



Factors Associated with Confirmed and Unconfirmed Autism Spectrum Disorder Diagnosis in Children Volunteering for Research

Susanne W. Duvall¹ · Rachel K. Greene² · Randi Phelps^{3,4} · Tara M. Rutter⁵ · Sheila Markwardt⁶ · Julia Grieser Painter⁵ · Michaela Cordova⁷ · Beth Calame⁵ · Olivia Doyle⁸ · Joel T. Nigg⁹ · Eric Fombonne⁹ · Damien Fair¹⁰

Accepted: 21 March 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Purpose Diagnostic accuracy of autism spectrum disorder (ASD) is crucial to track and characterize ASD, as well as to guide appropriate interventions at the individual level. However, under-diagnosis, over-diagnosis, and misdiagnosis of ASD are still prevalent.

Methods We describe 232 children ($M_{Age} = 10.71$ years; 19% female) with community-based diagnoses of ASD referred for research participation. Extensive assessment procedures were employed to confirm ASD diagnosis before study inclusion. The sample was subsequently divided into two groups with either confirmed ASD diagnoses (ASD+) or unconfirmed/inaccurate diagnoses (ASD-). Clinical characteristics differentiating the groups were further analyzed.

Results 47% of children with community-based ASD diagnoses did not meet ASD criteria by expert consensus. ASD+ and ASD- groups did not differ in age, gender, ethnicity, or racial make-up. The ASD+ group was more likely to have a history of early language delays compared to the ASD- group; however, no group differences in current functional language use were reported by caregivers. The ASD+ group scored significantly higher on ADI-R scores and on the ADOS-2 algorithm composite scores and calibrated severity scores (CSSs). The ASD- group attained higher estimated IQ scores and higher rates of psychiatric disorders, including anxiety disorder, disruptive behavior, and mood disorder diagnoses. Broadly, caregiver questionnaires (SRS-2, CCC-2) did not differentiate groups.

Conclusion Increased reported psychiatric disorders in the ASD- group suggests psychiatric complexity may contribute to community misdiagnosis and possible overdiagnosis of ASD. Clinician-mediated tools (ADI-R, ADOS-2) differentiated ASD+ versus ASD- groups, whereas caregiver-reported questionnaires did not.

Keywords Autism spectrum disorder · Diagnosis · Accuracy · Community-based diagnosis · Comorbidity

✉ Susanne W. Duvall
duvall@ohsu.edu

- ¹ Departments of Pediatrics and Psychiatry, Institute on Development and Disability, Center for Development and Child Rehabilitation, Oregon Health & Science University, 707 SW Gaines St, Portland, OR 98239, USA
- ² Departments of Pediatrics, Institute on Development and Disability, Center for Development and Child Rehabilitation, Oregon Health & Science University, 707 SW Gaines St, Portland, OR, USA
- ³ Staff Psychologist in the Division of Psychology, Phoenix Children's Hospital, 1919 E Thomas Rd, Phoenix, AZ, USA
- ⁴ Department of Child Health, University of Arizona College of Medicine, 475 N 5th St, Phoenix, AZ, USA
- ⁵ Department of Psychiatry, School of Medicine, Oregon Health & Science University, 3181 S.W. Sam Jackson Rd, Portland, OR, USA

- ⁶ Biostatistician III, Biostatistics and Design Program, Oregon Health & Science University, 3181 SW Sam Jackson Rd, Portland, OR 97217, USA
- ⁷ San Diego State University/UC San Diego Joint Doctoral Program in Clinical Psychology, 6363 Alvarado Ct, San Diego, CA, USA
- ⁸ Department of Psychiatry, Oregon Health & Science University, 3181 SW Sam Jackson Rd, Portland, OR 97217, USA
- ⁹ Department of Psychiatry, Oregon Health & Science University, 3181 SW Sam Jackson Rd, Portland, OR 97239, USA
- ¹⁰ College of Education and Human Development, Department of Pediatrics, Medical School University of Minnesota, Masonic Institute for the Developing Brain, Professor, Institute of Child Development, 2025 E. River Parkway 7962A, Minneapolis, MN 55414, USA

The prevalence of autism spectrum disorder (ASD), a neurodevelopmental condition impacting social communication and restricted and repetitive behaviors, has been found to be increasing over recent decades, with current US estimates of 1 in 36 in 8-year-old children and higher rates in males than females (Maenner et al., 2023; Center for Disease Control and Prevention (CDC), 2014 & 2023). The diagnosis of ASD is based on behavioral assessments, and, despite improvements in diagnostic reliability over the last 25 years, there remain issues of both under and over-diagnosis (Hill et al., 2015; Skellern et al., 2005; Fombonne, 2023; Zuckerman et al., 2013) that affect clinical management, access to services, population rates, and health/educational costs. Autistic individuals are also at-risk for a variety of struggles in learning, vocational attainment, and independent living (CDC, 2023). Increased services accessibility and awareness of ASD seems to be driving higher rates of diagnosis than ever before. Although identifying autistic individuals early (true positives) has promise for improving outcomes, there is increasing concern about misdiagnosis and overdiagnosis of ASD (false positives), which could negatively impact treatment and trajectories for these children and decrease our ability to accurately study the disorder at the population level (Lucinao et al., 2014; Fombonne et al., 2021).

Despite a higher degree of awareness, there is still a problem of late diagnosis, as the average age at diagnosis in recent surveys is 4–5 years of age (Hill et al., 2015; van't Hof et al., 2021; Maenner et al., 2023). In some underserved groups, such as racial or ethnic minorities, non-English speaking or low-income families, ASD is either diagnosed later or underdiagnosed as shown by significantly lower early identification rates in the US in CDC surveys (La Roche et al., 2018; Zuckerman et al., 2013); although, the most recent prevalence estimates indicate that this gap in identification of ASD in racial and ethnic minority groups is shrinking to some degree (Maenner et al., 2021, 2023). However, autistic Black children in this US sample of children aged 8 were significantly more likely to be identified as having intellectual disability, suggesting that perhaps Black children with the highest needs are being identified, thereby missing a subset of Black children with ASD with relatively lower support needs. Misdiagnosis is also common in individuals with more complex psychiatric profiles, including those with prominent mood concerns, severe ADHD, personality disorders, schizophrenia and anxiety, among others (Luciano et al., 2014; Havdahl et al., 2016, Greene et al., 2022).

Conversely, concerns regarding overdiagnosis have also been noted (Graf et al., 2017; Fombonne, 2018, 2023). Overdiagnosis may be related to increased awareness of ASD in individuals with average or above IQ, provision of

access to education services under a school, but not medical, eligibility of ASD, and overdiagnosis of ASD in individuals with other delays or co-occurring psychological diagnoses that could result in social difficulties (Hill et al., 2015; Fombonne, 2023, Van Schalkwyk et al., 2015). In fact, a study of Australian child psychiatrists and pediatricians indicated that when uncertain about an ASD diagnosis 58% had erred on the side of diagnosing ASD to qualify children for special educational supports (Skellern et al., 2005). There are few studies that have systematically examined the extent of misdiagnosis in either direction or the factors that influence it. This is a key gap in the field as early and accurate diagnosis and treatment has shown to be beneficial in improving quality of life and functional outcome in children with ASD. Additionally, our ability to study ASD in large community samples and at the population level is hampered by inaccurate diagnoses. Overidentification also has the possibility of taxing an already strained disability service system.

As part of screening for inclusion in a larger neuroimaging study, we evaluated, with research reliable diagnostic procedures, a high number of children who presented to the study with a prior diagnosis of ASD. This provided an opportunity to estimate the proportion of accurate/inaccurate diagnoses and to identify correlates of misdiagnosis among children presenting with a prior community ASD diagnosis. The specific aims of the study were:

Aim 1: to estimate the number and frequency of false positive ASD diagnoses observed in a research referred sample; and.

Aim 2: to compare clinical characteristics and sources of information between children with confirmed and unconfirmed ASD diagnosis in order to identify the origins of this discrepancy.

Methods

Study Design

Participants were drawn from the screening visit that was conducted as the first visit in an ongoing study examining neuroimaging correlates of children with ASD and ADHD. The basic study design is a case control comparison involving multi-method assessments of two groups: (1) children with an ASD diagnosis confirmed as part of the research evaluation [ASD+] and (2) children with community diagnoses of ASD who did not to meet ASD diagnostic criteria within the research evaluation (ASD-).

