### Title
The first five minutes: Initial impressions during autism spectrum disorder diagnostic evaluations in young children




### Journal
Autism Research, 14(9)

### ISSN
1939-3792



### Authors
Wieckowski, Andrea Trubanova
Marchena, Ashley
Algur, Yasemin
et al.



### Publication Date
2021-09-01

### DOI
10.1002/aur.2536

Peer reviewed

# The first five minutes: Initial impressions during autism spectrum disorder diagnostic evaluations in young children

Andrea Trubanova Wieckowski*[1], Ashley de Marchena*[2], Yasemin Algur[3], Lashae Nichols[1], Sherira Fernandes[1], Rebecca P. Thomas[4], Leslie A. McClure[3], Sarah Dufek[5], Deborah Fein[4], Lauren B. Adamson[6], Aubyn Stahmer[5], Diana L. Robins[1]

Published in *Autism Research, 2021, pgs 1-12*

**Affiliations:** [1] A.J. Drexel Autism Institute, Drexel University, Philadelphia, PA, [2] Department of Behavioral and Social Sciences, University of the Sciences, Philadelphia, PA, [3] Department of Epidemiology & Biostatistics, Drexel University, Philadelphia, PA, [4] Department of Psychological Sciences, University of Connecticut, Storrs, CT, [5] Department of Psychiatry and Behavioral Sciences, University of California, Davis MIND Institute, Sacramento, CA, [6] Department of Psychology, Georgia State University, Atlanta, GA

* Andrea Trubanova Wieckowski and Ashley de Marchena should be considered joint first authors.

**Corresponding Author:**
Andrea Trubanova Wieckowski
atw64@drexel.edu

**Acknowledgments:**

We thank all the children and their families for participating in the study. In addition, we thank members of the research team who participated in data collection, specifically Sarah Nanovic and Katherine Sand for their management of the data. We also thank our funding sources, Eunice Kennedy Shriver National Institute of Child Health and Human Development grant R01HD039961, and the National Institute of Mental Health, R01MH115715.

**Compliance with Ethical Standards:**

Drs. Robins and Fein are co-owners of M-CHAT LLC, which licenses use of the M-CHAT in electronic products. No royalties were received for any of the data presented in the current study. Dr. Robins sits on the advisory board of Quadrant Biosciences Inc. The other authors have indicated they have no potential conflicts of interest to disclose. The research involved human participants. The study protocol was approved by the Institutional Review Board at each data

collection site for R01HD039961 grant, and at Drexel University for the multi-site study for R01MH115715 grant.

**Lay Summary**

When children come in for an autism evaluation, clinicians often form early impressions – before doing any formal testing – about whether the child has autism. We studied how often these early impressions match the final diagnosis, and found that clinicians could not easily rule *out* autism (many children who initially appeared not to have autism were ultimately diagnosed), but were generally accurate ruling *in* autism (when a child appeared to have autism within five minutes, they were almost always so diagnosed).

## Abstract

Diagnosticians report that autism spectrum disorder (ASD) is *immediately* apparent in some, but not all, children ultimately diagnosed. Clinicians' initial diagnostic impressions have implications for ASD early detection, yet the literature raises questions about their accuracy. This study explores diagnostic impressions of ASD specialists made within the first five minutes of meeting a young child and investigates factors associated with the match between initial impressions and final diagnoses. Participants were children (*n*=294, aged 12-53 months) referred for an ASD evaluation as part of multi-site ASD screening studies. After five minutes observing each child, clinicians with expertise diagnosing ASD recorded if they thought the child would meet criteria for ASD following a complete evaluation, and recorded their confidence in this impression. Clinicians' initial impressions matched the final diagnosis in 81% of cases. Ninety-two percent of cases initially thought to have ASD met criteria following a full evaluation; however, 24% of cases initially thought *not* to have ASD also met criteria, suggesting a high miss rate. Clinicians were generally confident in their initial impressions, reporting highest confidence for children initially thought correctly not to have ASD. ASD behavioral presentation, but not demographic characteristics or developmental level, were associated with matching initial impression and final diagnosis, and confidence. Brief observations indicating ASD should trigger referral to intervention services, but are likely to under-detect positive cases and should not be used to rule *out* ASD, highlighting the need to incorporate information beyond initial clinical impression.


Keywords: Autism Spectrum Disorder; Early Detection; Diagnosis; Toddlers; Initial Impression; Clinician Confidence in Diagnosis

The first five minutes: Initial impressions during autism spectrum disorder diagnostic evaluations in young children

## Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder for which differential diagnosis often requires substantial expertise, especially when diagnosing toddlers (e.g., Allaby & Sharma, 2011; Camarata, 2014). A reliable ASD diagnosis can be made prior to three years of age (Charman & Baird, 2002; Eaves & Ho, 2004; Guthrie et al., 2013). However, the average age of ASD diagnoses in the United States is above four years (Christensen et al., 2018), and data indicate a significant lag between initial parent concerns and an eventual ASD diagnosis (Crane et al., 2016). Similarly, even when evaluated at earlier ages, some children with ASD are not diagnosed until 36 months of age (e.g., Ozonoff et al., 2015), indicating need for follow-up on early concerns. Delayed diagnosis leads to delays in receiving ASD-specific early intervention, which may negatively affect the child's development (Fountain et al., 2012; Kaboski et al., 2017). Barriers to accessing an ASD evaluation include a shortage of experts, clinicians with specialized training in diagnosis of ASD, and limited availability of time-consuming evaluations (e.g., Gordon-Lipkin, Foster, & Peacock, 2016). The delayed diagnostic process is also associated with significant stress, time, and financial cost for families (e.g., Crane et al., 2016). Therefore, the need to reduce delays in detection, and ultimately specialized treatment of ASD, is an important public health goal.

The diagnostic evaluation process for ASD involves integrating information from multiple sources (i.e., caregiver report, observation, and standardized tests; Kim & Lord, 2012; Risi et al., 2006) to create an accurate picture of the child's developmental history and symptom presentation. In addition to rigorous evaluation of assessment data, clinicians also rely on clinical judgment, which is strengthened and refined with experience. For example, Klin and colleagues (2000) explored interrater reliability of ASD diagnosis using DSM-IV criteria for Autistic Disorder. They found a symptom checklist approach was helpful for less experienced clinicians; however, for more experienced clinicians, agreement was higher when they used clinical judgement to make decisions, suggesting that experienced clinicians identify additional features beyond those captured by symptom checklists.

Individuals with ASD can be differentiated from neurotypical individuals even by non-experts. For example, naïve raters can reliably differentiate adults and children with ASD from neurotypical peers using gestalt judgments such as awkwardness, approachability, and likelihood of friendship (Grossman, 2015; Sasson et al., 2017). In these studies, raters discriminated groups with a large effect size, across modalities, including audio, video, and even static images. Similar experiences are reported by ASD experts. In a recent study exploring "frank" presentation in ASD (i.e., unmistakable or immediate behavioral presentation), de Marchena and Miller (2017) found that experienced clinicians estimate 40% of the ASD population has a frank presentation. These frank features are often detected rapidly, within the first few minutes of an interaction, although the exact time varied widely.

