# RESEARCH ARTICLE

# The Use of Eye Tracking as a Biomarker of Treatment Outcome in a Pilot Randomized Clinical Trial for Young Children with Autism

Jessica Bradshaw ⬤, Frederick Shic ⬤, Anahita N. Holden, Erin J. Horowitz, Amy C. Barrett, Tamsin C. German, and Ty W. Vernon

There is a pressing need for objective, quantifiable outcome measures in intervention trials for children with autism spectrum disorder (ASD). The current study investigated the use of eye tracking as a biomarker of treatment response in the context of a pilot randomized clinical trial of treatment for young children with ASD. Participants included 28 children with ASD, aged 18–48 months, who were randomized to one of two conditions: Pivotal Response Intervention for Social Motivation (PRISM) or community treatment as usual (TAU). Eye-tracking and behavioral assessment of developmental functioning were administered at Time 1 (prior to randomization) and at Time 2 (after 6 months of intervention). Two well-established eye-tracking paradigms were used to measure social attention: social preference and face scanning. As a context for understanding relationships between social attention and developmental ability, we first examined how scanning patterns at Time 1 were associated with concurrent developmental functioning and compared to those of 23 age-matched typically developing (TD) children. Changes in scanning patterns from Time 1 to Time 2 were then compared between PRISM and TAU groups and associated with behavioral change over time. Results showed that the social preference paradigm differentiated children with ASD from TD children. In addition, attention during face scanning was associated with language and adaptive communication skills at Time 1 and change in language skills from Time 1 to Time 2. These findings highlight the importance of examining targeted biomarkers that measure unique aspects of child functioning and that are well-matched to proposed mechanisms of change. *Autism Res* 2019, 00: 1–15. © 2019 International Society for Autism Research, Wiley Periodicals, Inc.

**Lay Summary:** Biomarkers have the potential to provide important information about how and why early interventions effect positive change for young children with ASD. The current study suggests that eye-tracking measures of social attention can be used to track change in specific areas of development, such as language, and points to the need for targeted eye-tracking paradigms designed to measure specific behavioral changes. Such biomarkers could inform the development of optimal, individualized, and adaptive interventions for young children with ASD.

## Introduction

The number of randomized clinical trials (RCTs) that test the efficacy of early intervention for autism spectrum disorder (ASD) has increased substantially in the past decade, with at least 40 RCTs published since 2010 [for review, see French & Kennedy, 2018]. While this research shows promise in documenting effective interventions for ASD, effect sizes for most RCTs remain relatively small. It is unlikely that this is due to minimally effective interventions, and may be more likely attributable to the persistent challenge of identifying outcome measures that are sufficiently sensitive to change in the core symptoms of ASD [McConachie et al., 2015]. Moreover, the significant

heterogeneity of outcome measures used across treatment studies hinders our ability to aggregate findings and develop a full understanding of true intervention effectiveness and optimal outcome measures [Bolte & Diehl, 2013; Cunningham, 2012; French & Kennedy, 2018; Spence & Thurm, 2010]. For this reason, investigators have begun to search for sensitive, objective, and quantifiable biomarkers of treatment outcome for children with ASD. Eye-tracking technology is one such viable biomarker due to its accessibility (noninvasiveness, affordability, and ease of use) for children with ASD [Boraston & Blakemore, 2007; Kim et al., 2014; Sasson & Elison, 2012; Shic, 2013]. The present study examined the utility of two eye-tracking paradigms as a biomarker of

treatment response in the context of a pilot RCT that evaluated a specific pivotal response treatment (PRT) intervention protocol: the Pivotal Response Intervention for Social Motivation (PRISM; Vernon et al., 2019) in young children with ASD.

Measuring treatment outcome in ASD has proven quite complex due to the heterogeneity of autism symptomology and developmental skills, as well as the vast array of potential treatment targets for children with ASD. Many treatments focus on core symptoms of ASD, which can refer to verbal and nonverbal communication, language and pragmatics, social skills, and/or restricted and repetitive behaviors. These symptoms can be measured with specific instruments that evaluate a narrow set of abilities, or with broad-based measures that cover the full range of developmental abilities and autism symptoms. Common outcome measures used in clinical trials for ASD include assessments of: autism symptomology (e.g., *Autism Diagnostic Observation Schedule*, Lord et al., 2012), language ability (e.g., *Mullen Scales of Early Learning*, Mullen, 1995; *MacAurther-Bates Communicative Development Inventories*, Fenson et al., 2007), social competencies (e.g., *Social Responsiveness Scale*, Constantino, 2012; *Social Communication Questionnaire*, Rutter, Bailey, & Lord, 2003), cognitive functioning (e.g., *Mullen Scales of Early Learning*), and adaptive behavior (e.g., *Vineland Adaptive Behavior Scales*, Sparrow, Cicchetti, & Saulnier, 2016) [McConachie et al., 2015]. Yet these currently available assessment tools were not necessarily designed to be sensitive to treatment-related change, especially over relatively short periods of time (i.e., 3–12-month intervention trials). While the development of such behavioral measures is under way [e.g., Fletcher-Watson & McConachie, 2017; Grzadzinski et al., 2016; Mazurek et al., 2018], there are no currently accepted "gold standard" tools for measuring treatment outcome in clinical trials for ASD. Such standards would allow for pooled findings across studies and advance science in the development of the most effective treatments [Dawson, 2017].

Many systems used for tracking treatment outcomes are based on parent report via interviews or questionnaires or expert clinical assessment. However, recent research has highlighted the presence of a striking placebo effect in ASD populations: even when outcome evaluators are masked to treatment condition, significant improvements are still reported in a surprisingly substantial number of participants assigned to control conditions [Bradshaw et al., 2017; Jones, Carberry, Hamo, & Lord, 2017; Masi, Lampit, Glozier, Hickie, & Guastella, 2015]. This, in addition to the cost of training and maintaining fidelity on clinically based measures, leads us to believe that neither questionnaires nor clinical phenotyping are optimal outcome measures.

In an effort to improve sensitivity and objectivity of outcome measures in clinical trials for ASD, researchers have been working to identify biomarkers that are easily accessible, noninvasive, and associated with behavioral targets and/or change mechanisms. A biomarker is defined as the measurement of a biological parameter that can be associated with treatment response in order to gain a mechanistic understanding of differences in clinical response [Biomarkers Definitions Working Group, 2001]. Many behavioral measures rely on an aggregate of skills that span multiple developmental domains (e.g., attention, inhibition, and receptive/expressive language). In contrast, biomarkers may provide a measure of specific constructs that are proposed to underlie observable behavior. They may also corroborate evidence from behavioral measures, which may provide stronger evidence for treatment efficacy in general. Such biomarkers can help to individualize treatment, recognize early and late responders, and identify specific mechanisms of behavioral change [McPartland, 2016].

Recent studies have begun to test measures of brain activity and looking behavior as potential biomarkers of treatment response. Yang et al. [2016, 2017] successfully used a well-established biological motion fMRI task to predict response to two different treatments for ASD. The first study evaluated child response to 16 weeks of PRT designed to improve social initiations and responsivity [Yang et al., 2016]. The second study evaluated response to virtual reality social cognition training, a 5-week behavioral intervention for adults with ASD designed to improve emotion recognition and theory of mind [Yang et al., 2017]. While the sample sizes of these studies were relatively small, results found strong associations between pretreatment brain activation to biological motion and behavioral change in the treatment target (emotion recognition for young adults and autism symptom severity as measured by the Social Responsiveness Scale for children), suggesting that brain response to this task is tapping into neurobiological readiness to respond to treatment [Yang et al., 2016, 2017].

Electroencephalography markers of treatment response have similarly been identified. Children with ASD who participated in a high-intensity treatment program consisting of 10 hr per week of the Early Start Denver Model for 2 years exhibited greater cortical activation and normalized neural signatures when viewing faces at posttreatment compared to children who participated in community treatment as usual (TAU) [Dawson et al., 2012]. Specifically, greater cortical activation was associated with fewer social pragmatic problems, and higher activation was correlated with better social communication, as measured with the parent report PDD Behavioral Inventory [Cohen, Schmidt-Lackner, Romanczyk, & Sudhalter, 2003]. In a RCT of a low-intensity 10-week intervention for 6–12-month-old infants at risk for ASD, Jones, Dawson, Kelly, Estes, and Jane Webb [2017] found improved neurophysiological response to faces compared

to the TAU group, with large effects. In addition to these neurobiological markers of treatment response, eye-tracking measures of social attention have been discussed as particularly powerful indicators of treatment response, especially for interventions that are designed to improve social engagement and motivation [Dawson, Bernier, & Ring, 2012; Murias et al., 2018; Umbricht et al., 2017]. In a clinical trial of the effect of vasopressin on social cognition in adults with ASD, eye tracking was used to demonstrate a large effect of administration of vasopressin on increased preference for biological motion [Umbricht et al., 2017].

Overall, studies suggest that visual and neurobiological measures of social attention and social sensitivity may predict response to treatment [Yang et al., 2016, 2017] and may be improved as a result of treatment [Dawson, Jones, et al., 2012; Jones, Dawson, et al., 2017; Umbricht et al., 2017]. This suggests feasibility in the utility of biomarkers to track treatment response while also highlighting a significant gap in the literature for how eye-tracking technology, which has been highly informative in the study of ASD in toddlers and young children, can be applied to biomarker research in the context of clinical trials for young children with ASD.

