Child and Family Characteristics Moderate Agreement between Caregiver and Clinician Report of Autism Symptoms

Emily Neuhaus, Theodore P. Beauchaine, Raphael A. Bernier, and Sara J. Webb

Rates of autism spectrum disorder (ASD) and age at first diagnosis vary considerably across the United States and are moderated by children's sex, race, ethnicity, and availability of services. We additionally suggest that degree of caregiver–clinician agreement on ASD symptoms may play a role in ASD assessment. Since gold standard ASD assessment integrates caregiver-reported developmental history with clinician observations, differential agreement between reporters across demographic groups may contribute to a host of detrimental outcomes. Here, we investigate whether caregiver–clinician agreement on ASD symptoms varies according to child and family characteristics. Comprehensive data from 2,759 families in the Simons Simplex Collection were analyzed. Linear models were created with caregiver reports predicting clinician reports, and moderating effects of child characteristics and family factors were examined. Poorer reporter correspondence was observed when children had higher IQ scores, stronger adaptive behavior, and more behavioral difficulties. Greater disagreement was also associated with African American racial status (for younger children), lower household income, and paternal social difficulties (for older children). Children's biological sex did not moderate caregiver–clinician agreement. Marked disagreement between caregivers and clinicians could lead to suboptimal or insufficient intervention services and negative experiences for families throughout development. Such families may also be less likely to qualify for research studies, and therefore be underrepresented in the ASD literature. Modified assessment procedures may be required to improve assessment accuracy and family experiences. ***Autism Res** 2018, 11: 476–487.* © 2017 International Society for Autism Research, Wiley Periodicals, Inc.

Lay Summary: Evaluation of autism spectrum disorder (ASD) incorporates both caregiver and clinician perspectives of symptoms, and disagreement between these perspectives could lead to poorer outcomes for families. Using data from 2,759 families, we show that caregiver–clinician agreement on ASD symptoms is poorer for children with higher cognitive and adaptive skills, more behavioral difficulties, lower household income, and African American racial status. These children may be at higher risk for misdiagnosis, poorer family experiences during evaluations, and poorer representation in ASD research.

Keywords: autism spectrum disorder; diagnosis; autism diagnostic observation schedule

Introduction

Autism spectrum disorder (ASD) is an early-emerging neurodevelopmental disorder defined by deficits in social communication and patterns of restricted or repetitive behaviors or interests [American Psychiatric Association, 2013]. Although the average age of ASD diagnosis in the United States is 4–5 years [CDC, 2014; Shattuck et al., 2009], many parents report social or communication concerns as early as age 6 months [Bolton, Golding, Emond, & Steer, 2012]. Recent estimates place ASD prevalence at 1.5% of the population [Christensen et al., 2016], but both age and rates of ASD diagnoses vary tremendously across different demographic groups. Perhaps best documented are sex differences, with a male-to-female ratio of 4.5:1 in ASD prevalence [Christensen et al., 2016], and an older mean age of diagnosis among girls compared to boys [Shattuck et al., 2009]. Similarly, rates of ASD differ across racial and ethnic groups, with African American and Latino children being less likely to receive an ASD diagnosis than White children [Christensen et al., 2016]. To date, such discrepancies have been attributed to true differences in ASD prevalence across subgroups [Werling & Geschwind, 2013], disparities in access to appropriate evaluation services [Zuckerman et al., 2014], and phenotypic differences that lead to under-identification among some children [Hiller, Young, & Weber, 2014; Kopp & Gillberg, 2011].

Like many psychiatric diagnoses, ASD is defined by a constellation of behavioral symptoms observed over the course of development. Initial diagnosis and subsequent

From the Seattle Children's Research Institute, Center on Child Health, Behavior, and Development, Seattle, Washington (E.N., S.J.W.); The Ohio State University, Columbus, Ohio (T.P.B.); University of Washington, Psychiatry and Behavioral Science, Seattle, Washington (R.A.B., S.J.W.)

Received March 06, 2017; accepted for publication November 22, 2017

Address for correspondence and reprints: Emily Neuhaus, Seattle Children's Research Institute, PO Box 5371, MS CW8-6, Seattle, WA, 98145. E-mail: eneuhaus@uw.edu

Published online 18 December 2017 in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/aur.1907

evaluations of skills and deficits therefore rely upon observations and reports from multiple informants. Gold standard ASD evaluation integrates developmental information from caregivers with thorough observational assessments by trained clinicians [Ozonoff, Goodlin-Jones, & Solomon, 2005]. Underlying this process is an understanding that caregivers and clinicians provide both *unique* and *overlapping* information necessary for understanding individuals' unique profiles of strengths and difficulties [Risi et al., 2006]. Although perspectives across reporters may match closely for some families, others may experience more striking differences between perspectives—for example, if a clinician perceives a difficulty that a family views as normative, or vice versa. Such divergences between reporters' perspectives may place at least some families at risk for detrimental consequences across the lifespan. Perhaps most importantly, families for whom there is poorer agreement between reporter perspectives may be at elevated risk for (a) inaccurate diagnoses and inappropriate service provision; (b) poorer family experiences with healthcare professionals; and (c) underrepresentation in ASD research. Research in this area has been extremely limited to date, with very little exploration as to how convergence or divergence between caregivers' and clinicians' perspectives might influence individual and family outcomes. Thus, our understanding of this issue is far from complete.

Since demographic factors relate to ASD diagnosis, they may also relate to perceptions and interpretations of ASD-related behaviors more broadly, with some families more likely than others to experience significant divergences in perspectives. Two common and well-regarded measures of caregiver and clinician report of ASD symptoms are the Autism Diagnostic Interview–Revised [ADI-R; Lord, Rutter, & Le Couteur, 1994] and the Autism Diagnostic Observation Schedule [ADOS; Lord, Rutter, DiLavore, & Risi, 2003], respectively. Early work with these measures indicated fair agreement for both ASD symptom levels [Risi et al., 2006] and diagnoses [de Bildt et al., 2004], with correlations between caregiver and clinician reports in the range of $r = 0.6$–0.7 [Le Couteur, Haden, Hammal, & McConachie, 2008; Risi et al., 2006]. However, studies also reveal marked variability in agreement across research samples [Gray, Tonge, & Sweeney, 2008; Ventola et al., 2006], from a low of $r = 0.28$ to a high of $r = 0.95$ based on sample characteristics [Risi et al., 2006]. Child characteristics such as cognitive level appear to matter, as studies have suggested lower correlations between ADOS and ADI-R scores, and poorer sensitivity of clinical cut-offs on some measures, in samples with intellectual disability than in samples with average cognitive ability [Havdahl et al., 2016; Risi et al., 2006]. Behaviorally, significant ADHD symptoms appear to increase the age at which children are diagnosed with ASD [Yee & Millichap, 2015]. More generally, possible demographic moderators of reporter agreement include socioeconomic status and child race/ethnicity—both of which are associated with ability and desire to pursue ASD evaluations and the age at which an ASD diagnosis is received [Christensen et al., 2016; Emerson, Morrell, & Neece, 2016; Mazurek et al., 2014; Zuckerman et al., 2014].