Clinicians form diagnostic impressions early in the assessment process, and these impressions are likely to affect clinical decisions such as how – or *if* – an assessment is conducted. However, limited evidence on rapid clinical impressions suggest some concerns about their accuracy. For example, in a study examining brief observations of toddlers (Gabrielsen et al, 2015), one-third of whom were known to have ASD, expert raters were asked to decide whether they would refer each child for an ASD-specific evaluation after watching a

10-minute video sample of toddlers' ASD evaluations. Experts missed 39% of the ASD cases, suggesting that ASD is immediately evident in only a subset of toddlers with the diagnosis. In the present study, we aim to test the match between clinicians' initial, rapid diagnostic impressions, and their final diagnosis after completing a full assessment, in a sample of toddlers presenting for an ASD evaluation.

Finally, although diagnostic reliability for ASD is high (Freedman et al., 2013), diagnosis often involves *uncertainty*, which may be particularly salient in the context of ASD (Penner et al., 2017; Rogers et al., 2016; Unigwe et al., 2017). Indeed, a recent study found that only 60% of diagnoses made by highly qualified clinicians were made with complete certainty and that clinicians were more certain when identifying ASD than ruling it out (McDonnell et al., 2019). Clinician certainty varied based on a range of child and family factors; for example, children presenting with moderate (compared to high or low) levels of observable ASD symptoms, older children, children with public insurance, and those with higher IQ and adaptive behavior abilities were all diagnosed with *lower* confidence. Identifying factors associated with certainty regarding an ASD diagnosis may assist in detecting children who may have been missed, delaying diagnosis and treatment. For example, girls with ASD are under-identified (Kirkovski et al., 2013; Loomes et al., 2017), and pediatricians report girls are more difficult to diagnose (Penner et al., 2017). Child age and family socioeconomic status may similarly be associated with certainty as both very young and older children are more challenging to diagnose, and pediatricians report difficulty identifying ASD in the context of environmental stressors (Penner et al., 2017). Clinical factors, such as higher levels of ASD symptoms (Hedley et al., 2016; Penner et al., 2017), and lower IQ and adaptive functioning (Hedley et al., 2016), also influence diagnostic certainty.

The literature has explored factors associated with certainty in diagnosis of ASD *following* a full evaluation; the present study extends this work to clinicians' confidence in their initial diagnostic impressions. In this study, clinicians with expertise diagnosing ASD spent five minutes observing a toddler presenting for an ASD evaluation, across a range of clinical contexts (e.g., watching from an observation room, speaking with the caregiver with the child nearby, or directly interacting with the child). After this brief observation period, clinicians documented their initial diagnostic impression (ASD or not ASD) as well as their confidence in this impression, before completing the full evaluation. The aims of the current study were to: 1) characterize clinicians' diagnostic impressions made within the first five minutes of meeting a toddler being evaluated for ASD, 2) investigate the match between initial impressions and final diagnoses, 3) characterize clinicians' confidence in their initial impressions, and 4) investigate factors associated with clinicians' match and confidence in their impression of ASD diagnosis in toddlers.

## Methods

### Participants

Participants included children ($n$ = 294; 199 males) aged 12–53 months who participated in an ASD evaluation at a university-based ASD specialty clinic as part of two multi-site research studies examining early ASD screening and referral methods in California, Connecticut, Georgia, and Pennsylvania. Referrals for evaluation were made during pediatric well-child visits based on positive screening results on at least one standardized ASD or social communication screener or a provider's concern for ASD. All at-risk children were invited for an evaluation. Twenty-one of the diagnostic evaluations were completed in Spanish; all others in English. Children with missing initial impression data ($n$ = 24) were excluded. See Table 1 for participant demographics.

## Measures

### *Demographic Information*

Caregivers reported demographic information about their child and family, including age, sex, race, and maternal education as a surrogate for socioeconomic status.

### *Initial Impression and Confidence*

After observing the child for five minutes, clinicians – all ASD experts – documented their initial impression of whether they thought the child had ASD (i.e., ASD or non-ASD), and how confident they were regarding this impression on a scale of 1 to 5, with 1 being *not very confident*, 3 being *confident*, and 5 being *extremely confident*. Initial observation period contexts varied between participants: clinicians either observed the child from an observation room, spoke with the caregiver with the child nearby, or interacted with the child. All clinicians were autism specialists, comprised largely of licensed clinicians (psychologists [*n*=14], physician [*n*=1]); additionally, five non-licensed clinicians (two psychologists, one behavior analysist, one occupational therapist, and one social worker) were supervised by licensed psychologists. Clinicians completing the diagnostic evaluations were blind to referral source and screening results.

### *ASD Assessment*

Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al. 2012) is a standardized, validated assessment of ASD symptoms, with distinct modules selected based on age and language abilities. In this study, 274 children received the Toddler Module, 11 received Module 1, and 8 received Module 2. ADOS-2 Calibrated Severity Scores (ADOS-CSS; Gotham, Pickles, & Lord, 2009) allow for score comparison across modules; higher scores indicate greater severity of ASD symptomatology.

### *Cognitive/Developmental Assessment*

Mullen Scales of Early Learning (MSEL; Mullen, 1995), a standardized developmental assessment for children 0–68 months of age, assesses visual reception, fine motor skills, receptive language, and expressive language. The Early Learning Composite standard score was utilized in the current study as a measure of cognitive development.

***Adaptive Behavior Assessment***

The Vineland Adaptive Behavior Scales—2nd Edition (VABS-2; Sparrow et al., 2005) or 3rd Edition (VABS-3; Sparrow et al., 2016) semi-structured interview was completed with caregivers. Standard scores for four domains (communication, daily living skills, socialization, motor skills) generate an overall adaptive behavior composite score, utilized in primary analyses, with lower scores indicating less adaptive functioning.

**Procedure**

ASD evaluations were conducted in a clinic setting. Two clinicians worked together to complete testing with the child and interviews with the caregiver, in the same room. A junior clinician – typically a trainee – administered and scored some of the measures, and a senior clinician administered additional measures and integrated all available information to make a final diagnosis. The senior clinician contributed the initial impression score. The evaluation typically lasted 3 hours (2 hours of data collection, and 1 hour of clinical feedback). All final diagnoses were based on *International Classification of Diseases, tenth edition* (ICD-10) criteria, integrating information obtained across all available measures. ASD diagnoses included Childhood Autism, Atypical Autism and Asperger's Syndrome. Non-ASD diagnoses included Other Disorder of Psychological Development, Expressive Language Disorder, Receptive Language Disorder, Other Developmental Disorder of Speech and Language, Specific Developmental Disorder of Motor Function, or No Diagnosis. All procedures were approved by university Institutional Review Boards, and all caregivers provided informed consent for themselves and their children.

**Statistical Analysis**

Frequencies and percentages were calculated to provide descriptive information on the match between initial impression and final diagnosis. McNemar's test was used to assess match between initial impression and final diagnosis. The effects of initial impression and final diagnosis on clinician's confidence was tested using a two-way Analysis of Variance (ANOVA). Regression analyses were used to assess factors associated with match between initial impression and final diagnosis and confidence in initial impression (5-point Likert scale). Match between initial impression and final diagnosis was modeled using logistic regression. Exploratory analyses were conducted to assess age and sex differences in the odds of a match using chi-square tests. Confidence in initial impression was modeled using ordinal regression and the assumption of proportional odds was assessed. Each potential factor (i.e., age, sex, race, maternal education, initial impression, clinician's confidence, ADOS-2 CSS, VABS-2 score, MSEL score) was assessed individually, and likelihood ratio tests (LRT) were used to compare the fit of the full model (with all factors) to reduced models (including only clinical assessment factors). P-values < .05 were considered statistically significant. Statistical analyses were performed in IBM SPSS Statistics Version 25 and SAS 9.4 (SAS Institute, Cary NC).

