Supplementary Online Content

Hollander E, Jacob S, Jou R, et al. Balovaptan vs placebo for social communication in childhood autism spectrum disorder: a randomized clinical trial. *JAMA Psychiatry*. Published online July 6, 2022. doi:10.1001/jamapsychiatry.2022.1717

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Additional Methods

Changes in conduct of study or planned analyses

Changes in conduct of study

As of the LPLV (June 30, 2020), there were five amendments to the original Protocol BP30153, released on March 23, 2016.

Amendment 1, Protocol BP30153 version 2 (Dated August 2, 2016)

The main changes to the protocol were as follows:

- The age range was changed to 13–17 years for adolescents and 5–12 years for children, in response to a request from the Food and Drug Administration due to a change in the definition of adolescents and children.
- The pharmacokinetic (PK) schedule was revised so that the visit with the most intense PK sampling was moved from week 1 to week 2. This was to ensure that PK assessments were taken within the first cohorts of 24 adolescents and 24 children at steady state for all major metabolites.

Amendment 2, Protocol BP30153 version 3 (Dated January 30, 2017)

The main changes to the protocol were as follows:

- For first adolescent and first child cohorts, screening window was changed from –8 to –1 weeks prior to first dose. Screening window for subsequent cohorts in the main study remained at –8 to –3 weeks prior to first dose. This new screening window was designed to allow time to recruit the required number of patients for subsequent randomization in a short period of time.
- Patients who were obliged to stop dosing on or before week 8, due to lack of sufficient data to inform the Internal Monitoring Committee and Scientific Oversight Committee decision on final dose, were allowed to restart in the main study.

Amendment 3, Protocol BP30153 version 4 (Dated May 19, 2017)

The main change to the protocol was as follows:

- Exclusion criterion 9 was adjusted after re-assessment of safety concerns in the population studied in this trial. A body mass index (BMI) at or above the 95th percentile for the same age and sex (according to Centers for Disease Control and Prevention growth charts) was considered to be safe. In addition, according to feedback from the study sites, the limitation to the 95th percentile was preventing otherwise eligible patients from participation in the study. The 99th percentile was used instead, as it matched higher BMI values found in trial candidates for Study BP30153.

Amendment 4, Protocol BP30153 version 5 (Dated March 29, 2018)

The main changes to the protocol were as follows:

- The text was updated to specify that the primary endpoint will be assessed based on the change from baseline on the Vineland™-II Adaptive Behavior Scales, 2nd edition (Vineland™-II) two-domain composite (2DC) instead of Vineland™-II Adaptive Behavior Scales, 2nd Edition (Vineland™-II) composite standard score.
- The following secondary objectives were added:
    - Proportion of patients with ≥6-point improvement in the Vineland™-II 2DC score to evaluate clinically meaningful response.
    - Patient-reported Pediatric Quality of Life™ (PedsQL™) v4.0 Generic Core Scale after 12 weeks and 24 weeks of treatment.
    - Evaluate safety and tolerability of up to 76 weeks of treatment with balovaptan.
- The secondary objective was changed from “Change from baseline on the Vineland™-II composite standard score after 12 weeks and 24 weeks of treatment” to “Change from baseline on the Vineland™-II 2DC score after 12 weeks and 24 weeks of treatment”.

- A responder definition and related objectives were added.
- The following secondary objectives were removed and added to the exploratory objectives:
    - Change from baseline in behaviors as measured by Aberrant Behavior Checklist Lethargy/Social Withdrawal Subscale after 12 weeks and 24 weeks of treatment.
    - Change from baseline in repetitive behaviors as measured by Repetitive Behavior Scale-Revised after 12 weeks and 24 weeks of treatment.
- The following exploratory objectives were added:
    - To evaluate the effect of treatment with balovaptan on change from baseline to week 76 (52 weeks of open-label treatment) as measured by Vineland™-II 2DC score.
    - To evaluate the efficacy of treatment with balovaptan on change from week 24 to week 76 (52 weeks of open-label treatment) as measured by Vineland™-II 2DC score.
    - Proportion of patients with ≥4-point improvement in Vineland™-II 2DC score.
    - Proportion of patients with ≥8-point improvement in Vineland™-II 2DC score.
- An open-label extension (OLE) was added.
- The total duration of the study was updated from 39 weeks to 91 weeks, and the end of study was updated from 31 weeks to 83 weeks.

