Statistical Analysis Plan (SAP)

We acknowledge the Cambridge University Hospitals Clinical Trials Unit for the development of an SAP template (version CCTU/TPLV2), which was later modified by the Michigan Institute for Clinical & Health Research. Here, we additionally modified this template based on information provided in the Delphi consensus, "Guidelines for Content of Statistical Analysis Plans in Clinical Trials" published in the *Journal of the American Medical Association* (*1*).

TRIAL FULL TITLE	Intranasal Vasopressin Treatment in Children with Autism
Clinicaltrials.gov number	NCT03204786
SAP VERSION	2.0
SAP VERSION DATE	11/17/2023
TRIAL STATISTICIAN	Professor Joseph P. Garner, DPhil
TRIAL PRINCIPAL INVESTIGATORS	Professor Karen J. Parker, PhD Professor Antonio Y. Hardan, MD
SAP AUTHORS	Professor Joseph P. Garner, PhD Professor Karen J. Parker, PhD Professor Antonio Y. Hardan, MD
Primary Funder	National Institutes of Health (NIH)

1 SAP Version History

Version	Date(s)	Revision description
1.0	Grant application received at NIH: 08/05/2016; Grant start date: 07/01/2017	SAP written and included in a grant application (R01HD091972) to NIH.
2.0	11/17/2023	No conceptual changes were made to the original SAP. However, additional details are included here.

The study design and SAP were developed in the context of an NIH grant application. This grant was subsequently awarded, which enabled us to initiate the present clinical trial. The study design, safety measures, trial's primary outcome measure, trial's main secondary outcome measures, and statistical analysis strategy have not changed since 2016. Several other pre-specified secondary outcome measures were added to this trial and registered on Clinicaltrials.gov and detailed in the Study Protocol. While we continue to follow the original study design and SAP strictly, here we provide additional details required for a more formal SAP.

2 Table of Contents

3 Abbreviations and Definitions

η_p^2	Partial eta-squared
AICc	Akaike Information Criterion corrected
ASD	Autism Spectrum Disorder
BP	Blood Pressure
CARS-2	Childhood Autism Rating Scale, 2nd Edition
CGI	Clinical Global Impression
CO2	Carbon Dioxide
CONSORT	Consolidated Standards of Reporting Trials
COVID-19	Corona Virus Disease 2019
CSHQ	Child's Sleep Habits Questionnaire
DANVA-2	Diagnostic Analysis of Nonverbal Accuracy, 2nd Edition
DD	Drug-Drug
DDAVP	Desmopressin
DOTES	Dosage Record Treatment Emergent Symptom Scale
DSM-5	Diagnostic and Statistical Manual, 5th Edition
DSMB	Data Safety Monitoring Board
EEG	Electroencephalogram
EKG	Electrocardiogram
FDA	Food and Drug Administration
FERT	Facial Emotion Recognition Test
HR	Heart Rate
IQ	Intelligence Quotient
LSM ± SE	Least Squares Mean ± Standard Error
NEPSY-II	Developmental Neuropsychological Assessment, 2nd Edition
NIH	National Institutes of Health
OAS	Overt Aggression Scale
PD	Placebo-Drug
PedsQL	Pediatric Quality of Life
PHI	Protected Health Information
PP	Placebo-Placebo
PRAS-ASD	Parent Rated Anxiety Scale - Autism Spectrum Disorder
RBS-R	Repetitive Behavior Scale Revised
REDCap	Research Electronic Data Capture
REML	Restricted Maximum Likelihood
RMET	Reading the Mind in the Eyes Test
SAP	Statistical Analysis Plan
SCAS	Spence Children's Anxiety Scale
SPCD	Sequential Parallel Comparison Design
SRS-2	Social Responsiveness Scale, 2nd Edition
Temp	Temperature
Tx	Treatment
VABS-3	Vineland Adaptive Behavior Scales, 3rd Edition
WISC-V	Wechsler Intelligence Scale for Children, 5th Edition

4 Introduction

4.1 Preface

Autism spectrum disorder (ASD) is characterized by core social impairments which limit an individual's ability to form and maintain meaningful relationships (*2*). At present, antipsychotics are the only medication approved by the FDA to treat ASD, but they target only associated symptoms, have unfavorable side-effects, and do not treat ASD's core behavioral features. Developing new medications that specifically target social functioning will thus address an important unmet need (*3*).

A large body of research (reviewed in the Study Protocol) has shown that the neuropeptide arginine vasopressin (hereafter, vasopressin) plays a critical role in promoting social behavior and that experimental dysregulation of the vasopressin signaling pathway produces social deficits in animal models (*4-6*). Although intranasal vasopressin administration improves social cognition and memory in neurotypical individuals (*7-9*), no research had tested the effects of vasopressin treatment in people with ASD prior to our pilot vasopressin treatment trial. Several lines of evidence underscore the necessity of such research. For example, this finding is consistent with our research showing that socially impaired monkeys and people with ASD have significantly diminished cerebrospinal fluid vasopressin levels compared to relevant controls (*10-12*). Similarly, data from the first neuropeptide receptor mapping study of post-mortem primate brain tissue revealed that vasopressin V_{1a} receptors are widely distributed throughout the extended neural amygdala pathway, suggesting that vasopressin administration can directly target pathways known to regulate social functioning (*13*). Interestingly, vasopressin's effects are especially evident in male animals (*14, 15*), and given ASD's male-biased prevalence, vasopressin deficits may be particularly relevant to understanding the risk for, and treatment of, ASD.

