Title

Accuracy of initial diagnostic impressions of autism in toddlers and behaviors that inform these impressions.

Journal

Autism Research, 17(3)

Authors

Milan, Stephanie

Burke, Jeffrey

Barton, Marianne

et al.

Publication Date

2024-03-01

DOI

10.1002/aur.3088

Peer reviewed

HHS Public Access

Author manuscript

Autism Res. Author manuscript; available in PMC 2024 April 18.

Published in final edited form as:

Autism Res. 2024 March ; 17(3): 568–583. doi:10.1002/aur.3088.

Accuracy of initial diagnostic impressions of autism in toddlers and behaviors that inform these impressions

Rebecca P. Thomas, MA, MS[1], **Ashley de Marchena, PhD**[2], **Andrea Trubanova Wieckowski, PhD**[2], **Aubyn Stahmer, PhD**[3], **Stephanie Milan, PhD**[1], **Jeffrey D. Burke, PhD**[1], **Marianne L. Barton, PhD**[1], **Diana L. Robins, PhD**[2], **Deborah A. Fein, PhD**[1]

[1]Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut,

[2]A.J. Drexel Autism Institute, Drexel University, Philadelphia, Pennsylvania,

[3]Department of Psychiatry and Behavioral Sciences, University of California, Davis, MIND Institute, Sacramento, California

Abstract

Clinicians form initial impressions about a child's diagnosis based on behavioral features, but research has not yet identified specific behaviors to guide initial diagnostic impressions. Participants were toddlers (N=55, mean age 22.9 months) from a multi-site early detection study, referred for concern for ASD due to screening or parent/provider concern. Within five minutes of meeting a child, clinicians noted ASD or non-ASD impression, confidence in impression, and behaviors that informed their impression. These clinicians also determined final diagnoses for each child. When a child's final diagnosis was ASD (n=35), senior clinicians formed an initial impression of ASD in 22 cases (63%) but missed 13 cases (37%). When final diagnosis was non-ASD (n=20), senior clinicians made an initial impression of non-ASD in all cases (100%). Results were similar among junior clinicians. Senior and junior clinicians used the same behaviors to form accurate impressions of ASD and non-ASD: social reciprocity, nonverbal communication, and eye contact. Senior clinicians additionally used focus of attention when forming accurate impressions of ASD and non-ASD; junior clinicians used this behavior only when forming accurate non-ASD impressions. Clinicians' initial impressions of ASD are very likely to be consistent with final diagnoses, but initial impressions of non-ASD need follow-up. Toddlers who show all four atypical behaviors (social reciprocity, nonverbal communication, eye contact, focus of attention) might receive expedited ASD diagnoses. However, presence of apparently typical behaviors should not rule out ASD; for some children a longer evaluation is necessary to allow for more opportunities to observe subtle social behavior.

Lay Summary:

Clinicians form first impressions about a child's diagnosis based on observations of the child's behavior. Although 2/3 of toddlers with ASD were identifiable within the first five minutes of

Corresponding author: Rebecca P. Thomas, rebecca.p.thomas@uconn.edu.

Potential Conflict of Interest: Drs. Robins, Barton, and Fein are co-owners of M-CHAT LLC, which licenses use of the M-CHAT in electronic products. Dr. Robins sits on the advisory board of Quadrant Biosciences Inc. and she also serves on the program quality committee at Bancroft. The other authors have indicated they have no potential conflicts of interest to disclose.

observation, the remaining 1/3 would have been missed without a full evaluation. Clinicians frequently used the same four behaviors to decide on their first impressions of ASD and non-ASD: social reciprocity, nonverbal communication, eye contact, and focus of attention. Toddlers who show atypical behavior in all four areas might receive faster ASD diagnoses, which supports earlier access to autism services.

Keywords

autism spectrum disorder; diagnostic decision-making; diagnostic confidence; toddlers; initial impressions

Diagnoses of autism spectrum disorder (ASD) in children are based on the reduction of expected social and communication behaviors, and presence of restricted/repetitive behaviors. A combination of clinician observation of the child and parent report of current and historical developmental and adaptive functioning is used by clinicians to make a diagnosis. Although most parents or pediatricians raise concerns related to autism before the child is three years old, the time between first concern and formal diagnosis can extend as long as three years (Constantino et al., 2020; Martinez et al., 2018). One persisting question is whether brief behavioral observations could expedite the time to diagnosis (Wieckowski et al., 2021), leading to earlier intervention.

Prior research has established high inter-rater agreement for the presence of autism-related behaviors during 10–15 minute observations (Constantino et al., 2007, 2012; Gabrielsen et al., 2015). Research has also shown that clinicians across varying levels of training and experience rapidly form initial impressions about a child's diagnosis (de Marchena et al., 2023; de Marchena & Miller, 2017; Wieckowski et al., 2021), which may have direct implications for how children are referred and evaluated. A recent survey found that clinicians estimate that 40% of individuals with autism have a "frank" presentation of autism (de Marchena & Miller, 2017) that may be detected via specific behavioral markers, such as unusual body mannerisms or atypical prosody. The term "frank presentation" has been used to describe readily identifiable presentations, or profiles associated with unmistakable autistic features. Gabrielsen and colleagues (2017) found that children with autism made significantly more repetitive sounds and had more atypical quality of vocalizations than children with language disorders in a 10-minute observation. However, research has not examined how specific behaviors relate to accurate, or inaccurate, initial diagnostic impressions of ASD and non-ASD.

The *accuracy* of overall initial impressions of ASD and non-ASD in toddlers, however, has been examined. The initial clinical impression of ASD in children matched a final diagnosis of ASD in 92% of cases, whereas an initial impression of non-ASD matched a final diagnosis of non-ASD in 76% of cases (Wieckowski et al., 2021). Accordingly, the sensitivity of detecting ASD in an initial five-minute encounter was only moderate (0.64) whereas the specificity was high (0.96). Similar rates of sensitivity and specificity were found for initial diagnostic impressions among trainee clinicians (de Marchena et al., 2023) and for brief 10-minute observations used to refer a child for evaluation (Gabrielsen et al., 2017). Overall, these findings suggest that an initial impression of ASD was likely to match

a final diagnosis of ASD, whereas an initial impression of non-ASD often warranted further examination.

In addition to identifying whether initial impressions of ASD matched final diagnoses of ASD, research has examined the role of clinician certainty in diagnostic impressions and final diagnoses (McDonnell et al., 2019; Wieckowski et al., 2021). McDonnell and colleagues (2019) found greater clinician certainty in diagnoses of ASD than non-ASD, which contrasts with Wieckowski and colleagues' (2021) finding that clinicians had the highest confidence in their initial impression when both their initial impression and final diagnosis for a child were non-ASD. One potential explanation for this discrepancy is the age of the children: the sample in Wieckowski et al. (2021) had a mean age of 21 months of age, while the sample in McDonnell et al. (2019) had an average age of 43 months. It may be that the behavioral presentation of ASD is clearer in older children than in those under two years of age.

Existing research, however, has not reached a consensus on the impact of child age on confidence in diagnosis or accuracy of impression. While some studies have found no relationship between child age and clinician confidence in diagnosis (Hedley et al., 2016), a recent study by McDonnell et al. (2019) identified a negative correlation between child age and clinician certainty, with more certainty for diagnoses made in children *under* 24 months. However, for accuracy of clinician impression, Gabrielsen et al. (2017) found that clinicians made more correct ASD referral decisions for children *over* 24 months of age (mean age: 24.7 months) than for younger children (mean age: 20.2 months) based on 10-minute observations. These results suggest that accuracy and confidence in diagnostic judgments may differ by child age.

Child sex, race, parental education, and markers of socioeconomic status apparently do not relate to confidence in initial impression or confidence in final diagnosis (McDonnell et al., 2019; Wieckowski et al., 2021). Behavioral features of autism (as measured by the ADOS-2), but not cognitive level, were associated with the accuracy of initial impression (Wieckowski et al., 2021). In contrast, prior research identified less clinician certainty in diagnoses of ASD for children who were older (mean age: 43 months) *and* with higher cognitive functioning (McDonnell et al., 2019), suggesting an interaction between age and developmental functioning.

Research suggests that clinicians form initial impressions about a child's diagnosis based on specific behavioral features (de Marchena & Miller, 2017); confidence in initial impression and diagnosis may differ based not only on diagnosis (ASD vs. non-ASD) but also on clinical characteristics (McDonnell et al., 2019; Wieckowski et al., 2021). Which specific behaviors guide accurate and inaccurate initial impressions of ASD and non-ASD in toddlers is not known. This question has implications for a theoretical understanding of the heterogeneity of ASD, specifically the question of whether children for whom clinicians form accurate initial diagnostic impressions show similar behaviors during brief observations, and whether the accuracy of initial impressions differs based on age or developmental delay.