The present study examines two eye-tracking measures of social attention: social preference and face scanning. These two paradigms were chosen because they are related to core deficits of ASD (social engagement) and previous literature has found looking patterns to distinguish between ASD and typically developing (TD) toddlers and young children [Chawarska & Shic, 2009; Pierce et al., 2016; Pierce, Conant, Hazin, Stoner, & Desmond, 2011]. The social preference paradigm is a forced-choice measure of attentional preference for social versus geometric stimuli and has been studied as an early diagnostic biomarker for ASD [Pierce et al., 2011, 2016]. When presented with a dynamic social stimulus and a dynamic geometric stimulus side-by-side, toddlers with ASD are observed to attend to the geometric stimulus for a significantly greater proportion of time than TD toddlers. In these studies, a stronger geometric preference was also associated with lower scores on measures of cognition, language, and autism severity. Similar results have been observed in older children with ASD using adapted versions of this social preference paradigm [Shaffer et al., 2017].

The face-scanning paradigm consists of the serial presentation of static images of faces. Research suggests that attention allocation during face scanning is a context-dependent developmental process and much attention has been given to the amount of time spent looking to the mouth and eyes [Chawarska, Macari, & Shic, 2012; Pelphrey et al., 2002; Shic, Macari, & Chawarska, 2014; Speer, Cook, McMahon, & Clark, 2007; Tenenbaum, Shah, Sobel, Malle, & Morgan, 2013]. Increased mouth looking is observed in TD infants and toddlers during active language learning [Kubicek et al., 2013; Lewkowicz & Hansen-Tift, 2012; Tenenbaum, Amso, Abar, & Sheinkopf, 2014], while

increased eye looking is found in individuals after 12 months of age, presumably due to language mastery and an interest in extracting social information [Lewkowicz & Hansen-Tift, 2012]. In ASD, however, mouth looking during static face scanning has been found to be attenuated compared to TD individuals, both in young children and adults, and increased mouth looking is associated with increased facial encoding [Chawarska & Shic, 2009; Pelphrey et al., 2002]. It has been suggested that decreased mouth-looking is possibly due to decreased attention and responsivity to speech and language [Chawarska & Shic, 2009], yet this hypothesis has not been tested directly. In an extension of this work, the current study chose to examine mouth-to-eyes ratio during presentation of static faces as a possible measure of clinical heterogeneity within ASD, similar to what has been done with dynamic faces [Norbury et al., 2009; Tenenbaum et al., 2014; Tenenbaum, Sobel, Sheinkopf, Malle, & Morgan, 2015].

While eye-tracking studies of social engagement have proven useful in advancing our understanding of social attention and communication skills in toddlers with ASD, only recently has eye-tracking technology been explored as a biomarker of treatment response [Shic, 2016; Umbricht et al., 2017]. For an intervention that targets social motivation as a method for improving verbal and nonverbal communication, eye-tracking measures of social attention would be particularly useful [Dawson, Bernier, & Ring, 2012]. The link between social attention and social motivation is predicated on the idea that looking behavior reflects selective information gathering and valuation of information. While TD infants prefer social over nonsocial stimuli from birth, [Farroni, Csibra, Simion, & Johnson, 2002], children with ASD demonstrate diminished attentional preference to social stimuli [for review, see Guillon, Hadjikhani, Baduel, & Rogé, 2014]. This pattern of reduced social attention is thought to be associated with reduced sensitivity to social reward and attenuated social motivation, which may underlie early social deficits [Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012]. At present, it remains unknown how eye-tracking measures of social attention may serve as an indicator of change over time as a result of behavioral intervention that targets social motivation for in children with ASD.

Social attention biomarkers of treatment response may be especially beneficial in the case of treatment "packages," where many strategies are employed to improve a broad range of behaviors. PRT is one such packaged intervention that can be categorized under the umbrella of naturalistic developmental behavioral intervention [NDBI; Koegel & Koegel, 2006; Schreibman et al., 2015]. The goal of PRT is to improve developmental functioning through strategies that leverage existing child motivation. Historically, PRT has been studied using single case research designs to demonstrate efficacy in improving specific behaviors (e.g., social initiations, language, and

eye contact). In the context of large-scale RCTs in which labor-intensive behavioral coding of multiple participants at multiple time points can be prohibitive, biomarkers can provide an extraordinarily efficient method for tracking treatment response.

In this study, in the context of a pilot RCT of a treatment model for toddlers with ASD, we administered two eye-tracking measures of social attention in addition to a battery of standardized assessments that evaluate autism symptomology, cognition, language, and adaptive behavior prior to randomization and treatment onset (Time 1) and following 6 months of either the PRISM model (a version of PRT) or TAU (Time 2). Measures of social attention were chosen to closely match the treatment targets and expected outcomes of the PRISM model. PRISM targets social engagement and communication for children with ASD and has been shown to increase social attention [Koegel, Vernon, & Koegel, 2009; Vernon, Koegel, Dauterman, & Stolen, 2012] in young children and toddlers with ASD. Therefore, we chose two eye-tracking paradigms that measure social attention and have been shown to be associated with ASD symptomology and language. We then tested how these eye-tracking measures of social attention: (a) differentiate toddlers with ASD from TD toddlers at Time 1; (b) relate to concurrent measures of cognition, language, and autism severity in toddlers with ASD at Time 1; and (c) correlate with clinical change from Time 1 to Time 2. In line with previous treatment research examining PRT with toddlers and young children with ASD [Bradshaw, Koegel, & Koegel, 2017; Hardan et al., 2015; Koegel et al., 2009; Vernon et al., 2012; Yang et al., 2016], 6 months of intervention was determined to be a sufficient length of time for observing treatment effects. This is consistent with other early intervention models [Kasari, Gulsrud, Paparella, Hellemann, & Berry, 2015; Schertz, Odom, Baggett, & Sideris, 2013] and several studies of NDBIs for toddlers [for review, see Bradshaw, Steiner, Gengoux, & Koegel, 2015].

We hypothesized that this study would replicate previous research showing that social preference and face scanning patterns differentiate the ASD and TD groups and are related to baseline clinical measures. Specifically, we hypothesized that, in line with previous research [Pierce et al., 2016; Shaffer et al., 2017], children with ASD would show an attenuated social preference compared to TD children and that social preference would be negatively associated with autism severity. Second, we hypothesized that children with ASD would show increased mouth looking compared to TD participants, similar to Chawarska and Shic [2009]. We hypothesized that mouth looking would be positively associated with language skills, similar to what has been demonstrated in dynamic speaking faces [Norbury et al., 2009]. Finally, we hypothesized that change in social attention patterns would be associated with behavioral change from Time 1 to Time 2. For example, an increase in social preference from Time 1 to Time 2 would be associated with a reduction in autism severity from Time 1 to Time 2.

## Methods
### Participants

Participants included children between 18 and 48 months of age with a confirmed diagnosis of ASD ($N = 28$; 24 males) and TD children ($N = 23$; 14 males). Participants were recruited through email and social media announcements, direct mailings, and print advertisements, as well as through direct communication with state regional centers, early childhood educators, and pediatricians. TD participants were required to have no first- or second-degree relatives with ASD, no pre- or perinatal complications, and no known visual or hearing impairments. All ASD participants were administered a comprehensive diagnostic evaluation to confirm diagnosis of ASD. Based on this evaluation, participants were included if they: (a) obtained a score in the *mild-to-moderate* or *moderate-to-severe* concern range on the Autism Diagnostic Observation Schedule – Toddler Module [ADOS-T; Luyster et al., 2009] or a classification of *autism* on Modules 1 or 2 of the Autism Diagnostic Observation Schedule, Second Edition [ADOS-2; based on Lord et al., 2012] and (b) met criteria for an ASD diagnosis based on DSM-5 (APA, 2013) and expert clinical judgment by a licensed clinical psychologist. Children with comorbid medical or psychiatric conditions (e.g., Down syndrome, Rett's disorder) or known visual or hearing impairment were excluded from participation. A total of 28 participants were then randomly assigned to one of two treatment conditions: the PRISM treatment model ($N = 16$) or waitlist with TAU ($N = 12$). Five participants (four from the PRISM group and one from the TAU group) withdrew from the study prior to the Time 2 visit, resulting in 23 ASD participants who were seen at both Time 1 and Time 2 (see Fig. 1). The research study was approved by the University of California, Santa Barbara Institutional Review Board and all participants provided informed consent prior to participating in any study procedures.