Participants

Participants included 232 children aged 7–12 years with historical ASD diagnoses who sought to be research participants in a neuroimaging study. Study recruitment occurred through posters in the community and on the hospital campus and the University Hospital Autism clinical program, community support groups and outreach, and targeted mailing and emails to patients with ICD-9/ICD-10 codes of ASD in their electronic health record. Inclusion criteria for the study included: an existing diagnosis of ASD; fluent in English; able to see and hear adequately for study completion; no major head trauma, no diagnoses of intellectual disability, schizophrenia, seizures or tic disorder; no medication needs that would be incompatible with washout; and no MRI contraindications.

Procedures

Potential participants completed a telephone screener regarding demographic variables, inclusion criteria, and ASD diagnosis history. Participants who passed telephone screening were then scheduled for an initial study visit. Participants provided consent and assent in compliance with IRB requirements. At the initial visit, participants completed cognitive assessments, behaviorally-based autism testing (Autism Diagnostic Observation Schedule-Second Edition (ADOS-2); Lord et al., 2012a, b) and caregiver based standardized clinical interview for symptoms of ASD (Autism Diagnostic Interview-Revised (ADI-R); Rutter et al., 2003). Caregivers completed questionnaires to provide the child's developmental and medical history, social-emotional and behavioral adjustment, and language functioning. Due to the heavy time burden of completing ADI-R interviews and the significant proportion of children who did not ultimately meet ASD study inclusion criteria following the first visit, study procedures switched to a streamlined initial visit that postponed the ADI-R until a later visit. Thus, 188 participants completed both the ADI-R and ADOS-2. All 232 participants completed the ADOS-2. The ADOS-2 and ADI-R were administered by one of three doctoral level clinical psychologists who all attained research reliability. All ADOS-2 assessments were videotaped for subsequent review. All data were collected between 2011 and 2018, and importantly, before this study was planned; therefore, all scores and ratings were obtained in a manner that was strictly blind to the hypotheses of this study.

Measures

Autism Diagnostic Observation Schedule-Second Edition (ADOS-2); Lord et al., 2012a, b; Gotham et al., 2009). One

of three research reliable clinical psychologists administered and scored the ADOS-2 Module 3 (for verbally fluent children) to all participants. The ADOS-2 is a 45–60 min standardized semi-structured observational measure of social, communication, rigidity and restricted interests that includes interacting with an examiner across several activities and contexts. A recent meta-analysis examining the ADOS-2 estimated sensitivity ranging from 0.89 – 0.92 and specificity ranging from 0.81 – .85 (Lebersfeld et al., 2021).

Autism Diagnostic Interview-Revised (ADI-R); Rutter et al., 2003). One of three research reliable doctoral-level clinical or developmental psychologists administered and scored this comprehensive caregiver interview. Questions focus on current and lifetime behaviors related to ASD symptoms, including but not limited to social communication, language, gestures, sensory processing, behavioral rigidity, restricted interests and repetitive behaviors. Reliability is strong and a recent meta-analysis examining the ADI-R estimated sensitivity of 0.75 and specificity of 0.82 when examined and pooled across both clinical and research samples (Cicchetti et al., 2008; Lebersfeld et al., 2021). Notably, this study reported higher ADI-R specificity in research versus clinical samples (Research = 0.85, Clinical = 0.72).

Wechsler Intelligence Scale for Children, 4th Edition (WISC-IV); Wechsler, 2003). A clinical psychologist administered a reliable and valid group of cognitive subtests (Block Design, Information, and Vocabulary) of the WISC-IV to estimate full scale IQ (estimated FSIQ; Sattler & Dumont, 2004).

Social Responsiveness Scale, 2nd Edition (SRS-2); Constantino & Gruber, 2012). Caregivers completed the SRS-2, a 65-item measure used to assess autism symptom severity. Each question is answered using a 4-point Likert scale. The SRS-2 has been previously shown to reliably distinguish individuals with ASD from individuals with other psychiatric diagnoses (Constantino et al., 2003; Constantino & Todd, 2000).

Psychiatric Symptoms and Diagnoses

We relied on a combination of questionnaires and clinical diagnostic approach. Co-occurring diagnoses were evaluated with a best estimate clinical diagnosis approach after review of ADOS-2 videotapes and other observational data, of medical and educational records as well as of parent diagnostic interviews and completed questionnaires when available.

The Multidimensional Anxiety Scale for Children, 2nd Edition (MASC 2); March, 2012). The MASC 2 is a multi-rater 50-item, 4-point Likert-type scale used to assess the presence and severity of various symptoms of anxiety (i.e., Physical Symptoms, Harm Avoidance, Social Anxiety,

Separation/Panic, and Obsessive and Compulsive behaviors) in children 8 to 19 years old. The MASC-2 demonstrates good discriminant validity (March, 2012). It was completed by caregivers.

Children's Communication Checklist, 2nd Edition (CCC-2; Bishop, 2006). The CCC-2 is a 70-item caregiver report measure used to assess communication skills related to overall speech, vocabulary, sentence structure, and social/pragmatic language skills in children ages 4–16 years old. Dolata et al. (2022) have recently shown that pragmatic language scores of the CCC-2 are highly predictive of autism diagnosis and of the prominence of autistic features.

Procedure for Determining Group Membership

A clinical expert team review was conducted for each potential participant to implement a best-estimate clinical diagnosis procedure for ASD status. To do so, from the three licensed psychologists who administered the assessments a team of at least one licensed psychologist and one licensed child psychiatrist (also certified trainer to the ADI-R and the ADOS-2) reviewed videos of each participant's ADOS-2, clinical, school and medical record information (including previous testing) and results from all research assessments described above (as well as additional structured interviews and questionnaires about mental health that were included in the parent study). These clinicians had research reliability on the ADOS-2 and extensive autism specific clinical experience, they then independently rated the presence/absence of ASD, and their degree of certainty. Each rater was asked the degree to which they were certain a subject had ASD (0 = certain no ASD diagnosis, 10 = certain of ASD diagnosis). Only in cases of disagreement regarding the presence/absence of ASD did the experts confer during the process. A unanimous clinical expert consensus was required for inclusion as ASD+ by DSM-IV and DSM-5 criteria. Agreement was high with Pearson's correlation coefficients for each pair of raters (4 pairs in total) for ratings of certainty of ASD diagnosis ranging from 0.96 to 0.99 (depending on the rating pair). The team also identified co-occurring psychiatric disorders using all information available.

Two groups were compared: individuals confirmed to have ASD by our group of experts (ASD+) and individuals who were determined not to meet criteria for ASD, despite reporting a prior history of ASD diagnosis in the community (ASD-).

Statistical Analysis

Descriptive statistics were used to characterize participant demographics, caretaker-reported behaviors, additional diagnoses, medication use, and clinical instruments for the

overall cohort and according to ASD diagnosis (ASD- vs. ASD+); frequencies and percentages for categorical variables and means and standard deviations for continuous variables were calculated. To test for differences across groups in each of our collected variables, we used chi-square or Fisher's exact tests for categorical variables and t-tests for continuous variables. All analyses were performed in Stata/SE 15.1.

Data used in this study were collected at the initial eligibility visit (described above), and as noted, due to the heavy time burden of completing all interviews/questionnaires and the significant proportion of children who did not meet ASD study inclusion criteria following the first visit, study procedures switched to a streamlined initial visit. Thus, there were participants that did *not* complete the ADI-R ($n=44$), WISC-IV subtests ($n=55$) or MASC ($n=97$), and this was disproportionately among participants in the ASD- group. To assess the potential bias introduced by these protocol changes, we assessed if those ASD- participants who did not complete all assessments differed in demographics and ADOS-2 scores from ASD- participants who did complete all assessments. We compared medians and interquartile ranges for continuous ADOS-2 sub-scores using Mann-Whitney U tests and compared frequencies and percentages for categorical variables using chi-square tests. No statistically significant difference was found between the 2 groups. We did not use any further statistical or imputation techniques to handle missing data, as it would have been outside the scope of this descriptive study. Whether a participant was enrolled in special education was added to the comprehensive phone screener around 2013 after recruitment began, and because of this, 123 participants had missing values for this characteristic. All other response missingness was non-systematic and not associated with meeting ASD criteria. We report missingness as table footnotes or in the table itself.

