapeutic Chemical (ATC) class level 1 (Anatomical main group), and stratified by Pharmacological subgroup (level 3) for nervous system medications.
- Protocol deviations
    - A listing of all randomized study participants with one or more protocol deviations. The listing will include protocol violation category, and a description of the violation.
- Early drug withdrawals
- Early study terminations
- Study drug compliance
    - The number of expected doses of SRX246 or placebo received will be summarized as a continuous variable. Study participants will be deemed as “compliant” if the expected percentage of pills taken falls between 75% and 125% of expected pills.
    - A listing of all randomized study participants with their summary compliance information will also be provided.

Figure 1. CONSORT Diagram



Baseline demographics and clinical characteristics will also be summarized by treatment group (SRX246 120 mg vs. SRX246 160 mg vs. placebo) with respect to important demographic characteristics. Distribution of numeric and categorical variables will be tabulated by treatment group and overall. Numeric variables will be summarized by the mean, median, standard deviation, minimum, and maximum by treatment group and overall. Categorical variables will be tabulated by proportions or percentages. Variables that will be summarized include:

- Baseline Demographics
  - Gender
  - Race
  - Ethnicity
  - Age
  - Years of Education
- Baseline Clinical Characteristics
  - Cohen-Mansfield Aggression Inventory
  - Unified Huntington's Disease Rating Scale (UHDRS) Total Score
  - Irritability Scale – Participant
  - Irritability Scale – Study Partner
  - Total Functional Capacity
  - Irritability Frequency (UHDRS)
  - Irritability Severity (UHDRS)
  - Aggression Frequency (UHDRS)
  - Aggression Severity (UHDRS)
  - Suicidal Ideation
  - CAG Repeat Number

6. ANALYSIS POPULATIONS

For the primary and secondary outcomes, consistent with the recent JAMA Guide to Statistics and Medicine (Detry, 2014), all analyses will be performed consistent with the intent-to-treat (ITT) principle, meaning all study participants will be analyzed as-randomized.

The proportion of randomized study participants completing each visit will be summarized as in Table 1 below.

Table 1: Completeness of Study Data by Visit

Visit	Placebo (N = XX)	120 mg (N = XX)	160 mg (N = XX)
0	XX (XX%)	XX (XX%)	XX (XX%)
2	XX (XX%)	XX (XX%)	XX (XX%)
4	XX (XX%)	XX (XX%)	XX (XX%)
6	XX (XX%)	XX (XX%)	XX (XX%)
8	XX (XX%)	XX (XX%)	XX (XX%)
10	XX (XX%)	XX (XX%)	XX (XX%)
12	XX (XX%)	XX (XX%)	XX (XX%)

7. PRIMARY ANALYSIS

Primary Objective: *The primary objective is to evaluate the tolerability of SRX246.*

The first primary objective of this study is to evaluate the tolerability of SRX246 versus placebo. The Go/No-Go decision for further consideration of SRX246 will be based on testing tolerance among the placebo and treatment groups using a single-tailed test. Because this is an early-phase study (phase IIa), the primary objective will be met by conducting a non-inferiority test (Blackwelder, 1982) of the proportion of study participants who complete the study while active on their assigned intervention arm among the placebo group and each of the treatment groups. Separate tests will be conducted for each of the treatment arms (i.e., a separate test for placebo vs. 120 mg and placebo vs. 160 mg) with a significance level of 0.025 for each test in order to apply an alpha-correction to control for multiple comparisons.

In traditional comparative analysis, the typical null hypothesis (H_0) states that two groups have equal values for their respective parameters, such as the means of the groups. That is, H_0 states that there is no difference between the parameters of the groups. The alternative hypothesis (H_A) states that the two groups are unequal regarding the parameter (i.e., the difference between the parameters is not 0). Failure to reject H_0 means we cannot rule out equal group parameters, and rejection of H_0 indicates the parameters might be unequal.