## Results

**Clinicians' Initial Impressions of Diagnosis**

Clinicians' initial impressions matched final diagnoses (Figure 1) in the majority of cases (overall accuracy = 81%). For toddlers for whom the clinicians' initial impression was ASD, the initial impression matched the final diagnosis 92% of the time. For toddlers for whom initial impression was non-ASD, initial impression matched final diagnosis 76% of the time. Sensitivity of detecting ASD within the first 5 minutes (defined as indicating ASD impression when a child had a final ASD diagnosis; true positives/all final ASD cases) was moderately low (.64); however, specificity (i.e., detecting non-ASD cases in the first five minutes; true negatives/all final non-ASD cases) was high (.96). Although the overall match between initial impression and final diagnosis was high, there was nevertheless a statistically significant change between the proportion of initial impression of ASD and final diagnosis of ASD (32% to 46%; $p < .001$).

**Relationship Between Initial Impression, Final Diagnosis, and Confidence in Clinicians' Initial Impressions**

Across both diagnostic groups, clinicians on average indicated feeling 'confident' in their initial impression ($M = 3.35$ $SD = 1.17$). When considering the relationship between initial impression and final diagnosis on clinicians' confidence, clinicians overall had higher confidence for the non-ASD final diagnostic group ($M = 3.71$, $SD = .99$) compared to the ASD group ($M = 2.93$, $SD = 1.22$; $F(1, 282) = 7.94$, $p = .005$, $\eta_p^2 = .027$). The main effect of initial impression on confidence, however, was not significant ($F(1, 282) = 0.45$, $p = .503$, $\eta_p^2 = .002$). As seen in Figure 2, the effect of final diagnosis on confidence varied by initial impression ($F(1, 282) = 10.73$, $p = .001$, $\eta_p^2 = .037$), with highest confidence when clinicians indicated non-ASD initial impression and final diagnosis was non-ASD ($M = 3.74$, $SD = 0.99$), and lowest confidence when initial impression was non-ASD and final diagnosis was ASD, indicating a mismatch ($M = 2.35$, $SD = 0.90$). Figure 3 portrays the mean match between initial impression and final diagnosis for all levels of confidence, depicting the linear increase in match with the increase in clinician's confidence. In cases for which the clinician was "not very confident," ($n =$

18) the match rate was 56% ($n$ = 10); in contrast, when the clinician was "extremely confident" ($n$ = 54), the match rate was 100%.

**Factors Associated with Match Between Initial Impression and Final Diagnosis**

The LRT indicated that the full model, including the demographic variables, fit significantly better than the model with only clinical assessment and confidence factors, although the regression coefficients of the clinical assessment factors did not change notably between the full and reduced models (see Table 2). Sex and age did not significantly affect the odds of a match between initial impression and final diagnosis.

Table 2 provides the results of logistic regression models assessing the relationship between demographic variables, clinical assessment measures, clinicians' confidence and match between initial impression and final diagnosis. Only clinicians' confidence ($\beta$ = 0.84, $p$ < .001) and the ADOS-2 CSS ($\beta$ = -0.18, $p$ = .03) were significantly associated with matching impression and diagnosis. A one point increase in confidence was associated with increased odds of match (Odds Ratio (OR) = 2.32 [95% CI 1.66, 3.24]). Additionally, each one-point increase in ADOS-2 CSS decreased the odds of match (OR = 0.83 [95% CI 0.71, 0.98]) when considering the entire sample, suggesting negative association between match and autism severity. However, follow-up analyses stratified by initial impression indicated that the relationship between match and autism severity (as measured by ADOS-2 CSS) was negative only when the initial impression was non-ASD (OR = 0.20 [95% CI 0.12, 0.34]). When the initial impression *was* ASD, autism severity was positively associated with match (OR = 2.60 [95% CI 1.47, 4.58]), suggesting that when clinicians initially thought a child had autism, this impression was more likely to match the final diagnosis when the child presented with more severe ASD features.

An examination of domain-specific ADOS-2 CSS separately—Social Affect (SA), Restricted and Repetitive Behavior (RRB)—led to an attenuation of magnitude of effect ($\beta$ = -0.076 for SA, $\beta$= -0.14 for RRB, and $\beta$ = -0.182 for total), with a greater attenuation for ADOS SA, but neither was significant, indicating that considering these domains together may be more informative.

**Factors Associated with Clinicians' Confidence in Initial Impression**

The LRT indicated that the full model with all nine demographic and clinical assessment variables fit significantly better than the model with only clinical assessment variables. The results from the full ordinal regression model, including the interaction between clinicians' impression and ADOS-2 scores are provided in Table 3. Only the ADOS-2 CSS ($\beta$ = -0.46, $p$ < .0001), clinicians' 5-minute impression ($\beta$ = -5.10, $p$ < .0001) and the interaction between these two variables ($\beta$ = 0.862, $p$ < .0001) remained significantly associated with clinicians' confidence in their initial impression. Stratified analyses indicated a stronger association between clinicians' initial ASD impression (vs. Non-ASD) and confidence when ADOS-2 CSS was high (6 or more) versus low (5 or less; OR = 2.64 [95% CI 1.10, 6.34] versus 0.32 [95% CI 0.09, 1.13]).

## Discussion

This study investigated the utility of brief expert clinical observations for predicting an ASD diagnosis in young children. Overall, the sensitivity of initial impression made by clinicians with specialized training diagnosing autism was low (64%), supporting our hypothesis that very brief observations, even by experts, are insufficient for screening/surveillance, and when conducted in the absence of diagnostic evaluation would lead to a high false negative rate. In contrast, the specificity of initial impression was high, indicating a low false positive rate, and suggesting that *in some children* autism is readily observable by experts.

As the focus of the current study was the match between clinicians' initial impressions with children's ultimate diagnoses, we looked separately at the group of children who initially gave an impression of ASD compared to those who did not. When a clinician did *not* initially think that a toddler had ASD, there was still a 24% chance that toddler would ultimately be diagnosed with ASD after the evaluation. These results are similar to Gabrielsen and colleagues (2015), who found that when asked whether to make a referral for an ASD-specific evaluation following a 10-minute video observation, one third of toddlers thought not to need a referral eventually received an ASD diagnosis. The implications of these findings are clear: clinicians cannot rely on brief behavioral observations alone to *rule out* ASD. Caregiver concerns, pediatrician concerns, and positive developmental screeners *must* be followed up, even if an experienced clinician does not have the impression a child is on the spectrum after a brief observation.

But what happens when a clinician initially thinks a toddler *does* have ASD? In the current study, 92% of children initially thought to have ASD met criteria for the diagnosis after a complete evaluation. These findings are consistent with the observational study described above (Gabrielsen et al., 2015) in which 88% of the toddlers who appeared to have ASD after a brief evaluation ultimately met criteria. Overall, these findings suggest that clinicians can feel confident that children who give an initial clinical impression of ASD, in the context of a positive screener, are likely to end up with a diagnosis.