The goals of the proposed study are (1) to replicate findings of the overall accuracy of senior and junior clinicians' initial impression of ASD and non-ASD, as well as the sensitivity and specificity of initial impression and confidence in these impressions in a new sample, (2) to explore other child characteristics (Developmental Quotient (DQ), age, sex, race) that may relate to the accuracy of initial impressions of ASD and non-ASD, (3) to identify the specific behaviors that contribute most to accurate and inaccurate ASD and non-ASD initial impressions, and (4) to compare senior and junior clinicians' accuracy and the behaviors used to form initial impressions.

We hypothesized that (1) sensitivity and specificity values among junior and senior clinicians would replicate those of Gabrielsen and colleagues (2017), Wieckowski and colleagues (2021), and de Marchena et al. (2023); (2) senior and junior clinicians would be more confident in accurate (than inaccurate) initial impressions (de Marchena et al., 2023; Wieckowski et al., 2021); (3) child sex and race would not relate to accuracy of initial diagnostic impression (McDonnell et al., 2019; Wieckowski et al., 2021), whereas higher DQ and younger age would lead to more inaccurate impressions of non-ASD (replicating Gabrielsen et al., 2017); and (4) the subset of diagnostically-relevant behaviors that guide accurate initial impressions of ASD and non-ASD for junior and senior clinicians would include social reciprocity and other observable ASD behaviors, such as motor mannerisms and unusual prosody (Gabrielsen et al., 2017); and junior clinicians would use fewer behaviors than senior clinicians to form initial impressions.

Methods

Participants and Procedure

Participants were toddlers in a multi-site study on early detection and intervention in ASD; all toddlers were referred for ASD concerns after positive screening and/or surveillance (McClure et al., 2021). This study extends work from the investigative team (de Marchena et al., 2023; Wieckowski et al., 2021), but its sample of children does not overlap with the prior samples. Initial Diagnostic Impression Checklists were completed for 55 children (final diagnoses: ASD: n=35; non-ASD: n=20). The non-ASD group included children with No Diagnosis (n=4), Language Delays (n=9), and Global Developmental Delay (n=7).

The ASD and non-ASD groups were not significantly different on race, ethnicity, maternal education, chronological age, or sex ratio (see Table 1). The mean age of the sample was 24.1 months (SD = 8.5 months, range = 15.2–53.1 months). The sample was diverse in race and ethnicity. The ASD group had significantly more impairment than the non-ASD group on all standardized measures (Vineland Adaptive Behavior Scales-3rd edition [VABS-3], Mullen Scales of Early Learning [MSEL]) except motor and daily living skills (Table 1).

Evaluation Procedure

All participants received a diagnostic evaluation at one of three sites (Drexel University, UC Davis, University of Connecticut). IRB approval was provided by Drexel; other sites relied on Drexel's IRB. Due to the COVID-19 pandemic, adaptations for six evaluations conducted between March 2020 and the end of 2022 included masking (parents, clinicians,

and children to the extent children tolerated masks), and physical distance. Excluding these children from analyses did not change the results. Clinicians abided by CDC protocols for COVID Personal Protective Equipment (PPE) and wore face masks and clear face shields to protect their eyes. Physical distancing was not standardized as clinicians were encouraged to maintain safe distances from other individuals in the room while completing an evaluation that was as close as possible to the normed standardization.

Evaluations were conducted by teams of senior and junior clinicians. Nine of the 13 senior clinicians were licensed PhD-level psychologists; others were a physician, a social worker, an occupational therapist, and a PhD/BCBA. Senior clinicians ranged in years of ASD diagnostic experience from 10 to 30+ years. Junior clinicians were graduate students in clinical psychology PhD programs or post-baccalaureate or post-master's research assistants. The senior clinician made the final Clinical Best Estimate diagnosis with input from the junior clinician. Impressions were considered accurate when they were consistent with final (Clinical Best Estimate) diagnosis that integrated clinical interviews with parents, comprehensive history, observation, and direct testing of the child with standardized assessments of cognitive functioning, play, and communication; inaccurate impressions differed from final diagnosis.

Measures

Demographic Information—Caregivers completed a comprehensive history form including information on child demographics, including age, sex, race, ethnicity, and socioeconomic status. The history form also collected information on pregnancy and birth history, developmental history (early milestones), and intervention/educational history as relevant.

Initial Diagnostic and Behavioral Impressions

Initial Diagnostic Impressions.: Clinicians completed an Initial Diagnostic Impressions Checklist (©Thomas, Wieckowski, Stahmer, Barton, Robins, Fein 2023; Appendix A) after the initial five minutes of meeting and interacting with the child (settings included waiting room, the walk to the evaluation room, exploring the evaluation room). Although clinicians were asked to indicate the setting used to form their initial impression, this information was only completed for three children (and all selected "exploring the evaluation room"); therefore data on the setting is considered missing. Due to the study design (see McClure et al., 2021), clinicians were unaware of whether the child was referred due to positive screen or pediatrician surveillance and had limited access to information about the child prior to testing, other than that the child was referred for possible ASD. In addition to providing their diagnostic impression (ASD or non-ASD), clinicians indicated their confidence in their impression (1–5 Likert Scale from 1 – not very confident – to 5 – extremely confident). After completing the evaluation, the clinician determined the child's final diagnosis and rated their confidence in the final classification (ASD or non-ASD) on a 1–5 Likert scale.

Contributing Behaviors.: The checklist included a list of behaviors in seven domains: social reciprocity, nonverbal communication strategies, eye contact, motor mannerisms, prosody and vocalizations, facial expressions, and focus of attention. Each item included

an atypical and typical option. For analysis, each behavior was classified as “atypical,” “typical,” or “not judged.” Clinicians only selected a typical or atypical behavior if it contributed to their initial diagnostic impression; otherwise they left that item blank. For example, the options for social reciprocity were “impaired responses to social reciprocity” or “good interactions with others.” These behaviors were based on constructs in the “frank” autism presentation (de Marchena & Miller, 2017) and discussions with clinical teams at participating sites about the behaviors they use most readily in forming diagnostic impressions. A review process among participating clinical teams narrowed down the list of behaviors based on majority consensus. The original Initial Diagnostic Impressions Checklist reported on in prior work (e.g., de Marchena et al., 2023; Wieckowski et al., 2021) was adapted for the current study to include the behaviors.

Cognitive Assessment—The Mullen Scales of Early Learning (MSEL; Mullen, 1995) is a norm-referenced, standardized instrument assessing cognitive and motor abilities in children up to 68 months. Four MSEL subscales were collected: Visual Reception, Fine Motor, Expressive Language, and Receptive Language. Developmental quotients (DQs) were calculated for each child to avoid potential floor effects for the Early Learning Composite (ELC), using the average age equivalent from Visual Reception, Fine Motor, Expressive Language, and Receptive Language divided by chronological age × 100 (Bishop et al., 2015).

ASD Assessment—The Autism Diagnostic Observation Schedule – 2nd Edition (ADOS-2; Lord et al., 2012) is a semi-structured, standardized assessment of communication, social interaction, and play skills designed to evaluate symptoms of ASD and produce scores for Social Affect, Restrictive and Repetitive Behavior and a total score. Fifty participants received the ADOS-2 Toddler Module, two participants received the ADOS-2 Module 1 and three received the ADOS-2 Module 2. Six participants did not have ADOS-2 scores entered due to COVID-19 procedures (child and evaluator masked), although the observation was used to gather clinical information about the child and contribute to the final diagnosis. Severity of ASD symptoms (Total, Social Affect, Restricted and Repetitive Behaviors) was measured with the ADOS-2 Calibrated Severity Scores (CSS; Gotham et al., 2009; see Table 1).

Adaptive Behavior Assessment—The Vineland Adaptive Behavior Scales, 3rd Edition, Parent Interview Form (VABS-3; Sparrow et al., 2016) is a semi-structured parent interview of adaptive behavior in the domains of Communication, Daily Living, Socialization, and Motor Skills. An overall Adaptive Behavior Composite (ABC) is a standard score derived from the domain-level standard scores and was used to characterize the sample.