On the basis of this body of evidence, we previously tested the effects of 4-week intranasal vasopressin treatment in children with ASD in a quadruple-blinded, randomized, placebo-controlled pilot trial (*16*). Vasopressin was overall well tolerated with minimal side effects in this sample (N=30). Importantly, vasopressin treatment improved social abilities in children with ASD as assessed by parent ratings on the Social Responsiveness Scale, 2nd Edition (SRS-2), the trial's primary outcome measure. Vasopressin treatment also improved social abilities as measured by clinician impression and child performance on social cognition assessments. Finally, vasopressin treatment diminished anxiety symptoms and some repetitive behaviors. Most of these findings were more pronounced when we accounted for pre-treatment vasopressin concentration in blood.

Here we seek to replicate and extend the pilot trial's findings in an 8-week quadruple-blinded, randomized, and placebo-controlled intranasal vasopressin treatment trial. Our randomization target is N=100 children, aged 6 to 17 years. Our primary outcome measure is improvement in child social abilities as assessed by parent ratings on the SRS-2. We will also test the safety and tolerability of vasopressin treatment, and whether pre-treatment blood vasopressin levels are a personalized predictor of treatment efficacy. Finally, we will test whether vasopressin treatment improves ASD symptoms as assessed by clinician impression, additional parent report measures, and child performance on assessments of social cognition and communication. We predict that vasopressin treatment will improve social abilities in children with ASD, and that vasopressin will be well-tolerated, in keeping with our pilot vasopressin treatment trial data.

4.2 Scope of the Analyses

The goal of this investigation is to test whether intranasal vasopressin treatment improves social abilities in children with ASD (N=100), aged 6 to 17 years. The overall design, methods, and efficacy and safety outcome measures of the present trial are based on our now published pilot vasopressin treatment trial

(N=30) findings (*16*). However, several improvements in the present trial were made in light of our experiences in the pilot trial. In addition to using a Sequential Parallel Comparison Design (SPCD), as described below, we expanded the trial age range to 17 years to assess the efficacy of vasopressin treatment in adolescents. We also expanded the IQ cut-off to 40 and above to optimize the generalizability of our study findings.

5 Study Objectives and Outcome Measures

5.1 Study Objectives

As with our prior pilot vasopressin treatment trial, this trial's primary outcome measure will be assessed by change from baseline in blinded parent ratings on the SRS-2. Secondary outcome measures consist of drug safety and tolerability, biological predictors of treatment response, and symptom improvement assessed by clinician ratings, parent ratings, and child performance on social cognition and communication assessments.

5.2 Outcome Measures

The Clinicaltrials.gov registration and the Study Protocol elaborate on each of these outcome measures. Here we briefly list them according to primary, secondary, and other pre-specified outcome measures. The listed primary and secondary outcome measures were included in the pilot vasopressin treatment trial. Some new secondary outcome measures were added to the current trial, and appear in the other pre-specified outcome measures section.

Primary Outcome Measure:

1. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Total scores during treatment. [Time Frame: 4-week; 8-week]

Secondary Outcome Measures:

2. Change from baseline in Clinical Global Impression (CGI) scores during treatment.
[Time Frame: 4-week; 8-week]
3. Change from baseline on Reading the Mind in the Eyes Test (RMET) during treatment.
[Time Frame: 4-week; 8-week]
4. Change from baseline on the Facial Emotion Recognition Test (FERT) during treatment.
[Time Frame: 4-week; 8-week]
5. Change from baseline in parent rated Repetitive Behavior Scale Revised (RBS-R) scores during treatment.
[Time Frame: 4-week; 8-week]
6. Change from baseline in parent rated Spence Children's Anxiety Scale (SCAS) during treatment.
[Time Frame: 4-week; 8-week]
7. Change from baseline on electrocardiogram (EKG) P Duration during treatment.
[Time Frame: 4-week; 8-week]
8. Change from baseline on electrocardiogram (EKG) PR Interval during treatment.
[Time Frame: 4-week; 8-week]
9. Change from baseline on electrocardiogram (EKG) QRS Interval during treatment.
[Time Frame: 4-week; 8-week]
10. Change from baseline on electrocardiogram (EKG) QT Interval during treatment.
[Time Frame: 4-week; 8-week]
11. Change from baseline on blood clinical labs (Sodium) during treatment.
[Time Frame: 4-week; 8-week]
12. Change from baseline on blood clinical labs (Potassium) during treatment.
[Time Frame: 4-week; 8-week]
13. Change from baseline on blood clinical labs (Chloride) during treatment.
[Time Frame: 4-week; 8-week]
14. Change from baseline on blood clinical labs (CO2) during treatment.

[Time Frame: 4-week; 8-week]
15. Change from baseline on blood clinical labs (Anion Gap) during treatment.
[Time Frame: 4-week; 8-week]
16. Change from baseline on blood clinical labs (Glucose) during treatment.
[Time Frame: 4-week; 8-week]
17. Change from baseline on blood clinical labs (Creatinine) during treatment.
[Time Frame: 4-week; 8-week]
18. Change from baseline on blood clinical labs (Urea Nitrogen) during treatment.
[Time Frame: 4-week; 8-week]
19. Change from baseline on blood clinical labs (Calcium) during treatment.
[Time Frame: 4-week; 8-week]
20. Change from baseline on blood clinical labs (Osmolality) during treatment.
[Time Frame: 4-week; 8-week]
21. Change from baseline on urine clinical labs (Osmolality) during treatment.
[Time Frame: 4-week; 8-week]
22. Change from baseline on vital signs (systolic blood pressure) during treatment.
[Time Frame: 4-week; 8-week]
23. Change from baseline on vital signs (diastolic blood pressure) during treatment.
[Time Frame: 4-week; 8-week]
24. Change from baseline on vital signs (pulse) during treatment.
[Time Frame: 4-week; 8-week]
25. Change from baseline on height during treatment.
[Time Frame: 4-week; 8-week]
26. Change from baseline on weight during treatment.
[Time Frame: 4-week; 8-week]
27. Change from baseline on the Dosage Record Treatment Emergent Symptom Scale (DOTES) during treatment.
[Time Frame: 2-week, 4-week; 6-week, 8-week]
28. Change from baseline in Overt Aggression Scale (OAS) during treatment.
[Time Frame: 2-week, 4-week; 6-week, 8-week]
29. Baseline vasopressin concentration predicting primary and secondary behavioral outcome measures.
[Time Frame: 4-week; 8-week]