The finding that demographics were unrelated to diagnostic match is somewhat unexpected, and may reflect the expertise of the clinicians. Developmental level, as measured by both the MSEL and the VABS, was also unrelated. This suggests that clinicians' initial impressions of autism in young children are orthogonal to global impairment, indicating that expert clinicians are able to differentiate symptoms related to ASD from those related to developmental delay when forming an initial impression of ASD. That is, clinical expertise appears to seamlessly integrate observable diagnostic features with observable features relevant to a child's developmental level.

Finally, the relationship between match and ADOS-2 severity scores varied based on the clinician's initial impression; higher ADOS-2 scores (more severe presentation) were associated with higher odds of match when the clinician's initial impression was ASD, but associated with lower odds of match when the initial impression was non-ASD. In other words, when the child's ASD-specific behavioral presentation was consistent with the clinician's initial impression, the odds of match were higher.

A secondary goal of this study was to investigate factors related to clinicians' *confidence* in their initial impressions. Clinicians were generally confident in their initial impressions, and the degree of clinicians' initial confidence corresponded with both the likelihood of match between initial and final diagnosis, and with the variance associated with that likelihood. Accordingly, when clinicians were "extremely confident," the match rate was 100%; when clinicians were "not very confident," the chances of match were closer to 50-50. These findings suggest that more weight can be given to the initial impression when the clinician is especially confident. Additionally, clinicians in this study were more confident in their initial impressions of children who were *not* ultimately diagnosed with ASD, reflecting the complexities of the ASD phenotype in many children, even for experienced clinicians.

Similarly, clinicians' confidence was driven entirely by autism-specific factors: initial impression of ASD and the child's ADOS-2 severity score, as well as the interaction between these two variables. Unsurprisingly, confidence was higher when initial impression was ASD and ADOS-2 severity scores were high, and when initial impression was non-ASD and ADOS-2 severity scores were low. Clinician confidence was not associated with child demographics or developmental level. Overall, this suggests that expert clinicians' initial impressions of autism are specific to the presentation of autism itself, and unlikely to be driven by general developmental level or child demographics. These findings are inconsistent with McDonnell and colleagues (2019), who found that clinicians were less confident in their final diagnosis when children had stronger cognitive and adaptive behavior skills, when children were older, and when children were on public insurance. This discrepancy likely reflects the much higher levels of complexity and integration of information involved in making a final clinical diagnosis relative to forming an initial impression – the former reflecting a more controlled, deliberative process, and the latter reflecting a more spontaneous appraisal, or a form of pattern recognition. This discrepancy may also reflect the difference in age between our sample (mean age: 21 months) and theirs (mean age: 43 months), as behavioral presentation often becomes more complex as children get older (Kim et al., 2018).

**Clinical Implications**

The most important clinical takeaway from the current study is that brief observations are *not* sufficient to rule out autism. Even expert clinicians with specialized training diagnosing autism gave an autism diagnosis to nearly one out of four toddlers they initially thought did not have autism. Although we did not systematically probe in the current study how such children behaved, it may well be that some children who show interest during the first several minutes of a clinical interaction, such as offering or responding to a greeting, or making eye contact with the clinician, may not sustain a high level of social engagement throughout the evaluation. In these cases, critical social interaction skills may need to be specifically probed (e.g., joint attention), observed over a longer timeframe (e.g., initiating or sustaining social interactions with others), or elicited in the context of competing nonsocial demands for weaknesses to become evident. We note that these examples are hypothetical and their relationship to match between initial impression and final diagnosis was not directly tested in the current study. It is thus imperative that clinicians incorporate information from multiple sources, including parent concerns, results of screening questionnaires, and complementary data from an evaluation, to support or refute their initial impression. Assuming that the final diagnoses were more accurate than the initial impressions, it is encouraging that the clinicians were willing to consider all information gathered in the evaluation and change their diagnostic decision. It is possible, however, that the procedure of making and recording initial impressions actually influences the final diagnosis in the direction of agreeing with one's initial impression.

These findings are particularly critical in light of recent findings that among toddlers screening positive for autism on the M-CHAT at a well visit, *less than half* were given *any* of the referrals recommended by the American Academy of Pediatrics at the time of screening (e.g., to early intervention or for further evaluation; Monteiro et al., 2019; Wallis et al., 2020). Referrals for ASD specialty evaluation are particularly low; referral rates for further evaluation *among children who screen positive for ASD risk on the M-CHAT* are as low as 11% (Wallis et al., 2020) to 31% (Monteiro et al., 2019). The reasons for these failures to refer are poorly understood and a critical research need; review of records suggest some potential causes are a "wait and see" approach, clinician attribution of the positive screen result to another neurodevelopmental disorder or speech delay (an attribution which is likely based in part on a brief behavioral observation), and parents declining a suggested referral (Monteiro et al., 2019).

In the case of toddlers who appeared to have autism within the initial five-minute period, initial impression matched the final diagnosis approximately nine times out of ten. When expert clinicians suspect autism after a brief clinical observation, this information should be sufficient justification to begin early intervention, even before a diagnostic evaluation can be completed. Provider concerns following a brief evaluation should be taken very seriously, even in cases where parents do not endorse symptoms. In contrast, if a provider does not initially observe behavior indicative of autism, then they should rely primarily on other information such as parent concerns and standardized measures. Any delay in beginning intervention can have negative consequences (e.g., Dimian, Symons, & Wolff, 2020); thus, it should be a priority to refer children to services who are likely to need them.

These findings can be translated to improved clinical care for toddlers with autism. For example, brief observations, completed by a clinician with ASD training, can be combined with response inventories, completed by a parent, and children who show ASD risk from either approach can then be referred directly to intervention. A low risk finding on one approach should therefore not override a referral if another approach indicates an ASD risk. The goal of this approach would not be to circumvent or replace a full evaluation, but rather to expedite the process of getting a population of children who are *very* likely to need intervention services into early intervention as soon as possible while they wait for a comprehensive evaluation. Our results show that six out of the seven children for whom clinicians indicated impression of ASD, but who were not diagnosed with ASD following an evaluation, received a diagnosis of other developmental delay, indicating a need for intervention. This is especially critical given the long waiting lists for autism specialty evaluations.

There are multiple ways to expedite the process of entering early intervention. A first impression approach could be advantageous in that it capitalizes on pattern detection skills. However, such an approach is highly unstructured, and may only be valid in clinicians with substantial expertise in autism. Structured brief behavioral observations, such as the Screening Tool for Autism in Toddlers and Young Children (STAT; Stone, Conrood, & Ousley, 2000), the Childhood Autism Rating Scale, Second Edition, based on patient observation (CARS-2$^{obs}$; Sanchez & Constantino, 2020) or the Rapid Interactive Screening Test for Autism in Toddlers (RITA-T; Choueira and Wagner, 2015), though not included in our study, are intriguing tools for formalizing first impressions, and could potentially be combined efficiently with parent report via screeners. South Carolina has implemented an effective statewide process consistent with this approach: a two-tiered screening process in which children who screened positive on the M-CHAT were *automatically* referred for a STAT, and those whose performance on the STAT indicated risk as well were automatically eligible to begin early intensive behavioral intervention. This approach led to a five-fold increase in the number of children eligible to receive services before age three. Consistent with our findings, the false positive rate for ASD was very low; following a comprehensive evaluation, only 2.5% of the sample was later found not to meet criteria for a diagnosis, suggesting that a large number of children who would have met criteria were missed prior to this process being implemented (Rotholz et al., 2017). The need to become more adaptable during the autism diagnostic process is especially salient when clinicians interact with families remotely rather than in-person.

