2020

Diagnosis and Screening of Autism Spectrum Disorder in Clinical Populations

Jenna Brooke Lebersfeld
University of Alabama at Birmingham

DIAGNOSIS AND SCREENING OF AUTISM SPECTRUM DISORDER IN CLINICAL POPULATIONS

by

Jenna Brooke Lebersfeld

SARAH E. O'KELLEY, COMMITTEE CHAIR
FRED J. BIASINI
KRISTI C. GUEST
MARIA I. HOPKINS
SYLVIE MRUG

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2020

DIAGNOSIS AND SCREENING OF AUTISM SPECTRUM DISORDER IN CLINICAL POPULATIONS

JENNA BROOKE LEBERSFELD

MEDICAL/CLINICAL PSYCHOLOGY PROGRAM

ABSTRACT

Screening and diagnosis of children with autism spectrum disorder (ASD) are crucial for these individuals to receive appropriate ASD-focused intervention as early as possible. Using ASD-specific strategies in early intervention leads to improvement across skill areas. The *Autism Diagnostic Observation Schedule, Second Edition* (*ADOS-2*) and the *Autism Diagnostic Interview, Revised* (*ADI-R*) are two gold-standard diagnostic instruments used in comprehensive ASD evaluations. These measures have high accuracy in research settings, but evidence of accuracy in clinical settings is less robust. They also require significant training and expertise to administer, and comprehensive ASD evaluations are in high demand. Therefore, it is important to optimize the diagnostic process to ensure that children at-risk for ASD have access to timely evaluations.

The first study, a systematic literature review and meta-analysis, showed that these measures had high accuracy overall, with the *ADOS-2* performing better than the *ADI-R*. The *ADI-R* was less accurate in clinical settings compared with research settings, but results were variable for the *ADOS-2*. The second study evaluated the accuracy of these measures in a tertiary care center and showed that the *ADOS-2* had high accuracy predicting final clinical diagnosis. The *ADI-R* was additive at times, such as when evaluating older children who may have more complex clinical presentations. In the third study, the clinical utility of using two screening instruments, the *Modified Checklist for Autism in Toddlers, Revised* (*M-CHAT-R*) and the *Communication and Symbolic*

Behavior Scales, Infant-Toddler Checklist (*CSBS-ITC*), in the triage process in a tertiary care center was investigated. Both screening measures had high sensitivity and low specificity in this setting. The *CSBS-ITC* predicted both the ASD diagnostic measures and final clinical diagnosis. The *M-CHAT-R* predicted the *ADOS-2*, but results for the *ADI-R* and final clinical diagnosis were mixed.

Overall, these studies shed light into the clinical utility of screening and diagnostic measures, particularly the accuracy of the *ADOS-2* in ASD evaluations and the additive nature of the use of the *CSBS-ITC* in screening and clinic assignment. Future research should continue to investigate ways to optimize access to screening, evaluation, diagnosis, and appropriate intervention for children at-risk for ASD.

Keywords: autism spectrum disorder, diagnosis, screening, ADOS-2, ADI-R

DEDICATION

To my incredibly supportive and loving family

TABLE OF CONTENTS

LIST OF TABLES

Table *Page*

SYSTEMATIC REVIEW AND META-ANALYSIS OF THE CLINICAL UTILITY OF THE *ADOS-2* AND THE *ADI-R* IN DIAGNOSING AUTISM SPECTRUM DISORDERS IN CHILDREN

AGREEMENT OF THE *ADOS-2* AND *ADI-R* IN A CLINICAL SAMPLE OF CHILDREN REFERRED FOR AN AUTISM SPECTRUM DISORDER EVALUATION

CLINICAL UTILITY OF TWO SCREENING MEASURES FOR AUTISM SPECTRUM DISORDER: *M-CHAT-R* AND *CSBS-DP-ITC*

LIST OF FIGURES

Figure *Page*

INTRODUCTION

Background

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects one in every 59 children (Baio et al., 2018). Children with ASD have difficulty with social communication skills, including pragmatic language and peer interactions, as well as restricted interests and repetitive patterns of behavior, as specified in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychiatric Association, 2013). Long-term outcomes for individuals with ASD vary widely, with some living independently in adulthood and others requiring support throughout their lives (Fein et al., 2013; Gray et al., 2014). In addition to higher cognitive, language, and social abilities, one of the primary predictors of independent living and overall better quality of life includes receiving ASD-specific interventions as early as possible in childhood (Helt et al., 2008).