Non-inferiority test differs from the traditional analysis in that H_0 states that the difference between the parameters of the groups is equal to or greater than a threshold value, δ, which is a meaningful cutoff value that is usually based on clinical relevance or previous research results. Conversely, H_A states that the difference is less than δ. The test of non-inferiority is inherently one-tailed, as H_0 and H_A are directional hypotheses. Focus is on the difference in the proportion of participants meeting the study definition for tolerability in each group. Let us denote the population proportion of study participants meeting the study definition for tolerability in the placebo group as $\pi_{placebo}$, and the population proportion of participants meeting the study definition for tolerability in the 120 mg group as π_{120mg} (a similar proportion is defined for the 160 mg group). Consider the non-inferiority hypotheses for the comparison of the placebo and 120 mg groups:

$$H_0: \pi_{placebo} - \pi_{120mg} \geq \delta$$

vs.

$$H_A: \pi_{placebo} - \pi_{120mg} < \delta.$$

For the purposes of this study, the non-inferiority margin was set to $\delta = 0.5$, which is the same value used in a series of clinical trials sponsored by the Huntington Study Group (HSG, 1998; HSG, 2003; HSG, 2004; HSG/TETRA-HD Investigators, 2009). Using this threshold means that when H_0 is rejected, $\pi_{120mg} - \pi_{placebo} < 0.5$. Or, equivalently, the tolerability in the treated group is less than 50% lower than the tolerability in the placebo group, which is taken to be a negligible (non-meaningful) difference in tolerability. Thus, non-inferiority of tolerability within the margin is consistent with the rejection of the null hypothesis and will constitute a Go decision for further consideration of that dose of SRX246, perhaps in a phase IIb study to detect signs of activity. Conversely, failure to reject the null hypothesis would imply that there could be tolerability concerns and will result in a No-Go decision for that dose. Note that rejection of H_0 does not necessarily mean there is strict equivalence among the groups (i.e., does not imply that the difference is 0). Rather, any tolerability difference is below the acceptable 50% threshold.

The statistical test of H_0 will be performed using a one-sided CI for the difference of the proportions based on exact methods. Exact methods have known sampling distributions in contrast to asymptotic methods that have approximate sampling distributions (Agresti, 2003). Exact methods are preferable because their confidence intervals (CIs) have a minimum coverage probability, especially for small sample sizes (the minimum is $1 - \alpha$). The CI to be used corresponds to inverting a family of exact tests, with the 97.5% CI ($\alpha = 0.025$) including all the values that are not rejected at the .025 level in the corresponding test. Though an exact CI has a minimum coverage probability of $1 - \alpha$, its length can be too long, which means the associated inverted test has low statistical power. To address this issue, and to ensure that H_0 is evaluated using the highest possible statistical power, a computer-intensive method will be used to select the shortest one-sided CI for testing (Wang, 2010).

The exact testing assumes binomial distributions for the variables. A study participant is classified as meeting the definition for tolerability if they complete the study on assigned treatment, or as a non-completer if they do not finish the trial (withdrawal for any reason) or have to withdraw from assigned treatment before the end of the study. The number of participants meeting this definition in the placebo group and each treatment group (e.g., 120 mg) is binomially distributed. There is an analytic formula for the joint probability function, which forms the basis of the exact test.

Suppose that $\theta = \pi_{placebo} - \pi_{120mg}$, so that H_0: $\theta \geq 0.5$ and H_A: $\theta < 0.5$. We reject H_0 if the one-tailed $1 - \alpha$ CI for θ does not contain 0.5. When the proportions are equal, $\theta = 0$, and when tolerability is higher in the placebo group, $\theta > 0$. Therefore, we are interested in the lower one-sided CI with range [-1 , UCI], where UCI is the upper CI limit. If UCI < 0.5, then we reject H_0. If UCI ≥ 0.5, then we do not reject H_0. In summary,

UCI < 0.5, Reject H_0;

UCI ≥ 0.5, Do Not Reject H_0.

For each analysis, we set $\alpha = 0.025$, which means H_0 is tested with a one-sided $1 - \alpha = 0.975$ CI.