Following from this discussion, our goal was to explore potential child- and family-level moderators of agreement between caregiver and clinician reports of ASD symptoms. We operationalize caregiver report with the Autism Diagnostic Interview - Revised [ADI-R; Lord et al., 1994], and clinician report with the Autism Diagnostic Observation Schedule [ADOS; Lord et al., 2003]. Although diagnostic processes in the clinical and research communities vary and may not routinely include both measures, these instruments are standardized, well-researched, and frequently used in both clinical and research settings. Thus, they represent a solid foundation from which to explore caregiver–clinician agreement.

Method

Participants

Data were obtained through the Simons Simplex Collection (SSC), a national consortium spanning 12 sites across the United States [Fischbach & Lord, 2010]. The study protocol was approved by each site's human subjects division, and all families provided informed consent prior to participating. Families were enrolled if they had one child with ASD between ages 4 and 18 years, and no history of diagnosed or suspected ASD in their immediate or extended family. Exclusionary criteria included a nonverbal mental age below 18 months, presence of known genetic conditions (e.g., Fragile X), histories of neurological disease or significant head injury, significant sensory or motor impairment, extensive pregnancy or birth complications, gestational age below 36 weeks at birth, birth weight under 2,000 g, and a primary language other than English.

This procedure yielded a sample of $N = 2{,}759$ children and adolescents (375 female) with a mean age of 108.3 months ($SD = 42.8$, range 48–216). Research reliable clinicians diagnosed children and adolescents with ASD using CPEA criteria [Lainhart et al., 2006], which includes the ADOS, ADI-R, and expert clinical judgment, with rigorous reliability procedures within and between SSC sites. Severity of symptoms varied across the sample, with a mean calibrated severity score [Gotham, Pickles, & Lord, 2009] of 7.39 ($SD = 1.7$, range 4–10). Self-reported racial/ethnic backgrounds were as

Table 1. Descriptive Statistics for Participating Families

	Mean (SD)	Range
Child variables		
Age (months)	108.3 (42.8)	48–216
ADI-R current behavior total score	27.47 (10.1)	1–60
ADOS total score	15.19 (5.2)	7–28
ADOS calibrated severity score	7.44 (1.7)	4–10
Verbal IQ score	78.04 (31.3)	5–167
Nonverbal IQ score	84.54 (26.2)	9–161
Full scale IQ score	81.17 (28.0)	7–167
Vineland adaptive behavior composite	73.14 (12.1)	27–115
CBCL behavior problems total	62.35 (9.1)	27–92
Parent variables		
Maternal age at child's birth (years)	31.35 (4.96)	16.08–45.25
Paternal age at child's birth (years)	33.50 (5.72)	17.08–57.58
Maternal BAPQ	85.89 (20.9)	36–169
Paternal BAPQ	96.38 (21.8)	40–179
Maternal SRS-ARV	29.56 (20.4)	0–135
Paternal SRS-ARV	29.78 (22.7)	0–172

ADI-R, Autism Diagnostic Interview, Revised; ADOS, Autism Diagnostic Observation Schedule; CBCL, Child Behavior Checklist; BAPQ, Broader Autism Phenotype Questionnaire; SRS-ARV, Social Responsiveness Scale - Adult Research Version.

follows: African American (4.0%), Asian (4.0%), Native American or Hawaiian (0.3%), White (78.5%), other (4.5%), more than one race (7.8%). An additional 0.8% declined to disclose their racial/ethnic identity. See Table 1 for participant characteristics.

Measures

Diagnostic measures. Consistent with SSC data collection protocols, families typically completed the ADI-R and ADOS during a single visit and so reflect concurrent assessment of the child's ASD symptoms by both reporters. Both the ADI-R and ADOS demonstrate high sensitivity and specificity in identifying ASD [Gotham, Risi, Pickles, & Lord, 2007; Lord et al., 1997].

Caregiver reports of ASD symptoms were assessed with the ADI-R, a semi-structured interview that evaluates current and historical difficulties in domains of social development, communication, and restricted or repetitive behaviors and interests. It should be noted that, as a semi-structured interview, the ADI-R does require some interpretation by clinicians regarding caregivers' comments and observations. The ADI-R yields scores in each of three domains: social, communication, and restricted/repetitive behavior and interests. Scores across domains were summed for a total ADI-R score using the published "current behavior" algorithm. The number of items contributing to this algorithm differs slightly for children ages 4 years through 9 years, 11 months versus children ages 10 years and older. To account for this, ADI-R total scores were computed using age-appropriate algorithms (younger group $n = 1{,}783$, older group $n = 976$). Within each age group, we also summed subscores for the social and communication domains to yield a single score for social-communication difficulties. This procedure allowed for ADI-R total scores, as well as domain-specific social-communication and restricted/repetitive subscores.

Clinician report of ASD symptoms was measured with the ADOS, a semi-structured play-based assessment of the same three areas of functioning. For purposes of this study, ADOS total scores were computed using the revised algorithm [Gotham et al. 2007], which sums 14 items across domains of communication, reciprocal social interaction, and restricted/repetitive behaviors. Although the precise items contributing to the algorithm score vary across modules, the number of contributing items is equivalent across them. This algorithm also yields subscores reflecting social-communication difficulties (social affect subtotal) and restricted/repetitive behaviors.

Child characteristics. Child characteristics were identified as potential moderators of caregiver–clinician agreement. Demographic factors included children's age at assessment, biological sex, birth order, and caregiver-reported race (with sufficient sizes to examine effects for families reporting African American, Asian, and White race). Behavioral characteristics included full-scale IQ, assessed with the Differential Abilities Scale, 2nd Edition [Elliott, 2007], the Wechsler Abbreviated Scale of Intelligence [Wechsler, 1999] or Mullen Scales of Early Learning [Mullen, 1997]; adaptive behavior, assessed with the composite standard score from the Vineland Adaptive Behavior Scales, 2nd Edition [Vineland-II; Sparrow, Cicchetti, and Balla, 2005]; and behavioral difficulties, assessed with the Total Problems *T*-score from the age-appropriate version of the Child Behavior Checklist [Achenbach, 1991].

Family characteristics. Family characteristics included annual household income (dichotomized below/above $80,000), parents' ages at child's birth, and parents' education (dichotomized below/above college completion). Finally, parents' social functioning was assessed with the Broader Autism Phenotype Questionnaire [Hurley, Losh, Parlier, Reznick, & Piven, 2007], which yields a self-report estimate, and with the total score from the Social Responsiveness Scale—Adult Research Version [SRS-ARV; Constantino & Todd, 2005] as reported by the parent's partner. With the exception of annual household income and parental education, all variables were entered as continuous variables.