Other Pre-specified Outcome Measures:

30. Change from baseline in Vineland Adaptive Behavior Scales, 3rd edition (VABS-3) - Social Skills and Relationships Domain during treatment.
[Time Frame: 4-week; 8-week]
31. Change from baseline in parent rated Pediatric Quality of Life (PedsQL) inventory scores during treatment.
[Time Frame: 4-week; 8-week]
32. Change from baseline on Eye Gaze Assessment (eye tracking) during treatment.
[Time Frame: 4-week; 8-week]
33. Change from baseline on the Developmental Neuropsychological Assessment, 2nd edition (NEPSY-II) Theory of Mind Test during treatment.
[Time Frame: 4-week; 8-week]
34. Change from baseline the Diagnostic Analysis of Nonverbal Accuracy, 2nd edition (DANVA-2) Child Voices Prosody Test during treatment.
[Time Frame: 4-week; 8-week]

35. Change from baseline in parent rated Stanford Social Motivation Scale (also known as the Stanford Social Dimensional Scale) Total scores during treatment.
[Time Frame: 4-week; 8-week]
36. Change from baseline in Parent Rated Anxiety Scale - Autism Spectrum Disorder (PRAS-ASD) score during treatment.
[Time Frame: 4-week; 8-week]
37. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Social Avoidance Factor Score during treatment.
[Time Frame: 4-week; 8-week]
38. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Emotion Recognition Factor Score during treatment.
[Time Frame: 4-week; 8-week]
39. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Interpersonal Relatedness Factor Score during treatment.
[Time Frame: 4-week; 8-week]
40. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Insistence on Sameness Factor Score during treatment.
[Time Frame: 4-week; 8-week]
41. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Repetitive Mannerisms Factor Score during treatment.
[Time Frame: 4-week; 8-week]
42. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Attachment and Affiliation Factor Score during treatment.
[Time Frame: 4-week; 8-week]
43. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Non-facial Communication Production Factor Score during treatment.
[Time Frame: 4-week; 8-week]
44. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Facial Communication Production Factor Score during treatment.
[Time Frame: 4-week; 8-week]
45. Change from baseline in parent rated Social Responsiveness Scale, 2nd edition (SRS-2) Mental States Understanding Factor Score during treatment.
[Time Frame: 4-week; 8-week]
46. Change from baseline in parent rated Child's Sleep Habits Questionnaire (CSHQ) Score during treatment.
[Time Frame: 4-week; 8-week]
47. Change from baseline on spectral power in the alpha, theta, and gamma frequencies as measured by electroencephalogram (EEG) during treatment.
[Time Frame: 4-week; 8-week]

6 Study Methods

6.1 General Study Design and Plan

In addition to vasopressin treatment efficacy and tolerability, data from our prior pilot vasopressin treatment trial revealed a placebo response consistent with findings from many other clinical trials conducted in patients with ASD (*17-19*), as well as in patients with neuropsychiatric disorders (*20*). To tackle this issue, here we will employ an SPCD.

SPCD is a highly innovative approach to blinded randomized placebo-controlled trials (*20-24*). SPCD is a powerful clinical trial design compared to standard parallel comparison designs or crossover designs when high placebo response rates are expected. SPCD is thus useful in overcoming core clinical trial design challenges, and is rapidly gaining acceptance as the gold-standard approach to clinical trial design. This powerful methodological approach is also in keeping with NIH's emphasis on scientific rigor and reproducibility.

SPCD involves two phases. Phase 1 comprises an initial drug vs. placebo parallel comparison, whereby the sample size in each group is scaled to the expected placebo response rate. In Phase 2, the placebo group is divided into two groups, with half of the participants continuing on placebo and half switching to drug. Participants randomized in Phase 1 to the drug group continue on drug in Phase 2. Thus, in this clinical trial, participants will be assigned to one of three pre-determined groups: Drug-Drug (DD); Placebo-Placebo (PP); or Placebo-Drug (PD).

The advantage of the SPCD design lies in the statistical analysis. Phase 1 can be analyzed as a classic parallel comparison, in which the PP and PD groups are pooled, as they have only received placebo at this point, and recognizing that placebo response rates will be similar in participants receiving placebo and drug. The inherent problem with parallel designs is that there is no objective way to identify placebo responders within the drug-treated group. This is where Phase 2 is critical. Within the PP and PD groups, placebo responders are readily identified upon completion of Phase 1, and are eliminated from further comparison in Phase 2, just as if Phase 1 had been a double-blind lead-in to identify placebo responders. Phase 2 is thus analyzed by comparing only the Phase 1 placebo non-responders, and by comparing the improvement during Phase 2 as if it was a parallel comparison between placebo and drug (which it is). The initial sample sizes are chosen so that the number of placebo non-responders in the PD group (i.e., the sample size for the PD group in the PD vs. PP Phase 2 comparison) is equal to the sample size for the DD group. Finally, the data from both Phase 1 and 2 for all participants in the PP and DD groups can be used to investigate treatment response over the course of the entire study. This ingenious experimental design therefore facilitates three comparisons in a single clinical trial: a straightforward, classic parallel comparison (in Phase 1); a lead-in and placebo responder