An additional consideration is that using the usual asymptotic test can be inaccurate with smaller sample sizes, especially when $\pi_{placebo}$ (or π_{120mg}) are close to 1.0, as is anticipated in this study (Fagerland, 2011). Computer intensive exact methods can have better performance under these conditions (Santner, 2007). The shortest lower one-sided CI will be computed with a computer-intensive ranking-based method (Wang, 2010). Wang's method appears to have superior performance to other approaches, and software is readily available (Wang, 2012). The method will be computed with the BinomCI() function of the ExactCIdiff package (Shan & Wang, 2013) for the R computing software.

The results of the tolerability analysis will be summarized graphically, with a bar plot showing the proportion difference and one-sided CI for each group comparison. The results will also be depicted as in Table 2 below.

Table 2. Proportion of Participants Meeting Study Definition for Tolerability

Group	Placebo (N = XX)	120 mg (N = XX)	160 mg (N = XX)
Number (%)	XX (XX%)	XX (XX%)	XX (XX%)
One-Sided CI for Difference	N/A	[-1 , XX%]	[-1 , XX%]
Decision	N/A	Tolerable (Yes/No?)	Tolerable (Yes/No?)

8. SECONDARY ANALYSIS

Secondary Objective: *The secondary objective is to evaluate the safety of SRX246.*

The safety objective will be met by conducting a non-inferiority test of the proportion of study participants with AEs or SAEs among the placebo group and each of the treatment groups (separate tests for placebo vs 120 mg and placebo vs 160 mg). The safety analysis will be conducted in a similar manner as the tolerability analysis. Each study participant will be classified as having experienced a treatment-emergent AE or SAE (event = "Yes") or not (event = "No"). Suppose the proportion of events in the placebo group is now denoted as $\pi_{placebo}$ and the proportion of events in the 120 mg group is now π_{120mg}, with $\theta = \pi_{placebo} - \pi_{120mg}$. The same onesided CI will be used to test the same null hypothesis (H_0: $\theta \geq \delta$) with the same threshold ($\delta = 0.5$) as in the tolerability analysis. Each treatment group (120 mg, 160 mg) will be compared to the placebo group using $\alpha = 0.025$, meaning a $1 - \alpha = 0.975$ one-sided CI will be used for each test. Results will be summarized as shown in Table 3 below.

Table 3. Summary of Safety Analyses

	Placebo (N = XX)	120 mg (N = XX)	160 mg (N = XX)
Any AE	XX (XX%)	XX (XX%)	XX (XX%)
Any Treatment-Related AE	XX (XX%)	XX (XX%)	XX (XX%)
Any Serious AE	XX (XX%)	XX (XX%)	XX (XX%)
Early Drug Withdrawals (Any)	XX (XX%)	XX (XX%)	XX (XX%)
Early Drug Withdrawals (Due to AE)	XX (XX%)	XX (XX%)	XX (XX%)
Early Terminations (Any)	XX (XX%)	XX (XX%)	XX (XX%)
Early Terminations (Due to AE)	XX (XX%)	XX (XX%)	XX (XX%)
Deaths (Any)	XX (XX%)	XX (XX%)	XX (XX%)
Deaths (Due to AE)	XX (XX%)	XX (XX%)	XX (XX%)

In addition to an overall comparison, this hypothesis will be repeated for assessing the percentage of study participants having at least one treatment-emergent AE within each MedDRA system organ class (SOC). If there are specific differences within any specific SOC, then additional tests will compare differences across groups for specific MedDRA preferred terms in order to further explore the cause of the observed differences. Finally, the subset of treatment-related SAEs will be analyzed in the same manner described above.