Results

Demographic Characteristics of the Study Sample

Of the 232 participants, 123 were labeled ASD+ (e.g., meeting full diagnostic criteria for ASD in the comprehensive research criteria clinical evaluation), whereas 109 were labeled ASD- (e.g., participants who indicated they had existing diagnoses of ASD who did not meet diagnostic criteria by expert consensus). ASD+ and ASD- groups did not differ in age ($p=0.61$; Total $M=10.71$ years, $SD=2.31$ years), gender ($p=0.20$; Total % Male = 81.5%), or ethnicity ($p=0.86$; Total % Hispanic/Latinx = 13.5%; see Table 1). Additionally, groups had commensurate racial

Table 1 Characteristics of study cohort by ASD diagnosis

| Characteristic | Total (<i>n</i> =232) | ASD- (<i>n</i> =109) | ASD+ (<i>n</i> =123) | Test statistic | p-value ¹ |
|---|---------------------------|--------------------------|--------------------------|-----------------------------|----------------------|
| <i>M</i> _{Age} at visit (years), (SD) ² | 10.71 (2.31) | 10.62 (2.32) | 10.78 (2.31) | t(226)=-0.51 | 0.61 |
| Gender (male), <i>n</i> (%) | 189 (81.5%) | 85 (78.0%) | 104 (84.6%) | X ² (1,232)=1.65 | 0.20 |
| Hispanic/Latinx, <i>n</i> (%) ² | 27 (13.5%) | 12 (13.0%) | 15 (13.9%) | X ² (1,200)=0.03 | 0.86 |
| Race, <i>n</i> (%)² | | | | X ² (5,196)=7.02 | 0.22 |
| White | 155 (79.1%) | 67 (73.6%) | 88 (83.8%) | | |
| Black/African American | 4 (2.0%) | 3 (3.3%) | 1 (1.0%) | | |
| American Indian/Alaska Native | 7 (3.6%) | 4 (4.4%) | 3 (2.9%) | | |
| Asian | 4 (2.0%) | 4 (4.4%) | 0 (0.0%) | | |
| Hawaiian/Pacific Islander | 2 (1.0%) | 1 (1.1%) | 1 (1.0%) | | |
| More than one race | 24 (12.2%) | 12 (13.2%) | 12 (11.4%) | | |
| Family income, <i>n</i> (%)² | | | | X ² (2,164)=3.20 | 0.20 |
| <\$35,000 | 43 (26.2%) | 20 (24.1%) | 23 (28.4%) | | |
| \$35,000 to <\$75,0000 | 70 (42.7%) | 41 (49.4%) | 29 (35.8%) | | |
| \$75,000 and over | 51 (31.1%) | 22 (26.5%) | 29 (35.8%) | | |
| Age at initial autism diagnosis (years), mean (SD) ² | 4.95 (2.67) | 5.14 (2.80) | 4.79 (2.57) | t(190)=0.89 | 0.37 |
| State, <i>n</i> (%)² | | | | X ² (3,213)=2.02 | 0.57 |
| Oregon | 158 (74.2%) | 76 (74.5%) | 82 (73.9%) | | |
| Washington | 53 (24.9%) | 25 (24.5%) | 28 (25.2%) | | |
| Idaho | 1 (0.5%) | 1 (1.0%) | 0 (0.0%) | | |
| Canada | 1 (0.5%) | 0 (0.0%) | 1 (0.9%) | | |
| Child grade, <i>n</i> (%)² | | | | X ² (3,203)=1.19 | 0.76 |
| Home schooled | 8 (3.9%) | 3 (3.2%) | 5 (4.5%) | | |
| Elementary school | 116 (57.1%) | 54 (58.1%) | 62 (56.4%) | | |
| Middle school | 61 (30.0%) | 26 (28.0%) | 35 (31.8%) | | |
| High school | 18 (8.9%) | 10 (10.8%) | 8 (7.3%) | | |

Note. Information provided in Table 1 is per caregiver report in phone screening. *Abbreviation:* *M*_{Age}, Mean age; ASD, autism spectrum disorder; SD, standard deviation

¹ p-values are from chi-square (X²) tests for categorical variables and t-test for continuous variables

² Missing values: Age of onset, *n*=40; state, *n*=19; age at visit, *n*=4; child grade, *n*=29; Hispanic/Latinx, *n*=32; race, *n*=36; family income, *n*=68

makeups ($p=0.22$), totaling to 79.1% White, 2.0% Black, 3.6% American Indian/Alaska Native, 2.0% Asian, 1.0% Hawaiian/Pacific Islander, and 12.2% Multi-Racial across both groups. ASD+ and ASD- groups did not differ in terms of reported family income ($p=0.20$), though there was a high degree of missingness in this data point. A majority of the participants were residents of Oregon (74.2%), and the remainder were from Washington (24.9%), Idaho (0.5%), and Canada (0.5%). About half the children (51.1%) in the ASD+ group were receiving support through their schools through an individualized educational program (IEP) or 504 plan compared to approximately a quarter (24.2%) of the ASD- group ($p<0.01$) although caution is advised as only 109 participants had valid data for this variable.

Early Neurodevelopmental Features

On a caregiver developmental history questionnaire, the ASD+ group was more likely to have a caregiver endorse current language disorder or delays (33.0%) and a lifetime history of language or articulation differences (60.2%) compared to the ASD- group (14.9% and 43.8%; $p<0.01$, $p=0.02$, respectively; see Table 2). Additionally, on the ADI-R, development in the first 3 years was judged as more

delayed and more consistent with autism in the ASD+ group compared to ASD- patients. The reported age at which first phrases were used trended toward significance such that the ASD+ group was slightly older than the ASD- group at time of first phrases used (See Table 1).

Psychiatric & Developmental History

As the study began in 2011 DSM-IV was utilized for caregiver reports of previous diagnoses. All individuals invited to participate in the study were identified by their families as having ASD and received a prior diagnosis of Autistic Disorder (62.7%), Asperger's Syndrome (23.2%), or Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS; 14.1%), and no group differences were observed among these initial reported diagnoses (all p 's > 0.05).

Caregiver-Reported Diagnoses

The incidence of caregiver-reported lifetime psychiatric disorders (not including ASD) was higher in the ASD- sample ($p=0.022$), with 73.4% reporting no existing psychiatric diagnoses apart from ASD, 22.9% reporting 1, and 3.7% reporting 2 or more psychiatric diagnoses (see Table 3). By

Table 2 Early neurodevelopmental features, by ASD diagnosis

| Characteristic ¹ | Total (<i>n</i> =232) | ASD- (<i>n</i> =109) | ASD+ (<i>n</i> =123) | Test statistic | <i>p</i> -value |
|--|---------------------------|--------------------------|--------------------------|-----------------------|-----------------|
| Lifetime speech or language difficulties, <i>n</i> (%) | 110 (52.6%) | 42 (43.8%) | 68 (60.2%) | $\chi^2(1,209)=5.62$ | 0.02 |
| Current language disorder or delay, <i>n</i> (%) | 53 (24.5%) | 15 (14.9%) | 38 (33.0%) | $\chi^2(1,216)=9.61$ | <0.01 |
| Age (months) first noticed any concerns, <i>M</i> (<i>SD</i>) ² | 24.2 (18.7) | 24.8 (19.3) | 23.8 (18.5) | $t(179)=0.33$ | 0.74 |
| Age (months) of first single word, <i>M</i> (<i>SD</i>) | 26.1 (15.3) | 24.2 (15.0) | 27.3 (15.5) | $t(171)=-1.33$ | 0.18 |
| Age (months) of first phrases, <i>M</i> (<i>SD</i>) | 36.8 (19.0) | 33.8 (17.7) | 38.8 (19.6) | $t(175)=-1.75$ | 0.08 |
| Severity of atypicality at age 3 (0–4 scale), <i>n</i> (%) ² | | | | $\chi^2(4,185)=26.86$ | <0.01 |
| Development in the first 3 years of life clearly typical | 6 (3.2%) | 5 (6.9%) | 1 (0.9%) | | |
| Development possibly typical | 11 (6.0%) | 8 (11.1%) | 3 (2.7%) | | |
| Development probably atypical | 19 (10.3%) | 11 (15.3%) | 8 (7.1%) | | |
| Development definitely atypical but not unambiguously autistic | 63 (34.0%) | 30 (41.7%) | 33 (29.2%) | | |
| Development definitely abnormal and strongly indicative of autism | 86 (46.5%) | 18 (25.0%) | 68 (60.2%) | | |