Importantly, although we advocate for supporting children's entry into appropriate intervention as early as possible, full assessment should still be done. Autism is highly heterogeneous, and use of standardized measures is essential for determining functioning across multiple domains, including language and cognitive development, adaptive behavior skills, and interfering behaviors, that must inform treatment (Kanne & Bishop, 2020). Indeed, long-term treatment planning is not based on a diagnostic binary, but rather on each child's symptom presentation and needs, which cannot be determined in five minutes. Additionally, full assessment guards against providers without significant ASD training requesting unnecessary ASD-specific treatments without evidence, which places a strain on the system.

**Implications for Understanding Autism**

The current study also has implications for understanding the behavioral presentation of autism. Our findings suggest that a substantial subset of toddlers with ASD (about 60%) have a distinctive enough presentation that clinicians were able to observe ASD in as little as five minutes. These results are consistent with clinicians' reports that 40% of the greater ASD population may have a "frank" or "classic" behavioral presentation that is specific to autism and often quickly ascertained (de Marchena & Miller, 2017). The current study cannot speak directly to the phenomenon of frank presentations of autism because we did not directly compare participants whose ASD was and was not evident within the first five minutes. Future studies are needed to elucidate if frank presentations are indeed associated with any autism subtyping. In addition, what may be frank to an experienced clinician may not be obvious to those with limited experience; similarly, less experienced clinicians' initial impressions may match final diagnosis less consistently, two possibilities that can be addressed in future studies.

**Limitations and Future Directions**

Since we did not collect any information on the clinicians aside from their name and credentials, limited information was available about the clinicians; and indeed, clinician factors, in addition to child factors, are likely to influence both first impressions and confidence. Previous research suggests that clinical experience with autism influences both the speed of first impressions (de Marchena & Miller, 2017), and the reliability of the autism diagnosis itself (Klin et al., 2000), particularly when using a gestalt approach to diagnostic decision making. Future research can investigate clinician factors and how they relate to diagnostic certainty. In addition, analysis of certainty at the time of diagnosis may provide important information especially in regard to cases of mismatch between initial impression and final diagnosis. Intriguingly, diagnostic certainty is lower for children with known genetic abnormalities (Bishop et al., 2017), suggesting the value of investigating the link between genetic profiles, initial impressions, and clinician certainty.

Another limitation is that the initial impressions and final diagnoses were made by the same clinician. The robust phenomenon of confirmation bias suggests that it can be hard to change one's first impression. Thus, the relatively high match between initial and final impressions may reflect clinician reluctance to reject their initial impression in light of new evidence from the full evaluation. This is likely true to some extent; however, our data suggest that clinicians did revise their initial impressions in a substantial minority of cases. And indeed, confirmation bias is not the only form of motivation influencing our cognitive processes. The motivation to be *accurate* activates effortful cognitive processes such as attention to detail and complex information processing (Kunda, 1990) which are essential during a challenging task such as a diagnostic evaluation.

We note that the brief observations in both Gabrielsen et al. (2015) and the current study were conducted in clinical settings, and often in the context of structured testing; thus, it is unknown whether these findings would generalize to different, less structured, contexts. In addition, the administration of the ADOS-2 was non-standard, as parents completed interviews while also attending to the child, and therefore, scores should be interpreted in that context. Another potential limitation is that the setting for the initial impressions was not standardized. Clinicians essentially conducted their five-minute observations in a context that made the most sense for them clinically on a case-by-case basis. However, the lack of standardization may be testament to the clinical utility of the initial impression as a measure; it could easily be conducted – and is likely to generalize – to a wide range of settings.

## Conclusions

Overall, the results of the current study indicate that while expert clinicians' initial impressions made within the first five minutes of meeting children referred for an ASD evaluation generally matched children's final diagnoses after a full diagnostic evaluation, these brief observations are not sufficient to rule out autism. However, when clinicians with expertise in ASD suspect autism after only a brief observation, this information is highly reliable, and therefore could be utilized to refer children to begin early intervention as soon as possible, even before comprehensive evaluation can be completed, in order to avoid delays. Comprehensive assessment – including standardized evaluation of developmental skills – remains necessary to determine a child's strengths and weaknesses across domains, establish a more sophisticated diagnostic picture, and guide individualized treatment planning. Future research is needed to explore clinician factors in initial impressions, and to characterize the "frank" or "classic" presentation evident in some children with ASD.

References

Allaby, M. & Sharma, M. (2011). Screening for autism spectrum disorders in children below the age of 5 years. A draft report for the UK National Screening Committee.

Bishop, S. L., Farmer, C., Bal, V., Robinson, E. B., Willsey, A. J., Werling, D. M., ... & Thurm, A. (2017). Identification of developmental and behavioral markers associated with genetic abnormalities in autism spectrum disorder. *American Journal of Psychiatry*, *174*(6), 576-585. doi:10.1176/appi.ajp.2017.16101115

Camarata, S. (2014). Early identification and early intervention in autism spectrum disorders: Accurate and effective? *International Journal of Speech-Language Pathology, 16*(1), 1-10. https://doi-org.ezproxy2.library.drexel.edu/10.3109/17549507.2013.858773

Charman, T. & Baird, G. (2002). Practitioner review: Diagnosis of autism spectrum disorder in 2-and 3-year-old children. *Journal of Child Psychology and Psychiatry*, *43*(3), 289-305.

Choueiri, R., & Wagner, S. (2015). A new interactive screening test for autism spectrum disorders in toddlers. *The Journal of Pediatrics*, *167*(2), 460-466.

Christensen, D. L., Braun, K. V. N., Baio, J., Bilder, D., Charles, J., Constantino, J. N., ... & Lee, L. C. (2018). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveillance Summaries*, *65*(13), 1.

Crane, L., Chester, J. W., Goddard, L., Henry, L. A., & Hill, E. (2016). Experiences of autism diagnosis: A survey of over 1000 parents in the United Kingdom. *Autism*, *20*(2), 153-162.

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. https://doi-org.ezproxy2.library.drexel.edu/10.1002/aur.1706

Dimian, A. F., Symons, F. J., & Wolff, J. J. (2020). Delay to early intensive behavioral intervention and educational outcomes for a Medicaid-enrolled cohort of children with autism. *Journal of Autism and Developmental Disorders*. https://doi-org.ezproxy2.library.drexel.edu/10.1007/s10803-020-04586-1

Eaves, L. C., & Ho, H. H. (2004). The very early identification of autism: Outcome to age 41/2-5. *Journal of Autism and Developmental Disorders*, *34*(4), 367-378.

Estes, A., Zwaigenbaum, L., Gu, H., John, T. S., Paterson, S., Elison, J. T., ... & Kostopoulos, P. (2015). Behavioral, cognitive, and adaptive development in infants with autism spectrum disorder in the first 2 years of life. *Journal of Neurodevelopmental Disorders*, *7*(1), 1-10.

Fountain, C., Winter, A. S., & Bearman, P. S. (2012). Six developmental trajectories characterize children with autism. *Pediatrics*, *129*(5), e1112-e1120.