Analytic Approach

We chose to first examine simple correspondences between reports of ASD symptoms using a correlational

Table 2. Main and Interactive Effects of Putative Moderators on Caregiver–Clinician Agreement for the Younger Group

	Main effect of moderator		Main effect of ADI-R		Interactive effect		Variance
Moderator	β	t-value	β	t-value	β	t-value	Model adj. r^2
Child age	−0.081	−1.138	0.348	3.782***	0.045	0.422	0.156
Child sex	−0.081	−1.249	0.260	3.589***	0.183	1.905	0.156
Child birth order	0.109	1.566	0.441	8.927***	−0.102	−1.179	0.155
Child race/ethnicity							
African American	0.187	2.597**	0.401	18.068***	−0.142	−1.966*	0.157
Asian	0.038	0.586	0.392	17.509***	0.002	0.023	0.154
White	−0.167	−2.397*	0.321	6.84***	0.137	1.695	0.157
Child full scale IQ	−0.256	−3.526***	0.438	6.127***	−0.246	−3.029**	0.348
Child adaptive behavior	0.025	0.391	1.018	7.490***	−0.692	−5.885***	0.245
Child behavior problems	−0.003	−0.042	0.714	5.236***	−0.346	−2.135*	0.170
Household income	−0.041	−0.572	0.329	4.342***	0.088	0.903	0.154
Maternal age at birth	−0.079	−1.116	0.143	1.007	0.278	1.781	0.156
Paternal age at birth	−0.083	−1.220	0.174	1.383	0.248	1.773	0.155
Maternal education	−1.07	−1.529	0.254	3.254**	0.183	1.864	0.155
Paternal education	−0.033	−0.481	0.322	4.220***	0.098	0.991	0.154
Maternal BAPQ	−0.193	−2.807**	0.286	3.225**	0.152	1.337	0.165
Paternal BAPQ	−0.060	−0.842	0.390	3.858***	0.006	0.051	0.156
Maternal SRS-ARV	−0.044	−0.649	0.414	10.932***	−0.041	−0.528	0.159
Paternal SRS-ARV	0.014	0.217	0.436	12.572***	−0.110	−1.520	0.161

BAPQ, broader autism phenotype questionnaire (self-report; Hurley et al., 2007); SRS-ARV, social responsiveness scale—adult research version (partner-report; Constantino & Todd, 2005).

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

approach. This was followed by regression-based moderator analyses to test potential moderators of caregiver–clinician correspondence. Because the ADI-R Current Behavior algorithm incorporates differing numbers of items depending on children's ages, younger (ages 4:0–9:11) and older (10 years and older) groups were analyzed using separate but parallel approaches. For each age group, caregiver–clinician correspondence was assessed through a series of linear regression models in two steps:

Step 1: Individual moderator models. Each putative moderator (e.g., child IQ) was entered into a separate regression model along with the ADI-R total score and the Moderator × ADI-R interaction term as predictors of ADOS total scores. In this step, significant interaction terms indicate moderation of caregiver–clinician agreement.

Step 2: Combined moderators models. Interaction terms that were significant in Step 1 were included (along with their respective main effects) together in a combined regression model to compare their relative contributions in predicting ADOS total scores. In this second step, significant interaction terms indicate a moderator that maintains its effect above-and-beyond other significant effects.

Since our focus is on identifying factors that affect agreement between reporters, we focus primarily on moderating effects in which child/family characteristics interact with ADI-R Current scores to predict ADOS total scores. Consistent with this approach, we devote less attention to main effects (i.e., child/family characteristics that predict ADOS total scores but do not interact with ADI-R scores) within the following models.

Results

Correlations between Reports of ASD Symptoms

Caregiver- and clinician-reported ASD symptoms were correlated significantly for both younger, $r = 0.39$, $P < 0.001$, and older, $r = 0.40$, $P < 0.001$, participants. Domain-specific correlations between the ADI-R and ADOS suggested that the two reports were moderately correlated with respect to social-communication difficulties (younger: $r = 0.38$, $P < 0.001$; older: $r = 0.39$, $P < 0.001$), with smaller correlations with respect to restricted/repetitive behavior (younger: $r = 0.16$, $P < 0.001$; older: $r = 0.27$, $P < 0.001$). Comparisons of correlation strengths indicated that reporter correspondence was significantly greater for social-communication difficulties compared to restricted/repetitive behaviors in both the younger, $z = 7.28$, $P < 0.001$, and older, $z = 3.00$, $P < 0.01$, groups.

Moderation Analyses

Entered alone, caregiver reports of ASD symptoms on the ADI-R current scores predicted clinician assessments with the ADOS total score for both younger, $\beta = 0.39$,