Note. All information is per caregiver report in phone screening, apart from ADI-R specific items

Abbreviation: ASD, autism spectrum disorder; *M*, mean; *SD*, standard deviation

¹ Missing values: speech or language difficulties, *n*=23; current language disorder or delay, *n*=16; age first noticed, *n*=51; age first single word, *n*=59; age first phrase, *n*=55 age abnormality, *n*=46

² Age first noticed any concerns is from question 2 on ADI-R. Severity of abnormality at age 3 in this table is paraphrased here from formal ADI-R coding for “abnormality first evident” question 86: 0=development in the first 3 years of life clearly normal, 1=development possibly normal, 2=development probably abnormal, 3=development definitely abnormal but not unambiguously autistic, 4=development definitely abnormal and strongly indicative of autism

contrast, 87.0% of the ASD+ sample reported no existing psychiatric diagnoses apart from ASD, 9.8% reported 1, and 3.3% reported 2 or more psychiatric diagnoses. Groups had similar rates of all specific psychiatric disorders (all p 's > 0.05). Notably though, some specific diagnoses (e.g., adjustment disorder, trauma-related disorders, disruptive mood disorders, mood disorders) were reported rather infrequently in the current sample; therefore, those results should be interpreted with caution. Groups did not differ in prevalence of other non-psychiatric diagnoses such as prior developmental delay ($p=0.27$) or sensory processing disorder ($p=0.26$). No caregivers reported diagnoses of psychosis (a study exclusion criteria) or tic disorders. The ASD- group trended towards elevated levels of ADHD, though this difference fell short of statistical significance ($p=0.06$).

Research Team-Provided Diagnoses

The research evaluation team then provided psychiatric and developmental diagnoses based on their evaluation findings and all available data from the broader parent study (structured clinical interviews, medical, clinical and school

reports). The incidence of research team-provided psychiatric disorders (not including ASD) was elevated in the ASD- sample ($p<0.01$), with 38.5% meeting criteria for no psychiatric diagnoses, 44.7% meeting for 1, and 13.9% meeting for 2 or more psychiatric diagnoses (see Table 3). Alternatively, 64.2% of the ASD+ sample met criteria for no other psychiatric diagnoses apart from ASD, 30.9% met for 1 additional diagnoses, and 4.9% met for 2 or more psychiatric diagnoses. Comparatively higher rates of anxiety ($p<0.01$), ADHD ($p<0.02$), and disruptive behavior disorders ($p<0.02$) were found within the ASD- group whereas participants demonstrated no significant differences in rates of adjustment disorders, trauma-related disorders, mood disorders, psychosis, Tourette Syndrome or tic disorders, developmental delay, language disorder, or sensory processing disorder (all p 's > 0.05).

Medication Use

Of all psychotropic and non-psychotropic medications, stimulants were the most commonly used medications for both ASD+ (16.4%) and ASD- samples (17.2%). There

Table 3 Psychiatric and developmental diagnoses, by informant and ASD diagnostic confirmation status

| | Caregiver-Reported Prior Diagnoses | | | | Research Team-Provided Diagnoses | | | |
|---|------------------------------------|------------|-------------|----------------------|----------------------------------|------------|-------------|----------------------|
| | Total | ASD- | ASD+ | p-value ¹ | Total | ASD- | ASD+ | p-value ¹ |
| Autism Spectrum Disorders, n (%) | | | | | | | | |
| Autistic Disorder | 138 (62.7%) | 56 (54.9%) | 82 (69.5%) | 0.08 | - | 0 (0%) | 118 (98.3%) | - |
| Asperger Syndrome | 51 (23.2%) | 28 (27.5%) | 23 (19.5%) | | - | 0 (0%) | 1 (0.8%) | - |
| PDD-NOS | 31 (14.1%) | 18 (17.7%) | 13 (11.0%) | | - | 0 (0%) | 1 (0.8%) | - |
| Psychiatric Disorders², n (%) | | | | | | | | |
| Anxiety | 14 (6.0%) | 9 (8.3%) | 5 (4.1%) | 0.18 | 21 (9.1%) | 17 (15.6%) | 4 (3.3%) | <0.01 |
| Adjustment disorder | 1 (0.4%) | 0 (0.0%) | 1 (0.8%) | 0.35 | 1 (0.4%) | 0 (0.0%) | 1 (0.8%) | 0.35 |
| Trauma | 1 (0.4%) | 0 (0.0%) | 1 (0.8%) | 0.35 | 1 (0.4%) | 0 (0.0%) | 1 (0.8%) | 0.35 |
| ADHD | 32 (13.8%) | 20 (18.3%) | 12 (9.8%) | 0.06 | 92 (39.7%) | 52 (47.7%) | 40 (32.5%) | <0.02 |
| Disruptive behavior | 2 (0.9%) | 2 (1.8%) | 0 (0.0%) | 0.13 | 8 (3.4%) | 7 (6.4%) | 1 (0.8%) | <0.02 |
| Mood disorder | 3 (1.3%) | 2 (1.8%) | 1 (0.8%) | 0.49 | 3 (1.3%) | 3 (2.8%) | 0 (0.0%) | 0.06 |
| Psychosis | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | | 1 (0.4%) | 1 (0.9%) | 0 (0.0%) | 0.29 |
| Tourette Syndrome or tic disorder | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | | 5 (2.2%) | 2 (1.8%) | 3 (2.4%) | 0.75 |
| Non-Psychiatric Disorders, n (%) | | | | | | | | |
| Developmental delay | 15 (6.5%) | 5 (4.6%) | 10 (8.1%) | 0.27 | 15 (6.5%) | 8 (7.3%) | 7 (5.7%) | 0.61 |
| Language disorder | 11 (4.7%) | 10 (9.2%) | 1 (0.8%) | <0.01 | 67 (28.9%) | 32 (29.4%) | 35 (28.5%) | 0.88 |
| Sensory processing | 11 (4.7%) | 7 (6.4%) | 4 (3.3%) | 0.26 | 1 (0.4%) | 1 (0.9%) | 0 (0.0%) | 0.29 |
| Frequency of Diagnoses | | | | | | | | |
| # psychiatric ³ | | | | | | | | |
| 0 | 187 (80.6%) | 80 (73.4%) | 107 (87.0%) | 0.02 | 121 (52.2%) | 42 (38.5%) | 79 (64.2%) | <0.01 |
| 1 | 37 (16.0%) | 25 (22.9%) | 12 (9.8%) | | 90 (38.8%) | 52 (47.7%) | 38 (30.9%) | |
| 2+ | 8 (3.5%) | 4 (3.7%) | 4 (3.3%) | | 21 (9.1%) | 15 (13.9%) | 6 (4.9%) | |
| # non-psychiatric ³ | | | | | | | | |
| 0 | 198 (85.3%) | 89 (81.7%) | 109 (88.6%) | 0.31 | 151 (65.1%) | 70 (64.2%) | 81 (65.9%) | 0.32 |
| 1 | 31 (13.4%) | 18 (16.5%) | 13 (10.6%) | | 79 (34.1%) | 37 (33.9%) | 42 (34.2%) | |
| 2+ | 3 (1.3%) | 2 (1.8%) | 1 (0.81%) | | 2 (0.9%) | 2 (1.8%) | 0 (0%) | |

Abbreviation: ADHD, attention deficit/hyperactivity disorder; ASD, autism spectrum disorder; SD, standard deviation

¹ p-values are from chi-square tests or Fisher's exact for categorical variables

² Trauma includes (OSTSRD, other specific trauma or stressor related disorder; PTSD, post-traumatic stress disorder)

³ Total number of psychiatric or non-psychiatric diagnoses, excluding ASD

were no differences in rate of medication use ($p=0.69$) or type of medications used across groups.