Amendment 5, Protocol BP30153 version 6 (Dated December 19, 2018)

The main changes to the protocol were as follows:

- The study design was changed to a single dose (10 mg equivalent balovaptan) compared with placebo in accordance with a randomization ratio of 1:1 of balovaptan 10 mg adult-equivalent: placebo.
- Adolescent patients who had been discontinued because of lack of dose confirmation (per protocol prior to week 8) were to be replaced in the study. The total sample size of the study was increased to 340 patients.
- The initial starting dose (main study) was changed following review of available safety and PK data from the study. A table was added to outline updated starting doses for adolescents and children, aged 8–17 years.
- The screening window duration was reduced for all new patients.
- The frequency of troponin T monitoring was increased, whereas frequency of creatine phosphokinase (CPK) assessments was decreased. Creatine Kinase of Heart Muscle type was added as descriptive cardiac biomarker to timepoints of CPK measurements. For elevated cardiac biomarkers, stopping rules were added with regard to troponin I, troponin T, N-terminal pro–B-type natriuretic peptide levels. Additional eligibility criteria of confirmed elevated cardiac biomarkers at screening were also introduced.
- Capillary blood sampling (introduced in version 5 of this protocol) was removed.

eAppendix 2. Procedures

Study visits occurred at screening, baseline (randomization visit; day 1), every 2 weeks through to week 24, plus follow-up visits at weeks 26 and 30 for participants who did not transition into the OLE or upon early withdrawal (EW). PK blood sampling occurred at weeks 8, 12, and 24. Vineland™-II and the patient-reported PedsQL™ were assessed at baseline, and at weeks 12 and 24. Aberrant Behavior Checklist, Repetitive Behavior Scale-Revised, Ohio Autism Clinical Global Impression-Severity (CGI-S) , and CGI-Improvement (CGI-I), were assessed throughout the trial at each site visit. All clinical outcome assessments were completed on an electronic device. When possible, assessments were administered by the same trained rater/clinician and completed by the same caregiver/parent of the participant. To ensure consistency of rating, audio recordings of the Vineland™-II administration were reviewed by a centralized facility for accuracy of administration and scoring. Safety was assessed throughout the trial at each visit by occurrence, type, and intensity of adverse events, determined using the Adverse Event Severity Grading Scale. Vital signs were monitored at each site visit prior to and post study drug administration, and physical and neurologic examinations were conducted at baseline, EW, weeks 6, 12, 18, and 24. Hematologic examinations were conducted at baseline and weeks 4, 8, 16, 20, and EW; electrocardiogram assessments were performed at baseline, weeks 6, 24, and EW; and troponin samples were taken at weeks 4, 8, 16, and 20. The Columbia-Suicide Severity Rating Scale (C-SSRS) was assessed at baseline, and at weeks 6, 12, 18, 24, and EW. Full serum chemistry examinations (consisting of sodium, potassium, chloride, bicarbonate, glucose [fasting for at least 4 hours at screening only], urea, creatinine, phosphate, calcium, total and direct bilirubin, alkaline phosphatase, alanine transaminase, aspartate transaminase, protein, albumin, urate, lactate dehydrogenase, creatine phosphokinase, total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglycerides) were conducted at

baseline and weeks 12, 24, and EW. Urinalysis examinations were performed at baseline, weeks 6, 12, 18, 24, and EW.

eAppendix 3. Study Design

The initial PK study investigation to confirm the age-adjusted balovaptan doses resulting in plasma exposures equivalent to 4 mg daily and 10 mg daily in adults for children and adolescents in the following aged groups: 5–7 years; 8–12 years; 13–14 years; and 15–17 years. Enrollment was staggered, starting with adolescents.

Participants who completed the 24-week, double-blind treatment period could participate in the OLE period, in which they would receive open-label 10 mg adult-equivalent balovaptan irrespective of previous treatment allocation. Balovaptan or placebo was administered orally, once-daily for 24 weeks.

A sample size of 240 individuals with autism spectrum disorder (ASD) (80 per treatment arm), providing evaluable data at 24 weeks, was planned. That sample size was derived to ensure 80% power to detect as statistically significant, at a one-sided, 5% significance level, a difference between each active dose and placebo with an effect size of at least 0.4. No adjustment for multiple doses was performed. Considering a withdrawal rate of around 15–20%, it was planned to recruit approximately 300 participants overall. Following study design change to a single dose (10 mg adult-equivalent balovaptan) compared with placebo, in accordance with a randomization ratio of 1:1 of balovaptan 10 mg adult-equivalent:placebo, a sample size of 80 participants per treatment arm (balovaptan 10 mg adult-equivalent and placebo) was required, for a total sample size of approximately 160 participants with ASD with evaluable data at 24 weeks. To maintain the number of evaluable participants, the sample size was increased to approximately 340 participants overall.

