HHS Public Access

Author manuscript

J Child Psychol Psychiatry. Author manuscript; available in PMC 2021 May 12.

Published in final edited form as:

J Child Psychol Psychiatry. 2021 January ; 62(1): 97–106. doi:10.1111/jcpp.13242.

Sex Differences in Scores on Standardized Measures of Autism Symptoms: A multi-site integrative data analysis

Aaron J. Kaat[1], **Amy M. Shui**[2], **Sheila S. Ghods**[2], **Cristan A. Farmer**[3], **Amy N. Esler**[4], **Audrey Thurm**[3], **Stelios Georgiades**[5], **Stephen M. Kanne**[6], **Catherine Lord**[7], **Young Shin Kim**[2], **Somer L. Bishop**[2]

[1]Feinberg School of Medicine, Northwestern University, Chicago, IL

[2]Department of Psychiatry, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA

[3]Pediatrics & Developmental Neuroscience Branch, National Institute of Mental Health, Bethesda, MD

[4]University of Minnesota, Minneapolis, MN

[5]McMaster University, Hamilton, ON, Canada

[6]University of Missouri Thompson Center, Columbia, MO

[7]University of California, Los Angeles, Los Angeles, CA

Abstract

Background: Concerns have been raised that scores on standard measures of autism spectrum disorder (ASD) symptoms may differ as a function of sex. However, these findings are hindered by small female samples studied thus far. The current study evaluated if, after accounting for age, IQ, and language level, sex affects ASD severity estimates from diagnostic measures among children with ASD.

Methods: Data were obtained from eight sources comprising 27 sites. Linear mixed-effects models, including a random effect for site, were fit for 10 outcomes (ADOS domain-level calibrated severity scores, Autism Diagnostic Interview–Revised [ADI-R] raw scores by age-based algorithm, and raw scores from the two indices on the Social Responsiveness Scale [SRS]). Sex was added to the models after controlling for age, NVIQ, and an indicator for language level.

Results: Sex significantly improved model fit for half of the outcomes, but least square mean differences were generally negligible (effect sizes [ES] < 0.20), increasing to small-to-moderate in adolescence (ES < 0.40). Boys received more severe RRB scores than girls on both the ADOS and

Corresponding author: Somer L. Bishop, UCSF Weill Institute for Neurosciences, 401 Parnassus Avenue, San Francisco, CA 94143, Somer.Bishop@ucsf.edu, Phone: 415-502-3349, Fax: 415-502-3556.

Conflict of interest:

Dr. Bishop has received royalties from the ADOS-2. All profits from her research are donated to charity.

Dr. Lord has received royalties from the ADI-R, ADOS, and ADOS-2. All profits from her research are donated to charity.

Ms. Shui, Ms. Ghods, and Drs. Kaat, Farmer, Georgiades, Thurm, Kanne, Esler, and Kim report no biomedical financial interests or potential conflicts of interest.

ADI-R (age 4+ algorithm) and girls received more severe scores than boys on both SRS indices, which emerged in adolescence.

Conclusions: This study combined several available databases to create the largest sample of girls with ASD diagnoses. We found minimal differences due to sex beyond other known influences on ASD severity indicators. This may suggest that, among children who ultimately receive a clinical ASD diagnosis, severity estimates do not systematically differ to such an extent that sex-specific scoring procedures would be necessary. However, given the limitations inherent in mostly clinically-ascertained samples, future research must address questions about systematic sex differences among children or adults who do not receive clinical diagnoses of ASD. Moreover, while the current study helps resolve questions about widely used diagnostic instruments, we could not address sex differences in phenotypic aspects outside of these scores.

Keywords

Sex differences; Autism spectrum disorder; Restricted and repetitive behavior; Social impairment

Introduction

During the past decade there has been growing concern that standard diagnostic assessment procedures for ASD may be biased against girls and women (Ratto et al., 2018). Substantial resources have been dedicated to understanding phenotypic differences between the sexes, in order to improve identification of girls and women with ASD (Charman et al., 2017), especially those with higher cognitive and/or language abilities (Howe et al., 2015). However, concerns remain that commonly used diagnostic instruments may yield systematically lower scores for girls and women referred for assessment of ASD (Tillmann et al., 2018). This could happen for multiple reasons; available measures may be biased toward detecting ASD among male individuals (Constantino & Charman, 2012; Dworzynski, Ronald, Bolton, & Happé, 2012; Kirkovski, Enticott, & Fitzgerald, 2013), or females may actually have less pronounced symptoms of ASD (Lai et al., 2011; Rynkiewicz et al., 2016). A hypothesis that girls and women with ASD may be more skilled than boys and men at hiding their ASD-related impairments (e.g., “camouflaging”) has further fueled these concerns, resulting in calls for sex-specific revisions to widely-used instruments (Constantino & Charman, 2016; Lai, Lombardo, Auyeung, Chakrabarti, & Baron-Cohen, 2015).

Given that the discrepant sex ratio is one of the most consistently replicated findings in the clinical and epidemiological literature on ASD (see Loomes, Hull, & Mandy, 2017), questions about phenotypic differences between male and female individuals with ASD are not new. However, early studies were significantly hindered by lack of access to sufficiently large numbers of girls and women with ASD. This resulted in inconsistent and sometimes conflicting findings from relatively small, clinically-ascertained samples. The field has recently benefitted from large-scale efforts to aggregate phenotypic, genetic, and other types of data across sites (Fischbach & Lord, 2010). This has yielded larger numbers of female participants with ASD, and a clearer picture of sex differences among individuals with ASD diagnoses has started to emerge. Several studies have shown that girls with ASD tend to receive lower scores on measures of restricted and repetitive behaviors (RRBs) (Charman et