Behavioral Observations of Autism Symptoms

The ASD+ sample scored significantly higher on the ADOS-2 in regard to the restricted and repetitive behavior total score ($p < 0.01$), social affect total score ($p < 0.01$), and ADOS-2 total score ($p < 0.01$; see Table 4). The ASD+ sample also had elevated CSSs ($M=7.64$, $SD=1.45$) compared to the ASD- group ($M=2.50$, $SD=1.93$; $p < 0.01$). Groups did not differ, however, in regard to other observed behaviors during the ADOS (i.e., "E-codes") including overactivity/agitation ($p=0.22$), tantrums, aggression, negative or disruptive behavior ($p=0.54$), or anxiety ($p=0.86$). Overall, 100% of the ASD+ sample received an ADOS-2 diagnostic classification of at least "Autism Spectrum", with 95% meeting the more stringent classification of "Autism". This is in contrast to the ASD- sample, in which 14.7% received an ADOS-2 diagnostic classification of at least "Autism

Spectrum", with 11.0% meeting for the "Autism" classification. The ASD+ group exhibited CSS scores within or exceeding the 6–10 "Autism" range ($M=7.64$; $SD=1.45$), whereas the CSSs of the ASD- group were well below in the "Nonspectrum" CSS of 1–3 range ($M=2.50$; $SD=1.93$).

Caregiver-Reported Autism Symptoms (SRS-2 and ADI-R)

Results from the caregiver-completed SRS-2 revealed no significant differences in any domain or the SRS-2 total score between the ASD+ or ASD- samples, with both showing high mean scores on this instrument in the total scores and all subtest scores (see Table 4). The ADI-R showed statistically significant group differences, with the caregivers of the ASD+ sample consistently reporting higher levels of ASD symptoms across both item-level and total scores. Overall, 88.6% of the ASD+ group had clinically elevated ADI-R total scores, in comparison to 56.2% of ASD- group.

Table 4 Autism instrument scores, by ASD diagnostic confirmation status

| | ASD diagnosis | | | p-value ¹ |
|---|------------------|-----------------|-----------------|----------------------|
| | Total (N=232) | ASD- (n=109) | ASD+ (n=123) | |
| ADOS-2 | N=230 | n=107 | n=123 | |
| Overactivity/Agitation score | 0.98 (0.83) | 0.91 (0.83) | 1.04 (0.83) | 0.22 |
| Tantrums, Aggression, Negative or Disruptive Behavior score | 0.10 (0.37) | 0.08 (0.31) | 0.11 (0.41) | 0.54 |
| Anxiety score | 0.13 (0.37) | 0.13 (0.39) | 0.12 (0.35) | 0.86 |
| Restricted and Repetitive Behavior total score | 2.37 (1.77) | 1.21 (1.06) | 3.37 (1.66) | <0.01 |
| Social Affect total score | 6.80 (4.67) | 3.28 (2.99) | 9.85 (3.59) | <0.01 |
| ADOS Total Score (Social Affect + Restricted and Repetitive Behavior) | 9.17 (5.67) | 4.50 (3.50) | 13.23 (3.73) | <0.01 |
| Calibrated Severity Score | 5.26 (3.07) | 2.50 (1.93) | 7.64 (1.45) | <0.01 |
| ADOS Diagnosis Classification, n (%) | | | | |
| Non-spectrum | 91 (39.2%) | 91 (83.5%) | 0 (0%) | <0.01 |
| Autism spectrum | 9 (3.9%) | 4 (3.7%) | 5 (4.1%) | |
| Autism | 130 (56.0%) | 12 (11.0%) | 118 (95.9%) | |
| SRS | N=213 | n=96 | n=117 | |
| Social Awareness t-score | 74.13 (11.42) | 73.36 (12.09) | 74.76 (10.85) | 0.38 |
| Social Cognition t-score | 72.78 (11.03) | 73.78 (11.49) | 71.97 (10.62) | 0.23 |
| Social Motivation t-score | 69.79 (12.13) | 69.75 (12.14) | 69.82 (12.18) | 0.97 |
| Restricted Interests and Repetitive Behaviors t-score | 79.19 (11.72) | 79.50 (13.09) | 78.93 (10.52) | 0.73 |
| Social Communication and Interaction t-score | 76.28 (10.79) | 76.26 (11.27) | 76.29 (10.44) | 0.98 |
| SRS-2 total t-Score | 77.69 (10.77) | 77.74 (11.46) | 77.64 (10.22) | 0.95 |
| ADI-R | N=188 | n=74 | n=114 | |
| Qualitative Abnormalities in Reciprocal Social Interactions score | 18.44 (6.39) | 15.19 (6.73) | 20.52 (5.21) | <0.01 |
| Qualitative Abnormalities in Non-Verbal Communication score | 8.23 (3.68) | 6.38 (3.77) | 9.43 (3.09) | <0.01 |
| Qualitative Abnormalities in Verbal Communication score | 14.75 (5.31) | 11.70 (5.43) | 16.73 (4.19) | <0.01 |
| Restricted, Repetitive, and Stereotyped Behavior score | 4.95 (2.43) | 3.82 (2.12) | 5.68 (2.34) | <0.01 |
| Reciprocal social interaction criterion met, n (%) | 168 (89.8%) | 57 (78.1%) | 111 (97.4%) | <0.01 |
| Qualitative abnormalities in communication criterion met, n (%) | 166 (88.3%) | 55 (74.3%) | 111 (97.4%) | <0.01 |
| Restricted, repetitive and stereotyped patterns criterion met, n (%) | 158 (84.0%) | 53 (71.6%) | 105 (92.1%) | <0.01 |
| Abnormality evident at or before 36 months criterion met, n (%) | 185 (98.4%) | 71 (95.9%) | 114 (100.0%) | 0.03 |
| All ADI-R criteria met, n (%) | 142 (75.9%) | 41 (56.2%) | 101 (88.6%) | <0.01 |

Values reported are means and standard deviations unless noted otherwise

¹ p-values are from chi-square tests for categorical variables and t-test for continuous variables

Cognitive Profiles

On the WISC-IV subtests administered, the ASD- sample demonstrated relatively greater performance on both the information ($p < 0.01$) and vocabulary ($p < 0.01$) subtests, but not the block design subtest ($p = 0.29$; see Table 5). Relatedly, individuals in the ASD- sample also obtained significantly higher estimated full scale IQ's ($M = 106.57$, $SD = 19.08$) compared to ASD+ group ($M = 97.64$, $SD = 21.59$; $p < 0.01$) although both group means fell in the average range of function.

Caregiver-Reported Anxiety Symptoms

Caregivers of ASD+ participants reported comparatively elevated symptoms of anxiety across a number of domains on the MASC-2, including the following indices and domains: the Anxiety Disorder Index ($p = 0.01$), Humiliation/Rejection ($p < 0.01$), Physical Symptoms ($p = 0.01$),

Table 5 Cognitive and Clinical instrument scores, by ASD diagnostic confirmation status

| | ASD diagnosis | | | p-value |
|---|----------------------------|---------------------------|---------------------------|---------|
| | Total (<i>N</i> = 232) | ASD- (<i>n</i> = 109) | ASD+ (<i>n</i> = 123) | |
| WISC-IV | <i>N</i> = 177 | <i>n</i> = 65 | <i>n</i> = 112 | |
| Block Design, scaled score | 11.02 (3.34) | 11.37 (3.37) | 10.82 (3.33) | 0.29 |
| Information, scaled score | 9.90 (3.78) | 10.71 (3.73) | 9.42 (3.74) | 0.03 |
| Vocabulary, scaled score | 10.21 (3.95) | 11.29 (3.74) | 9.56 (3.95) | < 0.01 |
| Estimated Full Scale IQ (FSIQ) | 100.92 (21.10) | 106.57 (19.08) | 97.64 (21.59) | < 0.01 |
| MASC | <i>N</i> = 135 | <i>n</i> = 52 | <i>n</i> = 81 | |
| MASC-2 total t-score | 57.70 (11.86) | 54.48 (10.65) | 59.84 (12.20) | 0.01 |
| CCC2 | <i>N</i> = 216 | <i>n</i> = 108 | <i>n</i> = 118 | |
| Standard GCC Score | 74.05 (11.39) | 75.01 (11.53) | 73.25 (11.25) | 0.26 |
| SIDI: Social Interaction Difference Index | -7.87 (9.17) | -8.81 (9.40) | -7.09 (8.94) | 0.17 |

Values reported are means and standard deviations unless noted otherwise

Separation Anxiety/Phobias ($p = 0.04$), Social Anxiety ($p = 0.01$), Tense/Restless ($p = 0.01$), and the MASC-2 total score ($p = 0.01$; see Table 5).