### Procedures

Participants with ASD completed two visits: the first initial visit prior to randomization (Time 1) and a second visit 6 months later (Time 2). At each of the two visits, participants were administered a battery of standardized clinical assessments and a battery of eye-tracking paradigms. The visit at each time point took place over a period of 2 days and breaks were provided throughout the day as needed. Assessments were administered by a licensed clinical psychologist or a clinical psychology doctoral student who was supervised by a licensed psychologist. Clinicians were masked to treatment condition at the Time 2 visit.
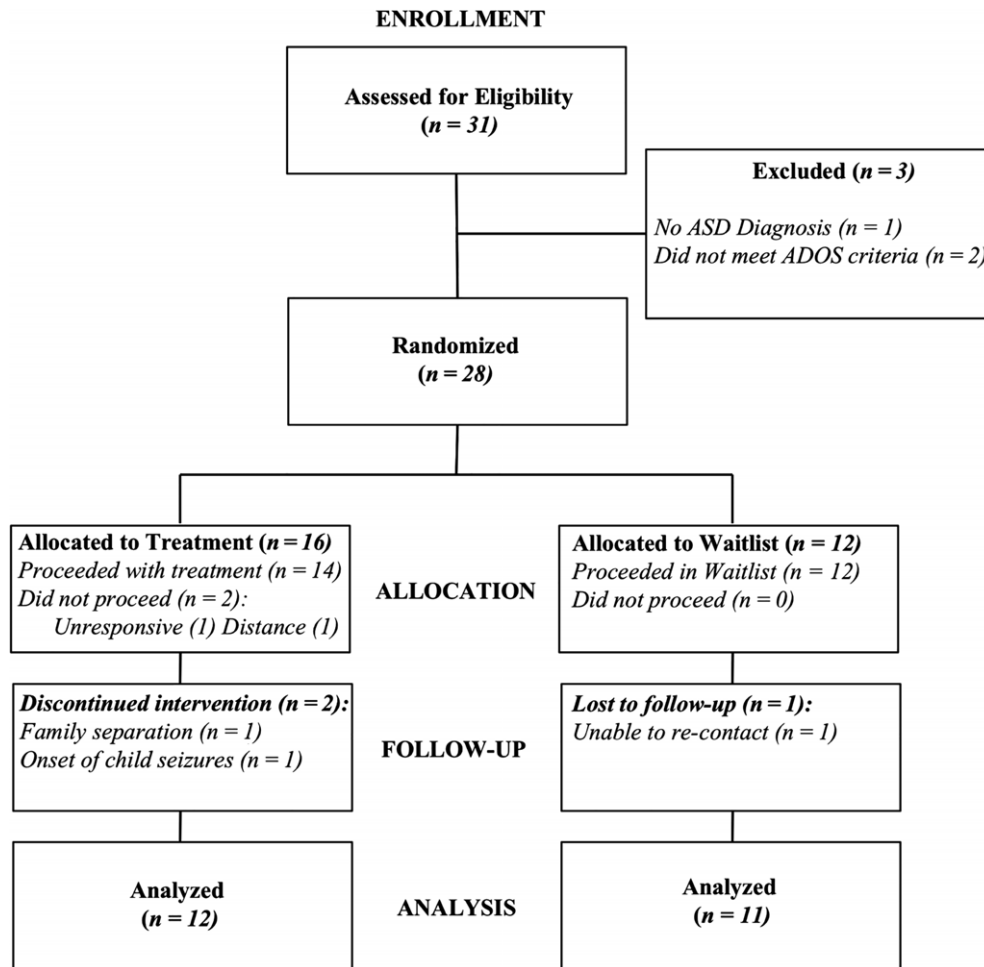
**Figure 1.** Consort diagram for randomized clinical trial. Source: Reprinted from Vernon et al. 2019 with permission from Springer Nature.

Participants in the TD group completed a single initial visit (Time 1) during which eye-tracking and demographic data were collected.

*Clinical Measures*

***Cognitive development.*** The Mullen Scales of Early Learning [Mullen, 1995] is a standardized developmental measure that provides *t*-scores and age equivalences for five domains of development: Visual Reception, Gross Motor, Fine Motor, Receptive Language, and Expressive Language. Verbal and nonverbal developmental quotients (DQ) can be calculated using the participant's chronological age and age equivalence for each of the verbal and nonverbal domains [Chawarska & Shic, 2009]. For this study, verbal DQ (VDQ) was calculated by summing the expressive and receptive language age equivalences and dividing by chronological age × 100. The same method was used to calculate nonverbal DQ (NVDQ) with the visual reception and fine motor domains.

***Adaptive behavior.*** The Vineland Adaptive Behavior Scales, Second Edition [Sparrow, Cicchetti, & Balla, 2005] is a semi-structured caregiver interview that assesses the child's everyday adaptive functioning. It contains four broad domains: Socialization, Communication, Daily Living Skills, and Motor Skills. Standard scores have a mean of 100 and standard deviation of 15. The socialization and communication domains were used in analyses for the current study.

***Autism symptomology.*** The ADOS-2 is a semi-structured standardized observational assessment of social communication, restricted interests, and repetitive behaviors for individuals aged 12 months to adulthood [Lord et al., 2012]. One of three modules (Toddler, 1, or 2) was administered as appropriate based on the child's chronological age and verbal abilities. Calibrated severity scores (CSS) are used to equate ADOS social affect, restricted/repetitive behavior, and total scores across all modules [Gotham, Pickles, & Lord, 2009]. The total CSS ranges from 1 to 10 and is associated with an ADOS classification of "non-spectrum" (score

of 1–3), "autism spectrum disorder" (score of 4–5), or "autism" (score of 6–10).

***Language skills.*** The PLS-5 is a developmental language assessment that evaluates both auditory comprehension and expressive communication skills of children from birth to 7 years [Zimmerman, Steiner, & Pond, 2011]. The total score was used in analyses for this study and is represented in a standard score (mean of 100, standard deviation of 15).

### Intervention

Participants randomly assigned to the PRISM condition participated in 10 hr per week of intervention for 6 months: 8 hr of direct clinician-delivered treatment and 2 hr of parent coaching. All intervention sessions occurred in the child's home or in the community. The intervention was delivered by clinical psychology doctoral students or undergraduate students who were trained to procedural fidelity on implementation of the PRISM model [Vernon et al., 2019]. All interventionists received weekly supervision from a licensed clinical psychologist.

The PRISM model is based on the core principles of PRT [Koegel & Koegel, 2006] and uses empirically supported naturalistic developmental behavioral treatment strategies [Schreibman et al., 2015] to set up language learning opportunities. The PRISM model uses a three-part contingency (antecedent, behavior, and consequence) and fosters social engagement through exclusive use of social reinforcement, high affect bids, and noncontingent exposure. In this approach, adults are required to create a motivating social activity that is only possible through their active participation [Koegel et al., 2009; Vernon et al., 2012]. In the PRISM treatment model, the clinician or parent initially provided non-contingent (free) access to a potential social activity of interest. Social activities included the parent as an active necessary component of the activity, for example, singing songs, tickling, swinging the child, or jumping on a trampoline with the child. After the child indicated signs of interest and engagement in a social activity, the adult then (a) created a language trial using an animated high-affect bid, (b) waited for a verbal response attempt from the child, and (c) immediately reinforced the language attempts through engagement in a carefully constructed social activity. The adult intervention providers were not permitted to simply deliver or permit access to a desired toy or item. Instead, they were required to create a motivating social activity that is only possible through their active participation. This requirement was fulfilled by observing a child's existing interests, analyzing the sensory appeal of these preferred activities, and creating interactions that leveraged the use of these preexisting but historically nonsocial interests [Koegel et al., 2009;

Vernon et al., 2012]. Participants randomly assigned to the PRISM condition began treatment immediately after the Time 1 evaluation.

Participants randomly assigned to the TAU condition were not provided any intervention by the research team from Time 1 to Time 2 and began PRISM treatment after the Time 2 evaluation. Participants in this condition accessed intervention services on their own and data were collected on the type and dose of intervention the TAU participants received during the waitlist period. All participants randomized to the TAU condition received an assortment of early interventions that included applied behavior analysis, speech therapy, and enrollment in special needs preschool. Data on specific treatment targets were not collected, but given the description of services it was likely that these children received intervention targeting behavior, speech and language, and pre-academic skills. No TAU participants received PRT or the PRISM model during the waitlist period. The number of hours per week of intervention for the TAU group ranged from 2 to 25 and was not associated with child severity, with children receiving a mean of 10 hr per week.

### Eye-Tracking Measures

***Apparatus and stimuli.*** Visual scanning patterns were recorded using a SensoMotoric Instruments iView X-Version 2.8 eye-tracking system with a sampling rate of 120 Hz and analyzed using BeGaze software (Sensomotoric Instruments, 2014). Stimuli were displayed on a 22″ widescreen LCD monitor (1680 × 1050 pixels). Toddlers were placed in a car seat located 75 cm away from the monitor with their eye level at the center of the monitor. Each participant viewed a set of five different paradigms, two of which are the focus of the present study: face scanning and social preference. A five-point calibration validation procedure was used prior to the onset of each paradigm and child friendly videos were played in between paradigms to maintain child engagement.

Face scanning stimuli consisted of six color static images of affectively positive female faces selected from the Karolinska Directed Emotional Faces database [Lundqvist et al., 1998]. Each image measured 530 × 720 pixels and was displayed for 5 sec on a light gray background. The stimuli were presented in the same order for each participant.