al., 2017; Frazier, Georgiades, Bishop, & Hardan, 2014; Hartley & Sikora, 2009; Knutsen, Crossman, Perrin, Shui, & Kuhlthau, 2019; Lai et al., 2011; Mandy et al., 2012; Supekar & Menon, 2015; Szatmari et al., 2012; Wilson et al., 2016), but findings remain inconsistent with regard to social-communication deficits. In addition, even when samples are relatively large, ascertainment continues to pose a critical problem. For example, if at least a portion of girls with higher cognitive and language abilities are at risk of not being identified until later in life (Begeer et al., 2013; Giarelli et al., 2010; Howe et al., 2015), then they would not be represented in samples recruited during early childhood. A recent meta-analysis (e.g., Van Wijngaarden-Cremers et al., 2014) directly considered the issue of ascertainment by evaluating findings from 20 studies of sex differences. Overall, findings indicated no major differences in core symptoms between males and females; however, the potential limitation of certain girls (or boys) with ASD not being included in studies hinders meta-analyses similarly as individual studies. At the same time as increasing attention has been focused on possible sex differences, a number of studies have found that scores on these measures are affected by factors such as age, cognitive ability, and behavior problems (Havdahl et al., 2016; Hus, Bishop, Gotham, Huerta, & Lord, 2013). Thus, in some cases, more severe scores may reflect demographic or developmental characteristics rather than actual symptom severity (Bishop et al., 2019). These findings are critical to consider in relation to questions of sex differences, especially given the consistently replicated finding that females with ASD are underrepresented at the higher end of IQ and overrepresented at the lower end (Lai et al., 2011; Mandy et al., 2012; Skuse et al., 2009). Sex differences in IQ distributions in ASD likely result from a combination of biological and social-contextual factors. At the low end of IQ, females with ASD are more likely to have highlypenetrant genetic mutations which are also associated with intellectual disability (Iossifov et al., 2014; Sanders et al., 2015). At the high end, females who are more successful at “camouflaging” may fail to be diagnosed with ASD at all (Lai et al., 2017; Livingston & Happé, 2017). Regardless of the factors driving IQ differences in males and females with ASD, it is essential to consider the possibility that the effects of IQ and sex on measures of ASD symptoms are confounded. Except for recent notable examples (Charman et al., 2017; Tillman et al., 2018), previous studies have not been adequately powered to consider multiple developmental variables, or have not accounted for these factors simultaneously. Thus questions remain about whether any observed differences in scores are actually attributable to sex.

The current study was initiated to determine whether—after accounting for age, nonverbal IQ (NVIQ), and language level—scores on ASD symptom measures differ systematically between boys and girls with a best-estimate clinical diagnosis of ASD. Further understanding of this issue is necessary to inform proper use and future revisions of ASD diagnostic instruments, including the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005), the Autism Diagnostic Interview–Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003), and the Autism Diagnostic Observation Schedule (ADOS/ADOS-2; Lord et al., 2000; Lord, Rutter, et al., 2012). Importantly, as data were limited to scores on three widely used measures, our analyses could not address sex differences in aspects of phenotype that may not be reflected in scores from these measures.

Methods

Participants

Data were obtained from eight sources (see Supplementary Table 1 for a description of the data sources): 20 clinics from the Autism Treatment Network (ATN), Center for Autism and the Developing Brain/University of Michigan Autism and Communication Disorders Center (CADB), University of Minnesota (UMN), Korean Epidemiological Cohort (KOR), National Institute of Mental Health (NIMH), Simons Simplex Collection (SSC), UCSF STAR Center (STAR), and the Pathways study (PATH). At the time of data collection, written informed consent and/or assent was obtained from all participants in accordance with IRB approved protocols at each site. Sample size varied by analysis insofar as participants were administered different measures depending on age, language ability, date of assessment (e.g., some participants were assessed prior to publication of the SRS), and site/study-specific practices (e.g., the ADI-R was not administered at all ATN sites). Across all analyses, 8,985 individuals met the overall inclusion criteria of: age at ADOS assessment 12 months to less than 18 years, best-estimate clinical diagnosis of ASD (determined from all information obtained during the assessment but not contingent on scores from any one measure), and NVIQ and ADOS data available from the same study visit. If a participant had more than one visit that was eligible for inclusion, a single visit was chosen at random. Table 1 show participant demographics overall for any participant who provided data to any outcome, and supplementary Table S2 provides demographics by analysis and site.

Measures

Cognitive Assessment.—NVIQ was assessed with a test selected based on clinician judgment and the child's age and ability. The majority (80%) of children were administered either the Differential Ability Scales-II (Elliott, 2007), the Mullen Scales of Early Learning (MSEL; Mullen, 1995), or the Stanford-Binet, Fifth Edition (SB-5; Roid, 2003). Standard scores for NVIQ were used whenever possible. In cases where standard scores could not be calculated, ratio NVIQ scores were derived using age equivalent scores (Bishop, Guthrie, Coffing, & Lord, 2011; Farmer, Golden, & Thurm, 2015).

Autism Diagnostic Observation Schedule, Second Edition (ADOS-2).—The ADOS-2 consists of five modules (Toddler Module, Modules 1-4), one of which is selected based on the child's expressive language and chronological age (Lord, Rutter, et al., 2012). The 10-point calibrated severity score (CSS; Gotham, Pickles, & Lord, 2009) provides a measure of overall autism symptom severity, accounting for the age and language level of the child, with higher scores representing greater severity. Separate domain calibrated scores for Social Affect (SA) and Restricted and Repetitive Behaviors (RRB) (Hus, Gotham, & Lord, 2014) were used in our analyses.

When using the ADOS as the outcome-of-interest, several exclusionary criteria were implemented: Children who received the Toddler Module (n=279) or Module 4 (n=215) were excluded from analysis due to small sample sizes. Participants who were younger than 31 months of age and who received Module 1 (n=432) were also excluded because they would have received the Toddler Module, had it been available when tested. Three children

ages < 36 months with Module 3 were excluded based on recommendations from the test developers (Lord, Rutter, et al., 2012). None of these exclusionary rules was enforced for the other outcome (e.g., the ADI-R or SRS).

Autism Diagnostic Interview – Revised (ADI-R).—The ADI-R is a standardized, semi-structured caregiver interview designed to assess developmental and behavioral aspects of ASD (Lord, Rutter, & Le Couteur, 1994; Rutter et al., 2003). Raw total scores were calculated for each of the current behavior algorithm domains: Social Interaction (range 0-30); Communication (range Verbal 0-26, Nonverbal 0-14); and RRB (range 0-12), where higher scores represent greater abnormality. It is important to note that the ADI-R was not originally developed to index severity (see Hus & Lord, 2013), though these raw total scores are commonly employed as continuous measures of ASD symptoms in research (e.g., Meilleur & Fombonne, 2009; Seltzer et al., 2003). Raw scores were analyzed for all domains except for Communication, where the proportion of maximum possible score ([POMP] range 0-100) was used, given the different number of applicable items (Cohen, Aiken, & West, 1999). These scores were considered separately for children under and over 4, as there are different algorithms available for children under 4 years (Kim, Thurm, Shumway, & Lord, 2013). Focus of analyses was on ADI-R current algorithm scores in order to approximate current autism severity and account for current age and cognitive abilities in analyses of score differences. The diagnostic algorithm (which uses ever/abnormal 4 to 5 scores), was also evaluated. Similar results were found, though these analyses were limited by ceiling effects on all domains.