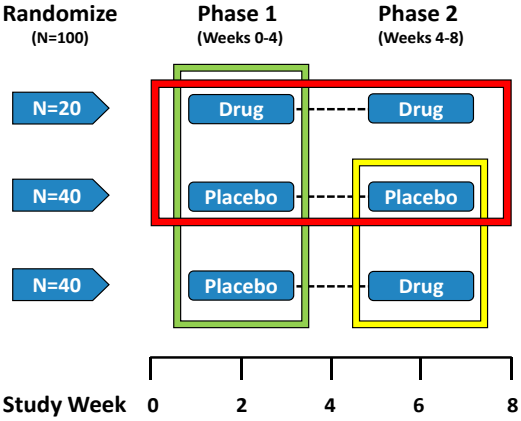

Figure 1. *Participants are randomized to one of three groups (Drug-Drug; Placebo-Placebo; or Placebo-Drug), comprised of two Phases. Sequential Parallel Comparison Design facilitates three comparisons, depicted here by colored boxes. The green box represents a classic 4-week parallel comparison during Phase 1. The yellow box represents a 4-week parallel comparison during Phase 2, using only participants who were placebo non-responders in Phase 1. The red box represents an extended 8-week time course parallel comparison (Phase 1 + Phase 2).*

excluded parallel comparison (in Phase 2); and an extended time course parallel comparison (Phase 1 + Phase 2). Furthermore, by combining these analyses, power is far greater than conducting each of these three comparisons as separate trials.

For this trial (see Figure 1), Phase 1 and Phase 2 will each last four weeks. The rationale for selecting 4-week phases is that we observed a significant vasopressin treatment response within this period of time in our pilot trial. In the pilot trial, we observed a bimodal distribution of SRS-2 Total Scores (change from baseline). Thus, we were able to distinguish placebo responders and non-responders in the prior trial with a 10-point improvement or more on the SRS-2 Total Score. Given that we are working with a different aged population here, we will adopt a similar approach, i.e., we will check for bimodality and identify a cut-off to distinguish placebo responders from non-responders. Following SPCD as outlined above, placebo responders will be excluded from the statistical comparison of PP and PD groups in Phase 2. Given a placebo response rate in our pilot data of approximately 50%, we aim to randomize N=20 individuals to DD, N=40 individuals to PD, and N=40 individuals to PP. This anticipates that roughly N=20 participants will be identified as placebo lead-in non-responders each in the PP and PD groups. The 4-week Phase 1 comparison thus will comprise N=20 DD vs. N=80 PD+PP participants; the 4-week Phase 2 comparison will comprise N=20 PP vs. N=20 PD placebo non-responder participants; and the 8-week extended time course comparison will comprise N=20 DD vs. N=40 PP participants. The Restricted Maximum Likelihood (REML) mixed model analysis developed for SPCD combines evidence from all three of these comparisons to generate an omnibus test for the primary outcome measure of the trial (*21*).

6.2 Sample Size

For details see the description of SPCD in the section above. The target randomized sample size is N=100 in a 1:2:2 allocation ratio. This number was determined based on custom SPCD power considerations and secondarily confirmed by Mead's Rule for factorial designs (*25*). Study drop-out is anticipated and included in these calculations.

6.3 Inclusion-Exclusion Criteria and General Study Population

Inclusion Criteria

All participants will meet the following:

a) Medically healthy outpatients between 6 and 17 years of age;
b) Diagnostic and Statistical Manual 5th edition (DSM-5) criteria for Autism Spectrum Disorder on the basis of clinical evaluation, confirmed with the Autism Diagnostic Interview Revised and Autism Diagnostic Observation Schedule, 2nd edition or Childhood Autism Rating Scale, 2nd Edition (CARS-2);
c) Males and females;
d) IQ of 40 and above;
e) Rating of 4 or higher on the Social Communication domain of the CGI - Severity;
f) SRS-2 Total score of 70 and above;
g) Care provider who can reliably bring participant to clinic visits, provide trustworthy ratings, and interact with participant on a regular basis;
h) Stable concomitant psychotropic medications or medications potentially affecting vasopressin for at least 4 weeks (with the exception of fluoxetine, 6 weeks);
i) No planned changes in psychosocial and biomedical interventions during the trial;
j) Willingness to provide blood samples and ability to participate in key study procedures (i.e., diagnostic assessments and laboratory safety measurements).

Exclusion Criteria

Participants will be excluded if one or more of the following are met:

a) DSM-5 diagnosis of schizophrenia, schizoaffective disorder, or psychotic disorder;
b) Regular nasal obstruction or nosebleeds;
c) Unstable medical conditions such as migraine, asthma attacks, or seizures, and significant physical illness (e.g., serious liver disease, renal dysfunction, or cardiac pathology);
d) Clinically significant abnormal EKG reading;
e) History of hypersensitivity to vasopressin, its analogs, or compounding preservatives (e.g., chlorobutanol);
f) Evidence of a genetic mutation known to cause ASD or intellectual disability (e.g., Fragile X Syndrome); or metabolic or infectious etiology for ASD on the basis of medical history, neurologic history, and available tests for inborn errors of metabolism and chromosomal analysis;
g) Significant hearing or vision impairments;
h) Habitually drinks large volumes of water;
i) Pregnant or sexually active females not using a reliable method of contraception;
j) Current use of any medications known to interact with vasopressin including: 1) carbamazepine (i.e., Tegretol); chlorpropamide; clofibrate; urea; fludrocortisone; tricyclic antidepressants (all of which may potentiate the antidiuretic effect of vasopressin when used concurrently); 2) demeclocycline; norepinephrine; lithium; heparin; alcohol (all of which may decrease the antidiuretic effect of vasopressin when used concurrently); 3) ganglionic blocking agents including benzohexonium, chlorisondamine, pentamine (all of which may produce a marked increase in sensitivity to the pressor effects of vasopressin);
k) Previous participation in a vasopressin clinical trial or current use of vasopressin;
l) Current use of desmopressin (DDAVP) or oxytocin.