ASD-Specific Early Intervention

Under the Individuals with Disabilities Education Act (IDEA, 2004), early intervention services for children from birth to three years old with developmental delay or another specific health condition are provided in every state. General early

intervention services improve the course of the disorder for individuals with ASD, but therapies *specifically* focused on improving the core symptoms of ASD are more effective compared to therapies for general developmental delays (Bryson, Rogers, & Fombonne, 2003). Effective ASD interventions address core ASD symptoms (i.e., social communication and restricted and repetitive behaviors and interests) as well as other features commonly seen in individuals with ASD including language delays, motor delays, behavioral difficulties, and sensory sensitivities (Wong et al., 2015). Therapy is provided individually or in group settings and can take place at home, in school, or in the community (Wong et al., 2015).

A multitude of interventions have been established as effective for targeting core symptoms of ASD in young children under age five years old (National Autism Center, 2015; National Collaborating Center for Mental Health, 2013). Social skills therapies can be used to improve social difficulties for individuals with ASD. Strategies include in-vivo and video modeling of appropriate social skills (Murray & Noland, 2012), story-based interventions (e.g., Social Stories™; Gray & Garand, 1993; Kokina & Kern, 2010) in which stories are written from a first-person perspective demonstrating appropriate behaviors in different social situations, and training parents and peers to implement strategies across settings (Beaudoin, Sébire, & Couture, 2014; Watkins et al., 2015). Behavioral interventions are often beneficial for children with ASD such as discrete trial teaching (DTT), applied behavior analysis (ABA; Lovaas, 1987), and early intensive behavioral intervention (EIBI; Reichow, Barton, Boyd, & Hume, 2012). These types of interventions focus on improving behavior using reward systems to encourage positive behaviors and decrease negative behaviors (Lovaas, 1987). Children with ASD often

thrive in this predictable environment and learn prerequisite skills necessary for future skill acquisition such as paying attention and following directions. Comprehensive programs which combine aspects of effective interventions to address many different areas of functioning have also been shown to be effective (National Autism Center, 2015; National Collaborating Center for Mental Health, 2013; Wong et al., 2015).

These ASD interventions have been shown to be more effective at earlier ages by improving a wide range of skills necessary for success in school and adulthood, including play skills, social skills, communication, cognition, nonverbal abilities, and motor skills (Bryson et al., 2003). It is crucial to identify children with ASD as early as possible and differentiate them from other children with global delays to implement effective ASD-focused therapies. In order to differentiate children with ASD from children with other neurodevelopmental disorders, children must receive a comprehensive evaluation for ASD. Additional benefits of ASD diagnosis include access to systems and programs as well as insurance coverage for certain specific of therapies (e.g., ABA) in certain states.

The benefits of early diagnosis are significant. School-age children who are diagnosed prior to three years of age have greater access to intervention services, better cognition, and required less support overall when compared with children diagnosed after age three (Clark, Vinen, Barbaro, & Dissanayake, 2018). Although families of most children with ASD reported having concerns with development by 36 months, only 42% had completed a comprehensive diagnostic evaluation for ASD by that time (Baio et al., 2018). Despite these early concerns, the median age of diagnosis is 52 months (Baio et al., 2018). Further, there is a dearth of opportunities for children and families to receive ASD evaluations across the country (Oswald, Haworth, Mackenzie, & Willis, 2017). It is

imperative to determine which high-risk children referred for evaluation would be best served through receiving a comprehensive ASD-focused evaluation as opposed to another type of evaluation (e.g., broad developmental, psychoeducational, behavioral, etc.,).

ASD Evaluation and Diagnosis

When evaluating a child for ASD, the best approach is to utilize a multidisciplinary team to evaluate all aspects of the child's development to determine areas of strengths and weaknesses as well as the most accurate diagnosis. Multidisciplinary team members may include a psychologist, speech-language pathologist, occupational therapist, physical therapist, and pediatrician, as well as other disciplines to evaluate areas of concern for the child and family. An ASD-specific evaluation should include an observational measure, most commonly the *Autism Diagnostic Observation Schedule, Second Edition* (*ADOS-2*; Lord et al., 2012), and an ASD-specific clinical interview such as the *Autism Diagnostic Interview-Revised* (*ADI-R*; Penner, Anagnostou, Andoni, & Ungar, 2017; Rutter, Le Couteur, Lord, & others, 2003).