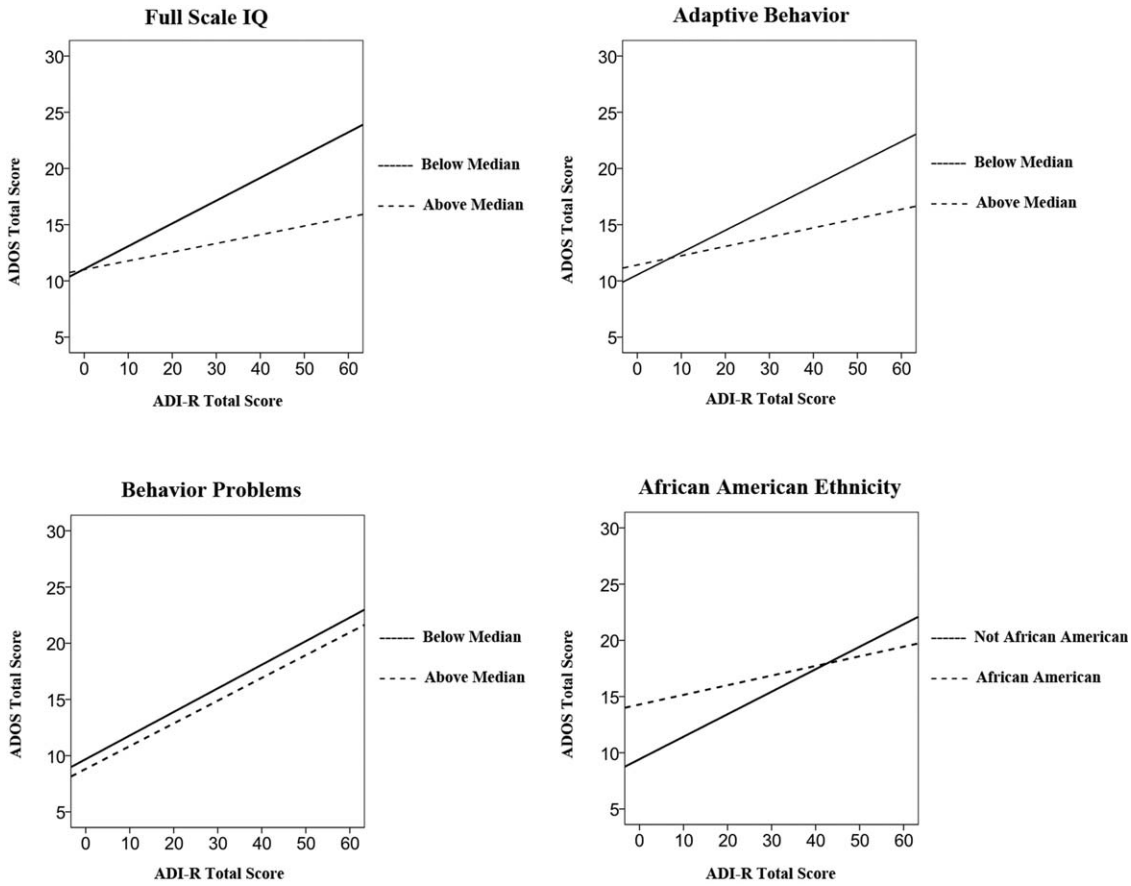

Figure 1. For younger children, caregiver–clinician agreement was strongest when children had lower IQ scores, lower adaptive behavior, fewer behavioral difficulties, and were not African American. Note that variables were entered as continuous moderators and dichotomized only for purposes of visual display.

$t = 27.7$, $P < 0.001$, and older children, $\beta = 0.40$, $t = 17.3$, $P < 0.001$. ADI-R current scores accounted for 15.4%, $F(1, 1{,}779) = 324.5$, $P < 0.001$, and 15.5%, $F(1, 896) = 165.97$, $P < 0.001$, respectively, of the variance in ADOS scores, suggesting influence from additional factors. These factors were subsequently examined through moderation analyses.

Moderators: Younger Children

Step 1: Individual moderators models. For the younger age group (4:0–9:11), we observed main effects of child IQ, $\beta = -0.26$, $t = -3.53$, $P < 0.001$, African American racial status, $\beta = 0.19$, $t = 2.60$, $P = 0.009$, White racial status, $\beta = -0.17$, $t = -2.40$, $P = 0.017$, and mother BAPQ scores, $\beta = -0.19$, $t = -2.81$, $P = 0.005$. Higher ADOS total scores (greater autism severity) were associated with lower IQ, parent-report of African American racial status, and lower maternal BAPQ scores (see Table 2).

Interactions were significant for IQ scores, $\beta = -0.25$, $t = -3.03$, $P = 0.002$, Vineland adaptive behavior scores, $\beta = -0.69$, $t = -5.89$, $P < 0.001$, CBCL total scores, $\beta = -0.35$, $t = -2.14$, $P = 0.03$, and African American racial status, $\beta = -0.14$, $t = -1.97$, $P = 0.049$. Thus, we observed moderating effects of child IQ, adaptive behavior, behavioral difficulties, and race, such that caregiver–clinician agreement of ASD symptom levels was strongest when children had lower IQ scores, lower adaptive behavior, fewer behavioral difficulties, and were not African American (see Fig. 1).

Step 2: Combined moderators model. For the younger group, the subsequent model containing IQ, adaptive behavior, behavior problems, and African American status plus their respective ADI-R interaction terms accounted for 36.1% of the variance in ADOS total scores, $F(9,1{,}774) = 112.4$, $P < 0.001$. For this model, the interaction between ADI-R scores and behavior problems remained significant, $\beta = -0.41$, $t = -2.71$, $P = 0.007$, as did the interaction between ADI-R and African American race, $\beta = -0.17$, $t = -2.62$, $P = 0.009$. Thus, for younger children, behavioral difficulties and African American race emerged as the most robust moderators of caregiver–clinician symptom agreement, with stronger agreement between caregivers

Table 3. Main and Interactive Effects of Putative Moderators on Caregiver–Clinician Agreement for the Older Group

	Main effect of moderator		Main effect of ADI-R		Interactive effect		Variance
Moderator	β	t-value	β	t-value	β	t-value	Model adj. r^2
Child age	−0.108	−1.105	0.192	1.046	0.232	1.123	0.155
Child sex	−0.132	−1.422	0.260	2.577*	0.193	1.408	0.156
Child birth order	−0.013	−0.126	0.329	4.213***	0.115	0.903	0.160
Race/ethnicity							
African American	−0.033	−0.347	0.386	12.407***	0.150	1.564	0.168
Asian	0.041	0.407	0.390	12.433***	0.024	0.238	0.158
White	0.023	0.239	0.478	7.001***	−0.161	−1.459	0.168
Child full scale IQ	−0.200	−2.156*	0.480	6.233***	−0.334	−3.393**	0.388
Child adaptive behavior	−0.036	−0.405	0.919	5.545***	−0.644	−4.493***	0.297
Child behavior problems	0.101	1.111	1.163	5.237***	−0.844	−3.288**	0.194
Household income	−0.173	−1.740	0.184	1.684	0.283	2.039*	0.158
Maternal age at birth	0.067	0.700	0.348	1.752	0.054	0.250	0.162
Paternal age at birth	0.004	0.045	0.191	0.962	0.231	1.072	0.166
Maternal education	−0.032	−0.338	0.369	3.608***	0.037	0.274	0.154
Paternal education	−0.079	−0.830	0.329	3.184**	0.094	0.681	0.154
Maternal BAPQ	−0.134	−1.380	0.319	2.427*	0.111	0.660	0.159
Paternal BAPQ	0.022	0.232	0.503	3.440**	−0.130	−0.745	0.156
Maternal SRS-ARV	−0.001	0.011	0.439	8.150***	−0.091	−0.806	0.160
Paternal SRS-ARV	0.075	0.784	0.485	9.619***	−0.216	−2.021*	0.169

BAPQ, broader autism phenotype questionnaire (self-report; Hurley et al., 2007); SRS-ARV, social responsiveness scale—adult research version (partner-report; Constantino & Todd, 2005).

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

and clinicians for children with fewer behavioral difficulties and without caregiver-reported African American ethnicity.

Moderators: Older Children and Adolescents

Step 1: Individual moderators models. For the older group (10:0–18:0), main effects were significant only for IQ, $\beta = -0.20$, $t = -2.16$, $P = 0.031$. Higher ADOS scores (greater severity) corresponded with lower IQ among older participants (see Table 3).

Interactions indicating moderating effects were significant for child IQ, $\beta = -0.33$, $t = -3.39$, $P = 0.001$, adaptive behavior, $\beta = -0.64$, $t = -4.49$, $P < 0.001$, behavior problems, $\beta = -0.84$, $t = -3.29$, $P = 0.001$, family household income, $\beta = 0.28$, $t = 2.04$, $P = 0.042$, and father social difficulties, $\beta = -0.22$, $t = -2.02$, $P = 0.044$. Caregiver–clinician agreement was strongest for children who had lower IQ scores, lower adaptive behavior, fewer behavioral difficulties, higher family income, and fathers with fewer social difficulties (see Fig. 2).