Freedman, R., Lewis, D. A., Michels, R., Pine, D. S., Schultz, S. K., Tamminga, C. A., Gabbard, G. O., Gau, S. S.-F., Javitt, D. C., Oquendo, M. A., Shrout, P. E., Vieta, E., & Yager, J. (2013). The initial field trials of DSM-5: New blooms and old thorns. *The American Journal of Psychiatry*, *170*(1), 1-5. https://doi-org.ezproxy2.library.drexel.edu/10.1176/appi.ajp.2012.12091189

Gabrielsen, T. P., Farley, M., Speer, L., Villalobos, M., Baker, C. N., & Miller, J. (2015). Identifying autism in a brief observation. *Pediatrics*, *135*(2), e330-e338. https://doi-org.ezproxy2.library.drexel.edu/10.1542/peds.2014-1428

Gordon-Lipkin, E., Foster, J., & Peacock, G. (2016). Whittling down the wait time: exploring models to minimize the delay from initial concern to diagnosis and treatment of autism spectrum disorder. *Pediatric Clinics*, *63*(5), 851-859.

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

Grossman, R. B. (2015). Judgments of social awkwardness from brief exposure to children with and without high-functioning autism. *Autism*, *19*(5), 580-587. https://doi-org.ezproxy2.library.drexel.edu/10.1177/1362361314536937

Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry*, *54*(5), 582-590.

Hedley, D., Nevill, R., Uljarević, M., Butter, E., & Mulick, J. A. (2016). ADOS-2 Toddler and Module 1 standardized severity scores as used by community practitioners. *Research in Autism Spectrum Disorders*, *32*, 84-95. https://doi-org.ezproxy2.library.drexel.edu/10.1016/j.rasd.2016.09.005

Kaboski, J., McDonnell, C. G., & Valentino, K. (2017). Resilience and autism spectrum disorder: Applying developmental psychopathology to optimal outcome. *Review Journal of Autism and Developmental Disorders*, *4*(3), 175-189.

Kanne, S. M., & Bishop, S. L. (2020). Editorial Perspective: The autism waitlist crisis and remembering what families need. *Journal of Child Psychology and Psychiatry*. https://doi-org.db.usciences.edu/10.1111/jcpp.13254

Kim, S. H., Bal, V. H., Benrey, N., Choi, Y. B., Guthrie, W., Colombi, C., et al. (2018). Variability in autism symptom trajectories using repeated observations from 14 to 36 months of age. Journal of the American Academy of Child Adolescent Psychiatry, 57(11), 837–848. https://doi.org/10.1016/j.jaac.2018.05.026.

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.

Kirkovski, M., Enticott, P. G., & Fitzgerald, P. B. (2013). A review of the role of female gender in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(11), 2584-2603. https://doi.org/10.1007/s10803-013-1811-1.

Klin, A., Lang, J., Cicchetti, D. V., & Volkmar, F. R. (2000). Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV Autism Field Trial. *Journal of Autism and Developmental Disorders*, *30*(2), 163-167. https://doi-org.ezproxy2.library.drexel.edu/10.1023/A:1005415823867

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498. https://doi-org.ezproxy2.library.drexel.edu/10.1037/0033-2909.108.3.480

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 56*(6), 466-474. https://doi.org/10.1016/j.jaac.2017.03.013.

Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2)*. Torrance, CA: Western Psychological Services.

Mandell, D. S., Wiggins, L. D., Carpenter, L. A., Daniels, J., DiGuiseppi, C., Durkin, M. S., ... & Shattuck, P. T. (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health*, *99*(3), 493-498.

McDonnell, C. G., Bradley, C. C., Kanne, S. M., Lajonchere, C., Warren, Z., & Carpenter, L. A. (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. https://doi-org.ezproxy2.library.drexel.edu/10.1007/s10803-018-3831-3

Miller, D. J., Spengler, E. S., & Spengler, P. M. (2015). A meta-analysis of confidence and judgment accuracy in clinical decision making. *Journal of Counseling Psychology*, *62*(4), 553-567. https://doi-org.ezproxy2.library.drexel.edu/10.1037/cou0000105

Monteiro, S. A., Dempsey, J., Berry, L. N., Voigt, R. G., & Goin-Kochel, R. P. (2019). Screening and referral practices for autism spectrum disorder in primary pediatric care. *Pediatrics, 144*(4), e20183326.

Mullen, E. M. (1995). *Mullen Scales of Early Learning (AGS ed.)*. Circle Pines, MN: American Guidance Service Inc.

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., ... & Zwaigenbaum, L. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*, *56*(9), 988-998.

Penner, M., King, G. A., Hartman, L., Anagnostou, E., Shouldice, M., & Hepburn, C. M. (2017). Community general pediatricians' perspectives on providing autism diagnoses in Ontario, Canada: A qualitative study. *Journal of Developmental and Behavioral Pediatrics, 38*(8), 593

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., Cook, E. H., Jr., Leventhal, B. L., & 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.ezproxy2.library.drexel.edu/10.1097/01.chi.0000227880.42780.0e

Robins, D. L., Fein, D., & Barton, M. L. (2009). Modified-Checklist for Autism in Toddlers, Revised, with Follow-up (M-CHAT-R/F). Self-published.

Rogers, C. L., Goddard, L., Hill, E. L., Henry, L. A., & Crane, L. (2016). Experiences of diagnosing autism spectrum disorder: A survey of professionals in the United Kingdom. *Autism*, *20*(7), 820-831. https://doi-org.ezproxy2.library.drexel.edu/10.1177/1362361315611109

Rotholz, D. A., Kinsman, A. M., Lacy, K. K., & Charles, J. (2017). Improving early identification and intervention for children at risk for autism spectrum disorder. *Pediatrics, 139*(2), e20161061.

Sanchez, M. J., & Constantino, J. N. (2020). Expediting clinician assessment in the diagnosis of autism spectrum disorder. *Developmental Medicine & Child Neurology*, *62*(7), 806-812. https://doi-org.ezproxy2.library.drexel.edu/10.1111/dmcn.14530

Sasson, N., Faso, D., Nugent, J., Lovell, S., Kennedy, D.P., & Grossman, R.B. (2017). Neurotypical peers are less willing to interact with those with autism based on thin slice judgments. *Scientific Reports, 7*(1), 1-10. https://doi.org/10.1038/srep40700

Sparrow, S.S., Cicchetti, D.V., & Saulnier, C.A. (2016). *Vineland Adaptive Behavior Scales, Third Edition (Vineland-3)*. San Antonio, TX: Pearson.

Sparrow, S.S., Cicchetti, D.V., & Balla, D.A. (2005). *Vineland Adaptive Behavior Scales, Second Edition (Vineland-II)*. New York, NY: Springer.

Stone, W. L., Coonrod, E. E., & Ousley, O. Y. (2000). Brief report: Screening tool for autism in two-year-olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders, 30*(6), 607.