ASD Measures

Autism Diagnostic Observation Schedule, Second Edition (*ADOS-2*; Lord et al., 2012)

The *ADOS-2* is a semi-structured, 45- to 60-minute observation and interaction session with an evaluator and the child which is used to aid in the diagnosis of ASD. There are five modules available, and module selection is based on both the age and verbal ability of the child.

ADOS-2 Module	Classifications	Levels of Concern	Comparison Score
Toddler Module	N/A	Little-to-No Mild-to-Moderate Moderate-to-Severe	N/A
Modules 1-3	Non-Spectrum Autism Spectrum Autism	N/A	1-2: Minimal-to-No Evidence 3-4: Low 5-7: Moderate 8-10: High
Module 4	Non-Spectrum Autism Spectrum Autism	N/A	N/A

Figure 1. *ADOS-2* results by measure.

For each module, 10 to 15 developmentally appropriate activities, interview questions, and conversation topics with specific presses are administered by the clinician to yield different levels of symptoms associated with ASD. Following administration, most items are coded on a scale from zero to three, with zero indicating typical development, one indicating some abnormalities but not necessarily indicative of ASD, two indicating definite ASD symptoms, and three indicating severe ASD symptoms to such an extent that they may have interfered with test administration. Items are coded based solely on the clinician's observation of the child during the *ADOS-2* administration, and no other information is considered for item scoring. Following coding of all items, algorithm items based on module, age, and language level are summed. All items which received a score of three are converted to two when summing the algorithm. All modules yield raw scores in two domains: social communication and restricted and repetitive behaviors, as well as total raw scores.

For Modules 1 through 4, total raw scores are used to determine diagnostic classifications of *autism*, *autism spectrum disorder*, or *non-spectrum*. Children under 30 months old without phrase speech are administered the Toddler Module, and the total raw score is used to determine a level of concern (i.e., *little-to-no, mild-to-moderate,* or *moderate-to-severe*) instead of a classification. For Modules 1, 2, and 3, a comparison score ranging from one to 10 can be calculated which indicates the level of ASD severity, with higher scores indicating greater levels of ASD symptomatology in comparison to children of the same age and language level. Comparison scores cannot be calculated for the Toddler Module or Module 4 based on the published clinical algorithms. Although comparison scores for these modules have been published in research studies, often called "standardized calibrated severity scores" in the literature (Esler et al., 2015; Hedley, Nevill, Uljarevi'c, Butter, & Mulick, 2016), these are not used clinically, and comparison scores for the *ADOS-2* Toddler Module or Module 4 were not utilized in these studies. Figure 1 displays the types of information that can be gleaned from each of the *ADOS-2* modules.

Autism Diagnostic Interview, Revised (*ADI-R;* Rutter et al., 2003)

The *ADI-R* is a semi-structured diagnostic interview given to a parent or caregiver by a trained clinician. The time of administration of the comprehensive *ADI-R* interview ranges from 90 to 150 minutes, and the interview consists of detailed questions about development and underlying behaviors associated with ASD and focuses on whether the child currently demonstrates ASD-specific behaviors and whether they exhibited certain ASD symptoms in the past (Rutter et al., 2003). Questions are asked about abnormalities in the areas of (A) reciprocal social interaction; (B) communication; (C) restricted,

repetitive, and stereotyped patterns of behavior; and (D) whether the abnormalities were evident at or before 36 months. This structure reflects DSM-IV criteria for Autistic Disorder. Each clinician-rated question is given a score from zero to three with a score of zero indicating no abnormalities, one indicating mild abnormalities but not necessarily indicative of ASD, two indicating definite abnormalities associated with ASD, and three indicating severe abnormalities. As with the *ADOS-2*, items with a score of three are converted to two when calculating the algorithm.

The *ADI-R* operates such that certain items in the area of communication are not asked for children who are nonverbal given that questions regarding use of language more advanced than the child's abilities would not be appropriate (e.g., using language for social chat, conversations, stereotyped language, etc.); therefore, two separate algorithms are used for verbal and nonverbal children. Additionally, some questions are asked only for children of a certain age based on the age-appropriate nature of the question (e.g., questions regarding play and friendships). A total of 34 to 40 items are included in the algorithm. Given the length of this measure, at this tertiary care clinic only the items which are used to calculate the final diagnostic algorithm are administered. Although non-algorithm items were not asked directly, given that clinicians administering *ADI-R* were research reliable and had considerable experience with the measure, it is believed that the accuracy of this approach is comparable to that of a complete research reliable administration.