The social preference paradigm contained six 5-sec videos that were modeled after the existing GeoPref stimuli [Pierce et al., 2011, 2016]. Each stimulus consisted of two silent videos, measuring 575 × 710 pixels, playing simultaneously side-by-side, 190 pixels apart. One video depicted a dynamic social scene with children moving and dancing and the other video depicted dynamic geometric movement similar to an animated screensaver. Videos were selected and modified by eye to equate lumination, color, size, and movement. Three videos

displayed the social scene on the right and three videos displayed the social scene on the left; each participant was shown the videos in the same alternating order.

**Dependent variables.** Regions of interest (ROI) for each paradigm were drawn using SMI BeGaze software. ROIs for the face scanning task consisted of the eyes, mouth, face (excluding hair), and background. In line with our hypotheses that mouth looking would differentiate ASD and TD participants and be associated with language skills, the primary dependent measure for this paradigm was *mouth-to-eyes ratio*, defined as the ratio of mouth looking to eye looking (mouth/(mouth + eyes)). Because the primary measure was mouth-to-eyes ratio, valid trials were defined as trials in which fixation duration on the eyes and mouth together amounted to at least 500 ms. This criterion was chosen in order to minimize loss of data while also providing sufficient looking data necessary to reflect a clear fixation and systematic scanning, especially in the case of static images which do not hold attention as well as dynamic stimuli. This criterion is similar to other studies utilizing similar types of static stimuli [Elsabbagh et al., 2013; Mercure et al., 2018; Oakes & Ellis, 2013]. Of note, modifying this valid looking criterion to 200 ms and 1000 ms did not impact the results of the study as they are described below.

ROIs for the social preference stimuli consisted of two rectangles that encompassed the social and geometric scenes. The primary dependent variable for this paradigm was *social preference* (social/(social + geometric)). Valid trials were defined as trials in which fixation duration on the social and geometric clips together amounted to at least 500 ms.

*Statistical Analysis*

Differences between ASD and TD looking behavior during the two eye-tracking tasks at Time 1 were evaluated using linear mixed-effects models. Diagnosis (ASD vs. TD) and trial were included as fixed effects, and trial and subject intercept were specified as random effects. If the effect of trial was significant, trials were further analyzed for the presence of outliers. Model coefficients were based on restricted maximum likelihood estimation. Associations between Time 1 eye-tracking performance and concurrent clinical characteristics within the ASD sample were evaluated using Spearman correlations.

Next, treatment condition was evaluated for a significant effect on change in eye-tracking performance in the ASD participants from Time 1 to Time 2. The effect of randomization condition (PRISM vs. TAU) on eye-tracking performance from Time 1 to Time 2 was modeled conditioned on all baseline values. This conditional joint response model shows increased tolerance of missing data as compared to analysis of covariance [Carpenter & Kenward, 2007]. In this model, trial was specified as a random factor and a random intercept was included for each subject. Model coefficients were based on restricted maximum likelihood estimation.

When evaluating how change in eye-tracking performance from Time 1 to Time 2 was associated with clinical change, all ASD participants were combined regardless of treatment condition. This approach was taken for several reasons. First, the primary aim of this pilot RCT was to evaluate feasibility of the PRISM model and was only powered to evaluate preliminary efficacy of the intervention (primary feasibility and efficacy findings are presented in Vernon et al., 2019). Second, all participants, regardless of randomization condition, were receiving behavioral intervention for core symptoms of ASD. Finally, combining both groups together helped to increase statistical power. Primary eye-tracking variables (mouth-to-eyes ratio and social preference) were aggregated across valid trials within participants at each time point such that each participant had a single mean score on both eye-tracking variables at each time point. A change score from Time 1 to Time 2 was then calculated for each participant by subtracting the Time 2 score from the Time 1 score such that a positive change score indicated an increase in the measure score from Time 1 to Time 2. This was done for both eye-tracking variables and all primary clinical measures (ADOS Total CSS, PLS-5 Total, NVDQ, VDQ, Vineland Communication, and Vineland Socialization). Spearman correlations were used to test for associations.

Statistical significance was assessed at the 0.05 level unless otherwise noted. All analyses were run using IBM SPSS Statistics Version 24 (IBM, 2016).

## Results

Participants with ASD had a mean age of 34.2 months (SD = 9.3) at baseline and were comparable to TD participants whose mean age was 30.1 months (SD = 8.7) ($t(49) = 1.61$, $P = 0.114$). However, the TD group had a significantly lower proportion of males ($N = 14$, 61%; $\chi^2 = 4.10$, $P < 0.05$). Baseline characteristics of participants with ASD, including baseline differences between children who completed PRISM and TAU, are presented in Table 1. Intercorrelations among clinical measures are presented in Table S1. At baseline, the PRISM group had significantly higher scores on the Vineland-II, but was comparable on all other clinical measures and demographic characteristics. There was no association between the number of hours of intervention the TAU group received during the waitlist period and autism severity ($r_s = 0.287$, $P = 0.393$).

*Differences in Social Attention between ASD and TD Participants*

The number of valid trials did not differ between ASD and TD participants for the face-scanning stimuli (ASD

**Table 1. Demographic and Clinical Characteristics for ASD Participants at Baseline**

| Measure | All ($n = 28$) | PRISM ($n = 12$) M | TAU ($n = 11$) M | Test statistic ($t$ or $X^2$) | P | Mean Dif | 95 CI for mean Dif Low | High |
|---|---|---|---|---|---|---|---|---|
| Age (months) | 34.21 (9.28) | 35.75 (9.31) | 34.45 (10.08) | 0.32 | 0.752 | 1.3 | −7.16 | 9.75 |
| Sex | | | | 0.49 | 0.484 | | | |
|     Male | 24 (86%) | 11 (92%) | 9 (82%) | – | – | – | – | – |
|     Female | 4 (14%) | 1 (8%) | 2 (18%) | – | – | – | – | – |
| Race/ethnicity | | | | 4.42 | 0.219 | | | |
|     White | 16 (57%) | 9 (75%) | 4 (36%) | – | – | – | – | – |
|     Latino | 6 (22%) | 2 (17%) | 3 (27%) | – | – | – | – | – |
|     Asian | 4 (14%) | 1 (8%) | 2 (18%) | – | – | – | – | – |
|     Multi-racial | 2 (7%) | 0 (0%) | 2 (18%) | – | – | – | – | – |
| ADOS-2 CSS | 7.14 (1.74) | 7 (1.48) | 7.18 (1.25) | −0.32 | 0.754 | 0.182 | −1.37 | 1.01 |
| Preschool Language Scales, 5th Edition | | | | | | | | |
|     Auditory comprehension | 74.96 (20.85) | 84.33 (19.19) | 69.64 (16.5) | 1.96 | 0.063 | 14.7 | −0.89 | 30.29 |
|     Expressive communication | 75.89 (16.6) | 81.5 (12.09) | 72.91 (14) | 1.58 | 0.129 | 8.59 | −2.72 | 19.91 |
|     Total score | 74.11 (18.9) | 81.58 (15.42) | 69.91 (15.48) | 1.81 | 0.085 | 11.67 | −1.75 | 25.09 |
| Mullen Scales of Early Learning | | | | | | | | |
|     Visual reception | 36.43 (14.43) | 42.92 (13.37) | 34.27 (13.76) | 1.53 | 0.142 | 8.64 | −3.12 | 20.41 |
|     Fine motor | 29.61 (9.46) | 33.25 (9.43) | 28.55 (9.27) | 1.21 | 0.242 | 4.71 | −3.41 | 12.82 |
|     Receptive language | 30.29 (12.98) | 36.17 (14.39) | 26.09 (9.83) | 1.94 | 0.066 | 10.08 | −0.71 | 20.87 |
|     Expressive language | 31.57 (13.51) | 35 (14.31) | 29.82 (11.7) | 0.95 | 0.355 | 5.18 | −6.22 | 16.58 |
| Vineland Adaptive Behavior Scales, 2nd Edition | | | | | | | | |
|     Communication | 73.41 (12.95) | 81 (9.88) | 66.73 (12.85) | 3 | 0.007** | 14.02 | 4.3 | 23.75 |
|     Daily living | 86.59 (10.99) | 92.45 (10.1) | 81.55 (10.22) | 2.51 | 0.020* | 10.46 | 1.79 | 19.12 |
|     Socialization | 79.7 (9.34) | 85.36 (10.08) | 72.45 (5.07) | 3.9 | 0.001** | 12.71 | 5.94 | 19.49 |
|     Motor skills | 83.15 (16.93) | 88.18 (22.75) | 88.18 (22.75) | 1.28 | 0.216 | 9.13 | −5.75 | 24 |

*Note.* Five participants withdrew from the intervention prior to the Time 2 visit. These participants are not included in the between-group baseline analyses. Between-group differences in sex and race/ethnicity were tested using Chi-square tests.