Social Responsiveness Scale (SRS).—The SRS was designed to be a continuous measure of ASD traits appropriate for individuals with and without ASD (Constantino & Gruber, 2005, 2012). Raw sum scores for the two DSM-5 indices were used in the current analysis: the Restrictive and Repetitive Behavior (RRB) Index (range 0 - 36) and the Social Communication/Interaction (SCI) Index (range 0 - 159). Higher SRS scores are meant to indicate more ASD-related behaviors. While the authors provide sex-specific T-scores on the SRS, we focused on the effect of sex among individuals already diagnosed with ASD. Thus we did not want to prematurely adjust for sex, and thus used raw scores instead.

Statistical Analysis

In order to maximize available person- and item-level data, we planned an integrative data analysis (IDA). IDAs do not represent one type of statistical analysis, but rather a family of methods using pooled item-level data (Curran & Hussong, 2009). Within this family, we utilized a parametric mixed-effects random IDA. We regressed each of the 10 outcome measures on covariates known to affect ASD severity scores (age, NVIQ, and an indicator of language level), and included a random intercept for site to account for potential differences across clinics. This is especially likely on measures such as the ADOS or ADI-R, where clinicians may be more similar to other clinicians at their site than those at different sites, even if they had the same research-reliable training. Indeed, site-effects in diagnostic practices have been reported previously (Lord, Petkova, et al., 2012). Sex was then added to the model to evaluate if it significantly improved modeling each outcome. The large number of available datasets in this study (20 individual ATN sites plus seven other data sources)

mitigates any limitations of the random effects IDA (versus a fixed effect IDA) (Brincks et al., 2018; Hussong, Curran, & Bauer, 2013). Intra-class correlation coefficients (ICCs) were calculated to estimate the amount of shared variance within site.

The model building process consisted of the following: first, scatterplots of the outcome scores with age were examined for curvilinear trends. Second (only if a curvilinear relationship was suggested), a set of regression models with complete pooling (i.e., no random effects) compared the fit of quadratic and log-linear age terms. Each set of models contained the appropriate age terms, the other covariates, and interactions. To indicate language level, ADOS outcomes used Modules 1-3, ADI-R used the verbal vs. non-verbal algorithm indicator (i.e., 0 vs. 1 or 2 on Item 30), and SRS used ADOS Module 1 compared to 2, 3, or 4 combined (i.e., to index minimally-verbal vs. verbal). R-squared values and significance of terms were assessed to determine the appropriate curvilinear age relationship. Third, a set of linear mixed-effects models was fit, first with only the covariates and then adding sex and sex-by-age interactions to determine if it improved model fit. Both sets of models started with fixed effects for mean-centered age and non-linear age (as appropriate), mean-centered NVIQ, the same indicator for verbal level (as specified above), and all interactions. We evaluated whether model fit was significantly improved by adding sex using the Akaike information criterion (AIC) and Bayesian information criteria (BIC) and statistical significance using the Satterthwaite method for mixed effects models (Burnham & Anderson, 2003; SAS, 2008). If the AIC and BIC suggested different models, the difference test between nested models was used, with a significant result rejecting the model with fewer parameters and thus suggesting including sex. Given the large sample, we also calculated least square mean differences using Sidak correction for familywise Type 1 error control. For ADOS and SRS analyses, age refers to age at ADOS assessment; for ADI-R analyses, it refers to age when the ADI was completed. For the majority of participants, the ADI-R and ADOS were completed within 3 months of each other, but for a minority (<10%) it was completed at an earlier or later date than the index-visit ADOS. Alpha was set to .05. Residuals were visually inspected for linearity, homogeneity, and normality via residual vs. predicted value plots, histograms, and Q-Q plots. Statistical analyses were performed using SAS/STAT version 9.4 and R version 3.6.1 with packages lme4 1.1-21 (Bates, Mächler, Bolker, & Walker, 2014) and emmeans 1.4.2 (Lenth, 2019).

Results

The first step of our model building process determined the optimal effect for age on the outcome of interest. A curvilinear relationship was suggested for the ADOS, ADI-R (ages 4+), and the SRS. A linear age relationship was suggested for the ADI-R Toddler algorithm. In the second step quadratic age had comparable or better R-squared values than those with log-transformed age for those with putative curvilinear age models. Thus, the ADOS, ADI-R age 4+, and SRS models included quadratic age terms, and the ADI-R under age 4 only had a linear age term; no models continued with the log-transformed age term. Our final models directly compared the inclusion or exclusion of sex on the outcome of interest in a random-effects IDA.

Table 2 summarizes the mixed effect models for each of the 10 outcomes. Supplementary Tables 3 provide model coefficients for all models, and supplementary figures 1.a-1.j graphically represent the sex effects on predicted scores. The addition of sex was statistically significant for five of 10 outcomes, which was consistently supported by the AIC, though the BIC only supported including sex-effects for the ADOS CSS-RRB.

Within the RRB domain, there were consistent mean differences for the ADOS CSS-RRB. Given the discontinuous nature of the ADOS CSS-RRB scores, we conducted two sensitivity analyses: 1) evaluating the natural logarithm of the outcome, and 2) treating the outcome as nominal and using multinomial regression (due to the complexity of this model, the random effect for site could not be modeled). Both sensitivity analyses, available from the authors, resulted in similar estimated sex effects. On the ADI-R RRB for ages 4+ there were sex effects at younger ages but not in adolescence. There was no evidence of a sex-effect on it under age 4. The SRS had no significant sex differences at ages 5 or 10, but some evidence for more severe RRBs for girls (contradicting the ADOS and ADI-R) among adolescents.

Within the Social Communication domain, there was no evidence of a sex effect on the ADI-R Social Interactions or Communication at any age. On the ADOS CSS-SA, the AIC and BIC provided contradictory support, with no significant sex effect at ages 3 or 15, but a slight difference with more severe males at age 7 years. On the SRS SCI, again the AIC and BIC provided contradictory support, and there were no significant differences at ages 5 or 10, but some evidence for girls having more severe scores at age 15 years.