6.4 Randomization and Blinding

As noted above, we will be using a 1:2:2 treatment allocation ratio. Participants will be randomized on the basis of sex (male and female) and age (6-8; 9-11; 12-14; 15-17 years) in blocks of 5. Participants will be randomly allocated to treatment in the order of entry into the study such that every 5 participants will result in fully balanced treatment allocation (i.e., 1:2:2). Each block of 5 allocates treatment in a different order. This information is stored in excel and a single unblinded investigator (Parker) runs the randomization table along with the study statistician (Garner).

The trial employs a quadruple blind (i.e., participant, care provider, investigator, outcome assessors).

6.5 Study Visits and Assessments

Study visits, assessments, and COVID-19 related protocol modifications are detailed in the Study Protocol. A brief overview of study visits and assessments is provided below. Please also see Table 1.

There will be a total of six visits during the blinded portion of this 8-week treatment trial: Visit 1 (screening), Visit 2 (baseline and randomization), Visit 3 (2-week dosing), Visit 4 (4-week dosing), Visit 5 (6-week dosing), and Visit 6 (8-week dosing). Between visits, families will be contacted weekly to monitor protocol adherence and to inquire about side effects.

The following procedures will be completed as part of Visit 1 to determine study eligibility: medical and psychiatric history, physical exam, concomitant medications, inclusion and exclusion criteria, diagnostic testing, IQ testing, SRS-2, EKG, vital signs screening, clinical chemistry labs, and urine and

blood osmolality. Blood aliquots will also be banked during Visit 1 for research purposes to establish baseline biological measures (including blood vasopressin concentration). Clinician-rated scales to establish baseline behavioral functioning will also be obtained.

During Visit 2, inclusion and exclusion criteria will be reviewed to evaluate the participant's continued eligibility. Behavioral measures including clinician-rated scales, parent-rated scales, and child performance on social cognition and communication assessments will be completed.

After the baseline assessments are completed, Phase 1 of the study will begin. Participants will be randomized in a 1:2:2 blinded fashion (stratified by sex and age) to one of three groups as described above: DD, PD, or PP. Each participant will intranasally administer vasopressin or placebo at home.

During Visit 3, participants will be evaluated on the primary outcome measure (i.e., SRS-2) and certain secondary outcome measures including drug safety and tolerability and clinician-rated scales.

Participants will be evaluated during Visit 4 on the primary outcome measure (i.e., SRS-2), and secondary outcome measures (i.e., drug safety and tolerability, clinician-rated scales, parent-rated scales, child performance on social cognition assessments). The EKG, vital signs, clinical chemistry labs, and urine and blood osmolality will be collected for safety monitoring. Blood aliquots will again be banked during Visit 4 for research purposes.

Immediately thereafter, participants will begin Phase 2 of the study, with identical procedures completed during Visits 5 and 6 as were completed at Visits 3 and 4, respectively.

After completion of the blinded 8-week treatment period, participants will be informed of their treatment group and participants in the PP group will have the option of participating in a 4-week open-label vasopressin treatment extension period as detailed in the Study Protocol.

For participants in all three treatment allocation groups, research staff will contact families four weeks after completion of their last vasopressin dose to ask them to complete the SRS-2.

Table 1. Visits and experimental measurements during the blinded portion of the clinical trial.

Measurement	Rationale	Performed by	Outcome Measure Type	Screening Visit 1	Baseline & Randomize Visit 2	Tx Week 2 Visit 3	Tx Week 4 Visit 4	Tx Week 6 Visit 5	Tx Week 8 Visit 6
Medical History	Eligibility	M.D.	N/A	X					
Psychiatric History	Eligibility	M.D.	N/A	X					
Inclusion/Exclusion Criteria	Eligibility	M.D.	N/A	X	X				
Autism Diagnostic Interview – Revised	Eligibility	Investigator	N/A	X					
Autism Diagnostic Observation Schedule, 2nd Ed. or Childhood Autism Rating Scale, 2nd Ed.	Eligibility	Investigator	N/A	X					
Stanford Binet, 5th Ed. or WISC-V	Eligibility	Investigator	N/A	X					
Social Responsiveness Scale, 2nd Ed. (SRS-2) Total Scores	Eligibility & Efficacy	Parent	Primary	X	X	X	X	X	X
Clinical Global Impressions Scale – Severity	Eligibility & Efficacy	M.D.	Secondary	X	X	X	X	X	X
Clinical Global Impressions Scale – Improvement	Efficacy	M.D	Secondary			X	X	X	X
Reading the Mind in the Eyes Test	Efficacy	Investigator	Secondary		X		X		X
Facial Emotion Recognition Test	Efficacy	Investigator	Secondary		X		X		X
Repetitive Behavior Scale – Revised	Efficacy	Parent	Secondary		X		X		X
Spence Children's Anxiety Scale	Efficacy	Parent	Secondary		X		X		X
Side-effects (DOTES)	Safety	M.D.	Secondary	X	X	X	X	X	X
Overt Aggression Scale	Safety	M.D.	Secondary	X		X	X	X	X
Electrocardiogram	Safety	M.D.	Secondary	X			X		X
Blood and Urine Samples	Safety	Clinical Lab	Secondary	X			X		X
Vital Signs (HR, BP, Temp); Height and Weight	Safety	M.D.	Secondary	X			X		X
Baseline Blood Vasopressin Level	Efficacy	Investigator	Secondary	X					
NEPSY-II Theory of Mind Test	Efficacy	Investigator	Secondary		X		X		X
DANVA, 2nd Ed.	Efficacy	Investigator	Secondary		X		X		X
Vineland Adaptive Behavior Scales, 3rd Ed. [Socialization Domain]	Efficacy	Parent	Secondary		X		X		X
PedsQL Inventory	Efficacy	Parent	Secondary		X		X		X
Stanford Social Motivation Scale	Efficacy	Parent	Secondary		X		X		X
Parent-rated Anxiety Scale - ASD	Efficacy	Parent	Secondary		X		X		X
Child Sleep Habits Questionnaire	Efficacy	Parent	Secondary		X		X		X
SRS-2 Factor Scores	Efficacy	Parent	Secondary	X	X	X	X	X	X