Unigwe, S., Buckley, C., Crane, L., Kenny, L., Remington, A., & Pellicano, E. (2017). GPs' confidence in caring for their patients on the autism spectrum: An online self-report study. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, *67*(659), e445-e452. https://doi.org/10.3399/bjgp17X690449

Wallis, K. E., Guthrie, W., Bennett, A. E., Gerdes, M., Levy, S. E., Mandell, D. S., & Miller, J. S. (2020). Adherence to screening and referral guidelines for autism spectrum disorder in toddlers in pediatric primary care. *PloS one*, *15*(5), e0232335. https://doi.org/10.1371/journal.pone.0232335

Table 1

*Sample demographic characteristics, overall and by final diagnosis*

| | | **Final Diagnosis** | |
|---|---|---|---|
| | **Total Sample** $n$ **= 294** | **ASD** $n$ **= 135** | **Non-ASD** $n$ **= 159** |
| | **N (%)** | **N (%)** | **N (%)** |
| **Sex** | | | |
| Male | 199 (67.69) | 103 (76.30) | 96 (60.38) |
| Female | 95 (32.31) | 32 (23.70) | 63 (39.62) |
| **Race**[1] | | | |
| White/Caucasian | 138 (46.94) | 56 (41.48) | 82 (51.57) |
| Black/African American | 65 (22.11) | 27 (20.00) | 38 (23.90) |
| Asian | 22 (7.48) | 16 (11.85) | 6 (3.77) |
| Bi- or multiracial | 30 (10.20) | 17 (12.59) | 13 (8.18) |
| Other | 15 (5.10) | 10 (7.41) | 5 (3.14) |
| Unknown | 24 (8.16) | 9 (6.67) | 15 (9.43) |
| **Ethnicity** | | | |
| Hispanic | 80 (27.21) | 35 (25.93) | 45 (28.30) |
| Non-Hispanic | 185 (62.93) | 87 (64.44) | 98 (61.64) |
| Unknown | 29 (9.86) | 13 (9.63) | 16 (10.06) |
| **Maternal Education**[2] | | | |
| Less than high school or GED | 17 (5.78) | 3 (2.22) | 14 (8.81) |
| High school/GED | 75 (25.51) | 43 (31.85) | 32 (20.13) |
| Some college, technical or trade school | 63 (21.43) | 24 (17.78) | 39 (24.53) |
| College degree | 77 (26.19) | 37 (27.41) | 40 (25.16) |
| Advanced degree | 58 (19.73) | 28 (20.74) | 30 (18.87) |
| Unknown | 4 (1.36) | 0 (0.00) | 4 (2.52) |
| | ***M (SD)*** | ***M (SD)*** | ***M (SD)*** |
| **Age in months** | 20.95 (5.73) | 22.60 (6.10) | 19.55 (5.00) |
| **ADOS-2 CSS Total** (n = 293)[3] | 4.88 (3.10) | 7.76 (1.87) | 2.41 (1.27) |
| **VABS** (n = 277)[3] | 83.46 (14.99) | 76.01 (12.39) | 90.45 (13.84) |
| **MSEL** (n = 384)[3] | 74.36 (18.73) | 65.36 (16.12) | 81.53 (17.57) |
| **Clinician Confidence in Initial Impression** (n = 286)[4] | 3.35 (1.17) | 2.93 (1.22) | 3.71 (0.99) |

[1] Child's race was dichotomized to White and Other (Non-White) for modeling

[2] Maternal education was dichotomized into Less than college degree and College degree or higher for analyses

[3] CSS ranges from 1-10 with higher values indicating greater severity. VABS and MSEL are standard scores, with a mean of 100 and standard deviation of 15. One participant's ADOS-2 CSS is invalid due to mobility concerns. MSEL scores for 10 toddlers, and VABS scores for 17 toddlers were not available.

[4] Clinicians' confidence in initial impression being ASD or non-ASD ranges from 1 to 5, with higher values indicating greater confidence. Confidence scores for 8 toddlers were not available.

Table 2

*Results from logistic regression assessing the match between clinicians' initial impression and final diagnosis*

| | Model 1 | | | | Model 2 | | | |
|---|---|---|---|---|---|---|---|---|
| **Variable** | **β (SE)** | **Wald** | ***p*** | **Odds (95% CI)** | **β (SE)** | **Wald** | ***p*** | **Odds (95% CI)** |
| Age | 0.033 (0.035) | 0.908 | .341 | 1.033 (0.966, 1.106) | | | | |
| Sex | | | | | | | | |
| Male | 0.240 (0.416) | 0.334 | .563 | 1.272 (0.563, 2.874) | | | | |
| Female | Ref. | | | | | | | |
| Race | | | | | | | | |
| White | 0.031 (0.388) | 0.006 | .936 | 1.031 (0.482, 2.208) | | | | |
| Racial minority | Ref. | | | | | | | |
| Maternal education | | | | | | | | |
| College or higher | 0.189 (0.378) | 0.250 | .617 | 1.208 (0.576, 2.536) | | | | |
| Less than college degree | Ref. | | | | | | | |
| Confidence | 0.840 (0.172) | 23.977 | <.0001 | 2.317 (1.655, 3.243) | 0.838 (0.163) | 26.351 | <.0001 | 2.312 (1.679, 3.184) |
| ADOS-2 | -0.182 (0.082) | 4.907 | .027 | 0.833 (0.709, 0.979) | -0.158 (0.077) | 4.245 | .039 | 0.854 (0.734, 0.992 |
| VABS | 0.004 (0.019) | 0.038 | .846 | 1.004 (0.967, 1.042) | 0.004 (0.018) | 0.055 | .815 | 1.004 (0.969, 1.041) |
| MSEL | 0.005 (0.014) | 0.112 | .738 | 1.005 (0.977, 1.033) | 0.005 (0.014) | 0.151 | .697 | 1.005 (0.979, 1.033) |

Table 3

*Results from ordinal logistic regression assessing clinicians' confidence in their initial impression and final diagnosis*

| | Model 1 | | | | Model 2 | | | |
|---|---|---|---|---|---|---|---|---|
| **Variable** | **β (SE)** | **Wald** | ***p*** | **Odds (95% CI)** | **β (SE)** | **Wald** | ***p*** | **Odds (95% CI)** |
| Age | -0.023 (0.021) | 1.192 | .275 | 0.977 (0.937, 1.019) | | | | |
| Sex | | | | | | | | |
| Male | -0.202 (0.259) | 0.608 | .436 | 0.817 (0.492, 1.358) | | | | |
| Female | Ref. | | | | | | | |
| Race | | | | | | | | |
| White | 0.213 (0.256) | 0.691 | .406 | 1.237 (0.749, 2.042) | | | | |
| Racial minority | Ref. | | | | | | | |
| Maternal education | | | | | | | | |
| College or higher | -0.178 (0.250) | 0.506 | .477 | 0.837 (0.513, 1.366) | | | | |
| Less than college degree | Ref. | | | | | | | |
| VABS | -0.002 (0.013) | 0.037 | .848 | 0.998 (0.973, 1.022) | -0.009 (0.012) | 0.585 | .444 | 0.991 (0.969, 1.014) |
| MSEL | -0.007 (0.009) | 0.632 | .427 | 0.993 (0.975, 1.011) | 0.001 (0.008) | 0.007 | .935 | 1.001 (0.984, 1.017) |
| ADOS-2 | -0.464 (0.073) | 40.562 | <.0001 | | -0.238 (0.059) | 16.461 | <.0001 | 0.788 (0.702, 0.884) |
| Impression | | | | | | | | |
| ASD | -5.096 (0.916) | 30.953 | <.0001 | | 0.634 (0.328) | 3.741 | .053 | 1.884 (0.992, 3.581) |
| Non-ASD | Ref. | | | | Ref. | | | |
| Impression and ADOS-2 | | | | | | | | |
| ASD x Severity | 0.862 (0.129) | 44.560 | <.0001 | | | | | |
| Non-ASD x Severity | Ref. | | | | | | | |
| **Stratified analyses** | | | | | | | | |
| Impression, low severity | | | | | | | | |
| ASD | | | | 0.324 (0.093, 1.126) | | | | 0.341 (0.103, 1.137) |

| | | |
|---|---|---|
| Non-ASD | Ref. | Ref. |
| Impression, high severity | | |
| ASD | 2.636 (1.095, 6.342) | 2.794 (1.246, 6.266) |
| Non-ASD | Ref. | Ref. |

**Figure Legends**

*Figure 1.* Flow chart indicating match (blue solid lines) and mismatch (red dashed line) between clinician's initial impressions and child's final diagnoses.