Results

Accuracy, Sensitivity, and Specificity of Initial Impressions: Senior Clinicians

An exact McNemar’s test indicated that there was a statistically significant match between senior clinicians’ initial impressions and final diagnoses (p<.001). When a child’s final diagnosis was ASD (n=35), senior clinicians formed an initial impression of ASD in 22

cases (sensitivity = 0.63; Figure 1a). Impressions after the first five minutes of observation of the child, therefore, missed 37% (n=13) of cases for whom the final diagnosis was ASD. When a child's final diagnosis was non-ASD (n=20), senior clinicians made an initial impression of non-ASD in all cases (specificity = 1.00). No child had initial impressions of ASD but a final diagnosis of non-ASD.

Of the children for whom senior clinicians formed an initial impression of ASD (n=22), the initial impression matched the final diagnosis in all cases (n=22), making the positive predictive value (PPV) in this sample 1.00. Of the children for whom senior clinicians formed an initial impression of non-ASD (n=33), the initial impression matched the final diagnosis in 60.6% of cases (n=20), making the negative predictive value (NPV) 0.61.

Accuracy, Sensitivity, and Specificity of Initial Impressions: Junior Clinicians

Of the 55 children included in the senior clinician sample, 46 also had initial diagnostic impression checklists completed by a junior clinician (ASD = 29, non-ASD = 17).

When a child's final diagnosis was ASD (n=29), junior clinicians formed an initial impression of ASD in 17 cases (sensitivity = 0.59; Figure 1b). Impressions after the first five minutes of observation of the child, therefore, missed 41% (n=12) of cases for whom the final diagnosis was ASD. When a child's final diagnosis was non-ASD (n=17), clinicians made an initial impression of non-ASD in 16 cases (specificity = 0.94). One child had initial impressions of ASD but a final diagnosis of non-ASD. Of the children for whom the junior clinicians formed an initial impression of ASD (n=18), the initial impression matched the final diagnosis in all but one case (n=17), making the positive predictive value (PPV) in this sample of junior clinicians 0.94. Of the children for whom clinicians formed an initial impression of non-ASD (n=28), the initial impression matched the final diagnosis in 57.1% of cases (n=16), making the negative predictive value (NPV) 0.57.

Comparison of Accuracy: Junior and Senior Clinicians

Senior and junior clinicians showed large agreement on their initial impressions for the 46 cases seen by both clinicians (κ = .601, p<.001) with 83.6% agreement. Chi-square tests indicated no significant differences in the accuracy of senior versus junior clinicians for children with final diagnoses of ASD ($\chi^2(1, N=26)=2.622$, p=.105). Of the children with final diagnoses of non-ASD who were seen by both a senior and junior clinician, clinicians formed accurate initial impressions of non-ASD in all but one case in which the junior clinician formed an inaccurate initial impression of ASD.

Confidence, Initial Impressions, and Final Diagnoses: Senior Clinicians

Considering only initial impressions, senior clinicians were not significantly different in confidence in initial impressions of ASD (M = 2.8, SD = 1.3) and non-ASD (M = 2.7, SD = 1.0; $t(52)$ = .099, p=.921, d=.028). When divided into three groups based on the consistency of impression with final diagnosis, there was a significant difference in confidence in initial impression, $F(2, 51)$ = 3.191, p = .049, $\eta 2$ = .111 (see Table 2). Post-hoc analyses indicated that senior clinicians were significantly more confident when forming *accurate* non-ASD impressions (M = 2.7, SD = 1.0) than *inaccurate* non-ASD impressions (M = 2.1, SD =

1.0, $t(29) = 1.785$, $p=.042$, $d=.658$). Senior clinicians were equally confident when forming accurate ASD impressions ($M = 3.1$, $SD = 1.4$) and accurate non-ASD impressions ($M = 2.7$, $SD = 0.99$, $t(40)=1.052$, $p=.299$, $d=.326$).

Collapsed across diagnostic group and accuracy of impression, a paired samples t-test indicated that senior clinicians had significantly higher confidence in final diagnoses than initial impressions ($t(33) = -6.321$, $p<.001$, $d=1.1$). Senior clinicians were equally confident in their final diagnoses of ASD ($M = 4.2$, $SD = 1.1$) and non-ASD ($M = 4.6$, $SD = 0.5$; $t(31) = -1.036$, $p=.308$, $d=.383$). When senior clinicians formed accurate non-ASD initial impressions, they were significantly more confident in their final diagnoses ($M = 4.55$, $SD = 0.522$) than when their initial impressions were inaccurate ($M = 3.50$, $SD = 1.291$; $t(13) = 2.322$, $p=.019$, $d=1.356$).

Confidence, Initial Impressions, and Final Diagnoses: Junior Clinicians

Considering only initial impressions, junior clinicians were not significantly different in confidence in initial impressions of ASD ($M = 2.8$, $SD = 1.1$) and non-ASD ($M = 2.6$, $SD = 1.2$; $t(43) = .516$, $p=.609$, $d=.159$). When divided into three groups based on the consistency of impression with final diagnosis, there was a significant difference in confidence in initial impression, $F(2, 43) = 4.664$, $p = .015$, $\eta 2 = .178$ (see Table 2). Post-hoc analyses indicated that junior clinicians were significantly more confident when forming *accurate* non-ASD impressions ($M = 3.5$, $SD = 0.9$) than *inaccurate* non-ASD impressions ($M = 1.6$, $SD = 0.7$, $t(23) = 5.459$, $p<.001$, $d=2.229$).

Collapsed across diagnostic group and accuracy of impression, a paired samples t-test indicated that junior clinicians had significantly higher confidence in final diagnoses than initial impressions ($t(27) = -6.004$, $p<.001$, $d=1.135$). Junior clinicians were equally confident in their final diagnoses of ASD ($M = 4.8$, $SD = .42$) and non-ASD ($M = 3.8$, $SD = 1.3$; $t(26) = 2.466$, $p=.021$, $d=.973$).

Child Characteristics

In logistic regression models, DQ was not a significant predictor of accuracy of impression (AOR = 1.039, 95% CI: .985–1.094, $p=.158$). Each MSEL subscale was entered into logistic regression models; none significantly predicted accuracy of impression (Expressive Language: AOR = .829, 95% CI: .660–1.041, $p = .106$; Receptive Language: AOR = 1.239, 95% CI: .972–1.581, $p = .084$; Visual Reception: AOR = .923, 95% CI: .720–1.184, $p=.530$; Fine Motor: AOR = .923; 95% CI: .810–1.308, $p=.814$).

Although the groups did not differ on the distribution of sex or race, exploratory analyses were conducted to identify whether the proportion of accurate and inaccurate impressions differed by diagnosis in these groups. Fisher's exact tests indicated no significant differences in accuracy of impression based on race ($p=.999$) or sex ($p=.750$).

In order to examine the effect of age, the sample was divided using a median split of the age at the time of the evaluation (*Median* = 21.5 months). Among the younger group ($n=30$), the average age was 20.1 months ($SD = 1.4$ months). When a younger child's final diagnosis was ASD ($n=17$), senior clinicians formed an initial impression of ASD in 58.8% of cases

(*n*=10). When a younger child's final diagnosis was non-ASD (*n*=13), senior clinicians made an initial impression of non-ASD in 100% of cases (*n*=13). An exact McNemar's test indicated that there was a statistically significant match between senior clinicians' initial impressions and final diagnoses in younger children (*p*=.016).

Among the older group (*n*=25), the average age was 29.0 months (SD = 10.7 months). When an older child's final diagnosis was ASD (*n*=18), senior clinicians formed an initial impression of ASD in 66.7% of cases (*n*=12). When an older child's final diagnosis was non-ASD (*n*=7), senior clinicians made an initial impression of non-ASD in 100% of cases (*n*=7). An exact McNemar's test indicated that there was a statistically significant match between senior clinicians' initial impressions and final diagnoses in older children (*p*=.031). A McNemar test indicated no significant difference in sensitivity between older and younger toddlers (.59 and .67, respectively; $p = .238$).

A one-way ANOVA revealed significant differences in ADOS-2, VABS-3 ABC, and MSEL DQ between the three accuracy groups (accurate non-ASD impressions, inaccurate non-ASD impressions, accurate ASD impressions; Table 2). Tukey's HSD test for multiple comparisons found that the accurate ASD impression group had significantly higher (meaning more severe) ADOS-2 scores than the accurate non-ASD impression group (*p*<.001) and the inaccurate non-ASD impression group (*p*=.012). Post-hoc Tukey's HSD revealed that the accurate ASD impression group had significantly lower VABS-3 ABC scores than the accurate non-ASD impression group (*p*<.001) but was not significantly different from the inaccurate non-ASD group (*p*=.536). The accurate non-ASD impression group had significantly higher VABS-3 ABC scores than the inaccurate non-ASD impression group (*p*=.027). The accurate ASD impression group had significantly lower MSEL DQs than the accurate non-ASD impression group (*p*=.001), whereas the two non-ASD impression groups did not differ on DQ (*p*=.117).