Tables S3.a-S3.b show the model coefficients for each model with and without sex. As expected, the known covariates (age, NVIQ, and verbal/nonverbal status) were related to most measures in the expected way. The second interesting finding was the effect of site. While we hypothesized that ICCs would be high for clinician-rated measures within site, we observed smaller ICCs for the ADI-R and ADOS (ICCs 0.03-0.09) than for the SRS (0.21 or 0.29), which was unexpected.

Supplementary Figures 1 graphically represent the results of these analyses. The predicted values are plotted as points, with a generalized additive model (GAM) smoother using shrunken cubic splines through predicted values by sex. Across the 10 figures, it is evident that the size of effects are generally small, but also that they vary by age.

Discussion

Despite multiple concerns being raised, previous sex-based comparisons of scores on standardized instruments have yielded inconsistent results. One explanation for this is the small sample size of these studies. Given that girls tend to comprise 20% of children diagnosed with ASD, the number of girls on which findings are based is usually quite small, and therefore the estimates of differences between girls and boys are not very precise. The variability in results can perhaps also be explained by differences in ascertainment methods across studies; if the diagnostic system is biased, then among individuals already diagnosed with ASD, sex effects will be minimized because inclusion will have been preferentially extended to the girls who are phenotypically similar to the boys. Findings of differences may

also be dependent on the age of the patient population; expert clinicians report observing greater sex differences in adolescents and adults with ASD (Jamison, Bishop, Huerta, & Halladay, 2017), which is consistent with findings from the current study. However, as mentioned previously, ascertainment and age of assessment may be heavily confounded. Thus, in the absence of longitudinal data, it is difficult to separate out actual age differences in how ASD manifests in boys vs. girls, from other phenotypic differences (e.g., in type or severity of symptoms) that drive age of identification.

While still limited to cross-sectional data, the current study attempted to resolve previous debates about sex differences in scores on three widely used measures of ASD symptomology (i.e., the SRS, ADI-R, ADOS-2) by aggregating a very large sample of females (n=1,463) and males (n=7,522) with clinical diagnoses of ASD. The large sample afforded us the opportunity to extend previous work by examining whether sex affected scores, *above and beyond* the known influences of developmental factors. This is a significant strength of our study, as there is a high chance of drawing erroneous conclusions about "sex" differences, or lack of, when compared groups of males and females differ on other phenotypic variables besides sex.

After accounting for age, NVIQ, and language level, sex had a statistically significant effect on some of the scores. Boys received higher raw scores on measures of restricted and repetitive behaviors according to both parent report (ADI-R RRB) and direct observation (ADOS-2 RRB CSS). This is consistent with several other published reports that girls tend to exhibit fewer repetitive behaviors than boys (Kreiser & White, 2014; Rubenstein, Wiggins, & Lee, 2015; Van Wijngaarden-Cremers et al., 2014). However, on the SRS RRB, differences emerged in adolescence, such that girls received more severe scores In addition, while social communication scores on the ADI-R and ADOS did not differ between boys and girls, differences on the SRS SCI also emerged in adolescence, with girls receiving more severe scores the SRS SCI. This stands in contrast to recent observations that girls exhibit *less* severe impairments in core social-communication behaviors (Lai et al., 2017; Sedgewick, Hill, Yates, Pickering, & Pellicano, 2016), and may possibly reflect gender socialization differences that are picked up more by the SRS. For example, parents (primarily mothers in this case) may have higher expectations of their girls when it comes to certain social-communication behaviors; perhaps "Is emotionally distant, doesn't show his or her feelings" and "Has good personal hygiene," are rated differently for boys than for girls. Further, several of the items on the SRS RRB concern *perceptions* of unusual behaviors/behaviors that are inappropriate to the social situation (e.g., "Is regarded by other children as odd or weird" and "Behaves in ways that seem strange or bizarre"). Thus, it is possible that while a sex difference in discrete, observable repetitive behaviors would be expected, this was mitigated by higher parental expectations for their daughters' social behaviors that are embedded within certain RRB items on the SRS.

Limitations

A major limitation of this study sample is that participants were drawn from databases of children with clinical diagnoses of ASD, and therefore did not include females (or males) with ASD who were *not* diagnosed. For this reason, we could not evaluate the sensitivity and

specificity of the diagnostic instruments, and we were only able to treat scores as indices of ASD symptom severity. This highlights a critical problem of relying on data from clinically-ascertained and diagnosed samples: we can only evaluate the instruments' performance on individuals who were actually identified. Thus, it is impossible to comment on the utility of standard instruments among individuals whose symptom presentation was too mild and/or different to be captured by standard diagnostic practices (Evans, Boan, Bradley, & Carpenter, 2018; Kim et al., 2011; Loomes et al., 2017). Further, although ASD diagnoses in the current study were based on clinician best estimate and not scores from any one diagnostic measure, these measures were nevertheless employed as part of the diagnostic assessment, and in some cases (i.e., for participants enrolled in ATN or SSC), inclusion criteria did require meeting ASD cut-offs on one or more of the diagnostic measures. As a result, it is possible that sex differences in scores were minimized by our ascertainment procedures and that our findings are generalizable only to youth who already have a diagnosis of ASD. Nonetheless, the consistent findings regarding RRBs across instruments and modes of data collection (observation and parent-interview) suggests some generalization of these results even given this limitation.

Conclusion

This work was motivated by our interest to better understand the extent to which standard measures of ASD symptoms might perform differently in males and females. By compiling the largest set of data from girls with ASD, we were sufficiently powered to consider developmental variables in the discussion of how girls may score differently on ASD symptom measures than boys. Our results indicated that girls receive less severe scores on parent-reported and clinician-administered measures of restricted and repetitive behavior (ADI-R and ADOS RRB). In adolescence, girls received more severe scores on a parent report questionnaire of ASD symptoms (SRS SCI and RRB). However, all of these differences were small (effect sizes [ES] generally < 0.20) and thus of minimal clinical significance.

While our findings of no, or very small, differences in scores do not readily support the need for sex-specific scoring of these instruments among children already diagnosed with ASD, it is still possible that some girls and women exhibit *different* ASD-related difficulties. This would indicate a need to continue to examine the content validity of diagnostic instruments and systems that potentially lack sensitivity to detect certain ASD symptoms, in order to ensure they adequately capture the range of symptoms characteristic of females with ASD. Considering differences in symptom presentation of other groups traditionally underrepresented in ASD research (e.g., ethnic minorities, rurally located families, families with lower SES, older adults with ASD) will also be critical in future measure development and revision efforts. As part of this work, Item Response Theory (IRT) based approaches will be particularly useful for understanding true differences in psychometric performance of ASD symptom measures in different subgroups.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

This work was supported by grants from the National Institute of Child Health and Human Development (R01HD093012 to Dr. Bishop and R01HD081199 to Dr. Lord), the National Institute of Mental Health (R01MH081873 to Dr. Lord), and the Autism Science Foundation (ASP-002 to Dr. Bishop). This work was also supported (in part) by the Intramural Research program of the NIMH (ZIC MH002961 to Dr. Thurm).