8.1. Additional Exploratory Analyses

A number of additional exploratory analyses are also planned to assess several measures of irritability and other behaviors, and clinical assessments for activity signal, but will not be included as part of the FSR. These additional analyses will include:

- To explore the activity of SRX246 over a period of 12 weeks compared to placebo using measures of irritability and other problem behaviors:
  - Aberrant Behavior Checklist (ABC)
  - Cohen-Mansfield Aggression Inventory (CMAI)
  - Problem Behaviors Assessment – Short Form (PBA-s)
  - Irritability Scale (IS)
  - Unified Huntington Disease Rating Scale (UHDRS)
- To explore the use of the following in irritable study participants with early symptomatic HD who are given SRX246 over a period of 12 weeks compared to placebo:
  - Clinical Global Impression (CGI)
  - HD Quality of Life (HD QoL)
  - Caregiver Burden Questionnaire

9. SAMPLE SIZE JUSTIFICATION

Required sample size was estimated for the primary objective of evaluating the tolerability of SRX246. As previously discussed, this objective will be met by conducting a non-inferiority test of the proportion of completers among the placebo group and each of the treatment groups (separate tests for placebo vs. 120 mg and placebo vs. 160 mg). As described in section 7, the non-inferiority test for the primary objective will be based on exact methods with a computer-intensive ranking procedure to ensure the highest statistical power (Wang, 2010). It is impractical to use the aforementioned procedure to estimate the required sample size. Therefore, an approximation was used based on the asymptotic test of the difference between two proportions.

Let us again denote the proportion of participants meeting the study definition for tolerability in the placebo group as $\pi_{placebo}$, and the proportion in the 120 mg group as π_{120mg}. The sample size for achieving 80% power for the placebo vs 120 mg groups can be calculated with the following equation,

$$n_{\text{placebo}} = \frac{(Z_{.025/2} + Z_{.80/2})^2 \cdot [\pi_{\text{placebo}} \cdot (1 - \pi_{\text{placebo}}) + \pi_{120\text{mg}} \cdot (1 - \pi_{120\text{mg}})]}{(\delta - \pi_{\text{placebo}} - \pi_{120\text{mg}})^2},$$

where δ is the threshold previously discussed, $n_{placebo}$ is the sample size for the placebo group, and the total for the entire study is $N = 3 \cdot n_{placebo}$.

The above equation indicates that a power analysis for this scenario requires estimates for the values of $\pi_{placebo}$, π_{120mg}, and δ. A power analysis was conducted using estimated values of the parameters obtained from previous research. Following the practice of the Huntington Study Group (HSG, 1998; HSG, 2003; HSG, 2004; HSG/TETRA-HD Investigators, 2009), the margin was set at $\delta = 0.50$ for all analysis. A difference of <50% between groups is considered acceptable because of the paucity of effective treatments for HD (HSG, 1998). For the remaining parameters, plausible values were based on published literature for relevant randomized clinical trials (RCTs) in HD (HSG, 1998; HSG, 2003; HSG, 2004; HSG/TETRA-HD Investigators, 2009; Kieburtz et al, 2010; Squitieri et al, 2013). Table 4 below shows completer information for placebo and treatment groups of six HD RCTs employing various drugs (for studies with multiple treatment groups, group with the smallest percentage of completers is shown). Placebo completion rates across the six studies ranged from 82%-97%, and the observed treatment group differences ranged from 0%-18%.

In light of the information in Table 4, sample size was computed for:

- $\pi_{placebo}$ = 0.80, 0.85, 0.90, 0.95, and
- $\pi_{placebo} - \pi_{treatment}$ = 0.01, 0.02,…, 0.19, 0.20.