**$P < 0.01$; *$P < 0.05$.

mean = 2.7, TD mean = 2.6; $t(48) = 0.14$, $P = 0.890$) or the social preference stimuli (ASD mean = 4.9, TD mean = 4.8; $t(49) = 0.51$, $P = 0.614$). Similarly, the total time looking to the screen during valid trials did not differ between the groups for the face scanning stimuli (ASD mean = 3.8 (1.1) sec, TD mean = 3.9 (1.1) sec; $t(48) = −0.48$, $P = 0.629$) or the social preference stimuli (ASD mean = 4.5 (1.0) sec, TD mean = 4.3 (1.0) sec; $t(244) = 1.39$, $P = 0.166$). The dependent measures from the two eye-tracking tasks were not associated with each other, suggesting that they are tapping into different components of social attention ($r = 0.02$, $P = 0.924$). Linear mixed effects models showed a significant effect of trial during the social preference task ($F_{1,272} = 10.63$, $P < 0.001$), but not for the face-scanning task. Upon further investigation, one of the six social preference trials emerged as a significant outlier compared to other trials and removal of this trial decreased this effect ($F = 6.54$, $P < 0.05$). Therefore, this trial was removed from all subsequent analyses, but trial remained a random factor in all models.

Linear mixed effects models revealed that diagnosis did not have a significant effect on mouth-to-eyes ratio at Time 1 ($F_{1,101} = 0.04$, $P = 0.84$). The model-based mean mouth-to-eyes ratio was 0.46 (SE = 0.07) for the TD group and 0.47 (SE = 0.07) for the ASD group (Fig. 2A). In contrast, diagnosis did have a significant effect on social preference at Time 1 ($F_{1,98} = 4.36$, $P < 0.05$; Fig 2B). The model-based mean social preference ratio for the TD group was 0.70 (SE = 0.03), significantly greater than that of the ASD mean ratio of 0.62 (SE = 0.03). However, both groups exhibited a social preference significantly greater than chance level of 0.5 (TD: $t(22) = 7.02$, $P < 0.001$; ASD: $t(27) = 4.17$, $P < 0.001$).

*Eye Tracking and Clinical Characteristics for ASD Participants at Time 1*

Associations between Time 1 eye tracking and clinical characteristics are presented in Table 2. Mouth-to-eyes ratio was found to be significantly associated with higher scores on the PLS-5, Vineland communication, and VDQ. There were no significant associations between social preference at Time 1 and concurrent autism symptomology, language, cognitive, or adaptive skills, but the association between social preference and Vineland socialization was approaching significance. Additionally, age was not associated with mouth-to-eyes ratio ($r_s = 0.14$, $P = 0.52$) or social preference ($r_s = 0.02$, $P = 0.93$).

*Associations between Eye Tracking and Clinical Characteristics over Time*

The linear mixed effects model, conditioned on baseline, revealed no significant effect of treatment condition on
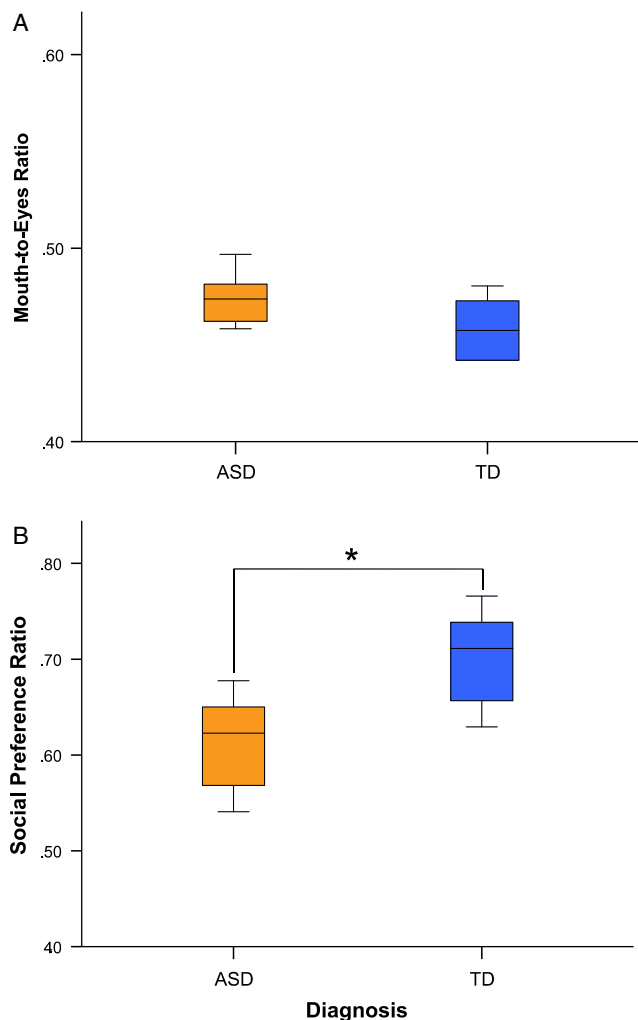
**Figure 2.** Boxplots of model-based estimates for mouth-to-eyes ratio (mouth/(mouth + eyes)) (**A**) and social preference ratio (social/(social + geometric)) (**B**) for ASD and TD participants at baseline. *$P < 0.05$.

mouth-to-eyes ratio from Time 1 to Time 2 ($F_{2,120} = 0.08$, $P = 0.92$). At Time 1, the model-based mean for all ASD participants was 0.54 (SE = 0.06), which increased to 0.55 (SE = 0.09) for participants randomized to PRISM and decreased to 0.51 (SE = 0.09) for participants randomized to TAU. There was also no significant effect of treatment condition on social preference from Time 1 to Time 2 ($F_{2,83} = 1.24$, $P = 0.23$). At Time 1, the model-based mean for all ASD participants was 0.63 (SE = 0.03), which decreased to 0.58 (SE = 0.04) for participants randomized to PRISM and increased to 0.65 (SE = 0.09) for participants randomized to TAU. As mentioned above, all ASD participants, regardless of randomization condition, received weekly intervention services for the 6-month period between Time 1 and Time 2 and all participants were combined in subsequent analyses. The intervention dose (average hours per week) across participants in both

conditions was not associated with change in either eye-tracking measures (mouth-to-eyes ratio: $r_s = 0.06$, $P = 0.80$; social preference ratio: $r_s = -0.16$, $P = 0.49$) or with change in autism symptomology (ADOS CSS: $r_s = 0.17$, $P = 0.43$). Associations between change in eye-tracking measures and change in clinical measures are presented in Table 3. Change in mouth-to-eyes ratio from Time 1 to Time 2 was significantly associated with change in the PLS-5 total score. That is, children who showed a decrease in looking to the mouth, compared to the eyes, from Time 1 to Time 2 also showed improved performance on the PLS-5 (see Fig. 3). There were no significant associations between change in social preference and change in clinical measures from Time 1 to Time 2.

## Discussion

The goal of this study was to contribute to the pressing need for viable objective treatment outcome measures that can be used in clinical trials for ASD. Despite decades of research supporting altered patterns of attention in children with ASD, only a few studies have tested eye tracking as a possible treatment outcome measure. In the context of a pilot RCT, this study examined the preliminary utility of two eye-tracking paradigms for assessing clinical change in children with ASD across a 6-month intervention period. Analyses revealed two important trends. First, results showed that preference for social stimuli in a forced-choice paradigm differentiated children with ASD from TD children at Time 1, but was not associated with Time 1 clinical characteristics, or with change in clinical features over time, for children with ASD. Second, it was found that attention to the mouth of a static face (mouth-to-eyes ratio) did not differ between ASD and TD children but, within the ASD group, a higher mouth-to-eyes ratio was associated with superior language skills at Time 1 and a decline in mouth looking was associated with greater gains in language from Time 1 to Time 2. These findings are generally consistent with, but also reveal notable departures from, prior research.

Children with ASD in the present study showed a significantly lower preference for social stimuli compared to TD participants, providing an independent replication of research originally conducted by Pierce et al. [2011, 2016]. This is particularly encouraging given that the stimuli used in this study were modeled after the original stimuli used in Pierce et al. [2011], but were independently designed. However, unlike Pierce et al. [2016], the present study did not identify statistically significant associations between social preference and clinical characteristics of the ASD sample. Rather, a moderate, though still nonsignificant, effect of social preference on adaptive socialization skills was found at Time 1. There are several possibilities for this observed lack of association. First, it is

**Table 2. Associations between Eye Tracking and Clinical Measures at Baseline for ASD Participants**

| Baseline clinical measure | Mouth-to-eyes ratio | | Social preference | |
|---|---|---|---|---|
| | Spearman's rho | P value | Spearman's rho | P value |
| ADOS CSS total | 0.213 | 0.317 | −0.126 | 0.524 |
| PLS-5 total | 0.562 | 0.004** | 0.153 | 0.436 |
| Vineland communication | 0.598 | 0.002** | 0.104 | 0.607 |
| Vineland socialization | −0.110 | 0.610 | 0.348 | 0.076 |
| NVDQ | 0.377 | 0.069 | 0.056 | 0.776 |
| VDQ | 0.457 | 0.025* | 0.111 | 0.575 |