The Autism Treatment Network activity was supported by Autism Speaks and conducted through the Autism Speaks Autism Treatment Network. This information or content and conclusions are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by Autism Speaks.

We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, 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, E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC population dataset described in this study by applying at https://base.sfari.org.

We are grateful to the parents and children who participate in the Pathways in ASD study, and to the current and past members of the study team, including principal and co-investigators (P. Szatmari, P. Mirenda, C. Waddell, L. Zwiagenbaum, J. Volden, T. Vaillancourt, T. Bennett, I. Smith, M. Elsabbagh, S. Georgiades, E. Duku, W. Ungar, C. Kerns). We are also grateful to our funders, the Canadian Institutes for Health Research, Kids Brain Health, Autism Speaks, the Sinneave Family Foundation, and Alberta Innovates Health Solutions.

Abbreviations:

ADI-R	Autism Diagnostic Interview – Revised
ADOS	Autism Diagnostic Observation Schedule
ASD	Autism Spectrum Disorder
NVIQ	Nonverbal Intelligence Quotient
RRB	Restricted and Repetitive Behavior
SA	Social Affect
SCI	Social Communication/Interaction
SRS	Social Responsiveness Scale

References

Bates D, Mächler M, Bolker B, & Walker S (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Begeer S, Mandell D, Wijnker-Holmes B, Venderbosch S, Rem D, Stekelenburg F, & Koot HM (2013). Sex differences in the timing of identification among children and adults with autism spectrum disorders. J Autism Dev Disord, 43(5), 1151–1156. doi:10.1007/s10803-012-1656-z [PubMed: 23001766]

Bishop SL, Farmer C, Kaat A, Georgiades S, Kanne S, & Thurm A (2019). The Need for a Developmentally Based Measure of Social Communication Skills. Journal of the American Academy of Child & Adolescent Psychiatry, 58(6), 555–560. doi:10.1016/j.jaac.2018.12.010 [PubMed: 31130206]

Bishop SL, Guthrie W, Coffing M, & Lord C (2011). Convergent validity of the mullen scales of early learning and the differential ability scales in children with Autism Spectrum Disorders. American Association on Intellectual and Developmental Disabilities, 116(5), 331–343. doi:10.1352/1944-7558-116.5.331

Brincks A, Montag S, Howe GW, Huang S, Siddique J, Ahn S, … Brown CH (2018). Addressing Methodologic Challenges and Minimizing Threats to Validity in Synthesizing Findings from Individual-Level Data Across Longitudinal Randomized Trials. Prev Sci, 19(Suppl 1), 60–73. doi:10.1007/s11121-017-0769-1 [PubMed: 28434055]

Burnham KP, & Anderson DR (2003). Model selection and multimodel inference: a practical information-theoretic approach: Springer Science & Business Media.

Charman T, Loth E, Tillmann J, Crawley D, Wooldridge C, Goyard D, … Buitelaar JK (2017). The EU-AIMS Longitudinal European Autism Project (LEAP): clinical characterisation. Molecular Autism, 8(1), 27. doi:10.1186/s13229-017-0145-9 [PubMed: 28649313]

Cohen J, Aiken LS, & West SG (1999). The Problem of Units and the Circumstance for POMP. Multivariate Behavioral Research, 34(3), 315–346. doi:10.1207/S15327906MBR3403_2

Constantino JN, & Charman T (2012). Gender Bias, Female Resilience, and the Sex Ratio in Autism. Journal of the American Academy of Child & Adolescent Psychiatry, 51(8), 756–758. doi:10.1016/j.jaac.2012.05.017 [PubMed: 22840545]

Constantino JN, & Charman T (2016). Diagnosis of autism spectrum disorder: reconciling the syndrome, its diverse origins, and variation in expression. Lancet Neurol, 15(3), 279–291. doi:10.1016/s1474-4422(15)00151-9 [PubMed: 26497771]

Constantino JN, & Gruber CP (2005). Social Responsiveness Scale (SRS). Los Angeles, CA: Western Psychological Services.

Constantino JN, & Gruber CP (2012). Social Responsiveness Scale, Second Edition. Los Angeles, CA: Western Psychological Services.

Curran PJ, & Hussong AM (2009). Integrative Data Analysis: The Simultaneous Analysis of Multiple Data Sets. Psychological methods, 14(2), 81–100. doi:10.1037/a0015914 [PubMed: 19485623]

Dworzynski K, Ronald A, Bolton P, & Happé F (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? J Am Acad Child Adolesc Psychiatry, 51(8), 788–797. doi:10.1016/j.jaac.2012.05.018 [PubMed: 22840550]

Elliott CD (2007). Differential Ability Scales - Second Edition. San Antonio, TX: The Psychological Corporation.

Evans SC, Boan AD, Bradley C, & Carpenter LA (2018). Sex/gender differences in screening for autism spectrum disorder: implications for evidence-based assessment. Journal of Clinical Child & Adolescent Psychology, 1–15. [PubMed: 28574780]

Farmer C, Golden C, & Thurm A (2015). Concurrent validity of the differential ability scales, second edition with the Mullen Scales of Early Learning in young children with and without neurodevelopmental disorders. Child Neuropsychol, 1–14. doi:10.1080/09297049.2015.1020775

Fischbach G, & Lord C (2010). The Simons simplex collection: A resource for identification of autism genetic risk factors. Neuron, 68(2), 192–195. doi:10.1016/j.neuron.2010.10.006 [PubMed: 20955926]

Frazier TW, Georgiades S, Bishop SL, & Hardan AY (2014). Behavioral and cognitive characteristics of females and males with autism in the Simons Simplex Collection. J Am Acad Child Adolesc Psychiatry, 53(3), 329–340 e321-323. doi:10.1016/j.jaac.2013.12.004 [PubMed: 24565360]

Giarelli E, Wiggins LD, Rice CE, Levy SE, Kirby RS, Pinto-Martin J, & Mandell D (2010). Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. Disabil Health J, 3(2), 107–116. doi:10.1016/j.dhjo.2009.07.001 [PubMed: 21122776]