Behaviors Used to Guide Initial Impressions: Senior Clinicians

Senior clinicians judged an average of 3.6 behaviors out of 7 possible options that contributed to their initial diagnostic impression for each child (*SD* = 2.1). When forming *accurate* initial impressions of ASD (*n*=22), senior clinicians frequently observed atypical behaviors within the first five minutes (Figure 2). Accurate initial impressions of ASD were associated with atypical social reciprocity (*n*=13; 59%), nonverbal communication (*n*=14; 64%), focus of attention (attention to objects versus people; *n*=12; 55%), and eye contact (*n*=12; 55%). Behaviors that were less frequently observed in accurate initial impressions of ASD were: atypical motor mannerisms (*n*=7; 32%), atypical prosody and vocalizations (*n*=4; 18%), and atypical facial expressions (*n*=6; 27%).

When forming *inaccurate* initial impressions of non-ASD (initial impression: non-ASD, final diagnosis: ASD; *n*=13; Figure 3), senior clinicians judged typical social reciprocity (*n*=4; 31%), typical eye contact (*n*=7; 54%), absent repetitive motor mannerisms (*n*=5; 38%), typical facial expressions (*n*=5; 38%), and typical focus of attention (*n*=3; 23%). In some children, clinicians judged atypical social reciprocity (*n*=3; 23%), atypical nonverbal communication (*n*=4; 31%), and atypical facial expressions (*n*=3; 23%), but still formed an initial impression of non-ASD.

When forming *accurate* initial impressions of non-ASD (*n*=20; Figure 4), senior clinicians judged typical behaviors of social reciprocity (*n*=12; 60%), nonverbal communication (*n*=11; 55%), focus of attention (*n*=10; 50%), and eye contact (*n*=13; 65%). Behaviors less frequently endorsed in accurate initial impressions of non-ASD were: typical facial expressions (*n*=9; 45%) and a lack of atypical motor mannerisms (*n*=7; 35%). However, senior clinicians observed atypical behaviors in several children in all behavior domains except for prosody and vocalizations.

Behaviors Used To Guide Initial Impressions: Junior Clinicians

Junior clinicians judged the same behaviors as senior clinicians most frequently when forming accurate initial impressions of ASD and non-ASD: nonverbal communication, social reciprocity, and eye contact. When forming *accurate* initial impressions of ASD (*n*=17), junior clinicians frequently observed atypical behaviors within the first five minutes (Figure 2). Accurate initial impressions of ASD were associated with atypical social reciprocity (*n*=6; 35%), nonverbal communication (*n*=7; 41%), eye contact (*n*=8; 47%), and atypical facial expressions (*n*=5; 29%). Interestingly, none of the junior clinicians rated atypical focus of attention in forming accurate ASD impressions. Behaviors that were less frequently observed in accurate initial impressions of ASD were: atypical motor mannerisms (*n*=3; 38%) and atypical prosody and vocalizations (*n*=1; 17%). When forming *inaccurate* initial impressions of ASD (initial impression: non-ASD, final diagnosis: ASD; *n*=1), junior clinicians judged typical social reciprocity and typical eye contact.

When forming *accurate* initial impressions of non-ASD (*n*=16), junior clinicians judged typical behaviors of social reciprocity (*n*=7; 44%), nonverbal communication (*n*=6; 38%), focus of attention (*n*=6; 38%), and eye contact (*n*=10; 63%). Behaviors less frequently endorsed by junior clinicians in accurate initial impressions of non-ASD were: typical facial expressions (*n*=4; 25%) and a lack of atypical motor mannerisms (*n*=4; 25%). When forming inaccurate initial impressions of non-ASD (*n*=12), junior clinicians endorsed atypical behaviors in individual cases.

Discussion

The goals of the current study were to (1) replicate findings on accuracy, sensitivity, and specificity of junior and senior clinicians' initial diagnostic impressions of ASD and non-ASD in toddlers evaluated for possible ASD, (2) examine confidence in diagnostic impressions based on accuracy of impression for junior and senior clinicians, (3) examine child characteristics that relate to the accuracy of initial impressions of ASD and non-ASD, (4) identify the subset of diagnostically-relevant behaviors that guide accurate and inaccurate initial impressions of ASD and non-ASD for junior and senior clinicians

In general, the current results replicated prior findings of high specificity and moderate sensitivity of initial impressions, similar to results of Wieckowski and colleagues (2021) and de Marchena and colleagues (2023). Of the toddlers for whom senior clinicians formed an initial impression of ASD, 100% were ultimately diagnosed with ASD (and all but one were ultimately diagnosed with ASD when taking into account junior clinicians' impressions). Although 2/3 of toddlers with ASD were identifiable within the first five minutes of

observation, the remaining 1/3 were incorrectly initially classified as non-ASD by the senior clinicians. These toddlers required a regular length evaluation to allow observation of ASD characteristics. The current findings also replicated prior work on accuracy of initial impressions among junior clinicians, or trainees, which found higher specificity (0.92) than sensitivity (0.56; de Marchena et al., 2023) A five-minute unstructured interaction is not sufficient to elicit the nuances of ASD symptoms in all cases. Therefore, initial diagnostic impressions should not be used to definitively rule out ASD, or sensitivity will be low.

Clinicians were generally confident in their initial impressions and final diagnoses. Although senior clinicians did not vary in confidence based on whether their impression or diagnosis was ASD or non-ASD, junior clinicians were significantly more confident in final diagnoses of ASD than non-ASD. Both senior and junior clinicians reported significantly higher levels of confidence for final diagnoses than impressions made in brief observations, which is to be expected given the much richer data available to form a final diagnosis (structured diagnostic and developmental testing, extended observation of child, integration of data from parent report). Although high confidence in a non-ASD impression may be likely to indicate an accurate impression, to maximize sensitivity, all children with non-ASD impressions warrant closer examination to determine final diagnosis. Both senior and junior clinicians were generally confident in accurate ASD impressions. Future research with larger sample sizes could better identify the relationship between confidence and accuracy in ASD impressions.

In the current sample, the overall accuracy of initial impressions was not substantially different for younger and older children, consistent with similar work with toddlers (i.e., Wieckowski et al., 2021). We note that our sample did not have much age variability and did not include school-age children or adolescents. Accuracy of initial impression also did not differ by race or sex, although these findings are limited by small sample sizes. Future research should examine whether and how these child and family characteristics relate to clinicians' impressions. Replication is especially relevant given the racial and ethnic disparities in the time to diagnosis for Black children with ASD (Aylward et al., 2021; Mandell et al., 2009; Williams, 2022). Additionally, future research should examine the behaviors that clinicians use when forming inaccurate diagnostic impressions in larger samples of female versus male children, given the possibly different presentation of ASD in very young boys and girls, which may contribute to delayed diagnoses in female individuals (Carter et al., 2007; Ros-Demarize et al., 2020).

Neither overall cognitive functioning (DQ) nor individual cognitive domains were significant predictors of accuracy of impression. The MSEL is a skills-based assessment, so it is not surprising that it was not a good predictor of accuracy of impression since the skills it assesses would likely not be observed in a brief observation. Additionally, cognitive abilities are not predictive of the presence of absence of ASD. However, this analysis clarifies that clinicians are not inaccurately forming impressions of ASD in children who have lower cognitive levels; while developmental levels may contribute to impression formation, they did not relate to accuracy. This suggests that clinicians relied more on observation of ASD-specific behaviors to form their impressions. It also may be that developmental level became clear during the evaluation and might have contributed to final diagnosis, but was not obvious in the first five minutes. It should be noted that the ASD

group had lower cognitive functioning than the non-ASD group. Although everyone in the current sample was walking independently, future research should examine how the severity of motor functioning impacts accuracy of diagnostic impressions. It will be important to replicate findings of behaviors used to guide initial diagnostic impressions in samples that are more closely matched on cognitive functioning.

When clinicians formed accurate non-ASD impressions, children were more likely to have higher DQs and higher levels of adaptive behavior. Clinicians were more confident in accurate non-ASD initial impressions than inaccurate non-ASD initial impressions and were ultimately more confident in their final diagnoses for those with accurate non-ASD initial impressions, compared to inaccurate non-ASD initial impressions. This suggests that clinician confidence in impression may be a useful marker of accuracy of impression when the impression is non-ASD.