Step 2: Combined moderators model. For the older age group, the combined model again contained child IQ, adaptive behavior, and behavior problem scores, but also family income and father social difficulties, plus their corresponding ADI-R interaction terms. This model accounted for 40.1% of the variance in ADOS scores, $F(11,840) = 52.1$, $P < 0.001$. Of the moderators entered, behavior problems, $\beta = -0.43$, $t = -1.79$, $P = 0.074$, and father social difficulties, $\beta = -0.19$, $t = -1.92$, $P = 0.055$, approached significance, but no moderators emerged as statistically significant when entered simultaneously. Thus, although agreement between reporters was moderated by child IQ, adaptive behavior, behavioral difficulties, family income, and father social difficulties, none of these moderators arose as significantly more influential than the rest when compared directly against one another.

Discussion

Our study is among the first to examine correspondence and divergence between caregiver and clinician reports of ASD symptoms. We identify several child and family characteristics that moderate agreement between these reporters' perspectives. For both younger (9 years and under) and older (10 years and older) children, caregivers and clinicians showed significant but modest correspondences between their reports of ASD symptoms, with caregiver reports accounting for approximately 15% of variance in clinician assessments. Caregiver–clinician agreement was moderated by a number of child and family factors that increased shared variance to approximately 40%. Moderating effects of child IQ, adaptive behavior, and behavioral difficulties were observed in both groups, such that children with lower IQs, lower adaptive behavior, or fewer behavioral difficulties had better correspondence

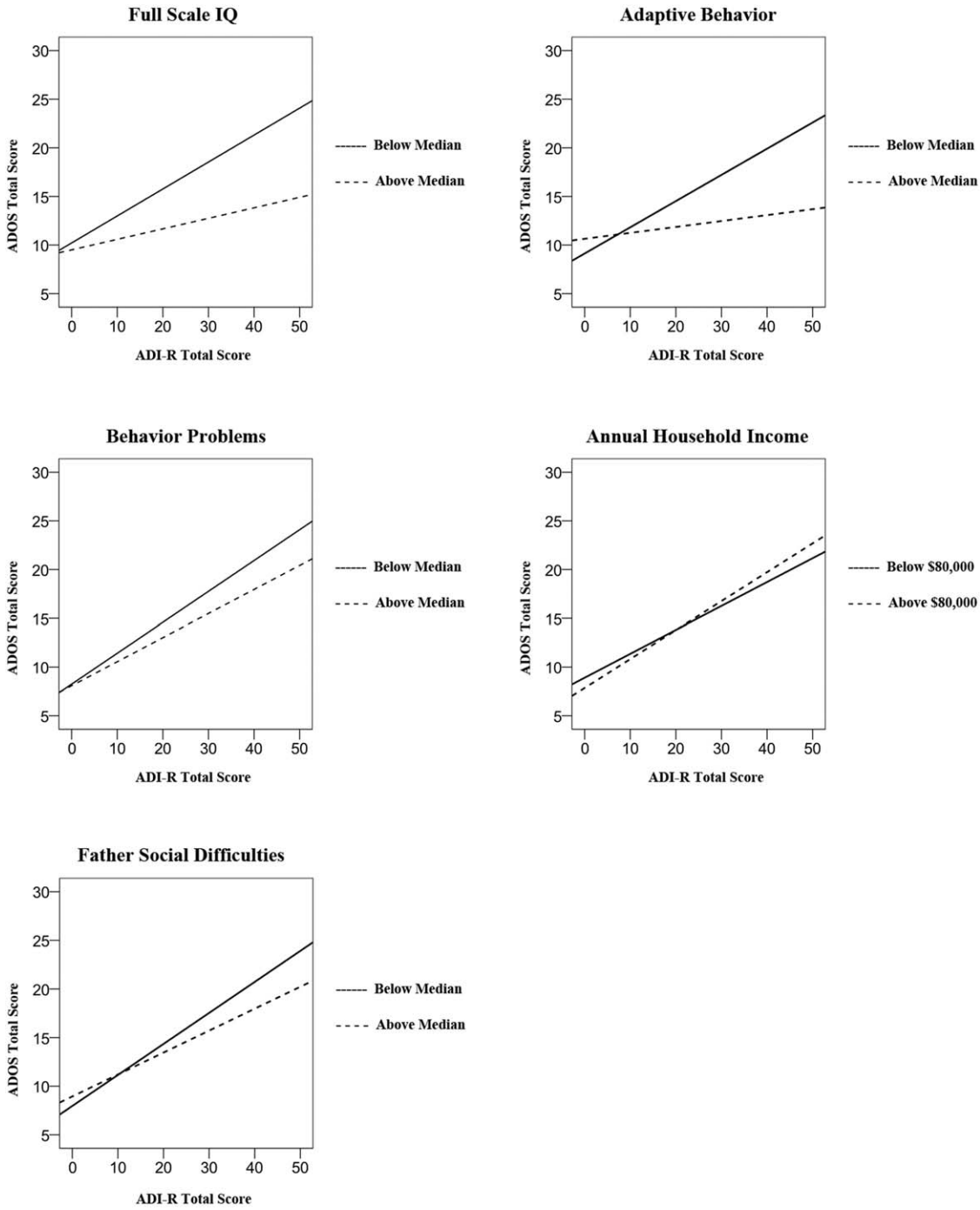

Figure 2. For older children, caregiver–clinician agreement was strongest when children had lower IQ scores, lower adaptive behavior, fewer behavioral difficulties, higher family income, and fathers with fewer social difficulties. Note that household income was entered as a dichotomous variable, while other moderators shown here were entered as continuous variables and dichotomized only for purposes of visual display.

between reporters. From a practical standpoint, children with the converse—higher IQs, stronger adaptive skills, and more behavioral concerns—were more likely to have poor agreement between caregiver reports and clinician assessments and consequently may be at elevated risk for detrimental outcomes. To the extent that our findings extend beyond symptom description to include ASD diagnostic outcomes, a point to which we return below, these findings are consistent with concerns that current assessment practices might miss "high-functioning" individuals with ASD (those with strong cognitive and adaptive skills) as well as those

with comorbid symptoms such as anxiety, inattention, or hyperactivity that might complicate differential diagnosis [Lai & Baron-Cohen, 2015; Yee & Millichap, 2015].