Gotham KO, Pickles A, & Lord C (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. J Autism Dev Disord, 39(5), 693–705. doi:10.1007/s10803-008-0674-3 [PubMed: 19082876]

Hartley SL, & Sikora DM (2009). Sex differences in autism spectrum disorder: an examination of developmental functioning, autistic symptoms, and coexisting behavior problems in toddlers. J Autism Dev Disord, 39(12), 1715–1722. doi:10.1007/s10803-009-0810-8 [PubMed: 19582563]

Havdahl KA, Hus Bal V, Huerta M, Pickles A, Øyen A, Stoltenberg C, … Bishop SL (2016). Multidimensional Influences on Autism Symptom Measures: Implications for Use in Etiological Research. Journal of the American Academy of Child & Adolescent Psychiatry, 55(12), 1054–1063.e1053. doi:10.1016/j.jaac.2016.09.490 [PubMed: 27871640]

Howe YJ, O'Rourke JA, Yatchmink Y, Viscidi EW, Jones RN, & Morrow EM (2015). Female Autism Phenotypes Investigated at Different Levels of Language and Developmental Abilities. J Autism Dev Disord, 45(11), 3537–3549. doi:10.1007/s10803-015-2501-y [PubMed: 26100851]

Hus V, Bishop SL, Gotham K, Huerta M, & Lord C (2013). Factors influencing scores on the social responsiveness scale. Journal of Child Psychology and Psychiatry, 54(2), 216–224. doi:doi:10.1111/j.1469-7610.2012.02589.x [PubMed: 22823182]

Hus V, Gotham KO, & Lord C (2014). Standardizing ADOS domain scores: separating severity of social affect and restricted and repetitive behaviors. J Autism Dev Disord, 44(10), 2400–2412. doi:10.1007/s10803-012-1719-1 [PubMed: 23143131]

Hus V, & Lord C (2013). Effects of child characteristics on the Autism Diagnostic Interview-Revised: implications for use of scores as a measure of ASD severity. J Autism Dev Disord, 43(2), 371–381. doi:10.1007/s10803-012-1576-y [PubMed: 22729382]

Hussong AM, Curran PJ, & Bauer DJ (2013). Integrative Data Analysis in Clinical Psychology Research. Annu Rev Clin Psychol, 9, 61–89. doi:10.1146/annurev-clinpsy-050212-185522 [PubMed: 23394226]

Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, … Wigler M (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature, 515(7526), 216–221. doi:10.1038/nature13908 [PubMed: 25363768]

Jamison R, Bishop SL, Huerta M, & Halladay AK (2017). The clinician perspective on sex differences in autism spectrum disorders. Autism, 21(6), 772–784. [PubMed: 28429618]

Kim SH, Thurm A, Shumway S, & Lord C (2013). Multisite study of new autism diagnostic interview-revised (ADI-R) algorithms for toddlers and young preschoolers. J Autism Dev Disord, 43(7), 1527–1538. doi:10.1007/s10803-012-1696-4 [PubMed: 23114567]

Kim YS, Leventhal BL, Koh YJ, Fombonne E, Laska E, Lim EC, … Grinker RR (2011). Prevalence of autism spectrum disorders in a total population sample. Am J Psychiatry, 168(9), 904–912. doi:10.1176/appi.ajp.2011.10101532 [PubMed: 21558103]

Kirkovski M, Enticott PG, & Fitzgerald PB (2013). A review of the role of female gender in autism spectrum disorders. J Autism Dev Disord, 43(11), 2584–2603. doi:10.1007/s10803-013-1811-1 [PubMed: 23525974]

Knutsen J, Crossman M, Perrin J, Shui A, & Kuhlthau K (2019). Sex differences in restricted repetitive behaviors and interests in children with autism spectrum disorder: An Autism Treatment Network study. Autism, 23(4), 858–868. doi:10.1177/1362361318786490 [PubMed: 30047281]

Kreiser NL, & White SW (2014). ASD in females: are we overstating the gender difference in diagnosis? Clinical child and family psychology review, 17(1), 67–84. [PubMed: 23836119]

Lai MC, Lombardo MV, Auyeung B, Chakrabarti B, & Baron-Cohen S (2015). Sex/Gender Differences and Autism: Setting the Scene for Future Research. Journal of the American Academy of Child and Adolescent Psychiatry, 54(1), 11–24. doi:10.1016/j.jaac.2014.10.003 [PubMed: 25524786]

Lai MC, Lombardo MV, Pasco G, Ruigrok AN, Wheelwright SJ, Sadek SA, … Baron-Cohen S (2011). A behavioral comparison of male and female adults with high functioning autism spectrum conditions. PLoS ONE, 6(6), e20835. doi:10.1371/journal.pone.0020835 [PubMed: 21695147]

Lai MC, Lombardo MV, Ruigrok AN, Chakrabarti B, Auyeung B, Szatmari P, … Consortium MA (2017). Quantifying and exploring camouflaging in men and women with autism. Autism, 21(6), 690–702. doi:10.1177/1362361316671012 [PubMed: 27899710]

Lenth R (2019). emmeans: estimated marginal means, aka least-squares means. R package v. 1.3. 4.

Livingston LA, & Happé F (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. Neuroscience & Biobehavioral Reviews, 80, 729–742. doi:10.1016/j.neubiorev.2017.06.005 [PubMed: 28642070]

Loomes R, Hull L, & Mandy WPL (2017). What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. Journal of the American Academy of Child & Adolescent Psychiatry, 56(6), 466–474. doi:10.1016/j.jaac.2017.03.013 [PubMed: 28545751]

Lord C, Petkova E, Hus VB, Gan W, Lu F, Martin DM, … Risi S (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. Arch Gen Psychiatry, 69(3). doi:10.1001/archgenpsychiatry.2011.148

Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, … Rutter M (2000). The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord, 30(3), 205–223. doi:0162-3257/00/0600-0205 [PubMed: 11055457]

Lord C, Rutter M, DiLavore PC, Risi S, Gotham KO, & Bishop SL (2012). Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1-4. Torrance, CA: Western Psychological Services.