**P < 0.01; *P < 0.05.

**Table 3. Associations between Change in Eye Tracking and Change in Clinical Measures for ASD Participants from Time 1 to Time 2**

| Δ Clinical measures | Δ Mouth-to-eyes ratio | | Δ Social preference | |
|---|---|---|---|---|
| | Spearman's rho | P value | Spearman's rho | P value |
| ADOS CSS total | 0.268 | 0.266 | 0.035 | 0.876 |
| PLS-5 total | −0.533 | 0.019* | 0.111 | 0.621 |
| Vineland communication | −0.213 | 0.413 | −0.271 | 0.248 |
| Vineland socialization | 0.175 | 0.501 | 0.228 | 0.335 |
| NVDQ | −0.435 | 0.063 | −0.319 | 0.148 |
| VDQ | −0.219 | 0.367 | −0.243 | 0.275 |

*P < 0.05.

possible that the purported mechanism we hoped this paradigm would tap into (decreased social motivation) is a highly conserved feature of ASD, but is not associated with the vast clinical heterogeneity observed within ASD. In this case, it could be argued that social preference is a good diagnostic biomarker of ASD, but not of treatment change. Similar to behavioral measures, it is important to identify and differentiate diagnostic biomarkers from biomarkers that are sensitive to change over time within diagnostic groups. It is also possible that phenotypic heterogeneity within ASD is less tightly linked to social motivation than we originally hypothesized. Perhaps an increased visual preference for social stimuli is not necessarily needed for children with ASD to learn new language or social skills. There are also few key differences in the present study and that of Pierce and colleagues that may help to explain some of the observed inconsistencies. The mean age of the ASD sample described in Pierce et al.
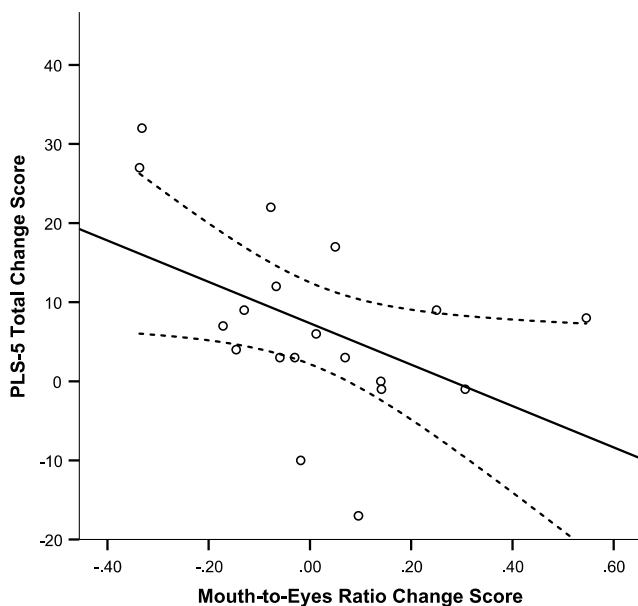


**Figure 3.** Scatterplot of change in PLS scores and change in mouth-to-eyes ratio from Time 2 to Time 1 for ASD participants with regression line and 95% confidence intervals.

[2011, 2016] was about 6–8 months younger than the ASD sample in the present study. In addition, the ascertainment method of the ASD sample in Pierce et al. [2011, 2016] made it likely that many participants were undergoing their first diagnostic evaluation at the time of the experiment. In contrast, recruitment efforts used in the present study focused on children with an ASD diagnosis and so many of the enrolled participants carried a diagnosis of ASD prior to study entry, increasing the likelihood that these children were already receiving some form of early intervention. Furthermore, the duration of the social preference paradigm in the initial study was twice as long in duration (~60 sec) than the paradigm used in this study (~30 sec), which may have impacted scanning patterns. Given these important methodological distinctions between the present study and Pierce et al. [2016], in addition to a much smaller sample size, this study can serve as a partial, independent replication of the potential utility of a social preference paradigm as a diagnostic biomarker.

Results showed that attention to the mouth during a static face scanning task emerged as a better indicator of clinical heterogeneity within the ASD group than the social preference task.

At Time 1, mouth-to-eyes ratio did not differentiate the diagnostic groups, but, as seen in older children and adults with ASD during presentation of dynamic stimuli [Norbury et al., 2009; Tenenbaum et al., 2014, 2015], it was positively associated with language and adaptive communication skills within the ASD group. The associations between mouth looking and language were stronger when looking at the PLS-5 than the language domains of the Mullen Scales of Early Learning. The PLS-5 is a measure designed specifically for identifying language disorders in preschoolers and it assesses basic vocabulary, language concepts, grammar, syntax, and semantics. It is possible that the PLS-5 is a more comprehensive measure of language development for children with ASD and so may have been more sensitive to the change in mouth looking observed from Time 1 to Time 2.

These findings align with previous research [Murias et al., 2018] that observed a relationship between clinical

phenotype and mouth looking during presentation of dynamic social stimuli (i.e., a person engaged in child-directed speech). It is important to consider these findings in light of the dynamic developmental changes inherent to face scanning patterns. Early on, at around 6 months of age, the eyes are the most captivating facial features for both infants who develop ASD as well as TD infants, possibly due to their perceptual salience [Shic et al., 2014]. As TD toddlers begin to acquire language, they shift their attention allocation from the eyes to the mouth region. Following mastery of basic verbal communication skills (i.e., phrase and early conversational speech) at around age 4, attention to the eyes increases once again and toddlers shift attention *between* the eyes and mouth more frequently [Shic, Chawarska, Bradshaw, & Scassellati, 2008]. This developmental transition in face scanning could be interpreted as the toddler's increasing appreciation for both eye and mouth regions as informative social communicative features, and their increasingly refined ability to fluidly integrate the two. In contrast, evidence suggests that ASD children do not make this same developmental transition [Shic et al., 2008]. Taken together, we conjecture that typical acquisition of language and social-communication skills may be characterized by increased mouth looking as well as increased attention shifts between the eyes and mouth. This theory is quite consistent with findings from the present study. While increased mouth-looking was associated with better language skills at Time 1, a decrease in mouth looking from Time 1 to Time 2 was associated with greater language improvements. In this framework, we hypothesize that decreases in mouth-to-eyes ratio for some participants with ASD were due to more sophisticated scanning patterns characterized by more transitions between the eyes and mouth, all resulting from improvements in language skills. Alternatively, it is possible that this may be a result of an unintended experimental effect in which children learned from the Time 1 presentation, albeit presented 6 months earlier, that the faces presented are not dynamic, speech-producing stimuli, and are therefore spending less time looking at the mouth expecting language.

The finding that randomization condition did not affect eye-tracking performance is worth discussion. The primary outcomes of this pilot RCT were feasibility and parent acceptability. In a preliminary efficacy analysis of this data [Vernon et al., 2019], participants who were randomized to the PRISM treatment condition showed significantly improved performance in the ADOS-2 CSS and PLS-5, while participants in the TAU condition did not show significant improvements on any measures. In contrast, the current study shows that changes in eye tracking from Time 1 to Time 2 did not significantly differ between participants in the PRISM and TAU conditions. The association between eye-tracking performance and improvement on the PLS-5, but not the ADOS, suggests

that the eye-tracking paradigms used in this study may have been helpful in tracking some treatment gains (i.e., language), but failed to capture the totality of observed child improvements (i.e., improvements in language *and* decreases in symptom severity). Vernon et al. [2019] observed that decreases in ADOS CSS from Time 1 to Time 2 in the PRISM condition had the highest effect size compared to all other clinical measures. The lack of association between eye-tracking performance and ADOS CSS at Time 1 may partially explain why eye tracking did not differ between randomization conditions, despite significant clinical improvement within the PRISM group. Additional research that includes a 6-month follow-up evaluation may reveal more robust differences.

This study has several limitations to note. First, our TD participants were only seen at Time 1 for collection of demographic and eye-tracking data, preventing a comparison of ASD and TD performance at Time 2. The lack of clinical data for our TD sample also prevents us from examining associations between eye tracking and clinical measures in a typical population. This analysis would be useful to determine whether these eye-tracking measures reflect broad developmental mechanisms and can be measured in non-ASD populations. In this study, we used composite measures of language from the PLS-5 and the Mullen. In light of research identifying differences in eye-tracking associations with receptive versus expressive language [Tsang, Atagi, & Johnson, 2018] and the unique profile of receptive versus expressive language skills in children with ASD [Hudry et al., 2010, 2014], future research should investigate these two domains of language separately. The small sample size of this study prevented us from examining heterogeneity within our ASD participants (e.g., comparing "responders" to "slow-responders") and identifying biomarkers of change for individual participants (e.g., conducting sensitivity and specificity analyses). However, our findings suggest that mouth looking should be further examined as a possible predictor of early response to treatment. In particular, mouth looking may serve different functions depending on the stimulus context (e.g., static vs. dynamic stimuli) and the language level of the child (e.g., preverbal vs. fluent). Future research should investigate the effect of stimulus by including dynamic faces and social scenes and incorporate repeated assessments across children at various stages in language learning. Additionally, eye-tracking paradigms that are associated with autism severity (i.e., ADOS) should be investigated and used in future clinical trials.