Outcome measures

Improvement on the CGI-I was defined as a score of 1 (“much improved”) or 2 (“very much improved”). Exploratory endpoints included change from baseline of PedsQL™ v4.0 Cognitive Functioning Scale and parent-reported Family Impact module, as well as change from baseline in behaviors as measured by the Aberrant Behavior Checklist Lethargy/Social Withdrawal Subscale after 12 weeks and 24 weeks.

Key safety endpoints were incidence, nature, and intensity of adverse events; incidence and nature of laboratory abnormalities, based on hematology, clinical chemistry, cardiac biomarkers, coagulation monitoring, and urinalysis testing; Tanner staging; physical examination (including body height and weight) and neurologic examination; and suicidality evaluated by the C-SSRS.

eAppendix 4. Statistical Analysis

A generalized estimating equation model was fitted including treatment, sex, and visit (week) as main effects, individual age and baseline CGI-S as covariates, and baseline-by-visit and treatment-by-visit interactions, with participant fitted as a repeated effect. The proportion of participants with a value of at least 2 (i.e. “much improved” or “very much improved”) on CGI-I was analyzed using logistic regression applying a similar model to the one described for CGI-S.

eAppendix 5. Rationale for Safety Analysis Population

The population used for safety analyses consisted of a subset of participants identified to address the primary efficacy study objective. This subset consisted of participants in the balovaptan 10 mg adult-equivalent (revised, age-adjusted dose) or the concurrently randomized placebo group in the corresponding age band and in the same randomization stage. Participants with dose adjustments or interruptions, or on a different dose than the final dose for their age group, were excluded. Data were summarized according to treatment classified as: placebo or balovaptan 10 mg adult-equivalent. Safety was evaluated for all participants (not shown) and there was no difference in safety profile between the total population and the primary efficacy analysis population.

eAppendix 6. Rationale for Age-Adjusted Dosing

A population PK modeling approach (non-linear, mixed-effect modeling) was used to characterize the PK and variability of balovaptan. The structural PK model used to estimate the initial age-based doses designed to achieve exposures equivalent to those in adults taking daily doses of balovaptan 4 mg or 10 mg was built based on prior information in adults. Key PK parameters (i.e. apparent clearance and volume of distribution) and variability were reported, and the influence of covariates (e.g. age, body size, genetic variability of CYP3A4) on the population PK parameters was investigated. Balovaptan is predominantly metabolized by CYP3A4, an enzyme located in the liver. The function of this enzyme is mature in those aged 5 years and over. Liver size (which is related to age) was the biggest predictor of exposure differences in pediatric participants in the initial

PK study. Following the initial PK study, doses were increased to produce exposures equivalent to the 4 mg and 10 mg doses used in adults (Supplement 3 p. 5)

eTable 1. Baseline Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2) Scores

	Balovaptan 10 mg (N = 86)	**Placebo (N = 81)**	**Total (N = 167)**
Module 2:			
Social affect total			
n	5	7	12
Mean (SD)	14.0 (3.9)	16.3 (2.1)	15.3 (3.1)
Median (min–max)	14.0 (8–18)	17.0 (13–18)	16.5 (8–18)
Restricted and repetitive behavior total			
n	5	7	12
Mean (SD)	4.2 (2.9)	4.3 (2.2)	4.3 (2.4)
Median (min–max)	4.0 (1–8)	4.0 (1–7)	4.0 (1–8)
Overall total (social affect total + restricted and repetitive behavior total)			
n	5	7	12
Mean (SD)	18.2 (5.2)	20.6 (3.8)	19.6 (4.4)
Median (min–max)	19.0 (10–23)	20.0 (14–24)	20.0 (10–24)
ADOS-2 comparison score			
n	5	7	12
Mean (SD)	8.2 (1.6)	8.9 (1.6)	8.6 (1.6)
Median (min–max)	9.0 (6–10)	10.0 (6–10)	9.0 (6–10)
Module 3:			
Social affect total			
n	53	50	103
Mean (SD)	12.4 (4.2)	12.8 (4.8)	12.6 (4.5)
Median (min–max)	12.0 (3–20)	12.5 (4–24)	12.0 (3–24)
Restricted and repetitive behavior total			
n	53	50	103
Mean (SD)	2.8 (2.0)	3.2 (1.8)	3.0 (1.9)
Median (min–max)	3.0 (0–8)	3.0 (0–7)	3.0 (0–8)
Overall total (social affect total + restricted and repetitive behavior total)			
n	53	50	103
Mean (SD)	15.2 (4.8)	16.0 (5.6)	15.6 (5.2)
Median (min–max)	15.0 (7–26)	15.0 (7–29)	15.0 (7–29)
ADOS-2 comparison score			
n	53	50	103
Mean (SD)	8.1 (1.8)	8.3 (1.8)	8.2 (1.8)
Median (min–max)	9.0 (4–10)	9.0 (4–10)	9.0 (4–10)
Module 4:			
Social affect total			
n	1	0	1
Mean (SD)	18.0		18.0
Median (min–max)	18.0 (18–18)		18.0 (18–18)
Restricted and repetitive behavior total			
n	1	0	1
Mean (SD)	21.0		21.0
Median; (min–max)	21.0 (21–21)		21.0 (21–21)
Overall total (social affect total + restricted and repetitive behavior total)			
n	1	0	1
Mean (SD)	9.0		9.0
Median (min–max)	9.0 (9–9)		9.0 (9–9)