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. J Autism Dev Disord, 24(5), 659–685. doi:10.1007/BF02172145 [PubMed: 7814313]

Mandy W, Chilvers R, Chowdhury U, Salter G, Seigal A, & Skuse D (2012). Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. J Autism Dev Disord, 42(7), 1304–1313. doi:10.1007/s10803-011-1356-0 [PubMed: 21947663]

Meilleur AA, & Fombonne E (2009). Regression of language and non-language skills in pervasive developmental disorders. J Intellect Disabil Res, 53(2), 115–124. doi:10.1111/j.1365-2788.2008.01134.x [PubMed: 19054269]

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

Ratto AB, Kenworthy L, Yerys BE, Bascom J, Wieckowski AT, White SW, … Anthony LG (2018). What About the Girls? Sex-Based Differences in Autistic Traits and Adaptive Skills. J Autism Dev Disord, 48(5), 1698–1711. doi:10.1007/s10803-017-3413-9 [PubMed: 29204929]

Roid GH (2003). Stanford-Binet Intelligence Scales, Fifth Edition. Itasca, IL: Riverside Publishing.

Rubenstein E, Wiggins LD, & Lee LC (2015). A Review of the Differences in Developmental, Psychiatric, and Medical Endophenotypes Between Males and Females with Autism Spectrum Disorder. Journal of Developmental and Physical Disabilities, 27(1), 119–139. doi:10.1007/s10882-014-9397-x [PubMed: 26146472]

Rutter M, Le Couteur A, & Lord C (2003). Autism Diagnostic Interview-Revised (ADI-R). Los Angeles, California: Western Psychological Services.

Rynkiewicz A, Schuller B, Marchi E, Piana S, Camurri A, Lassalle A, & Baron-Cohen S (2016). An investigation of the 'female camouflage effect' in autism using a computerized ADOS-2 and a test of sex/gender differences. Molecular Autism, 7(1), 10. doi:10.1186/s13229-016-0073-0 [PubMed: 26798446]

Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, … State MW (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron, 87(6), 1215–1233. doi:10.1016/j.neuron.2015.09.016 [PubMed: 26402605]

SAS. (2008). SAS/STAT ®9.2 User's Guide The MIXED Procedure. Cary, NC: SAS Institute Inc. Retrieved from http://support.sas.com/documentation/cdl/en/statugmixed/61807/PDF/default/statugmixed.pdf

Sedgewick F, Hill V, Yates R, Pickering L, & Pellicano E (2016). Gender Differences in the Social Motivation and Friendship Experiences of Autistic and Non-autistic Adolescents. J Autism Dev Disord, 46(4), 1297–1306. doi:10.1007/s10803-015-2669-1 [PubMed: 26695137]

Seltzer MM, Krauss MW, Shattuck PT, Orsmond G, Swe A, & Lord C (2003). The symptoms of autism spectrum disorders in adolescence and adulthood. J Autism Dev Disord, 33(6), 565–581. doi:10.1023/B:JADD.0000005995.02453.0b [PubMed: 14714927]

Skuse DH, Mandy W, Steer C, Miller LL, Goodman R, Lawrence K, … Golding J (2009). Social communication competence and functional adaptation in a general population of children: preliminary evidence for sex-by-verbal IQ differential risk. Journal of the American Academy of Child & Adolescent Psychiatry, 48(2), 128–137. [PubMed: 19106766]

Supekar K, & Menon V (2015). Sex differences in structural organization of motor systems and their dissociable links with repetitive/restricted behaviors in children with autism. Molecular Autism, 6(1), 50. [PubMed: 26347127]

Szatmari P, Liu XQ, Goldberg J, Zwaigenbaum L, Paterson AD, Woodbury-Smith M, … Thompson A (2012). Sex differences in repetitive stereotyped behaviors in autism: implications for genetic liability. Am J Med Genet B Neuropsychiatr Genet, 159B(1), 5–12. doi:10.1002/ajmg.b.31238 [PubMed: 22095612]

Tillmann J, Ashwood K, Absoud M, Bölte S, Bonnet-Brilhault F, Buitelaar J, … Canitano R (2018). Evaluating sex and age differences in ADI-R and ADOS scores in a large European multi-site sample of individuals with autism spectrum disorder. J Autism Dev Disord, 1–16. [PubMed: 28864932]

Van Wijngaarden-Cremers PJ, van Eeten E, Groen WB, Van Deurzen PA, Oosterling IJ, & Van der Gaag RJ (2014). Gender and age differences in the core triad of impairments in autism spectrum disorders: a systematic review and meta-analysis. J Autism Dev Disord, 44(3), 627–635. doi:10.1007/s10803-013-1913-9 [PubMed: 23989936]

Wilson CE, Murphy CM, McAlonan G, Robertson DM, Spain D, Hayward H, … Murphy DG (2016). Does sex influence the diagnostic evaluation of autism spectrum disorder in adults? Autism, 20(7), 808–819. doi:10.1177/1362361315611381 [PubMed: 26802113]

Key points

- Researchers and clinicians have increasingly questioned whether there are sex differences on measures used to determine ASD diagnosis and symptom severity.
- Previous investigations of how sex affects scores have been hindered by small groups of female participants, preventing adequate consideration of other variables known to affect scores on ASD symptom measures.
- Results of the current study indicated that among youth already diagnosed with ASD, select scores from the ADI-R, SRS, and ADOS-2 differ by sex, with girls receiving less severe scores, in the area of RRBs. However, effects of sex were small and likely of limited clinical significance.
- Clinicians must be sufficiently trained to recognize the widely varying presentations of individuals with ASD across the spectrum of age, IQ, language level, and ASD severity. Attending to how sex and other sociodemographic variables may affect symptom presentation is yet another component of best clinical practice.

Table 1.

Overall Participant Demographics

Characteristics	Overall	Female	Male
Age (months) (mean±SD (n), min-max)	82.9±45.9 (8985), 12-215	82.6±47.2 (1463), 13-215	83.0±45.7 (7522), 12-215
NVIQ (mean±SD (n), min-max)	79.0±26.8 (8985), 2-161	75.8±27.1 (1463), 2-153	79.6±26.7 (7522), 7-161
Male % (n)	84% (7522)		
White/Caucasian race % (n)	77% (6304)	76% (1023)	78% (5281)
Non-Hispanic ethnicity % (n)	90% (7308)	91% (1229)	89% (6079)
ADOS Module			
Toddler Module	3% (279)	3% (48)	3% (231)
Module 1 age <31 months	5% (432)	5% (75)	5% (357)
Module 1 age 31+ months	31% (2799)	33% (485)	31% (2314)
Module 2	21% (1925)	22% (329)	21% (1596)
Module 3	37% (3335)	33% (485)	38% (2850)
Module 4	2% (215)	3% (41)	2% (174)
Source (n)			
ATN	3332	543	2789
SSC	2756	374	2382
CADB	1894	372	1522
PATH	388	62	326
NIMH	261	42	222
KOR	163	40	123
UMN	103	15	88
STAR	85	15	70

SD = standard deviation, min = minimum, max = maximum, NVIQ = non-verbal intelligence quotient, ADOS = Autism Diagnostic Observation Schedule, ATN = Autism Treatment Network, CADB = Center for Autism and the Developing Brain/University of Michigan Autism and Communication Disorders Center, KOR = Korean Epidemiological Cohort, NIMH = National Institute of Mental Health, PATH = Pathways study, SSC = Simons Simplex Collection, STAR = UCSF STAR Center, UMN = University of Minnesota

Table 2.