*Figure 2.* Clinicians' confidence in initial impression (rated on a scale from 1 to 5) based on initial impression group (ASD vs. non-ASD) and child's final diagnosis (ASD vs. non-ASD). Solid lines indicate match between initial impression and final diagnosis whereas dashed lines indicate a mismatch.

*Figure 3.* Mean match (i.e., average of % of match cases within certainty) between initial impression and final diagnosis as a function of clinician's confidence in initial impression. Higher confidence scores indicate greater confidence (1 = 'not very confident;' 3 = 'confident;' 5 = 'extremely confident').

*Figure 1*:

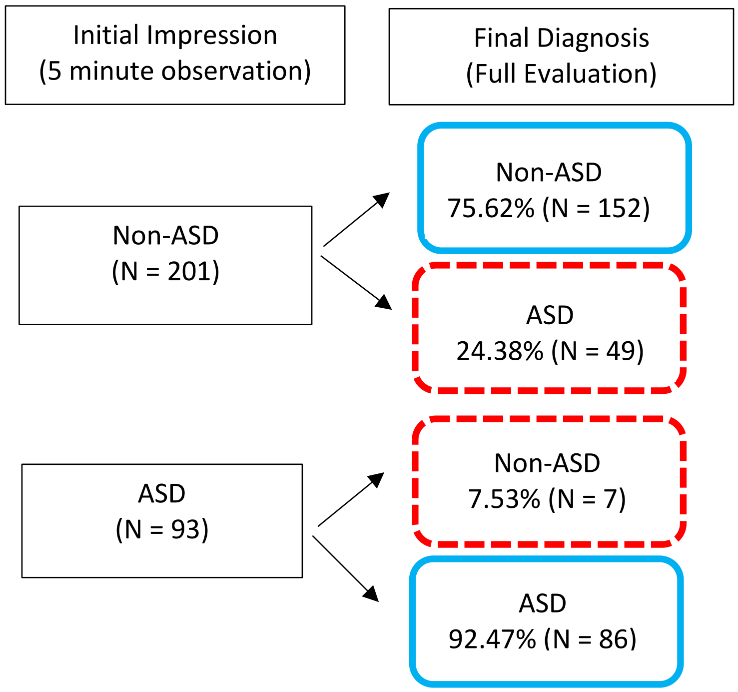

Initial Impression
(5 minute observation)
Final Diagnosis
(Full Evaluation)
Non-ASD
(N = 201)
Non-ASD
75.62% (N = 152)
ASD
24.38% (N = 49)
ASD
(N = 93)
Non-ASD
7.53% (N = 7)
ASD
92.47% (N = 86)

*Figure 2:*

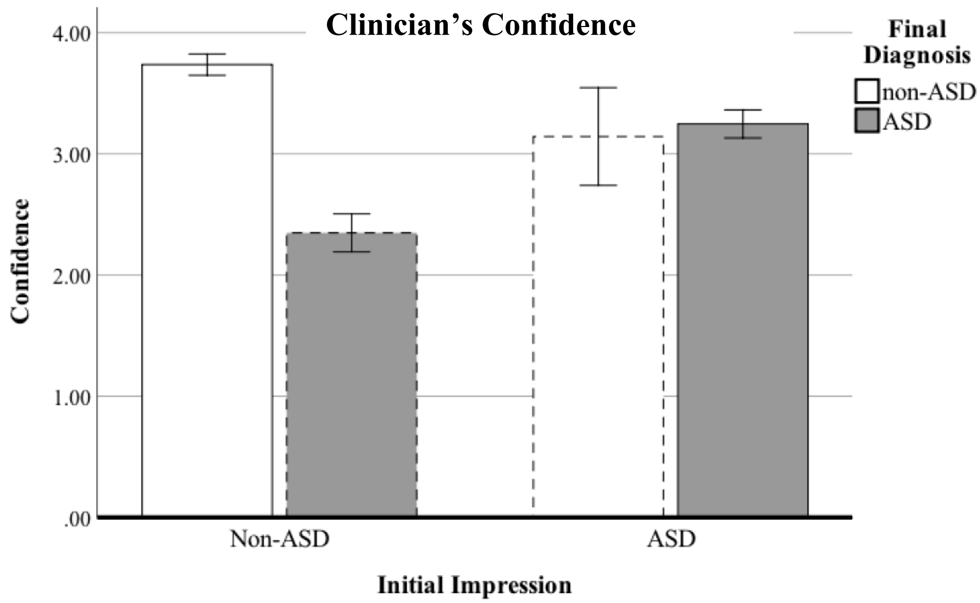
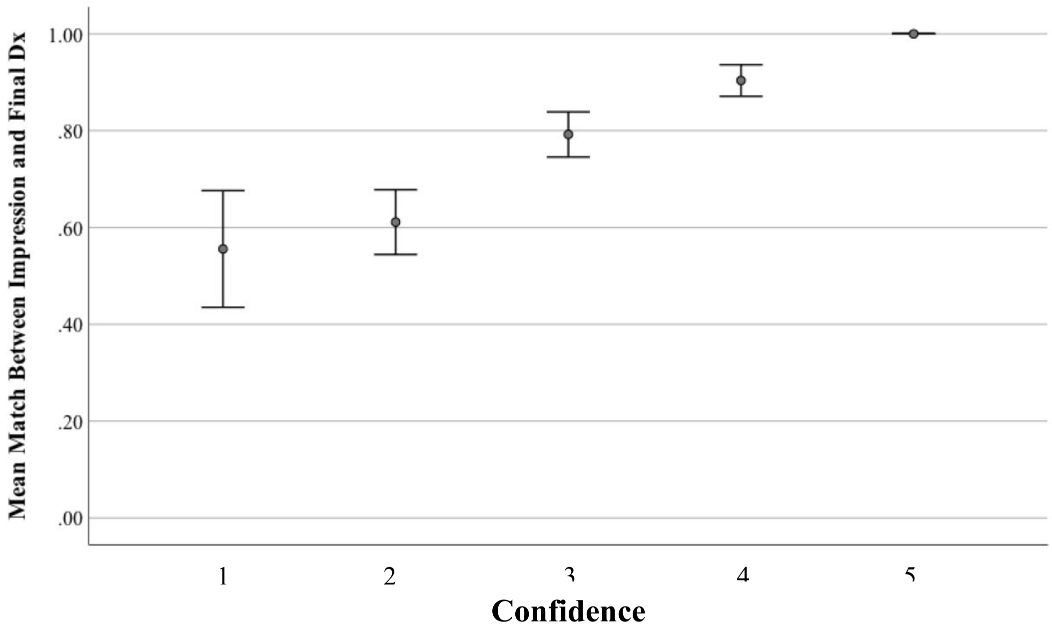



*Figure 3:*

n=18 n=54 n=77 n=83 n=54


Mean Match Between Impression and Final Dx
1.00
.80
.60
.40
.20
.00
1
2
3
4
5
Confidence