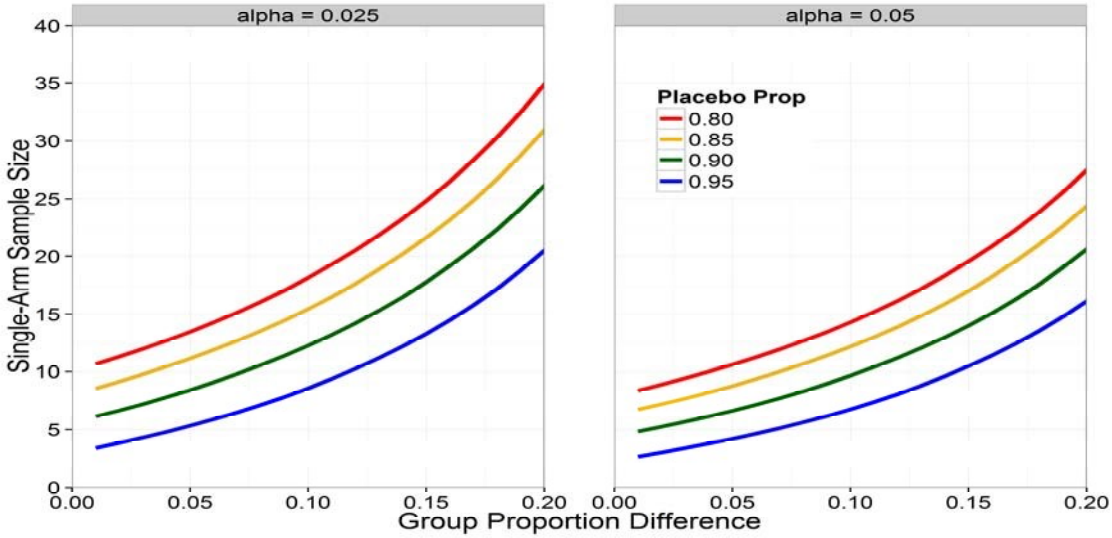
It should be noted that the minimum $\pi_{placebo}$ = 0.80 is less than any placebo value in Table 4, and the maximum difference ($\pi_{placebo} - \pi_{120mg}$) of .20 is greater than any observed difference in Table 4. Therefore, these values constitute reasonable bounds for assessing required sample size for this study. The non-inferiority test is single-tailed, and type I error rates of $\alpha = 0.05/2 = 0.025$ and $\alpha = 0.10/2 = 0.05$ were used to allow for two comparisons (each dose level versus placebo), with target power set at 80%.

Table 4. Tolerability Statistics from HD Clinical Trials.

Study	Drug	N	Placebo	Treatment	Difference
HSG (1998)	OPC-14117	64	94%	81%	13%
HSG (2003)	Riluzole	63	91%	87%	4%
HSG (2004)	Minocycline	60	96%	78%	18%
HSG/TETRA-HD (2009)	Tetrabenazine	84	97%	91%	+6%
Kieburtz et al (2010)	Latrepirdine	91	82%	87%	-5%
Squitieri et al (2013)	Pridopidine	353	86%	86%	0%

The results of the power analysis are shown in Figure 2 below. The figure shows the single-arm sample size ($n_{placebo}$) as a function of $\pi_{placebo}$ and the group proportion difference ($\pi_{placebo} - \pi_{120mg}$), paneled by α-level. In the left-hand panel, the upper red line shows that conducting two tests each at $\alpha = 0.025$ and 80% power requires approximately 36 participants per group for the smallest proposed control proportion ($\pi_{placebo} = 0.80$) and the largest proportion difference ($\pi_{placebo} - \pi_{treatment} = 0.20$). Therefore, we proposed a total sample size of $N = 3 \times 36 = 108$. The power analysis suggests that a total sample size of 2 x 36 = 72 will be used for each group comparison of completer proportions (120 mg vs placebo, 160 mg vs placebo). The margin used to derive this sample size ($\delta = 0.5$) allows the possibility of a relatively small proportion of completers in the treatment group. For example, if $\pi_{placebo} = 0.90$, H_0 will be rejected if $\pi_{placebo} > 0.04$, meaning there could be a minority of completers.

Figure 2. Sample Size for the Placebo Group as a Function of Effect Size, Placebo Proportion, and Alpha-Level (80% Power)



10. SAFETY MONITORING

10.1. Independent Medical Monitor

The PPI will appoint a Medical Safety Monitor (MSM). The MSM responsibilities will include independent review of safety data, including but not limited to review of all events that meet the regulatory definition of an SAEs in real-time upon notification via the NeuroNEXT Online Adverse Event Reporting System (AERS). In addition to performing real-time reviews of all SAEs (as described in section 4), the MSM will also receive aggregate, blinded reports of all AEs for review on a quarterly basis (or as requested). These aggregate reports will summarize all AEs by severity, attribution (expected or unexpected), and relationship to study drug procedures in a tabular form. The MSM will be responsible for providing a written summary of this review to the DCC. Should any concerns arise due to observed trends, the MSM will send a written recommendation to the DCC requesting that the report be forwarded to the Protocol PIs and/or the Data and Safety Monitoring Board (DSMB), as appropriate.