Caregiver-Reported Language Use

No caregiver-reported group differences in functional language use, as measured by the CCC-2, were observed in the current sample (p 's > 0.05; Table 5).

Discussion

To our knowledge, the current investigation is the first to utilize both the ADI-R and ADOS-2 to examine children who were ultimately excluded from ASD research but who presented with existing community-based ASD diagnoses. Strikingly, nearly half (47%) of the participants in the current study did not go on to meet rigorous diagnostic criteria for ASD in a research-based evaluation. Although investigations examining discrepancies between community- and research-based ASD diagnoses are scarce, our results build upon an investigation by Hausman-Kedem and colleagues (2018), in which nearly one-quarter of participants with a community-based ASD diagnosis were classified as non-autism spectrum based on ADOS-2 evaluation. A meta-analysis conducted by Lebersfeld and colleagues (2021) reported reduced specificity of the ADI-R and mixed findings regarding the accuracy of the ADOS-2 in clinical versus research settings. Given recent increased prevalence of ASD (Maenner et al., 2021, 2023; CDC, 2023) and widespread debates in the field about diagnosis, overdiagnosis, and concerns regarding interpretation of ADOS-2 and ADI-R scores in common clinical use (Bishop & Lord, 2023; Duvall et al., 2022; Fombonne, 2023), it is critical to understand possible over- and under-diagnosis of ASD and identify patterns or factors that may contribute to misdiagnosis of ASD.

The current data suggest few group differences regarding demographic characteristics (e.g., age, gender, race, ethnicity) in youth who met expert group consensus for confirmation of ASD and youth who were excluded from study consideration due to not meeting ASD criteria. When considering caregiver report tools in the current study, the SRS-2 was very high in both groups and did not differentiate groups, which may be related to the SRS-2's role as a screening instrument versus a diagnostic measure and underscores the importance of caution in interpreting report of ASD symptoms without comprehensive assessment and observation. This is also consistent with recent findings that have shown the SRS-2 to have decreased discriminant validity when used with psychiatrically complex patients, such as those with clinically significant anxiety (Capriola-Hall

et al., 2021; South et al., 2017) or ADHD symptomatology (Grzadzinski et al., 2016; Havdahl et al., 2016). In contrast with caregiver-derived reports, diagnostic tools that are examiner dependent (i.e., ADOS-2, ADI-R) revealed group differences that may inform diagnostic discrepancies. Specifically, higher ADOS-2 algorithm and CSSs were seen in ASD+ youth compared to ASD- youth. Results from the ADI-R also revealed differences in group membership on both item-level data and total scores, such that individuals classified as ASD+ had heightened ADI-R scores. Although the ADI-R showed group differences, it is important to note that, based on caregiver accounts, over half of the ASD- group also had elevated ADI-R total scores (56.2% of the ASD- group compared to 88.6% of the ASD+ group). Additionally, over 70% of the ASD- group obtained elevated scores on individual domain scores in the ADI-R for reciprocal social interaction, qualitative abnormalities in communication and restricted, repetitive and stereotyped behavior patterns. Of note, caregiver report of the timing of first ASD symptoms seems to provide limited value in differentiating groups as almost all in the ASD- group reported divergence in early development on the ADI-R (95.9%). In sum, this suggests that caregiver reports of ASD symptoms remain high in children who do not go on to meet diagnostic criteria for ASD. While inclusion of the ADI-R in addition to the ADOS-2 improves the accuracy of diagnostic outcomes (Risi et al., 2006; Ventola et al., 2006; Kim & Lord, 2012), in this study, when used alone, the ADOS-2 appears to provide more accurate differentiation between the team consensus categorizations than use of the ADI-R alone. In fact, through a meta-analysis, Lebersfeld and colleagues (2021) found the ADOS-2 to be the more stable and accurate assessment in both clinical and research settings when compared to the ADI-R.

No group differences emerged in caregiver-reported psychiatric diagnoses; however, the research evaluation team provided diagnoses of anxiety, ADHD, and disruptive behavior disorders relatively more frequently to ASD- than ASD+ participants. Thus, behaviors and clinical symptoms that the families, and perhaps community-based providers, may have conceptualized as ASD were felt to be better attributed to other psychiatric diagnoses. The ASD+ group demonstrated relatively elevated anxiety symptoms on a caregiver-reported measure of anxiety, which is consistent with previous research (e.g., Kent & Simonoff, 2017; Simonoff et al., 2008) and may reflect that the MASC-2 was more sensitive to rigidity in general (i.e., cognitive and/or behavioral), which can be associated with both anxiety and ASD.

It is also possible that the high rate of ASD- individuals within the current sample can be attributed, in part, to qualification for school-based services under the eligibility

classification of ASD which vary from state to state and over time. Youth who receive that classification at school may qualify with different criteria than DSM-5 (impact on learning as determining factor, less assessment for other conditions that may also result in social difficulties) and may not have completed a formal medical assessment (Laidler, 2005). It is possible that caregivers mistook an academic classification of ASD as a medical diagnosis and then, in some cases, carried this understanding forward to their medical providers, possibly resulting in the cognitive bias of diagnostic momentum (Streiner, 2021).

This distinction is particularly important given many widely cited prevalence studies of ASD rely on caregiver report (Blumberg et al., 2013) and school records (Maenner et al., 2021, 2023) of ASD diagnoses, suggesting that current rates may, in fact, overestimate the prevalence of ASD. Notably, a high percentage of individuals within the ASD- group were identified as having an ICD-9 code associated with ASD documented within their electronic medical record. Identifying ASD diagnoses from patient electronic medical records is a common recruitment strategy within ASD research and is also used regularly as an empirical outcome measure. For example, a large study investigating the accuracy of a commonly used ASD screening measure, the Modified Checklist for Autism in Toddlers with Follow-Up (M-CHAT/F), defined their outcome of later ASD diagnosis by the presence of prior visit diagnoses of ASD and/or ASD included within a patient's medical record "problem list" (Guthrie et al., 2019). While this approach enables researchers to examine ASD rates and the efficacy of screening instruments at a large scale without conducting individual ASD assessments for all participants, our current findings suggest using those metrics may falsely identify patients who do not truly meet diagnostic criteria for ASD, and it does not account for variable diagnostic stability (Wiggins et al., 2012; Woolfenden et al., 2012). Consequences of such misclassification can be more serious for etiological studies (e.g., genetics) or treatment efficacy studies. In clinical settings, over-identification of ASD may be following a similar trend of overdiagnosis across mental health conditions (especially in higher SES countries; Merten et al., 2017) and may further increase demand for already overextended ASD resource programs (e.g., county-based developmental disabilities and school-based services), thereby making it even more challenging for those who truly meet diagnostic criteria for ASD to access needed services (Pinals et al., 2022).

The current study has many strengths, including the use of a large sample of individuals who identified as having an existing diagnosis of autism spectrum disorder. Traditionally, in clinical settings, an ASD diagnosis is provided by one provider or a small interdisciplinary team of providers. The current study utilized a team of experts and

standardized assessment tools to come to diagnostic consensus to ensure accuracy. Most importantly, the current study addresses a critical gap in the literature regarding overdiagnosis and misdiagnosis. Our results highlight the importance of comprehensive assessment that includes utilizing an examiner-led standardized behavioral observation, such as the ADOS-2, in combination with developmental interviews and comprehensive record review to facilitate accurate diagnostic impressions, especially for individuals with age expected cognitive skills. Our results also suggest that while caregiver reports of autism spectrum symptoms remain critically useful in understanding an individual's unique strengths and challenges, overlapping traits or mimics may lead to inaccurate diagnoses. This may help to inform the utility and cost-effectiveness of instruments and measures employed within clinic settings.