Although senior clinicians had the option to consider behaviors in seven domains when forming initial diagnostic impressions of toddlers, they used the same four behaviors most frequently in accurate initial impressions of ASD and non-ASD: social reciprocity, nonverbal communication, eye contact, and focus of attention. Junior clinicians used the same three behaviors most frequently when forming accurate impressions: social reciprocity, nonverbal communication, and eye contact. Interestingly, no junior clinicians used focus of attention when forming accurate impressions. This suggests that attending to this nuanced behavior may require more years of training and clinical expertise to discriminate atypical versus typical focus of attention.

Some behaviors that were characteristic of ASD were endorsed less frequently, such as repetitive motor mannerisms, flat or atypical facial expressions, and atypical prosody and vocalizations. One potential reason for this may be that these behaviors were not as likely to appear in a brief observation of toddlers, as has been suggested by others (Gabrielsen et al., 2015; Hus et al., 2014; Wiggins et al., 2012). For example, there may have been insufficient language and vocalizations to judge prosody, and when children encountered a new environment, there may have been less tendency for repetitive motor behaviors; judgement of these behaviors may need to rely more on parent report or extended observation (Hus et al., 2014). Another related reason is that some behaviors may require specific prompts and engagement with task materials to be evident.

Prior research has found that primary care providers may under-refer for ASD evaluations even after children screen positive (Monteiro et al., 2019; Wallis et al., 2020; Wieckowski et al., 2022). Attending preferentially to the four behaviors used to form accurate impressions (social reciprocity, nonverbal communication, eye contact, focus of attention) might expedite diagnoses and access to services for children who otherwise would encounter lengthy waitlists or delayed timelines for evaluations. This does not preclude incorporating less commonly observed behaviors (e.g., hand as tool) when forming an initial impression. Future research should examine how primary care pediatricians form diagnostic impressions, such as during well-child visits, what behaviors they use to form these impressions, and whether these diagnostic impressions can be used by primary care providers to route children for diagnostic referrals. Additionally, future research should examine how these

initial impressions are formed in general pediatric populations not referred for ASD-specific concerns.

Approximately 37% of children who were ultimately diagnosed with ASD were misclassified as non-ASD after an initial five-minute impression. In over half of these children, clinicians judged the child as showing typical social reciprocity, eye contact, facial expressions, and focus of attention, as well as absence of repetitive motor mannerisms. Prior research has found that children with ASD may initially demonstrate observable, typical behavior (Coulter et al., 2021; Harris et al., 2021; Smith et al., 2017). These findings suggest that a child might show typical behaviors initially but demonstrate impairments in the sustained use of verbal and nonverbal communication strategies (Gabrielsen et al., 2015). The presence of an apparently typical behavior (e.g., a child who shows typical eye contact initially) should not be used as the basis for a non-referral, since a longer evaluation appears necessary to allow for more opportunities to observe subtle social engagement and nonverbal communication. There are at least two possible explanations for this finding. One is that some children initially make eye contact and respond socially due to increased arousal in a strange environment, but once the child became comfortable, the usual autistic behavior (e.g., avoidance of eye contact) emerged. Alternatively, the initial brief impression might happen to involve superficially typical behaviors (e.g., eye contact) but that after extended observation, a qualitative impression of atypicality emerged (e.g., eye contact is not used to modulate social interaction or to communicate). In the current sample, initial impressions of ASD matched final diagnoses in all cases. Although this is a small sample size, it suggests that toddlers who show all four of these specific atypical behaviors (social reciprocity, nonverbal communication, eye contact, focus of attention) might receive expedited ASD diagnoses and access to intervention, especially if these behaviors can be accurately judged by clinicians without autism specialization, such as general pediatricians.

Limitations

A notable limitation of these analyses is the small samples within each behavior on the Initial Diagnostic Impression Checklist, given that clinicians endorsed only those behaviors that contributed to initial impressions. Future research should identify the behaviors that clinicians use in forming initial diagnostic impressions in larger samples and with samples outside the toddler age range; these bases for judgment are likely to be quite different in older children, adolescents, and adults. Additionally, there were no children in the current sample with inaccurate ASD initial impressions (where a clinician formed an initial impression of ASD but ultimately diagnosed the child with non-ASD); the behaviors that might be used in such a decision would need to be examined in larger samples where presumably some such decisions would appear. It is most likely that the 100% agreement on non-ASD cases and 0% false positive rates in this small sample of senior clinicians are partly attributable to artifacts such as limited heterogeneity in presentation, or confirmation biases and lack of blinding. A related limitation is that the sample was a population referred for ASD concern, which limits generalizability of these findings to general pediatric populations. However, clinicians did not have additional information about the referral type (parent vs provider concern, elevated screener score), and the differential diagnosis between ASD and other developmental delays is clinically meaningful. Future research

should examine sensitivity and specificity of diagnostic impressions, and the behaviors used to guide these impressions, in non-ASD pediatric samples (e.g., general population seen by a pediatrician). In particular, the current sample does not include a large sample of children referred for other, non-ASD developmental concerns which would help clarify the specificity of behaviors used to form diagnostic impressions and the accuracy of initial impressions in distinguishing between ASD and other neurodevelopmental disabilities. We believe that the behaviors used to form accurate impressions of ASD would also be observed in non-referral samples, but this is an empirical question that future research should address.

A limitation related to study design is that the same clinician provided the initial diagnostic impression and the final diagnosis. The benefit of this design is its ecological validity: it mimicked the real-world setting of the same provider meeting the child, forming an initial impression, conducting the evaluation, and then determining a final diagnosis. However, this design also allowed for the potential for confirmation bias in clinicians' observations. Clinicians made diagnoses that did not match their final impressions in 37% of cases; it is possible that what the current study classifies as "accuracy of impression" may be partly attributable to confirmation bias. Clinically, it is important for providers to recognize the existence of confirmation bias when forming initial impressions and final diagnoses. Although there is a risk of confirmation bias given the study design, such bias is not the only heuristic that affects motivation and decision-making in complex clinical cases. Other processes, such as representativeness, availability, and over-confidence, are relevant when considering cognitive factors that influence decision-making (Fernández-Aguilar et al., 2022).

In this sample, the finding of parallel results with junior clinicians argues against simple confirmation bias being a major contributor to match rates. However, a limitation is that this sample of junior clinicians consisted of graduate students or research staff (e.g., post-baccalaureate research assistants) who were mostly trained by the senior clinicians in their lab or clinic. A completely different cohort of juniors from a different lab might have been in less agreement with senior clinicians based on different standards or practices of clinical training. Future research should more systematically examine how well junior clinicians from different training sites agree with each other, and with other senior clinicians' impressions and behaviors.

In terms of study design, a notable limitation is that data were not consistently collected on the setting used for the initial impression (e.g., waiting room versus walking with family to the evaluation room versus child exploring the evaluation room). The intention of this project was to closely model the variability of settings in which clinicians form diagnostic impressions in real-world evaluations. The downside of this type of standardization is that it might detract from the natural flow of conversation between clinicians and parents in the initial moments of meeting. However, the lack of standardization in the setting and interaction used for the initial impression is a limitation. Future research could standardize the setting (e.g., first five minutes after the family and clinician enter the room while toys are set up for ADOS-2 Free Play).

The ADOS-2 Toddler Module scoring algorithm for children between 12 and 20 months with few to no words was used for the majority of the sample. The limited vocalizations and speech of the children included in this sample likely contributed to clinicians infrequently judging prosody and vocalizations in forming initial diagnostic impressions. The accuracy of initial diagnostic impressions and the behaviors used to guide these impressions should be more closely studied in toddler and early childhood samples with a wider range of language profiles. Although prosody and vocalization may have appeared atypical during the remainder of the evaluation, an initial five-minute observation with toddlers was likely not sufficient to judge these behaviors in forming a diagnostic impression.

Another limitation is that nine out of 13 senior clinicians were PhD-level psychologists, which limits generalizability of the findings related to behaviors used to guide diagnostic impressions in a range of pediatric settings. These findings may not be relevant to settings in which the observers have much less experience with autism, such as in a primary care setting. Future research should compare accuracy of impressions and behaviors used by different disciplines (e.g., pediatricians versus psychologists), autism specialists versus non-specialists or experts, and variation by length of practice.