ADOS-2 comparison score			
n	1	0	1
Mean (SD)	48.0		48.0
Median (min–max)	48.0 (48–48)		48.0 (48–48)
Communication total			
n	24	22	46
Mean (SD)	5.2 (2.0)	4.9 (3.2)	5.0 (2.6)
Median (min–max)	5.5 (2–8)	4.0 (2–17)	5.0 (2–17)
Reciprocal social interaction total			
n	24	22	46
Mean (SD)	8.7 (2.3)	8.9 (4.7)	8.8 (3.6)
Median (min–max)	8.5 (5–14)	8.0 (5–27)	8.0 (5–27)
Communication and reciprocal social interaction total			
n	24	22	46
Mean (SD)	14.0 (3.9)	13.7 (7.8)	13.8 (6.0)
Median (min–max)	14.0 (7–21)	12.0 (7–44)	13.0 (7–44)
Imagination/creativity total			
n	24	22	46
Mean (SD)	1.0 (0.7)	1.0 (0.7)	1.0 (0.7)
Median (min–max)	1.0 (0–2)	1.0 (0–3)	1.0 (0–3)
Stereotyped behaviors and restricted interests total			
n	24	22	46
Mean (SD)	2.0 (1.7)	2.7 (2.0)	2.3 (1.9)
Median (min–max)	2.0 (0–7)	2.5 (0–6)	2.0 (0–7)

Abbreviations: ADOS, Autism Diagnostic Observation Schedule.
2nd Edition (ADOS-2) Scores. Module 2 for children of any age who use phrase speech but are not verbally fluent. Module 3 for verbally fluent children and young adolescents. Module 4 for verbally fluent older adolescents and adults. The ADOS-2 comparison score ranges from 1–10, where 1 indicates minimal-to-no evidence of autism-related symptoms and 10 indicates a high level of autism-related symptoms.

eTable 2. Summary of Participants Responding "Yes" During the 24-Week Double-Blind Period to the Columbia-Suicide Severity Rating Scale (C-SSRS)

	Balovaptan 10 mg (N = 86)	**Placebo (N = 81)**
Suicidal ideation or behavior (Item 1–10)		
n (%)	4 (4.7)[a]	5 (6.17)
Self-injurious behavior without suicidal intent		
n (%)	2 (2.3)[a]	3 (3.70)

Abbreviations: C-SSRS, Columbia-Suicide Severity Rating Scale.