The current study is not without its limitations. The sample comprised school age children seeking to participate in a neuroimaging study with an estimated FSIQ within the typically developing range as our participants required a certain level of communication in order to safely tolerate the scan without sedation; results may not apply in younger or more developmentally impaired individuals. The exclusion of children with medical complications precluding neuroimaging assessment or families reluctant to participate in this type of research may also impact our ability to extend findings to all individuals. Of note, data collection began in 2011 and spanned changes in diagnostic criteria and measure editions (e.g., WISC), additionally practices and trends in diagnosis have likely continued to shift over this time in both community and research settings. Thus, continued characterization of research samples (including children who volunteer for research with community-based diagnoses but do not meet ASD diagnostic threshold for inclusion) would be beneficial. Given the multiple study visits, changes in study protocol over the course of several years, and demanding nature of the study, missing data were unavoidable. While we do present inferential statistics in the form of bivariate analyses, we chose not to employ additional estimation strategies that could handle missing data, as the nature of this investigation was descriptive and not guided by a specific hypothesis. As noted, protocol changes reduced the administration of interviews/questionnaires, and this mainly affected the ASD- sample. Our examination of those impacted by this protocol change indicated these participants were no different than ASD- participants who fully participated; thus, providing some evidence that these protocol changes did not introduce bias in our estimates and the need to utilize additional statistical techniques. Despite this, missing data is a limitation and the implications related to ADI-R scores may be a particular consideration. Additionally, over 79% of participants in the current

study identified as Caucasian. Though this is generally representative of the geographic region in which the study was conducted, the homogeneity of race and ethnicity limits our understanding of race/ethnic differences in group membership and generalizability of findings. This is an important area for future research as currently the CDC notes that, although the racial gap in diagnosis of ASD appears to be narrowing for some select groups (e.g., Black Americans with co-occurring intellectual disability), early diagnosis remains poor across minority groups and disparities persist in Hispanic and bilingual groups broadly (Maenner et al., 2021, 2023). Similarly, socioeconomic status of participants remains largely unknown, further narrowing the conclusions that can be made. Despite these limitations, we believe the compensating strengths are unique in speaking to concerns regarding possible overdiagnosis of ASD given that nearly half of volunteers for this study presenting with school eligibility, community diagnoses and/or ASD documented in their medical chart ended up not meeting expert consensus for ASD using standardized tools.

Conclusions

Overall, clinicians conducting ASD evaluations should be comprehensive and not only consider using standardized examiner administered autism-specific measures to ensure higher diagnostic accuracy but also critically assess other possible psychiatric conditions that may mirror some ASD symptoms. Specifically, clinicians should consider the possible impact that other psychiatric disorders and symptoms (especially ADHD, anxiety and disruptive behaviors) may have on social communication and behaviors as caregiver reports of ASD symptoms remained high in many children who did not go on to meet diagnostic criteria for ASD. The bar for entering into research-based studies on ASD is high (expert consensus) and inclusion criteria are likely more stringent than the assessment methods commonly utilized in the community; thus, the utilization of data from medical records, community-based diagnoses and education-based eligibilities as proxies of ASD diagnosis in population-based studies should be considered cautiously.

Acknowledgements Approval was obtained from the institutional review board of Oregon Health & Science University. The procedures used in this study adhere to the tenets of the Declaration of Helsinki. The authors of this study would like to acknowledge all the members of Dr. Damien Fair's lab for their hard work on this study. We are also extremely grateful to all the families who took the time to participate in this study. This research was supported by the National Institute of Mental Health R01 MH096773 and R00MH091238 (Fair), the Simons Foundation #177894 (Nigg) and the Oregon Health & Science University Pediatric Biostatistics Pilot Program (Duvall). We acknowledge the editorial assistance of the Oregon Clinical & Translational Research Institute, which is supported by the National Center

for Advancing Translational Sciences, National Institutes of Health, through Grant Award Number UL1TR002369. Preliminary data were previously presented at the International Society for Autism Research Annual Meeting in Austin, TX, USA in 2022.

Declarations

Author Note Regarding Language The authors use both person-first and identity-first language interchangeably within the manuscript to respect diverse opinions within the autism community.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Conflict of interest The authors have no conflicts of interest, financial or non-financial disclosures to report.

References

- Bishop, D. (2006). *Children's communication Checklist—2 U.S. Edition*. The Psychological Corporation.
- Bishop, S. L., & Lord, C. (2023). Commentary: Best practices and processes for assessment of autism spectrum disorder – the intended role of standardized diagnostic instruments. *Journal Child Psychology and Psychiatry*, 64, 834–838. <https://doi.org/10.1111/jcpp.13802>.
- Blumberg, S. J., Bramlett, M. D., Kogan, M. D., Schieve, L. A., Jones, J. R., & Lu, M. C. (2013). *Changes in prevalence of parent-reported autism spectrum disorder in school-aged US children: 2007 to 2011*– (Vol. 2012). US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. (No. 65).
- Capriola-Hall, N. N., McFayden, T., Ollendick, T. H., & White, S. W. (2021). Caution when screening for autism among socially anxious youth. *Journal of Autism and Developmental Disorders*, 51(5), 1540–1549. <https://doi.org/10.1007/s10803-020-04642-w>.
- Center for Disease Control and Prevention (2023). Community Report on Autism.
- Cicchetti, D. V., Lord, C., Koenig, K., Klin, A., & Volkmar, F. R. (2008). Reliability of the ADI-R: Multiple examiners evaluate a single case. *Journal of Autism and Developmental Disorders*, 38(4), 764–770. <https://doi.org/10.1007/s10803-007-0448-3>.
- Constantino, J. N., & Gruber, C. P. (2012). The social responsiveness scale (2nd edition). Los Angeles: Western Psychological Services.
- Constantino, J. N., & Todd, R. D. (2000). Genetic structure of reciprocal social behavior. *American Journal of Psychiatry*, 157(12), 2043–2045. <https://doi.org/10.1176/appi.ajp.157.12.2043>.
- Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., & Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders*, 33(4), 427–433. <https://doi.org/10.1023/A:1025014929212>.
- Centers for Disease Control and Prevention (CDC) Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators (2014). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, 63(2), 1–21. <https://www.jstor.org/stable/24806108>.
- Dolata, J. K., Suarez, S., Calame, B., & Fombonne, E. (2022). Pragmatic language markers of autism diagnosis and severity. *Research in Autism Spectrum Disorders*, 94, 101970. <https://doi.org/10.1016/j.rasd.2022.101970>.
- Duvall, S., Armstrong, K., Shahabuddin, A., Grantz, C., Fein, D., & Lord, C. (2022). A road map for identifying autism spectrum disorder: Recognizing and evaluating characteristics that should raise red or pink flags to guide accurate differential diagnosis. *The Clinical Neuropsychologist*, 36(5), 1172–1207. <https://doi.org/10.1080/13854046.2021.1921276>.
- Fombonne, E. (2018). Editorial: The rising prevalence of autism. *Journal of Child Psychology and Psychiatry*, 59(7), 717–720. <https://doi.org/10.1111/jcpp.12941>.
- Fombonne, E. (2023). Editorial: Is autism overdiagnosed? *Journal of Child Psychology and Psychiatry*, 64, 711–714. <https://doi.org/10.1111/jcpp.13806>.
- Fombonne, E., MacFarlane, H., & Salem, A. C. (2021). Epidemiological surveys of ASD: Advances and remaining challenges. *Journal of Autism and Developmental Disorders*, 51(12), 4271–4290. <https://doi.org/10.1007/s10803-021-05005-9>.
- Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(5), 693–705. <https://doi.org/10.1007/s10803-008-0674-3>.
- Graf, W. D., Miller, G., Epstein, L. G., & Rapin, I. (2017). The autism epidemic ethical, legal, and social issues in a developmental spectrum disorder. *Neurology*, 88(14), 1371–1380. <https://doi.org/10.1212/WNL.0000000000003791>.
- Greene, R. K., Vasile, I., Bradbury, K. R., Olsen, A., & Duvall, S. W. (2022). Autism Diagnostic Observation schedule (ADOS-2) elevations in a clinical sample of children and adolescents who do not have autism: Phenotypic profiles of false positives. *The Clinical Neuropsychologist*, 36(5), 943–959.
- Grzadzinski, R., Dick, C., Lord, C., & Bishop, S. (2016). Parent-reported and clinician-observed autism spectrum disorder (ASD) symptoms in children with attention deficit/hyperactivity disorder (ADHD): Implications for practice under DSM-5. *Molecular Autism*, 7(1), 1–12. <https://doi.org/10.1186/s13229-016-0072-1>.
- Guthrie, W., Wallis, K., Bennett, A., Brooks, E., Dudley, J., Gerdes, M., & Miller, J. S. (2019). Accuracy of autism screening in a large pediatric network. *Pediatrics*, 144(4). <https://doi.org/10.1542/peds.2018-3963>.
- Hausman-Kedem, M., Kosofsky, B. E., Ross, G., Yohay, K., Forrest, E., Dennin, M. H., & Ward, M. J. (2018). Accuracy of reported community diagnosis of autism spectrum disorder. *Journal of Psychopathology and Behavioral Assessment*, 40(3), 367–375. <https://doi.org/10.1007/s10862-018-9642-1>.
- Havdahl, K. A., Bal, V. H., Huerta, M., Pickles, A., Øyen, A. S., Stoltenberg, C., & Bishop, S. L. (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. <https://doi.org/10.1016/j.jaac.2016.09.490>.
- Hill, A. P., Zuckerman, K., & Fombonne, E. (2015). Epidemiology of autism spectrum disorders. *Translational approaches to Autism Spectrum Disorder* (pp. 13–38). Springer International Publishing.
- Kent, R., & Simonoff, E. (2017). Prevalence of anxiety in autism spectrum disorders. In: *Anxiety in Children and Adolescents with Autism Spectrum Disorder*. Elsevier, pp5–32. <https://doi.org/10.1016/B978-0-12-805122-1.00002-8>.
- Kim, S. H., & Lord, C. (2012). Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Child Psychology and Psychiatry*, 53(2), 143–151. <https://doi.org/10.1111/j.1469-7610.2011.02458.x>.
- La Roche, M. J., Bush, H. H., & D'Angelo, E. (2018). The assessment and treatment of autism spectrum disorder: A cultural