Tx=Treatment; WISC-V=Wechsler Intelligence Scale for Children, 5th Ed.; DOTES=Dosage Record & Treatment Emergent Symptom Scale; HR=heart rate; BP=blood pressure; Temp=Temperature; DANVA=Diagnostic Analysis of Nonverbal Accuracy; PedsQL=Pediatric Quality of Life

7 General Analysis Considerations

7.1 Timing of Analyses

We will close the study after we have randomized a minimum of N=100 individuals but before the study medication donated to us by Endo Pharmaceuticals expires. All analyses will be performed after completion of the study (i.e., no interim analysis will be performed). Briefly, all study data will be dual-entered into REDCap and any discrepancies appropriately resolved. The database will be fully de-identified (i.e., all PHI removed). The database will then be exported into excel and provided to the study statistician (Garner).

7.2 Analysis Populations and Missing Data

For analysis of the primary and secondary efficacy data, the SPCD involves three different comparisons as noted above in section 7.1. Each comparison involves a different combination of treatment allocation and study phase/arm, as detailed above. All participants successfully completing the relevant study arm will be included. Any participants dropping out in the middle of a study arm will be excluded from the relevant analyses. Participants may drop out of their own volition or be dropped from analysis if they are found either during the study or following completion of it not to be sufficiently compliant with the Study Protocol. Therefore, we are not performing an Intention to Treat analysis.

For the analysis of safety data, all individuals with available data will be included. These data may be incomplete for participants who drop-out in the middle of a treatment arm, but will be analyzed to the full extent possible. This means that participants who are excluded from efficacy analyses may still be included in safety analyses.

Each participant's inclusion or exclusion status with regard to each analysis will be determined before analyses begin.

The REML mixed model approach detailed below explicitly deals with missing data. Thus, data imputation is not used in this statistical plan.

7.3 Interim Analyses and Data Monitoring

No interim efficacy analysis is planned. The study Data Safety Monitoring Board (DSMB) will review participant safety data on an ongoing basis. Should the DSMB deem an assessment of medication safety necessary during the trial, we will analyze the data according to their specifications.

7.4 Multiple Testing

In the present trial, we have one primary outcome measure. Other outcome measures are *a priori* described as secondary outcome measures.

The three different comparisons within SPCD pose three distinct and independent hypotheses and thus are subject to a family-wise error rate correction. Furthermore, the REML mixed model approach to SPCD combines all of these comparisons into a single omnibus test as noted above.

Similarly, the secondary outcome measures test distinctly different domains and therefore are independent hypotheses. Accordingly, it would be incorrect to apply a family-wise error rate correction to all of the secondary outcome measures.

Instead, to minimize the risk of false discovery from multiplicity (*26*), we will first test the total score for each instrument and then only test subscales if the total score is significant. Subscales will be Bonferroni-corrected for multiple comparisons. Post hoc tests will be performed as planned contrasts and further Bonferroni-corrected for multiple comparisons. We have employed this strategy in previous research, including in the prior pilot vasopressin treatment trial (*16*).

7.5 Summary of Study Data

A CONSORT flow diagram will be included in the resulting report similar to the flow diagram from our prior pilot vasopressin treatment trial (*16*), but modified for the present trial's design. This diagram will provide information on participant screening, enrollment, randomization, and completion.

In terms of the trial's outcome measures, REML mixed model analysis of an SPCD design produces LSM ± SE for each comparison. For the three different SPCD comparisons, we will report these values for each variable, the number of participants making up each estimate, the associated significance tests, and the associated effect sizes. This will follow the format of the previous pilot vasopressin treatment trial (*16*), using a similar figure and table presentation strategy to describe the trial's findings. See also section 8.10 below.

7.6 Efficacy Analyses

All study data will be double entered into REDCap, a secure database management system (*27*), as in our previous studies. Analyses will be conducted using SAS or SAS-JMP software packages. There are multiple approaches to SPCD statistical analysis (*20-22, 24*). However, the approach which yields the greatest power is to calculate the Phase 1 and Phase 2 treatment comparisons outlined above within a single REML repeated measures mixed model, and then combine and test these effects within that context, as detailed elsewhere (*21*). This approach uses the same techniques and is conceptually similar to analytical approaches we have employed before (*28-30*), including for generating our pilot vasopressin treatment trial data (*16*). Thus, we will use REML repeated measures mixed models to test the hypothesis that vasopressin treatment improves social abilities in children with ASD as assessed primarily by parent ratings on the SRS-2. This analytic approach also allows us to test the effect of our biological measures on treatment response. We will also adopt the same REML mixed model approach to test the hypothesis that pre-treatment biological measures are personalized predictors of the primary outcome measure, by testing for treatment-by-biological measure interactions, and by testing for compound biological signatures. One major advantage of REML mixed models, is that random effects and fixed effects exist in different design matrices. This allows continuous effects (such as biological values) that would otherwise be colinear with participant, to be included and explicitly tested as fixed effects. This approach is well recognized in mixed model theory and practice (*31*).