Beyond shared moderators, other factors were unique to our two age groups. For younger children, African American race was associated with lower agreement—an effect that remained over-and-above other moderators. Previous literature documents that African American children eventually diagnosed with ASD receive their diagnosis an average of 18 months later than White children [Mandell, Listerud, Levy, & Pinto-Martin, 2002], undergo more visits with clinicians before receiving a diagnosis [Mandell, Ittenbach, Levy, & Pinto-Martin, 2007], and more often receive a different diagnosis (e.g., ADHD) prior to ASD [Mandell et al., 2007]. Our finding of poorer symptom agreement between caregivers and clinicians suggests one mechanism through which these discrepancies arise, placing African American children with ASD at a disadvantage for identifying areas of need and obtaining appropriate service referrals. With respect to initial diagnosis, it is difficult to overstate the implications of these findings—overwhelming evidence underscores the critical importance of early and intensive intervention for children with ASD [Dawson et al., 2010; Rogers et al., 2014], and such delays in diagnosis (and thus ASD-specific care) have meaningful effects on trajectories and ultimate outcomes of children's social, communication, and cognitive skills [Dawson et al., 2010; Dawson et al., 2012].

A different set of family factors emerged as significant for older children and adolescents, perhaps suggesting an age-related shift in relative effects of different moderators. In our older subgroup, poorer agreement was associated with lower family income. This may reflect socioeconomic influences on parents' perspectives on behaviors suggestive of ASD. For instance, although speculative, families with fewer economic resources might have limited access to routine medical care that would otherwise shape their expectations for their child's development and thus their interpretation of behavior. Regardless of the cause, findings related to family income and child race are concerning, as they likely add a layer of complexity to disparities in accessing specialized evaluation, research, and support services for ASD, even at later ages.

We also observed an effect of fathers' social difficulties in our older group, such that better agreement emerged when fathers were more competent socially, but we did not observe a comparable effect for mothers. Interpretation of this effect is tentative without knowledge of which parent provided ADI-R data; in some families, mothers reported on both child and father characteristics, with potential common reporter variance. Nonetheless, findings regarding fathers' social difficulties present an interesting parallel to recent work revealing that fathers—but not mothers—underreport their own broader phenotype traits [Sasson, Faso, Parlier, Daniels, & Piven, 2014]. Such findings highlight the importance of incorporating multiple sources of information across home and school settings (e.g., teachers) into diagnostic decisions, as each reporter's social experiences and skills provide a unique interpretive lens for a child's behavior.

Child age, sex, and birth order did not moderate agreement between reporters. These null findings are surprising, particularly regarding sex. Since its recognition [e.g., Kanner, 1943], ASD has been diagnosed more frequently in boys than girls, leading some researchers to suggest that current referral and diagnostic processes may overlook girls with significant symptoms [Lai & Baron-Cohen, 2015], perhaps due to sex-specific ASD phenotypes [Kopp & Gillberg, 2011], patterns of comorbidity [Hiller et al., 2014], molecular genetic substrates [Halladay et al., 2015], or behavioral expectations [Shattuck et al., 2009]. Our current findings provide no evidence of an influence of sex in caregiver–clinician agreement.

Limitations

We suspect that null findings, particularly with regard to biological sex, are due in part to the nature of our sample. By design, the final SSC sample included only children/adolescents who met strict and rigorous research inclusion standards, meeting ASD diagnostic thresholds on both the ADOS and the ADI-R. This approach yields a sample with high confidence of true ASD, but also introduces a potential floor effect because the range of ASD severity is somewhat truncated. As a result, we likely underestimate moderating influences of some variables. Our findings are conservative in that sense, and a broader sample including children without an established ASD diagnosis might reveal additional moderators.

Similarly, because of sample characteristics, our findings cannot speak directly to factors affecting initial diagnosis of ASD, given that children with wider discrepancies between reporters' perspectives are less likely to have met the research inclusion diagnostic thresholds and consequently would have been excluded from the sample. Our analyses are best interpreted as relating to variability in symptom description within a sample that had strong agreement on child diagnosis. The implications for diagnostic outcomes are more speculative and will await exploration in additional, community-based samples where individuals might meet diagnostic criteria on one but not both measures. It is

also the case that, although frequently used in research settings, the ADI-R and the ADOS are not used universally in clinical settings, due in part to constraints on financial resources, training commitments, and administration time. Thus, we cannot determine whether our results generalize to other caregiver and clinician assessment tools that are more common to clinical settings, particularly those which vary by format (e.g., questionnaire vs. interview, written vs. verbal) or length.

The simplex nature of our sample may also affect generalizability of conclusions. Consistent with the goals of the SSC project more broadly [Fischbach & Lord, 2010], families in the SSC dataset have only one child with ASD, presumed to have ASD of de novo genetic etiology. Moderating factors may differ in multiplex families (those in which a second child has ASD), as parents' familiarity with ASD or family history of ASD might alter their perspective of their child's behavior and development, and so alter relative contributions of moderators. Future research with multiplex families will be particularly interesting given phenotypic differences in parents' social and communication skills across simplex versus multiplex families [e.g., Gerdts, Bernier, Dawson, & Estes, 2013].

Finally, we must consider possible effects of measurement issues on our findings. Divergence between caregiver and clinician reports could reflect differences in perception (as we discuss here), true variability in children's behavior across settings, or a combination of both, and our current analyses cannot differentiate between these possibilities. It may be that some children display greater variability in behavior from one setting to another, yielding caregiver and clinician reports that differ but are both accurate. This cannot be determined based on available data, although we emphasize that *both* child and parental characteristics appear to contribute to these differences. Similarly, although we use ADI-R scores for caregiver reports in this study, semi-structured interviews filter caregiver comments through clinician's interpretations. Finally, scoring procedures on both the ADOS and the ADI-R yield scores of 0, 1, or 2 for the majority of items, which are then summed according to empirically derived algorithms [Gotham et al., 2007]. Although higher algorithm scores are frequently interpreted as indicating ASD severity, correspondence between scores and ASD severity is not absolute, as higher totals may reflect a broader range of ASD symptoms rather than severity of symptoms. Our findings must be interpreted with this caveat in mind.

Implications

Because our data were derived from a research diagnostic process, the most proximal implication is for scientific knowledge of ASD. Families for whom reporters diverge on ASD symptoms may be less likely to meet study inclusion criteria and qualify for research studies and consequently may be poorly represented in the research literature. Many studies require participants to meet symptom thresholds on both caregiver-report and observational measures, and thus caregivers and clinicians must both indicate sufficient symptoms for study entry. If some subgroups within the population experience wider divergence in reporter perspective, they may be excluded systematically from research that could provide direct benefits to them, and provide more comprehensive understanding of ASD etiology and experience.

Extrapolating tentatively from our data into clinical diagnostic and referral processes, significant disagreement between caregivers and clinicians could promote a number of detrimental effects. Most immediately, marked disagreement in symptom reports at an initial evaluation could result in inaccurate diagnosis. For instance, a clinician may fail to recognize symptoms that are apparent to a caregiver, and may delay appropriate evaluation and supports [Zuckerman, Lindly, & Sinche, 2015]. Even years after initial diagnosis, vastly different views of children's skill deficits and strengths across caregivers and clinicians could result in inappropriate or inadequate supports and services that fail to meet individuals' medical, educational, or behavioral needs.