## Conclusions

The present study highlights the important distinction between diagnostic biomarkers and biomarkers that measure response to treatment. Attenuated social visual

engagement is a hallmark feature of ASD, and here we found that our proxy for this impairment (social preference) provided a diagnostic biomarker of ASD, consistent with previous research [Pierce et al., 2016]. The next step in this line of work is to incorporate other non-ASD samples that may have overlapping clinical phenotypes, such as individuals with global developmental delay, language disorder, and ADHD. However, social preference was not a useful measure of treatment change in our sample of young children with ASD. While the PRISM model did result in marked improvements in social and language skills significantly greater than that of the TAU group, our results suggest that the pathway to clinical change was not an increase in preference for social stimuli. Rather, our findings suggest the possibility that treatment may have led to decreases in mouth looking, possibly driven by increased attention shifting between the eyes and mouth. It is unknown, however, whether improvement in language led to decreased mouth looking or whether decreased mouth looking preceded improvements in language skills. Future research on biomarkers of treatment response should include multiple time points and large sample sizes in order to evaluate the causal relationship between treatment, biomarkers, and clinical change. In the case of treatment models hypothesized to improve social motivation, the development of stimuli that may be more sensitive to underlying social motivation should be explored. In this way, the study of biomarkers can enhance our understanding of precise treatment mechanisms, allowing us to individualize treatments and effect the greatest change in children with ASD.

## Acknowledgments

## Conflict of interest

All authors declare they have no conflicts of interest.

## References

Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clinical Pharmacology and Therapeutics, 69(3), 89–95. https://doi.org/10.1067/mcp.2001.113989.

Bolte, E. E., & Diehl, J. J. (2013). Measurement tools and target symptoms/skills used to assess treatment response for individuals with autism spectrum disorder. Journal of Autism and Developmental Disorders, 43(11), 2491–2501. https://doi.org/10.1007/s10803-013-1798-7.

Boraston, Z., & Blakemore, S.-J. (2007). The application of eye-tracking technology in the study of autism. The Journal of Physiology, 581(3), 893–898. https://doi.org/10.1113/jphysiol.2007.133587.

Bradshaw, J., Bearss, K., McCracken, C., Smith, T., Johnson, C., Lecavalier, L., … Scahill, L. (2017). Parent education for young children with autism and disruptive behavior: Response to active control treatment. Journal of Clinical Child & Adolescent Psychology, 47, 1–11. https://doi.org/10.1080/15374416.2017.1381913.

Bradshaw, J., Koegel, L., & Koegel, R. (2017). Improving functional language and social motivation with a parent-mediated intervention for toddlers with autism spectrum disorder. Journal of Autism and Developmental Disorders, 47(8), 2443–2458. https://doi.org/10.1007/s10803-017-3155-8.

Bradshaw, J., Steiner, A. M., Gengoux, G., & Koegel, L. (2015). Feasibility and effectiveness of very early intervention for infants at-risk for autism spectrum disorder: A systematic review. Journal of Autism and Developmental Disorders, 45(3), 778–794. https://doi.org/10.1007/s10803-014-2235-2.

Carpenter, J. R., & Kenward, M. G. (2007). Missing data in randomised controlled trials: a practical guide. Birmingham: Health Technology Assessment Methodology Programme. pp 199. http://researchonline.lshtm.ac.uk/id/eprint/4018500.

Chawarska, K., Macari, S., & Shic, F. (2012). Context modulates attention to social scenes in toddlers with autism. Journal of Child Psychology and Psychiatry, 53(8), 903–913. https://doi.org/10.1111/j.1469-7610.2012.02538.x.

Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. Journal of Autism and Developmental Disorders, 39(12), 1663–1672. https://doi.org/10.1007/s10803-009-0803-7.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. Trends in Cognitive Sciences, 16(4), 231–239. https://doi.org/10.1016/J.TICS.2012.02.007.

Cohen, I. L., Schmidt-Lackner, S., Romanczyk, R., & Sudhalter, V. (2003). The PDD behavior inventory: A rating scale for assessing response to intervention in children with pervasive developmental disorder. Journal of Autism and Developmental Disorders, 33(1), 31–45. https://doi.org/10.1023/A:1022226403878.

Constantino, J. N. (2012). Social responsiveness scale, second edition (SRS-2). Los Angeles, CA: Western Psychological Services.

Cunningham, A. B. (2012). Measuring change in social interaction skills of young children with autism. Journal of Autism and Developmental Disorders, 42(4), 593–605. https://doi.org/10.1007/s10803-011-1280-3.

Dawson, G. (2017). Questions remain regarding the effectiveness of many commonly used autism treatments. Pediatrics, 139(6), e20170730. https://doi.org/10.1542/peds.2017-0730.

Dawson, G., Bernier, R., & Ring, R. H. (2012). Social attention: A possible early indicator of efficacy in autism clinical trials. Journal of Neurodevelopmental Disorders, 4(1), 11. https://doi.org/10.1186/1866-1955-4-11.

Dawson, G., Jones, E., Merkle, K., Venema, K., Lowy, R., Faja, S., … Webb, S. J. (2012). Early behavioral intervention is associated with normalized brain activity in young children with autism. Journal of the American Academy of Child and Adolescent Psychiatry, 51(11), 1150–1159. https://doi.org/10.1016/j.jaac.2012.08.018.

Elsabbagh, M., Gliga, T., Pickles, A., Hudry, K., Charman, T., Johnson, M. H., & BASIS Team. (2013). The development of face orienting mechanisms in infants at-risk for autism. Behavioural Brain Research, 251, 147–154. https://doi.org/10.1016/j.bbr.2012.07.030.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. Proceedings of the National Academy of Sciences of the United States of America, 99, 9602–9605. https://doi.org/10.1073/pnas.152159999.

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., & Bates, E. (2007). The MacArthur-Bates communicative development inventories user's guide and technical manual (2nd) ed. Baltimore, MD: Paul H Brookes. Retrieved from: https://bpwf40lbo01.storage.googleapis.com/MTU1NzY2ODg0MQ==01.pdf

Fletcher-Watson, S., & McConachie, H. (2017). The search for an early intervention outcome measurement tool in autism. Focus on Autism and Other Developmental Disabilities, 32(1), 71–80. https://doi.org/10.1177/1088357615583468.

French, L., & Kennedy, E. M. M. (2018). Annual research review: Early intervention for infants and young children with, or at-risk of, autism spectrum disorder: A systematic review. Journal of Child Psychology and Psychiatry, 59(4), 444–456. https://doi.org/10.1111/jcpp.12828.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. Journal of Autism and Developmental Disorders, 39(5), 693–705. https://doi.org/10.1007/s10803-008-0674-3.

Grzadzinski, R., Carr, T., Colombi, C., McGuire, K., Dufek, S., Pickles, A., & Lord, C. (2016). Measuring changes in social communication behaviors: Preliminary development of the Brief Observation of Social Communication Change (BOSCC). Journal of Autism and Developmental Disorders, 46(7), 2464–2479. https://doi.org/10.1007/s10803-016-2782-9.

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. Neuroscience & Biobehavioral Reviews, 42, 279–297. https://doi.org/10.1016/j.neubiorev.2014.03.013.

Hardan, A. Y., Gengoux, G. W., Berquist, K. L., Libove, R. A., Ardel, C. M., Phillips, J., … Minjarez, M. B. (2015). A randomized controlled trial of Pivotal Response Treatment Group for parents of children with autism. Journal of Child Psychology and Psychiatry, 56(8), 884–892. https://doi.org/10.1111/jcpp.12354.

Hudry, K., Chandler, S., Bedford, R., Pasco, G., Gliga, T., Elsabbagh, M., … Charman, T. (2014). Early language profiles in infants at high-risk for autism spectrum disorders. Journal of Autism and Developmental Disorders, 44(1), 154–167. https://doi.org/10.1007/s10803-013-1861-4.

Hudry, K., Leadbitter, K., Temple, K., Slonims, V., McConachie, H., Aldred, C., … PACT Consortium. (2010). Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. International Journal of Language & Communication Disorders, 45(6), 681–690. https://doi.org/10.3109/13682820903461493.

Jones, E., Dawson, G., Kelly, J., Estes, A., & Jane Webb, S. (2017). Parent-delivered early intervention in infants at risk for ASD: Effects on electrophysiological and habituation measures of social attention. Autism Research, 10(5), 961–972. https://doi.org/10.1002/aur.1754.

Jones, R. M., Carberry, C., Hamo, A., & Lord, C. (2017). Placebo-like response in absence of treatment in children with autism. Autism Research, 10(9), 1567–1572. https://doi.org/10.1002/aur.1798.

Kasari, C., Gulsrud, A., Paparella, T., Hellemann, G., & Berry, K. (2015). Randomized comparative efficacy study of parent-mediated interventions for toddlers with autism. Journal of Consulting and Clinical Psychology, 83(3), 554–563. https://doi.org/10.1037/a0039080.

Kim, E. S., Friedlaender, L., Kowitt, J., Reichow, B., Naples, A., Gearty, G. V., … Shic, F. (2014). Development of an untethered, mobile, low-cost head-mounted eye tracker. In Proceedings of the symposium on eye tracking research and applications - ETRA '14 (pp. 247–250). New York, NY: ACM Press. https://doi.org/10.1145/2578153.2578209.

Koegel, R., & Koegel, L. (2006). Pivotal response treatments. Baltimore, MD: Paul H Brookes Publishing Co.