Data will be subjected to quality control tests (*32*) and transformed as appropriate to meet the assumptions of REML mixed models. REML mixed models are particularly powerful when matching factors are unbalanced (as is likely the case for sex), or when data are missing (e.g., if participants drop out of the study).

Because the SRS-2 is a parent-reported measure, we will collect SRS-2 scores at two pre-treatment time points to identify the reliability of an individual participant's scores. In our pilot study, this enabled us to use weighted analyses whenever parent-reported measures were assessed. Weighted analyses ideally use the inverse of the variance of a mean estimate as the weight (*33*), which we could obtain directly from our two pre-treatment SRS-2 scores. This approach is useful when there are a small number of unusually variable observations (as observed and implemented in our pilot trial). If we observe a small number of particularly variable ratings, we will adopt this approach in the present trial.

For covariates and interaction terms, we will follow a hypothesis-driven model design strategy guided by covariates and interaction terms used in the pilot vasopressin trial's statistical model (*16*). Explicitly, we do not use step-wise or other machine-driven model designs.

Accordingly, the model will also be examined for collinearity, non-orthogonality, considerations of marginality, and other potential sources of false negative and false positive influences. R^2 and AICc will be used to identify such problems. Covariates or interaction terms contributing to such issues will be removed following best practice (*26*). These procedures follow best practices for sensitivity analysis in clinical trials (*34*).

As in our pilot vasopressin treatment trial (*16*), secondary efficacy outcome measures (e.g., clinician impression (CGI - Impression, CGI - Severity), parent ratings (RBS-R, SCAS), and child performance on assessments of social cognition and communication abilities (RMET, FERT) will be analyzed using the same statistical model that is ultimately derived for the primary outcome measure, with the exception that the baseline, pre-treatment behavioral measure be replaced to match the outcome variable as appropriate. The same approach will also be used for other pre-specified outcome measures for both parent ratings (VABS-3 Socialization Domain; PedsQL inventory, PRAS-ASD; SRS-2 factor scores; CSHQ) and child performance on assessments (NEPSY-II Theory of Mind Test, DANVA-2).

7.7 Safety and Tolerability Analyses

Similar to our pilot vasopressin treatment trial (*16*), safety and tolerability will be evaluated using (i) the DOTES, (ii) change from baseline in electrocardiogram, clinical chemistry laboratory tests, vital signs (i.e., blood pressure, heart rate, temperature), and height and body weight, and (iii) aggressive behavior as measured by the OAS within the pre-specified SPCD comparisons above. REML mixed models for SPCD as described above will be used to test for any changes from baseline in these measures.

The same analyses will be used to provide summary statistics (i.e., LSM ± SE) to be included in the tables.

7.8 Reporting Conventions

We typically report all findings to four significant figures. P-values <0.0001 will be reported as "<0.0001." Some software generates these values, and some journals may require different standards, so significant figure reporting may vary accordingly. REML mixed models produce LSM ± SE estimates (i.e., more precise effect sizes controlled for all factors in the model, rather than simple arithmetic means). Effect sizes will be calculated and reported as η_p^2 (partial eta-squared), as appropriate for complex linear models. Equivalent Cohen's d will be provided for main effects where justifiable.

7.9 Quality Assurance of Statistical Programming

We will use the latest version of JMP Pro and the latest version of SAS for Windows. Initial analyses are performed in JMP and then converted to human-readable SAS code for further post hoc analyses and data sharing.

Standard practice of the study statistician (Garner) is to suffix every file by date, version number, and author. For example, a statistical JMP file might be labelled "Safety data analyses, 20200105, (2, JPG)," indicating that this version of the file was saved on January 5, 2020, and is the second version of it, authored by Dr. Joseph P. Garner. All the versions of these files are saved and new versions are created

every time a document is worked on. This produces a complete record of the analyses conducted for the trial.

Notes are likewise saved in word files of the same name and include rationale and statistical output.

7.10 Figures and Tables

We plan to follow a similar data presentation strategy for the present vasopressin treatment trial report as we did for our prior pilot vasopressin treatment trial report (*16*). We note here that the final figures and tables will be based on the scientific findings, guidelines of the journal where the trial's findings are ultimately published, and feedback from the journal editor(s) and referees on data presentation. Additionally, these considerations will also determine whether these figures and tables are presented in the main text or in the supplementary material section of the report.

Anticipated Figures:

Figure 1. Trial Design (similar to Figure 1 above)
Figure 2. CONSORT Flow Diagram
Figure 3. Trial main effects for primary and secondary behavioral outcome measures

Anticipated Tables:

Table 1. Participant visits and experimental measures
Table 2. Characteristics of participants in the trial
Table 3. Participants' stable concomitant medications by treatment allocation
Table 4. Change from baseline in behavioral outcome measures
Table 5. Reported adverse events during the trial assessed by the DOTES and OAS
Table 6. Change from baseline in the safety assessments during the trial