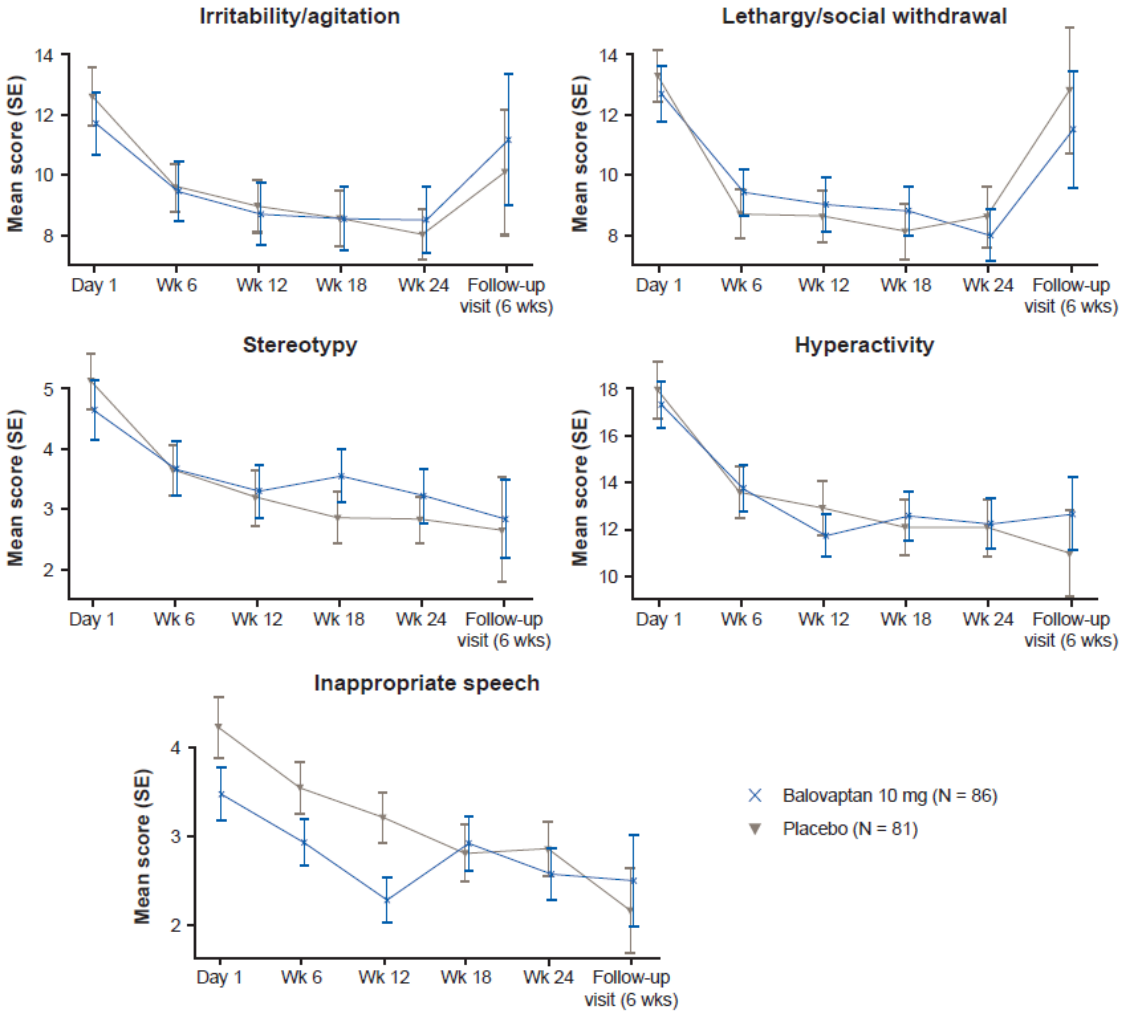
[a] One participant reported both suicidal ideation or behavior and self-injurious behavior without suicidal intent. The ten categories included in the C-SSRS are as follows: Category 1 – Wish to be Dead; Category 2 – Non-Specific Active Suicidal Thoughts; Category 3 – Active Suicidal Ideation With Any Methods (Not Plan), Without Intent to Act; Category 4 – Active Suicidal Ideation With Some Intent to Act, Without Specific Plan; Category 5 – Active Suicidal Ideation With Specific Plan and Intent; Category 6 – Preparatory Acts or Behavior; Category 7 – Aborted Attempt; Category 8 – Interrupted Attempt; Category 9 – Actual Attempt (non-fatal); Category 10 – Completed Suicide.