10.2. Data and Safety Monitoring Board

The monitoring of study participant safety and data quality will follow the NINDS Guidelines for Data and Safety Monitoring in Clinical Trials. A DSMB, appointed by NIH/NINDS, will meet at approximately six-month intervals (or as determined by the NINDS) to review partially unblinded study data provided by the study statistician. The DSMB will periodically review and evaluate the accumulated data for participant safety, adverse events, study conduct, and study progress. The DSMB may suggest changes to the protocol or consent form to the PPI as a consequence of AEs. The DSMB may also make recommendations to NINDS concerning continuation, modification, or termination of the study. The frequency and format of DSMB meetings and reports will be agreed upon prior to study enrollment.

REFERENCES

Agresti A (2003). “Dealing with discreteness: Making ’exact’ confidence intervals for proportions, differences of proportions, and odds ratios more exact,” *Statistical Methods in Medical Research*, 12:3-21.

Blackwelder WC (1982). “’Proving the Null Hypothesis’ in clinical trials,” *Controlled Clinical Trials*, 3: 345-353.

Detry MA and Lewis RJ (2014). “The intention-to-treat principle: How to assess the true effect of choosing a medical treatment,” *Journal of the American Medical Association*, 312: 85-86.

Fagerland MW, Lydersen S, and Laake P (2015). “Recommended confidence intervals for two independent binomial proportions,” *Statistical Methods in Medical Research*, 24: 224-254.

Huntington Study Group (1998). “Safety and tolerability of the free-radical scavenger OPC-14117 in Huntington’s disease,” *Neurology*, 50:1366-1373.

Huntington Study Group (2003). “Dosage effects of riluzole in Huntington’s disease: A multicenter placebocontrolled study,” *Neurology*, 61:1551-1556.

Huntington Study Group (2004). “Minocycline safety and tolerability in Huntington disease,” *Neurology*, 63:547549.

Huntington Study Group/TETRA-HD Investigators (2009). “Tetrabenazine as antichorea therapy in Huntington disease: An open-label continuation study,” *BMC Neurology*, 9: 62. Erratum in: *BMC* Neurology, 11: 18.
Santner TJ, Pradhan V, Senchaudhuri P, Mehta CR, and Tamhane A (2007). “Small-sample comparisons of confidence intervals for the difference of two independent binomial proportions,” *Computational Statistics & Data Analysis*, 51: 5791-5799.

Kieburtz K, McDermott MP, Voss TS, Corey-Bloom J, Deuel LM, Dorsey ER, Factor S, Geschwind MD, Hodgeman K, Kayson E, Noonberg S, Pourfar M, Rabinowitz K, Ravina B, Sanchez-Ramos J, Seely L, Walker F, Feigin A; Huntington Disease Study Group DIMOND Investigators. A randomized, placebo-controlled trial of latrepirdine in Huntington disease. Arch Neurol. 2010 67:154-60.Shan G and Wang W (2013). *ExactCIdiff: Inductive confidence intervals for the difference between two proportions*; https://CRAN.R-project.org/package=ExactCIdiff.

Squitieri, F., B. Landwehrmeyer, R. Reilmann, A. Rosser, J. Garcia de Yebenes, A. Prang, J. Ivkovic, J. Bright, Å. Rembratt. One-year safety and tolerability profile of pridopidine in patients with Huntington disease. Neurology Mar 2013, 80 (12) 1086-1094; DOI:10.1212/WNL.0b013e3182886965

Wang W (2010). “On construction of the smallest one-sided confidence interval for the difference of two proportions,” *The Annals of Statistics*, 38:1227-1243.

Wang W (2012). “An inductive order construction for the difference of two dependent proportions,” *Statistical and Probability Letters*, 82: 1623-1628.