8 References

1. C. Gamble *et al.*, Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA* **318**, 2337-2343 (2017).
2. A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders* A. P. Association, Ed., (Washington, DC, ed. 5th, 2013).
3. S. Webb, Drugmakers dance with autism. *Nature biotechnology* **28**, 772-774 (2010).
4. H. E. Albers, The regulation of social recognition, social communication and aggression: vasopressin in the social behavior neural network. *Hormones and behavior* **61**, 283-292 (2012).
5. I. F. Bielsky, S. B. Hu, K. L. Szegda, H. Westphal, L. J. Young, Profound impairment in social recognition and reduction in anxiety-like behavior in vasopressin V1a receptor knockout mice. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **29**, 483-493 (2004).
6. M. J. Paul *et al.*, Atypical Social Development in Vasopressin-Deficient Brattleboro Rats. *eNeuro* **3**, (2016).
7. J. K. Rilling *et al.*, Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men. *Psychoneuroendocrinology* **37**, 447-461 (2012).
8. A. J. Guastella, A. R. Kenyon, G. A. Alvares, D. S. Carson, I. B. Hickie, Intranasal arginine vasopressin enhances the encoding of happy and angry faces in humans. *Biol Psychiatry* **67**, 1220-1222 (2010).
9. A. J. Guastella, A. R. Kenyon, C. Unkelbach, G. A. Alvares, I. B. Hickie, Arginine Vasopressin selectively enhances recognition of sexual cues in male humans. *Psychoneuroendocrinology* **36**, 294-297 (2011).
10. O. Oztan, J. P. Garner, J. N. Constantino, K. J. Parker, Neonatal CSF vasopressin concentration predicts later medical record diagnoses of autism spectrum disorder. *Proc Natl Acad Sci U S A* **117**, 10609-10613 (2020).
11. O. Oztan *et al.*, Cerebrospinal fluid vasopressin and symptom severity in children with autism. *Annals of Neurology* **84**, 611-615 (2018).
12. K. J. Parker *et al.*, Arginine vasopressin in cerebrospinal fluid is a marker of sociality in nonhuman primates. *Sci Transl Med* **10**, (2018).
13. S. M. Freeman, K. Inoue, A. L. Smith, M. M. Goodman, L. J. Young, The neuroanatomical distribution of oxytocin receptor binding and mRNA in the male rhesus macaque (Macaca mulatta). *Psychoneuroendocrinology* **45**, 128-141 (2014).
14. J. T. Winslow, N. Hastings, C. S. Carter, C. R. Harbaugh, T. R. Insel, A role for central vasopressin in pair bonding in monogamous prairie voles. *Nature* **365**, 545-548 (1993).
15. C. S. Carter, Sex differences in oxytocin and vasopressin: implications for autism spectrum disorders? *Behav Brain Res* **176**, 170-186 (2007).
16. K. J. Parker *et al.*, A randomized placebo-controlled pilot trial shows that intranasal vasopressin improves social deficits in children with autism. *Sci Transl Med* **11**, (2019).
17. A. J. Guastella *et al.*, The effects of a course of intranasal oxytocin on social behaviors in youth diagnosed with autism spectrum disorders: a randomized controlled trial. *J Child Psychol Psychiatry* **56**, 444-452 (2015).
18. C. J. Yatawara, S. L. Einfeld, I. B. Hickie, T. A. Davenport, A. J. Guastella, The effect of oxytocin nasal spray on social interaction deficits observed in young children with autism: a randomized clinical crossover trial. *Mol Psychiatry*, (2015).
19. M. G. Aman *et al.*, Safety and Efficacy of Memantine in Children with Autism: Randomized, Placebo-Controlled Study and Open-Label Extension. *J Child Adolesc Psychopharmacol*, (2016).
20. M. Fava, A. E. Evins, D. J. Dorer, D. A. Schoenfeld, The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom* **72**, 115-127 (2003).

21. G. Doros, M. Pencina, D. Rybin, A. Meisner, M. Fava, A repeated measures model for analysis of continuous outcomes in sequential parallel comparison design studies. *Stat Med* **32**, 2767-2789 (2013).
22. Y. F. Chen, Y. Yang, H. M. Hung, S. J. Wang, Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. *Contemp Clin Trials* **32**, 592-604 (2011).
23. R. N. Tamura, X. Huang, An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. *Clin Trials* **4**, 309-317 (2007).
24. D. Rybin, G. Doros, M. J. Pencina, M. Fava, Placebo non-response measure in sequential parallel comparison design studies. *Stat Med* **34**, 2281-2293 (2015).
25. R. Mead, *The design of experiments: statistical principles for practical applications.* . (Cambridge University Press, Cambridge, England, 1988).
26. A. Grafen, R. Hails, *Modern Statistics for the Life Sciences* (Oxford University Press, Oxford, 2002).
27. P. A. Harris *et al.*, Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* **42**, 377-381 (2009).
28. D. S. Carson *et al.*, Cerebrospinal fluid and plasma oxytocin concentrations are positively correlated and negatively predict anxiety in children. *Mol Psychiatry* **20**, 1085-1090 (2015).
29. D. S. Carson *et al.*, Arginine Vasopressin Is a Blood-Based Biomarker of Social Functioning in Children with Autism. *PLoS One* **10**, e0132224 (2015).
30. K. J. Parker *et al.*, Plasma oxytocin concentrations and OXTR polymorphisms predict social impairments in children with and without autism spectrum disorder. *Proc Natl Acad Sci U S A* **111**, 12258-12263 (2014).
31. R. C. Littell, W. W. Stroup, R. J. Freund, *SAS for linear models.*, (SAS Institute, Cary, NC, 2002).
32. A. Grafen, R. Hails, *Modern Statistics for the Life Sciences.*, (Oxford University Press, Oxford and New York, 2002).
33. J. Neter, M. H. Kutner, C. Nachtscheim, W. Wasserman, *Applied Linear Statistical Models*. (McGraw-Hill, New York, ed. 4th, 1996).
34. L. Thabane *et al.*, A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Medical Research Methodology* **13**, 92 (2013).