Sex Differences on Measures of Restricted and Repetitive Behaviors and Social Communication in Children with ASD

Category	Outcome	Inclusion Criteria	Sample Composition		Model Selection			Marginal Mean (SE)			
					AIC	BIC	Deviance	Contrast	Boys	Girls	Difference (ES)
Restricted & Repetitive Behaviors	ADOS RRB CSS	Module 1 & Age >= 31 months; Mod 2 & Age >= 12 mo; Mod 3 & Age > 36 mo	6,711 Boys	Exclude	33516	33656	33476	Age 3 yrs:	7.6 (0.2)	7.4 (0.2)	0.2 * (0.12)
			1,280 Girls	Include	**33478**	**33639**	**33432**	Age 7 yrs:	7.3 (0.1)	6.9 (0.2)	0.3 *** (0.16)
			27 Sites (ICC=0.09)				ΔDev(3)=44.4 ***	Age 15 yrs:	7.7 (0.2)	6.9 (0.3)	0.8 *** (0.39)
	ADI-R RRB Current	Age < 48 months (Toddler Algorithm)	769 Boys	Exclude	**4078**	**4126**	**4058**	No Evidence for an Effect			
			174 Girls	Include	4078	413666	4054				
			9 Sites (ICC=0.15)				ΔDev(2)=4.5				
		Age >= 48 months	4016 Boys	Exclude	22173	**22263**	22145	Age 5 yrs:	4.8 (0.2)	4.3 (0.3)	0.5 *** (0.21)
			750 Girls	Include	**22156**	22265	**22122**	Age 10 yrs:	5.0 (0.3)	4.6 (0.3)	0.5 ** (0.18)
			14 Sites (ICC=0.06)				ΔDev(3)=23.0 ***	Age 15 yrs:	4.6 (0.4)	4.6 (0.4)	−0.0 (−0.01)
	SRS RRB	Age >= 48 months	2,965 Boys	Exclude	23190	**23276**	23162	Age 5 yrs:	15.3 (1.8)	14.6 (1.9)	0.7 (0.09)
			517 Girls	Include	**23180**	23285	**23146**	Age 10 yrs:	18.4 (1.9)	17.8 (1.9)	0.6 (0.08)
			6 Sites (ICC=0.29)				ΔDev(3)=15.3 **	Age 15 yrs:	17.5 (2.0)	19.7 (2.1)	−2.2 ** (−0.28)
Social Communication	ADOS SA CSS	Module 1 & Age >= 31 months; Mod 2 & Age >= 12 mo; Mod 3 & Age > 36 mo	6650 Boys	Exclude	32581	**32720**	32541	Age 3 yrs:	6.8 (0.1)	6.9 (0.1)	−0.2 (−0.08)
			1268 Girls	Include	**32577**	32737	**32531**	Age 7 yrs:	7.0 (0.1)	6.8 (0.1)	0.2 * (0.10)
			27 Sites (ICC=0.03)				ΔDev(3)=9.9 *	Age 15 yrs:	7.6 (0.2)	7.5 (0.3)	0.1 (0.05)
	ADI-R SI Current	Age < 48 months (Toddler Algorithm)	750 Boys	Exclude	**5499**	**5547**	**5479**	No Evidence for an Effect			
			170 Girls	Include	5502	5559	5478				
			8 Sites †				ΔDev(2)=1.2				
		Age >= 48 months	4009 Boys	Exclude	**28521**	**28612**	**28493**	No Evidence for an Effect			
			746 Girls	Include	28527	28637	28493				

Category	Outcome	Inclusion Criteria	Sample Composition		Model Selection AIC	BIC	Deviance	Contrast	Marginal Mean (SE) Boys	Girls	Difference (ES)
			14 Sites (ICC=0.04)				ΔDev(3)=0.5				
	ADI-R Communication POMP Current	Age < 48 months (Toddler Algorithm)	742 Boys	Exclude	**7769**	**7817**	**7749**		No Evidence for an Effect		
			168 Girls	Include	7772	7829	7748				
			8 Sites (0.002)				ΔDev(2)=1.6				
		Age >= 48 months	4016 Boys	Exclude	**40991**	**41082**	**40963**		No Evidence for an Effect		
			750 Girls	Include	40994	41104	40960				
			14 Sites (ICC=0.06)				ΔDev(3)=3.0				
	SRS SCI	Age >= 48 months	2,911 Boys	Exclude	30367	**30453**	30339	Age 5 yrs:	70.6 (4.5)	71.7 (4.8)	−1.1 (−0.05)
			506 Girls	Include	**30360**	30465	**30326**	Age 10 yrs:	75.9 (4.8)	76.9 (4.9)	−1.0 (−0.04)
			6 Sites (ICC=0.21)				ΔDev(3)=12.3 **	Age 15 yrs:	76.5 (5.5)	83.8 (5.7)	−7.4 *** (−0.32)

ICC is reported for the model including sex. For the ICC excluding sex, see Supplementary Tables 3a and 3b. Least-square means are calculated at pre-specified ages, averaging over the indicator variable for language status and over NVIQ, estimating contrasts at NVIQ of 70, 85, and 100.

See Tables S3a and S3b for list of terms in each model.

ADI-R = Autism Diagnostic Interview-Revised, ADOS = Autism Diagnostic Observation Schedule, ASD = autism spectrum disorder, CSS = Calibrated severity score, RRB = Restricted and Repetitive Behaviors, SA = Social Affect, SCI = Social Communication/Interaction, SRS = Social Responsiveness Scale

† Random-intercept model was singular, resulting in an ICC=0. To ensure robust results, the model was refit with fixed-effects for site and without any site effect, neither of which modified the conclusions of the model.

* $p<0.05$

** $p<0.01$

*** $p<0.001$