Koegel, R., Vernon, T., & Koegel, L. (2009). Improving social initiations in young children with autism using reinforcers with embedded social interactions. Journal of Autism and Developmental Disorders, 39(9), 1240–1251. https://doi.org/10.1007/s10803-009-0732-5.

Kubicek, C., de Boisferon, A. H., Dupierrix, E., Lœvenbruck, H., Gervain, J., & Schwarzer, G. (2013). Face-scanning behavior to silently-talking faces in 12-month-old infants: The impact of pre-exposed auditory speech. International Journal of Behavioral Development, 37(2), 106–110. https://doi.org/10.1177/0165025412473016.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. Proceedings of the National Academy of Sciences, 109(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule: ADOS-2. Torrance, CA: Western Psychological Services.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces—KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.

Masi, A., Lampit, A., Glozier, N., Hickie, I. B., & Guastella, A. J. (2015). Predictors of placebo response in pharmacological and dietary supplement treatment trials in pediatric autism spectrum disorder: A meta-analysis. Translational Psychiatry, 5(9), e640–e640. https://doi.org/10.1038/tp.2015.143.

Mazurek, M. O., Carlson, C., Baker-Ericzén, M., Butter, E., Norris, M., & Kanne, S. (2018). Construct validity of the Autism Impact Measure (AIM). Journal of Autism and Developmental Disorders, 1–13. https://doi.org/10.1007/s10803-018-3462-8.

McConachie, H., Parr, J. R., Glod, M., Hanratty, J., Livingstone, N., Oono, I. P., … Williams, K. (2015). Systematic review of tools to measure outcomes for young children with autism spectrum disorder. Health Technology Assessment, 19(41). Retrieved from), 1–506. http://resolver.ebscohost.com.pallas2.tcl.sc.edu/openurl?sid=google&auinit=H&aulast=McConachie&atitle=Systematic+review+of+tools+to+measure+outcomes+for+young+children+with+autism+spectrum+disorder&id=doi%3A10.3310%2Fhta19410&site=ftf-live.

McPartland, J. C. (2016). Considerations in biomarker development for neurodevelopmental disorders. Current Opinion in

Neurology, 29(2), 118–122. https://doi.org/10.1097/WCO.0000000000000300.

Mercure, E., Quiroz, I., Goldberg, L., Bowden-Howl, H., Coulson, K., Gliga, T., … MacSweeney, M. (2018). impact of language experience on attention to faces in infancy: Evidence from unimodal and bimodal bilingual infants. Frontiers in Psychology, 9. https://doi.org/10.3389/fpsyg.2018.01943.

Mullen, E. (1995). Mullen scales of early learning. Circle Pines, MN: American Guidance Service.

Murias, M., Major, S., Davlantis, K., Franz, L., Harris, A., Rardin, B., … Dawson, G. (2018). Validation of eye-tracking measures of social attention as a potential biomarker for autism clinical trials. Autism Research, 11(1), 166–174. https://doi.org/10.1002/aur.1894.

Norbury, C. F., Brock, J., Cragg, L., Einav, S., Griffiths, H., & Nation, K. (2009). Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. Journal of Child Psychology and Psychiatry, 50(7), 834–842. https://doi.org/10.1111/j.1469-7610.2009.02073.x.

Oakes, L. M., & Ellis, A. E. (2013). An eye-tracking investigation of developmental changes in infants' exploration of upright and inverted human faces. Infancy, 18(1), 134–148. https://doi.org/10.1111/j.1532-7078.2011.00107.x.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. Journal of Autism and Developmental Disorders, 32(4), 249–261. Retrieved from:. http://www.ncbi.nlm.nih.gov/pubmed/12199131.

Pierce, K., Conant, D., Hazin, R., Stoner, R., & Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. Archives of General Psychiatry, 68(1), 101–109. https://doi.org/10.1001/archgenpsychiatry.2010.113.

Pierce, K., Marinero, S., Hazin, R., McKenna, B., Barnes, C. C., & Malige, A. (2016). Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. Biological Psychiatry, 79(8), 657–666. https://doi.org/10.1016/j.biopsych.2015.03.032.

Rutter, M., Bailey, A., & Lord, C. (2003). Social Communication Questionnaire (SCQ). Torrance, CA: Western Psychological Services.

Sasson, N. J., & Elison, J. T. (2012). Eye tracking young children with autism. Journal of visualized experiments: JoVE, (61), 3675. https://doi.org/10.3791/3675.

Schertz, H. H., Odom, S. L., Baggett, K. M., & Sideris, J. H. (2013). Effects of joint attention mediated learning for toddlers with autism spectrum disorders: An initial randomized controlled study. Early Childhood Research Quarterly, 28(2), 249–258. https://doi.org/10.1016/J.ECRESQ.2012.06.006.

Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., … Halladay, A. (2015). Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder. Journal of Autism and Developmental Disorders, 45(8), 2411–2428. https://doi.org/10.1007/s10803-015-2407-8.

Shaffer, R. C., Pedapati, E. V., Shic, F., Gaietto, K., Bowers, K., Wink, L. K., & Erickson, C. A. (2017). Brief report: Diminished gaze preference for dynamic social interaction scenes in youth with autism spectrum disorders. Journal of Autism and Developmental Disorders, 47(2), 506–513. https://doi.org/10.1007/s10803-016-2975-2.

Shic, F. (2013). Eye-tracking. In Encyclopedia of autism spectrum disorders (pp. 1208–1213). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-1698-3_1474.

Shic, F. (2016). Eye tracking as a behavioral biomarker for psychiatric conditions: The road ahead. Journal of the American Academy of Child & Adolescent Psychiatry, 55(4), 267–268. https://doi.org/10.1016/j.jaac.2016.02.002.

Shic, F., Chawarska, K., Bradshaw, J., & Scassellati, B. (2008). Autism, eye-tracking, entropy. In 2008 7th IEEE International Conference on Development and Learning. Monterey, CA. (pp. 73–78). IEEE. https://doi.org/10.1109/DEVLRN.2008.4640808.

Shic, F., Macari, S., & Chawarska, K. (2014). Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. Biological Psychiatry, 75(3), 231–237. https://doi.org/10.1016/j.biopsych.2013.07.009.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). Vineland adaptive behavior scales, second edition (Vineland II) (2nd ed.). San Antonio, TX: Pearson.

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). Vineland adaptive behavior scales, third edition (Vineland-3) (3rd ed.). San Antonio, TX: Pearson.

Speer, L. L., Cook, A. E., McMahon, W. M., & Clark, E. (2007). Face processing in children with autism. Autism, 11(3), 265–277. https://doi.org/10.1177/1362361307076925.

Spence, S. J., & Thurm, A. (2010). Testing autism interventions: Trials and tribulations. Lancet (London, England), 375 (9732), 2124–2125. https://doi.org/10.1016/S0140-6736(10)60757-X.

Tenenbaum, E. J., Amso, D., Abar, B., & Sheinkopf, S. J. (2014). Attention and word learning in autistic, language delayed and typically developing children. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.00490.

Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2013). Increased focus on the mouth among infants in the first year of life: A longitudinal eye-tracking study. Infancy, 18(4), 534–553. https://doi.org/10.1111/j.1532-7078.2012.00135.x.

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. Journal of Child Language, 42(06), 1173–1190. https://doi.org/10.1017/S0305000914000725.

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. Journal of Experimental Child Psychology, 169, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002.

Umbricht, D., del Valle Rubido, M., Hollander, E., McCracken, J. T., Shic, F., Scahill, L., … Fontoura, P. (2017). A single dose, randomized, controlled proof-of-mechanism study of a novel vasopressin 1a receptor antagonist (RG7713) in high-functioning adults with autism spectrum disorder. Neuropsychopharmacology, 42(9), 1914–1923. https://doi.org/10.1038/npp.2016.232.

Vernon, T., Holden, A., Barrett, A., Bradshaw, J., Ko, J., McGarry, E., … German, T. (2019). A pilot randomized clinical trial of an enhanced pivotal response treatment approach for young children with autism: The PRISM model. Journal of Autism and

Developmental Disorders. https://doi.org/10.1007/s10803-019-03909-1.

Vernon, T., Koegel, R. L., Dauterman, H., & Stolen, K. (2012). An early social engagement intervention for young children with autism and their parents. Journal of Autism and Developmental Disorders, 42(12), 2702–2717. https://doi.org/10.1007/s10803-012-1535-7.

Yang, D., Allen, T., Abdullahi, S. M., Pelphrey, K. A., Volkmar, F. R., & Chapman, S. B. (2017). Brain responses to biological motion predict treatment outcome in young adults with autism receiving virtual reality social cognition training: Preliminary findings. Behaviour Research and Therapy, 93, 55–66. https://doi.org/10.1016/j.brat.2017.03.014.

Yang, D., Pelphrey, K. A., Sukhodolsky, D. G., Crowley, M. J., Dayan, E., Dvornek, N. C., … Ventola, P. (2016). Brain responses to biological motion predict treatment outcome in young children with autism. Translational Psychiatry, 6(11), e948–e948. https://doi.org/10.1038/tp.2016.213.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). Preschool language scale (5th ed.). San Antonio, TX: Pearson.

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Supporting Information Table S1.** Intercorrelations between Clinical Measures