Conclusions

Senior clinicians used the same four behaviors to form accurate impressions of ASD and non-ASD: social reciprocity, nonverbal communication, eye contact, and focus of attention. Toddlers who show all four atypical behaviors (social reciprocity, nonverbal communication, eye contact, focus of attention) might receive expedited diagnoses for ASD. However, the initial presence of apparently typical behaviors (e.g., typical eye contact) should not be used as the basis for a definitive judgment of non-ASD; for some children a longer evaluation is necessary to allow for more opportunities to observe and judge subtle social behavior.

The idea of prototypical or frank autism has been under consideration lately. In this sample of children referred for ASD concern, such a presentation exists for the majority (approximately 2/3) of toddler-aged children with ASD, where an initial impression of ASD is likely to be correct. However, the moderate sensitivity of impressions of ASD suggests that a substantial minority of ASD toddlers are missed by initial impression. These children require a lengthier and more in-depth evaluation to allow observation of ASD characteristics. An initial impression of non-ASD should not rule out ASD; these children require additional evaluation to confirm or dismiss this impression.

Financial Disclosure:

Drs. Robins, Fein and Barton are co-owners of M-CHAT LLC, which receives royalties from parties that license use of the M-CHAT in electronic products. No royalties were received for any data presented in the current study. Dr. Robins receives an honorarium for serving on the board of Quadrant Biosciences Inc. The other authors have indicated they have no financial relationships relevant to this article to disclose.

Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number R01MH115715 (D. Robins, PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix A: Initial Diagnostic Impressions Checklist (©Thomas, Wieckowski, Stahmer, Barton, Robins, Fein 2023)

COMPLETE AFTER FIRST 5 MIN WITH CHILD	***5 minute diagnostic impression***		***Confidence in 5-min impression***				
Clinician _____ Telehealth OR in-person eval Child wearing mask? Y/N	ASD	Non-ASD	1 Not very confident	2	3 Confident	4	5 Extremely Confident

Please check off only the items that contributed to your initial diagnostic impression within the first five minutes using any interactions of the child's:

- social reciprocity
  - impaired responses to social reciprocity
  - good interactions with others
- nonverbal communication strategies
  - limited or no initiations (ignores others)
  - some initiations to share interests (including pointing, showing, or gestures)
- eye contact
  - atypical quality or amount
  - good eye contact
- motor mannerisms
  - presence of atypical mannerisms (including repetitive object use)
  - lack of atypical mannerisms
- prosody and vocalizations
  - unusual vocal prosody and/or repetitive use of language
  - good prosody for level of language
- facial expressions
  - flat affect or unusual facial expressions
  - typical range of expressions
- focus of attention
  - primary attention to objects, minimal attention to people
  - attention to people, visual exploration of people in the room
- other:_____

What setting was used for the initial impression? (check one)

☐ waiting room ☐ walk to eval room ☐ kid exploring room ☐other: _____

Did the person completing this form have access to information about the child before the impression form was completed? ☐ yes ☐ no ☐ observed telehealth appointment ☐ completed telehealth appointment ☐ looked at TASI or other parent report ☐ other: _____

COMPLETE AFTER FINAL DIAGNOSIS	***Confidence in final diagnosis (circle)***				
Clinician_____	1	2	3	4	5
	Not very confident		Confident		Extremely Confident

References

Aylward BS, Gal-Szabo DE, & Taraman S (2021). Racial, Ethnic, and Sociodemographic Disparities in Diagnosis of Children with Autism Spectrum Disorder. Journal of Developmental and Behavioral Pediatrics : JDBP, 42(8), 682–689. 10.1097/DBP.0000000000000996 [PubMed: 34510108]

Bishop SL, Farmer C, & Thurm A (2015). Measurement of nonverbal IQ in autism spectrum disorder: Scores in young adulthood compared to early childhood. Journal of Autism and Developmental Disorders, 45(4), 966–974. 10.1007/s10803-014-2250-3 [PubMed: 25239176]

Carter AS, Black DO, Tewani S, Connolly CE, Kadlec MB, & Tager-Flusberg H (2007). Sex differences in toddlers with autism spectrum disorders. Journal of Autism and Developmental Disorders, 37(1), 86–97. 10.1007/s10803-006-0331-7 [PubMed: 17216333]

Constantino JN, Abbacchi AM, Saulnier C, Klaiman C, Mandell DS, Zhang Y, Hawks Z, Bates J, Klin A, Shattuck P, Molholm S, Fitzgerald R, Roux A, Lowe JK, & Geschwind DH (2020). Timing of the diagnosis of autism in african American children. Pediatrics, 146(3). 10.1542/PEDS.2019-3629

Constantino JN, Lavesser PD, Zhang Y, Abbacchi AM, Gray T, & Todd RD (2007). Rapid quantitative assessment of autistic social impairment by classroom teachers. Journal of the American Academy of Child and Adolescent Psychiatry, 46(12), 1668–1676. 10.1097/chi.0b013e318157cb23 [PubMed: 18030089]

Constantino JN, Zhang Y, Abbacchi AM, Calhoun A, Scofield F, & Grafeman SJ (2012). Rapid phenotyping of autism spectrum disorders: Inclusion of direct observation in feasible paradigms for clinical assessment. Neuropsychiatry, 2(3), 203–212. 10.2217/npy.12.28

Coulter KL, Barton ML, Robins DL, Stone WL, & Fein DA (2021). DSM-5 symptom expression in toddlers. Autism, 25(6), 1653–1665. 10.1177/13623613211000160 [PubMed: 33779335]

de Marchena A, & Miller J (2017). “Frank” presentations as a novel research construct and element of diagnostic decision-making in autism spectrum disorder. Autism Research, 10(4), 653–662. 10.1002/aur.1706 [PubMed: 27770496]

de Marchena A, Wieckowski AT, Algur Y, Williams LN, Fernandes S, Thomas RP, McClure LA, Dufek S, Fein D, Stahmer AC, & Robins DL (2023). Initial diagnostic impressions of trainees during autism evaluations: High specificity but low sensitivity. Autism Research, April, 1–7. 10.1002/aur.2933

Fernández-Aguilar C, Martín-Martín JJ, Minué Lorenzo S, & Fernández Ajuria A (2022). Use of heuristics during the clinical decision process from family care physicians in real conditions. Journal of Evaluation in Clinical Practice, 28(1), 135–141. 10.1111/jep.13608 [PubMed: 34374182]

Gabrielsen TP, Farley M, Speer L, Villalobos M, Baker CN, & Miller J (2015). Identifying Autism in a Brief Observation. Pediatrics, 135(2), e330–e338. 10.1542/peds.2014-1428 [PubMed: 25583913]

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. 10.1007/s10803-008-0674-3 [PubMed: 19082876]

Harris HK, Lee C, Sideridis GD, Barbaresi WJ, & Harstad E (2021). Identifying Subgroups of Toddlers with DSM-5 Autism Spectrum Disorder Based on Core Symptoms. Journal of Autism and Developmental Disorders, 51(12), 4471–4485. 10.1007/s10803-021-04879-z [PubMed: 33507459]

Hedley D, Brewer N, Nevill R, Uljarević M, Butter E, & Mulick JA (2016). The Relationship Between Clinicians' Confidence and Accuracy, and the Influence of Child Characteristics, in the Screening of Autism Spectrum Disorder. Journal of Autism and Developmental Disorders, 46(7), 2340–2348. 10.1007/s10803-016-2766-9 [PubMed: 26975451]

Hus V, Gotham K, & Lord C (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. Journal of Autism and Developmental Disorders, 44(10), 2400–2412. 10.1007/s10803-012-1719-1 [PubMed: 23143131]

Lord C, Rutter ML, DiLavore PC, Risi S, Gotham K, & Bishop SL (2012). Autism diagnostic observation schedule, 2nd edition (ADOS-2). In Manual (Part 1): Modules 1–4. Western Psychological Services.