More subtly, poor caregiver–clinician agreement may influence families' experiences with healthcare professionals and attitude toward supports throughout development. Limited existing research addressing discrepancies between reporters suggests that clinicians might prioritize their own ADOS observations over caregiver reports [Risi et al., 2006]. Evaluation for ASD is inherently stressful for families [Crane, Chester, Goddard, Henry, & Hill, 2015], and stress is likely exacerbated when caregivers and clinicians disagree about the presence or severity of ASD symptoms or proposed supports/interventions. Consistent with this, qualitative evidence reveals that pediatrician invalidation of parental concerns related to initial ASD diagnosis constitutes a marked source of stress and may affect parent mental health and trust in the provider [Zuckerman et al., 2014]. Furthermore, parents' experiences during the diagnostic process and satisfaction with its outcome likely influence their reception of the ASD diagnosis and their willingness to pursue treatment recommendations [Reed & Osborne, 2012].

Future Directions

Following from this discussion, we highlight a number of important avenues for further examination. First,

other potential moderators such as medical comorbidities or known genetic events should be considered, as families with ongoing medical or genetic issues may monitor and interpret their child's behavior differently from those without such awareness. In addition, caregiver characteristics such as mental health concerns (e.g., depression, anxiety), language proficiency, familiarity with child development, or comfort interacting with medical personnel more broadly might influence how caregivers perceive and describe strengths and concerns related to ASD. Recent research suggests that language match between parents and children may be an important consideration as well [Vanegas, Magaña, Morales, & McNamara, 2016], as parents who have a different primary language from their child (e.g., parents for whom Spanish is primary but whose children speak English as their primary language) may interpret or report their child's social and communication skills differently, tending to report greater skill and fewer deficits compared to clinician observation [Vanegas et al., 2016]. Findings such as this underscore the need to consider individual differences among family members in addition to family-level factors, as well as to consider how these influence instrument validity [Magaña & Smith, 2013; Vanegas et al., 2016].

From a practical standpoint, critical questions remain regarding the degree to which moderators of caregiver–clinician agreement influence referral and diagnosis, and alter ongoing care and treatment over the course of development. Despite their potential importance, these questions have not been addressed directly in the literature. Yet, for families, agreement with their healthcare providers—and variables affecting that agreement—may be critical to their reception of and adjustment to an ASD diagnosis, motivation to pursue recommended intervention services, and ability to find a medical home for their child. If agreement is systematically poorer for some families, they may be at greater risk of negative outcomes in these important areas. Identifying factors that predict such outcomes could help clinicians and researchers more easily identify families at risk and adjust their clinical approach accordingly. For example, clinicians might shift the relative weight they attribute to a family's report of their child's behavior versus their own observations, engage in additional conversation about the nature and course of ASD, prioritize transparency in their assessments (e.g., having parents observe assessments with a professional who can explain procedures in the moment), or obtain information from additional observers (e.g., teachers, coaches). Similarly, researchers might consider whether their inclusion protocols bias their samples against families with characteristics identified here, such as low household income or African American ethnicity, and develop strategies to mitigate inequities. Such strategies have the potential to improve clinical outcomes as well as representativeness in research, two critical goals as we continue to enrich our understanding of causes, correlates, and outcomes related to ASD.

Ethical Considerations

All participants provided written consent as approved by the University of Washington Human Subjects Division.

Acknowledgments

Funding was provided by a Meixner Autism Speaks Postdoctoral Fellowship (Neuhaus), Grant number: 8689, National Institutes of Health (MH10028, Webb/Bernier/Neuhaus), NIMH (MH100047, Bernier/Webb), and the National Institutes of Health (DE025980, Beauchaine). We also thank all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R.B., J. Constantino, E. Cook, E. Fombonne, D. Geschwind, E. Hanson, D. Grice, A. Klin, R. Kochel, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, and E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC population dataset described in this study (SSC distribution 15, obtained from SFARI Base) by applying at https://base.sfari.org.

References

Achenbach, T.M. (1991). Manual for the child behavior checklist/4–18 and 1991 profile. Burlington, VT: University of Vermont Department of Psychiatry.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Press.

Bolton, P.F., Golding, J., Emond, A., & Steer, C.D. (2012). Autism spectrum disorder and autistic traits in the Avon longitudinal study of parents and children: Precursors and early signs. Journal of the American Academy of Child and Adolescent Psychiatry, 51, 249–260.

Christensen, D.L., Baio, J., Braun, K.V., Bilder, D., Charles, J., …. Yeargin-Allsopp, M. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012. MMWR Surveillance Summaries, 65, 1–23.

Christensen, D.L., Bilder, D.A., Zahorodny, W., Pettygrove, S., Durkin, M.S., …Yeargin-Allsopp, M. (2016). Prevalence and characteristics of autism spectrum disorder among 4-year-old children in the autism and developmental disabilities monitoring network. Journal of Developmental and Behavioral Pediatrics, 37, 1–8.

Constantino, J.N., & Todd, R.D. (2005). Intergenerational transmission of subthreshold autistic traits in the general population. Biological Psychiatry, 57, 655–660.

Crane, L., Chester, J.W., Goddard, L., Henry, L.A., & Hill, E.L. (2015). Experiences of autism diagnosis: A survey of over 1000 parents in the United Kingdom. Autism, 20, 1–10.

Dawson, G., Jones, E.J.H., 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, 1150–1159.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., ...Varley, J. (2010). A randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. Pediatrics, 125, e17–e23.

De Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F., & Minderaa, R. (2004). Interrelationship between autism diagnostic observation schedule—generic (ADOS-G), autism diagnostic interview—revised (ADI-R), and the diagnostic and statistical manual of mental disorders (*DSM-IV-TR*) classification in children and adolescents with mental retardation. Journal of Autism and Developmental Disorders, 34, 129–137.

Elliott, C.D. (2007). Differential ability scales, 2nd edition: Introductory and technical handbook. San Antonio, TX: The Psychological Corporation.

Emerson, N.D., Morrell, H.E.R., & Neece, C. (2016). Predictors of age of diagnosis for children with autism spectrum disorder: The role of a consistent source of medical care, race, and condition severity. Journal of Autism and Developmental Disorders, 46, 127–138.

Fischbach, G.D., & Lord, C. (2010). The Simons Simplex Collection: A resource for identification of autism genetic risk factors. Neuron, 68, 192–195.

Gerdts, J.A., Bernier, R., Dawson, G., & Estes, A. (2013). The broader autism phenotype in simplex and multiplex families. Journal of Autism and Developmental Disorders, 43, 1597–1605.

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, 693–705.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The autism diagnostic observation schedule: Revised algorithms for improved diagnostic validity. Journal of Autism and Developmental Disorders, 37, 613–627.