eTable 3. Safety Evaluable Population Demographics and Baseline Characteristics

	Placebo (N = 112)	Low tertile (N = 57)	Medium tertile (N = 66)	High tertile (N = 66)	All active treatments (N = 196)
Male, n (%)	95 (84.8)	48 (84.2)	58 (87.9)	55 (83.3)	168 (85.7)
Mean age, years ± SD (median; min–max)	12.5 ± 3.0 (13; 5–17)	13.3 ± 2.3 (14; 7–17)	12.6 ± 3.0 (13; 5–17)	11.8 ± 3.3 (11; 5–17)	12.6 ± 2.9 (13; 5–17)
Age group, n (%)					
Children aged 2–11 years	40 (35.7)	11 (19.3)	19 (28.8)	35 (53.0)	66 (33.7)
Adolescents aged 12–17 years	72 (64.3)	46 (80.7)	47 (71.2)	31 (47.0)	130 (66.3)
Race, n (%)					
American Indian or Alaska native	0 (0.0)	1 (1.8)	0 (0.0)	0 (0.0)	1 (0.5)
Asian	5 (4.5)	3 (5.3)	7 (10.6)	3 (4.5)	13 (6.6)
Black or African American	4 (3.6)	1 (1.8)	4 (6.1)	5 (7.6)	10 (5.1)
Native Hawaiian or other Pacific Islander	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.5)	1 (0.5)
White	93 (83.0)	46 (80.7)	54 (81.8)	51 (77.3)	157 (80.1)
Multiple	9 (8.0)	5 (8.8)	1 (1.5)	5 (7.6)	12 (6.1)
Unknown	1 (0.9)	1 (1.8)	0 (0.0)	1 (1.5)	2 (1.0)
Ethnicity, n (%)					
Hispanic or Latinx	17 (15.2)	7 (12.3)	8 (12.1)	6 (9.1)	21 (10.7)
Not Hispanic or Latinx	93 (83.0)	49 (86.0)	58 (87.9)	59 (89.4)	173 (88.3)
Not stated	2 (1.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Unknown	0 (0.0)	1 (1.8)	0 (0.0)	1 (1.5)	2 (1.0)
BMI (Kg/m^2), mean ± SD (median; min–max)	20.76 ± 4.83 (19.69; 13.0–35.0)	21.28 ± 4.66 (20.28 14.4–33.9)	20.86 ± 4.28 (20.27; 13.8–33.7)	22.12 ± 5.16 (21.33; 13.8–34.3)	21.44 ± 4.73 (20.46; 13.8–34.3)
Participants with at least one ongoing medication, n (%)	91 (81.3)	53 (93.0)	52 (78.8)	63 (95.5)	175 (89.3)
Participants with at least one psychiatric comorbidity, n (%)	94 (83.9)	52 (91.2)	45 (68.2)	53 (80.3)	156 (79.6)
Mean WASI-II IQ ± SD (median; min–max)	99.6 ± 17.6 (101.5; 70–142)	97.6 ± 15.3 (96.0; 71–140)	97.4 ± 17.6 (97.5; 70–144)	97.8 ± 16.0 (96.0; 70–130)	97.7 ± 16.5 (96.0; 70–144)
CGI-S, n (%)					
CGI-S = 3[a]	1 (0.9)	0 (0.0)	0 (0.0)	1 (1.5)	1 (0.5)
CGI-S = 4	63 (56.3)	30 (47.6)	45 (67.2)	41 (62.1)	116 (59.2)
CGI-S = 5	43 (38.4)	30 (47.6)	20 (29.9)	19 (28.8)	69 (35.2)
CGI-S = 6	5 (4.5)	3 (4.8)	2 (3.0)	5 (7.6)	10 (5.1)

The Safety Evaluable Population was comprised of participants in the main study who received one dose of balovaptan grouped by PK exposure tertiles. Generally, the "Lower tertile" is comprised of participants who received 4 mg adult-equivalent balovaptan, the "Medium tertile" is comprised of participants who received 4 mg and 10 mg adult-equivalent balovaptan, and the "High tertile" is comprised of participants who received 10 mg adult-equivalent balovaptan, dependent on individuals' exposures.
Abbreviations: ABC, Adaptive Behavior Composite; ADOS, Autism Diagnostic Observation Schedule; BMI, body mass index; CGI-S, Clinical Global Impression-Severity; IQ, intelligence quotient; SRS-2, Social Responsiveness Scale, 2nd Edition; WASI-II, Wechsler Abbreviated Scale of Intelligence, 2nd Edition.
[a] Participants' baseline CSI-S scores met inclusion criteria of ≥4 at screening.

eFigure. Aberrant Behavior Checklist Mean Score



Abbreviations: SE, standard error; Wk(s), week(s).
Day 1: placebo, n = 81; balovaptan, n = 86; week 6: placebo, n = 73; balovaptan, n = 80; week 12: placebo, n = 68; balovaptan, n = 73; week 18: placebo, n = 61; balovaptan, n = 70; week 24: placebo, n = 62; balovaptan, n = 69; follow-up visit (6 weeks): n = 18; balovaptan, n = 20. Note that one participant was discounted in the balovaptan arm due to an error in baseline measurement data. There are fewer participants at follow-up visit (6 weeks) than other timepoints due to participants transitioning into the open-label extension.