Mandell DS, Wiggins LD, Carpenter LA, Daniels J, DiGuiseppi C, Durkin MS, Giarelli E, Morrier MJ, Nicholas JS, Pinto-Martin JA, Shattuck PT, Thomas KC, Yeargin-Allsopp M, & Kirby RS (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. American Journal of Public Health, 99(3), 493–498. [PubMed: 19106426]

Martinez M, Thomas KC, Williams CS, Christian R, Crais E, Pretzel R, & Hooper SR (2018). Family Experiences with the Diagnosis of Autism Spectrum Disorder: System Barriers and Facilitators of Efficient Diagnosis. Journal of Autism and Developmental Disorders, 48(7), 2368–2378. 10.1007/s10803-018-3493-1 [PubMed: 29453706]

McClure LA, Lee NL, Sand K, Vivanti G, Fein D, Stahmer A, & Robins DL (2021). Connecting the Dots: a cluster-randomized clinical trial integrating standardized autism spectrum disorders screening, high-quality treatment, and long-term outcomes. Trials, 22(1), 319. 10.1186/s13063-021-05286-6 [PubMed: 33934719]

McDonnell CG, Bradley CC, Kanne SM, Lajonchere C, Warren Z, & Carpenter LA (2019). When Are We Sure? Predictors of Clinician Certainty in the Diagnosis of Autism Spectrum Disorder. Journal of Autism and Developmental Disorders, 49(4), 1391–1401. 10.1007/s10803-018-3831-3 [PubMed: 30488151]

Monteiro SA, Dempsey J, Berry LN, Voigt RG, & Goin-Kochel RP (2019). Screening and Referral Practices for Autism Spectrum Disorder in Primary Pediatric Care. Pediatrics, 144(4), e20183326. 10.1542/peds.2018-3326 [PubMed: 31515298]

Mullen EM (1995). Mullen Scales of Early Learning. American Guidance Services.

Ros-Demarize R, Bradley C, Kanne SM, Warren Z, Boan A, Lajonchere C, Park J, & Carpenter LA (2020). ASD symptoms in toddlers and preschoolers: An examination of sex differences. Autism Research, 13(1), 157–166. 10.1002/aur.2241 [PubMed: 31747131]

Smith CJ, Rozga A, Matthews N, Oberleitner R, Nazneen N, & Abowd G (2017). Investigating the accuracy of a novel telehealth diagnostic approach for autism spectrum disorder. Psychological Assessment, 29(3), 245–252. 10.1037/pas0000317 [PubMed: 27196689]

Sparrow SS, Cicchetti DV, & Saulnier CA (2016). Vineland Adaptive Behavior Scales, Third Edition (Vineland-3). Pearson Clinical.

Wallis KE, Guthrie W, Bennett AE, Gerdes M, Levy SE, Mandell DS, & Miller JS (2020). Adherence to screening and referral guidelines for autism spectrum disorder in toddlers in pediatric primary care. PLoS ONE, 15(5), 1–17. 10.1371/journal.pone.0232335

Wieckowski AT, Marchena A, Algur Y, Nichols L, Fernandes S, Thomas RP, McClure LA, Dufek S, Fein D, Adamson LB, Stahmer A, & Robins DL (2021). The first five minutes: Initial impressions during autism spectrum disorder diagnostic evaluations in young children. Autism Research, 14(9), 1923–1934. 10.1002/aur.2536 [PubMed: 34021728]

Wieckowski AT, Zuckerman KE, Broder-Fingert S, & Robins DL (2022). Addressing current barriers to autism diagnoses through a tiered diagnostic approach involving pediatric primary care providers. Autism Research, 15(12), 2216–2222. 10.1002/aur.2832 [PubMed: 36254366]

Wiggins LD, Robins DL, Adamson LB, Bakeman R, & Henrich CC (2012). Support for a dimensional view of autism spectrum disorders in toddlers. Journal of Autism and Developmental Disorders, 42(2), 191–200. 10.1007/s10803-011-1230-0 [PubMed: 21448751]

Williams ZJ (2022). Race and Sex Bias in the Autism Diagnostic Observation Schedule (ADOS-2) and Disparities in Autism Diagnoses. JAMA Network Open, 5(4), e229503. 10.1001/jamanetworkopen.2022.9503 [PubMed: 35471575]

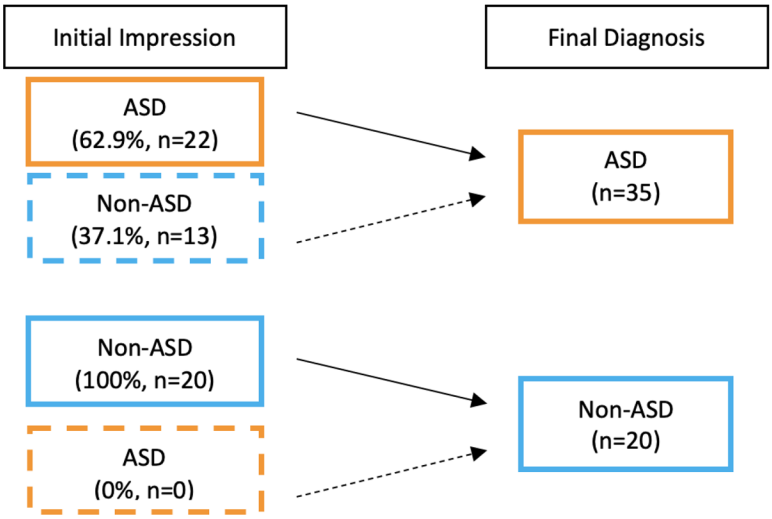

Figure 1a.

Accuracy of senior clinicians' initial impressions by diagnostic group. Dotted lines indicate a mismatch between initial impression and final diagnosis.

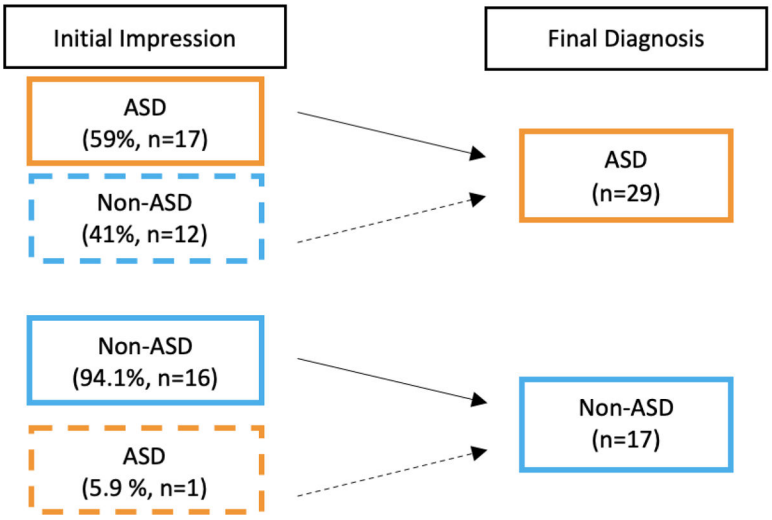

Figure 1b.

Accuracy of junior clinicians' initial impressions by diagnostic group. Dotted lines indicate a mismatch between initial impression and final diagnosis.

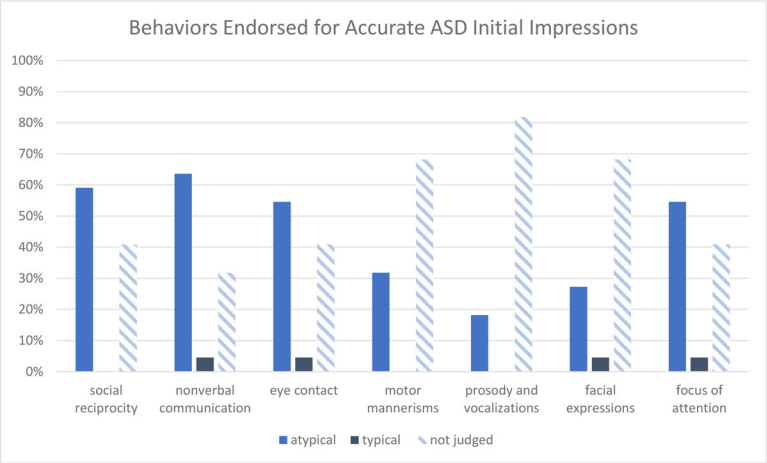

Figure 2.

Behaviors endorsed for senior clinicians' accurate ASD initial impressions (initial impression: ASD, final diagnosis: ASD; *n*=22).

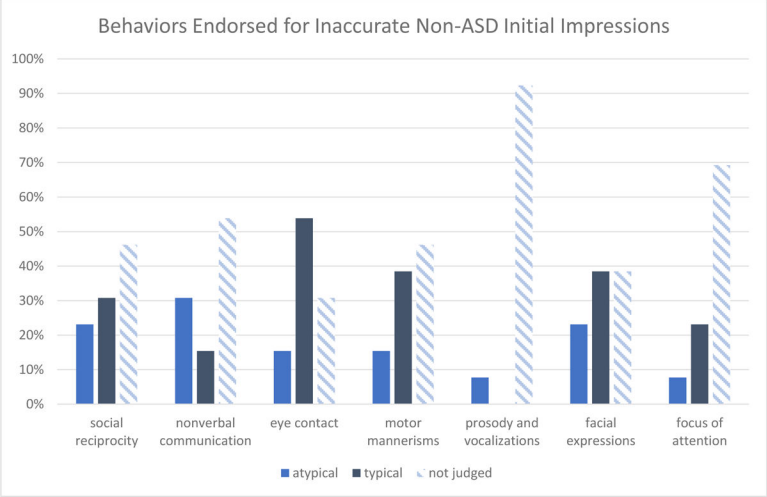

Figure 3.