Gray, K.M., Tonge, B.J., & Sweeney, D.J. (2008). Using the autism diagnostic interview—Revised and the autism diagnostic observation schedule with young children with developmental delay: Evaluating diagnostic validity. Journal of Autism and Developmental Disorders, 38, 657–667.

Halladay, A.K., Bishop, S., Constantino, J.N., Daniels, A.M., Koenig, K., Palmer, K., & Szatmari, P. (2015). Sex and gender differences in autism spectrum disorder: Summarizing evidence gaps and identifying emerging areas of priority. Molecular Autism, 6, 36.

Havdahl, K.A., Bal, V.H., Huerta, M., Pickles, A., Oyen, A., Stoltenberg, C., Lord, C., & Bishop, S.L. (2016). Multidimensional influences on autism symptom measures: Implications for use in etiological research. Journal of the American Academy of Child and Adolescent Psychiatry, 55, 1054–1063.

Hiller, R.M., Young, R.L., & Weber, N. (2014). Sex differences in autism spectrum disorder based on *DSM-5* criteria: Evidence from clinician and teacher reporting. Journal of Abnormal Child Psychology, 42, 1381–1393.

Hurley, R.S.E., Losh, M., Parlier, M., Reznick, J.S., & Piven, J. (2007). The broad autism phenotype questionnaire. Journal of Autism and Developmental Disorders, 37, 1679–1690.

Kopp, S., & Gillberg, C. (2011). The autism spectrum screening questionnaire (ASSQ)—revised extended version (ASSQ-REV): An instrument for better capturing the autism phenotype in girls? A preliminary study involving 191 clinical cases and community controls. Research in Developmental Disabilities, 32, 2875–2888.

Lai, M.-C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. Lancet Psychiatry, 2, 1013–1027.

Lai, M.-C., Baron-Cohen, S., & Buxbaum, J.D. (2015). Understanding autism in the light of sex/gender. Molecular Autism, 6, 24.

Lainhart, J.E., Bigler, E.D., Bocian, M., Coon, H., Dinh, E., Dawson, G., ... Volkmar, F. (2006). Head circumference and height in autism: A study by the collaborative program of excellence in autism. American Journal of Medical Genetics. Part A, 140, 2257–2274.

Le Couteur, A., Haden, G., Hammal, D., & McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardized assessment instruments: The ADI-R and the ADOS. Journal of Autism and Developmental Disorders, 38, 362–372.

Lord, C., Pickles, A., McLennan, J., Rutter, M., Bregman, J., Folstein, S., ... Minshew, N. (1997). Diagnosing autism: Analyses of data from the Autism Diagnostic Interview. Journal of Autism and Developmental Disorders, 27, 501–507.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview—Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders, 24, 659–685.

Lord, C., Rutter, M., DiLavore, P.C., & Risi, S. (2003). Autism diagnostic observation schedule manual. Los Angeles, CA: Western Psychological Services.

Magaña, S., & Smith, L.E. (2013). The use of the autism diagnostic interview—Revised with a Latino population. Journal of Autism and Developmental Disorders, 43, 1098–1105.

Mandell, D.S., Ittenbach, R.F., Levy, S.E., & Pinto-Martin, J.A. (2007). Disparities in diagnoses received prior to a diagnosis of autism spectrum disorder. Journal of Autism and Developmental Disorders, 37, 1795–1802.

Mandell, D.S., Listerud, J., Levy, S.E., & Pinto-Martin, J.A. (2002). Race differences in the age at diagnosis among medicaid-eligible children with autism. Journal of the American Academy of Child and Adolescent Psychiatry, 41, 1447–1453.

Mazurek, M.O., Handen, B.L., Wodka, E.L., Nowinski, L., Butter, E., Engelhardt, C.R. (2014). Age at first autism spectrum disorder diagnosis: The role of birth cohort,

demographic factors, and clinical features. Journal of Developmental and Behavioral Pediatrics, 35, 561–569.
Mullen, E.M. (1997). Mullen scales of early learning. Los Angeles, CA: Western Psychological Services.
Ozonoff, S., Goodlin-Jones, B.L., & Solomon, M. (2005). Evidence-based assessment of autism spectrum disorders in children and adolescents. Journal of Clinical Child and Adolescent Psychology, 34, 523–540.
Reed, P., & Osborne, L.A. (2012). Diagnostic practice and its impacts on parental health and child behaviour problems in autism spectrum disorders. Archives of Disease in Childhood, 97, 927–931.
Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., ... Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. Journal of the American Academy of Child and Adolescent Psychiatry, 45, 1094–1103.
Rogers, S.J., Vismara, L., Wagner, A.L., McCormick, C., Young, G., & Ozonoff, S. (2014). Autism treatment in the first year of life: A pilot study of infant start, a parent-implemented intervention for symptomatic infants. Journal of Autism and Developmental Disorders, 44, 2981–2995.
Sasson, N.J., Faso, D.J., Parlier, M., Daniels, J.L., & Piven, J. (2014). When father doesn't know best: Selective disagreement between self-report and informant report of the broad autism phenotype in parents of a child with autism. Autism Research, 7, 731–739.
Shattuck, P.T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D.S., Wiggins, L., ...Cuniff, C. (2009). The timing of identification among children with autism spectrum disorder: Findings from a population-based surveillance study. Journal of the American Academy of Child and Adolescent Psychiatry, 48, 474–483.
Sparrow, S., Cicchetti, D., & Balla, D. (2005). Vineland adaptive behavior scales (2nd ed.). Minneapolis, MN: Pearson Assessments.
Vanegas, S.B., Magaña, S., Morales, M., & McNamara, E. (2016). Clinical validity of the ADI-R in a US-based Latino population. Journal of Autism and Developmental Disorders, 46, 1623–1635.
Ventola, P.E., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J., ...Fein, D. (2006). Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. Journal of Autism and Developmental Disorders, 36, 839–847.
Wechsler, D. (1999). Wechsler abbreviated scale of intelligence. San Antonio, TX: The Psychological Corporation.
Werling, D.M., & Geschwind, D.H. (2013). Sex differences in autism spectrum disorders. Current Opinion in Neurology, 26, 146–153.
Yee, M.M., & Millichap, J.G. (2015). Relationship between age at diagnosis of ADHD and ASD. Pediatric Neurology Briefs, 29, 78.
Zuckerman, K.E., Sinche, B., Mejia, A., Cobian, M., Becker, T., & Nicolaidis, C. (2014). Latino parents' perspectives of barriers to autism diagnosis. Academic Pediatrics, 14, 301–308.
Zuckerman, K.E., Lindly, O.J., & Sinche, B.K. (2015). Parental concerns, provider response, and timeliness of autism spectrum disorder diagnosis. Journal of Pediatrics, 166, 1431–1439.