- examination. *Practice Innovations*, 3(2), 107–122. <https://doi.org/10.1037/pri0000067>.
- Laidler, J. R. (2005). US Department of Education Data on Autism Are Not Reliable for *Tracking Autism Prevalence Pediatrics*, 116(1): e120–e124. doi: <https://doi.org/10.1542/peds.2004-2341>.
- Lebersfeld, J. B., Swanson, M., Clesi, C. D., & O’Kelley, S. E. (2021). Systematic review and meta-analysis of the clinical utility of the ADOS-2 and the ADI-R in diagnosing autism spectrum disorders in children. *Journal of Autism and Developmental Disorders*, 51, 1–14. <https://doi.org/10.1007/s10803-020-04839-z>.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012a). *Autism diagnostic observation schedule: ADOS-2*. Western Psychological Services.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012b). *ADOS-2. Autism Diagnostic Observation schedule. Manual (part 1): Modules 1–4*. Western Psychological Services Los Angeles, CA.
- Luciano, C. C., Keller, R., Politi, P., Aguglia, E., Magnano, F., & Burti, L. (2014). Misdiagnosis of high function autism spectrum disorders in adults: An Italian case series. *Autism Open Access*, 4(131). <https://doi.org/10.4172/2165-7890.1000131>.
- Maenner, M. J., Shaw, K. A., Bakian, A. V., Bilder, D. A., Durkin, M. S., Esler, A., Furnier, S. M., Hallas, L., Hall-Lande, J., Hudson, A., Hughes, M. M., Patrick, M., Pierce, K., Poynter, J. N., Salinas, A., Shenouda, J., Vehorn, A., Warren, Z., Constantino, J. N., DiRienzo, M., & Cogswell, M. E. (2021). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. *Morbidity and Mortality Weekly Report. Surveillance summaries (Washington, D.C.: 2002)*, 70(11), 1–16. <https://doi.org/10.15585/mmwr.ss7011a1>.
- Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. V., Bilder, D. A., & Shaw, K. A. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveillance Summaries*, 72(2), 1. <https://doi.org/10.15585/mmwr.ss7202a1>.
- March, J. S. (2012). *Manual for the multidimensional anxiety scale for children (MASC 2)*. Multi-Health Systems.
- Merten, E. C., Cwik, J. C., Margraf, J., & Schneider, S. (2017). Overdiagnosis of mental disorders in children and adolescents (in developed countries). *Child and Adolescent Psychiatry and Mental Health*, 11(1), 5. <https://doi.org/10.1186/s13034-016-0140-5>.
- Pinals, D. A., Hovermale, L., Mauch, D., & Anacker, L. (2022). Persons with intellectual and developmental disabilities in the mental health system: Part 2. Policy and systems considerations. *Psychiatric Services*, 73(3), 321–328. <https://doi.org/10.1176/appi.ps.201900505>.
- 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 & Adolescent Psychiatry*, 45(9), 1094–1103. <https://doi.org/10.1097/01.chi.0000227880.42780.0e>.
- Rutter, M., Le Couteur, A., & Lord, C. (2003). *The ADI-R: Autism diagnostic interview revised: Manual (WPS)*. Western Psychological Services.
- Sattler, J. M., & Dumont, R. (2004). *Assessment of children: WISC-IV and WPPSI-III supplement*. J. M. Sattler Publishing Company.
- Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(8), 921–929. <https://doi.org/10.1097/CHI.0b013e318179964f>.
- Skellern, C., Schluter, P., & McDowell, M. (2005). From complexity to category: Responding to diagnostic uncertainties of autistic spectrum disorders. *Journal of Paediatrics and Child Health*, 41(8), 407–412. <https://doi.org/10.1111/j.1440-1754.2005.00634.x>.
- South, M., Carr, A. W., Stephenson, K. G., Maisel, M. E., & Cox, J. C. (2017). Symptom overlap on the SRS-2 adult self-report between adults with ASD and adults with high anxiety. *Autism Research*, 10(7), 1215–1220. <https://doi.org/10.1002/aur.1764>.
- Streiner, D. L. (2021). Decision-Making in Psychological Assessment. In *The Oxford Handbook of Personality Assessment, Second Edition*. <https://doi.org/10.1093/oxfordhb/9780190092689.013.32>.
- van’t Hof, M., Tisseur, C., van Berckeleer-Onnes, I., van Nieuwenhuyzen, A., Daniels, A. M., Deen, M., Hoek, H. W., & Ester, W. A. (2021). Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. *Autism*, 25(4), 862–873. <https://doi.org/10.1177/1362361320971107>.
- Van Schalkwyk, G. I., Peluso, F., Qayyum, Z., McPartland, J. C., & Volkmar, F. R. (2015). Varieties of misdiagnosis in ASD: An illustrative case series. *Journal of Autism and Developmental Disorders*, 45(4), 911. <https://doi.org/10.1007/s10803-014-2239-y>.
- 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(7), 839–847. <https://doi.org/10.1007/s10803-006-0128-8>.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition: Administrative and scoring manual*. The Psychological Corporation.
- Wiggins, L. D., Baio, J., Schieve, L., Lee, L. C., Nicholas, J., & Rice, C. E. (2012). Retention of autism spectrum diagnoses by community professionals: Findings from the autism and developmental disabilities monitoring network, 2000 and 2006. *Journal of Developmental and Behavioral Pediatrics*, 33(5), 387–395. <https://doi.org/10.1097/DBP.0b013e3182560b2f>.
- Woolfenden, S., Sarkozy, V., Ridley, G., & Williams, K. (2012). A systematic review of the diagnostic stability of autism spectrum disorder. *Research in Autism Spectrum Disorders*, 6(1), 345–354. <https://doi.org/10.1016/j.rasd.2011.06.008>.
- Zuckerman, K. E., Mattox, K., Donelan, K., Batbayar, O., Baghaee, A., & Bethell, C. (2013). Pediatrician identification of latino children at risk for autism spectrum disorder. *Pediatrics*, 132(3), 445–453. <https://doi.org/10.1542/peds.2013-0383>.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.