Behaviors endorsed for senior clinicians' inaccurate non-ASD initial impressions (intial impression: non-ASD, final diagnosis: ASD; n=13)

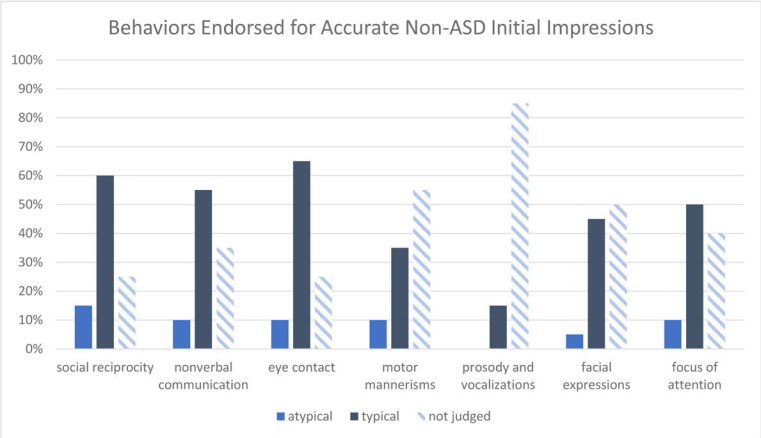

Figure 4.

Behaviors endorsed for senior clinicians' accurate non-ASD initial impressions (intial impression: non-ASD, final diagnosis: non-ASD; n=20)

Table 1.

Participant characteristics (n=55)

	Final Diagnosis			
	ASD ***n*** **= 35**	**Non-ASD** ***n*** **= 20**		***p***
	N (%)	**N (%)**		
Sex				.714
Male	21 (60.0%)	13 (65.0%)		
Female	14 (40.0%)	7 (35.0%)		
Race				.951
White	15 (51.7%)	10 (52.6%)		
BIPOC	14 (48.3%)	9 (47.4%)		
Racial Identities Within BIPOC				
Black	3 (8.6%)	4 (20.0%)		
Asian	2 (5.7%)	0 (0.0%)		
Bi- or multiracial	8 (22.9%)	3 (15.0%)		
Other	1 (2.9%)	2 (10.0%)		
Prefer not to answer	6 (17.1%)	1 (5.0%)		
Ethnicity				.960
Hispanic	12 (41.4%)	8 (42.1%)		
Non-Hispanic	17 (58.6%)	11 (57.9%)		
Maternal Education				.102
Less than high school or GED	0 (0%)	2 (10.0%)		
High school/GED	12 (34.3%)	2 (10.0%)		
Some college, technical or trade school	10 (27.5%)	7 (31.8%)		
College degree	6 (17.1%)	7 (35.0%)		
Advanced degree	7 (20.0%)	2 (10.0%)		
	M (SD)	***M (SD)***	***t*****(df)**	***p***
Age in months	23.5 (6.3)	25.3 (11.5)	$t(53) = -.781$	.438
ADOS-2 Total	7.9 (1.9)	2.7 (1.9)	$t(48) = 9.053$	*<.001*
ADOS-2 SA	7.8 (2.4)	2.9 (2.0)	$t(48) = 7.125$	*<.001*
ADOS-2 RRB	6.6 (2.9)	2.2 (2.4)	$t(48) = 5.354$	*<.001*
VABS-3 ABC	62.7 (10.8)	80.0 (17.5)	$t(51) = -4.452$	*<.001*
VABS-3 Communication	50.7 (18.7)	78.6 (18.7)	$t(51) = -5.227$	*<.001*
VABS-3 Socialization	66.6 (11.3)	87.3 (15.1)	$t(51) = -5.655$	*<.001*
VABS-3 Daily Living Skills	70.5 (15.2)	77.4 (26.6)	$t(51) = -1.195$	.238
VABS-3 Motor Skills	83.1 (11.5)	88.4 (18.1)	$t(51) = -1.295$	.201
MSEL ELC [a]	60.9 (11.4)	73.6 (18.1)	$t(51) = -3.136$	*.003*
MSEL Expressive Language [b]	11.1 (4.9)	18.0 (12.7)	$t(51) = -2.817$	*.007*
MSEL Receptive Language [b]	9.7 (4.7)	16.6 (12.6)	$t(50) = -2.857$	*.006*
MSEL Fine Motor [b]	17.5 (3.9)	21.4 (12.8)	$t(51) = -1.668$	.101

	Final Diagnosis			
	ASD ***n*** **= 35**	**Non-ASD** ***n*** **= 20**		***p***
	N (%)	**N (%)**		
MSEL Visual Reception [b]	15.1 (4.1)	21.1 (12.7)	$t(51) = -3.357$	*.015*
Developmental Quotient	59.7 (15.5)	77.2 (16.5)	$t(50) = -3.828$	*<.001*
Senior Clinician Confidence in Initial Impression	2.8 (1.3)	2.7 (1.0)	$t(52) = .099$	.921
Junior Clinician Confidence in Initial Impression	2.8 (1.1)	2.6 (1.3)	$t(43) = .516$	.609
Senior Clinician Confidence in Final Diagnosis	4.2 (1.1)	4.6 (0.5)	$t(31) = -1.036$	.308
Junior Clinician Confidence in Final Diagnosis	4.8 (0.4)	3.8 (1.3)	$t(26) = 2.466$	.021

[a] MSEL ELC is a standard score

[b] MSEL subscale scores are age equivalents

ADOS-2 Autism Diagnostic Observation Schedule, 2nd Edition; *SA* Social Affect; *RRB* Restricted and Repetitive Behaviors; *VABS-3* Vineland Adaptive Behavior Scales, 3rd Edition; *ABC* Adaptive Behavior Composite; *MSEL* Mullen Scales of Early Learning; *ELC* Early Learning Composite

Table 2.

Comparisons of child characteristics by impression groups (accurate ASD impression, accurate non-ASD impression, inaccurate non-ASD impression)

Variable	Accurate ASD-impression (*n*=22; *M*(*SD*))	Accurate non-ASD impression (*n*=20; (*M*(*SD*))	Inaccurate non-ASD impression (*n*=13; (*M*(*SD*))	F statistic	*p*	η^2
ADOS-2 SA	8.73 (1.78)	2.88 (2.03)	5.91 (2.54)	F (2,47) = 39.148	<.*001*	.625
ADOS-2 RRB	6.36 (3.26)	2.18 (2.35)	6.91 (2.12)	F (2,47) = 14.263	<.*001*	.378
ADOS-2 Total	8.50 (1.63)	2.71 (1.87)	6.55 (1.86)	F (2,47) = 52.36	<.*001*	.690
VABS-3 ABC	60.96 (9.96)	79.95 (17.50)	66.27 (12.06)	F (2,50) = 10.513	<.*001*	.296
MSEL DQ	59.96 (16.61)	77.24 (16.65)	65.16 (11.82)	F (2,49) = 8.464	<.*001*	.257
Senior Clinician Confidence in Impression	3.13 (1.36)	2.74 (0.99)	2.08 (1.0)	F (2,51) = 3.191	.049	.111
Junior Clinician Confidence in Impression	2.80 (1.1)	3.47 (0.92)	1.60 (.70)	F (2,43) = 4.664	*.015*	.178
Senior Clinician Confidence in Final Diagnosis	4.33 (1.03)	4.55 (0.52)	3.50 (1.29)	F (2,30) = 1.89	.169	.112
Junior Clinician Confidence in Final Diagnosis	4.89 (.33)	4.67 (0.50)	3.00 (1.20)	F (2,26) = 14.430	<.*001*	.526

ADOS-2 Autism Diagnostic Observation Schedule, 2nd Edition; *SA* Social Affect; *RRB* Restricted and Repetitive Behaviors; *VABS-3* Vineland Adaptive Behavior Scales, 3rd Edition; *ABC* Adaptive Behavior Composite; *MSEL* Mullen Scales of Early Learning; *DQ* Developmental Quotient