This measure yields a classification of *autism* or *not autism*. Each section of the algorithm has a raw score cut-off, and a child must meet or exceed the cut-off in all four

sections to receive a classification of *autism*. Failing to exceed any cut-off by just one point yields a classification of *not autism*.

Diagnostic Considerations

Following evaluations by all clinicians, the multidisciplinary team must take into account all available information gathered during the evaluation as well as family and environmental factors and utilize clinical expertise to arrive at a final consensus clinical diagnosis for each child (Risi et al., 2006). Combining information from multiple sources including information gathered from observation, parent report, and child and family history has been shown to be the most effective approach to ASD evaluation and is considered best practice (Kim & Lord, 2012a; Risi et al., 2006; Zander, Sturm, & Bölte, 2015).

Statistical Measurement of Accuracy

To determine the accuracy of different instruments including the *ADOS-2* and the *ADI-R*, statistical measures of accuracy (e.g., sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)), are used. For the purposes of evaluating the accuracy of measures when used in an ASD-focused clinical evaluation to predict final clinical diagnosis, sensitivity is the likelihood that a child with a clinical diagnosis of ASD will score in the ASD range on the measure, and specificity indicates the likelihood that a child without ASD will score in the non-ASD range on the measure. PPV is the likelihood that a child who received an ASD classification on a measure truly

has a diagnosis of ASD, and NPV is the likelihood that a child who scores in the non-ASD range on a measure will not receive a clinical ASD diagnosis (Figure 2).

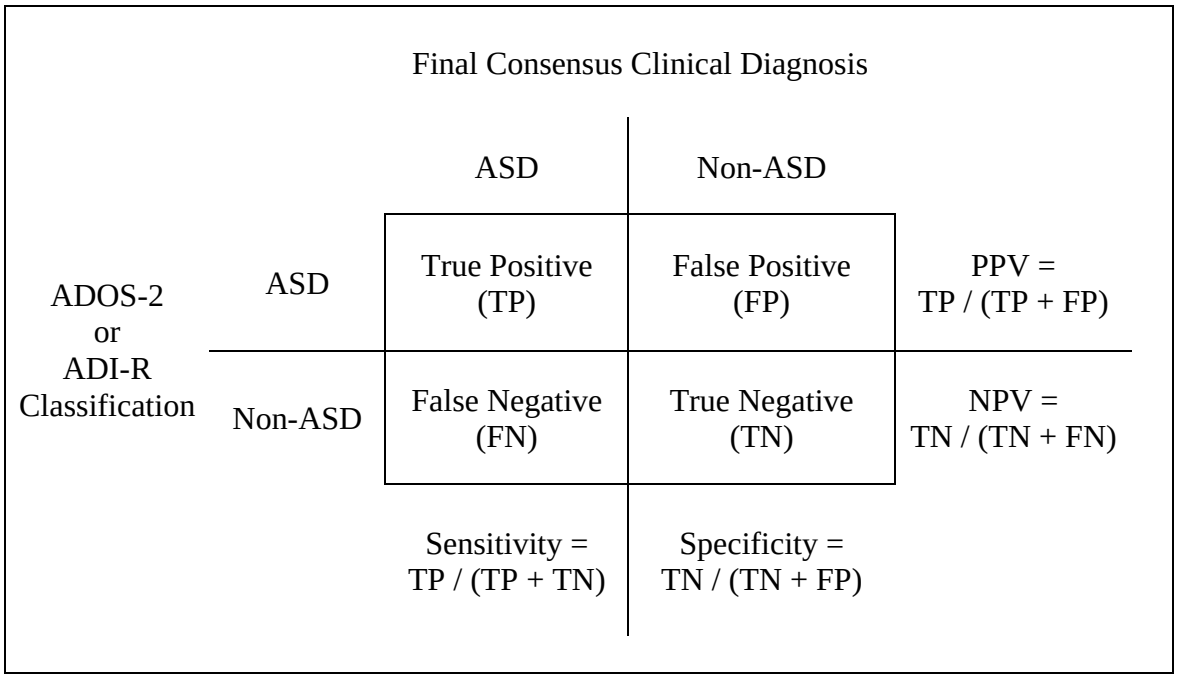

Figure 2. Measures of accuracy of ASD diagnostic tests.

Published *ADOS-2* sensitivity ranges from .60 to .95 and specificity ranges from .75 to 1.00 (Lord, Rutter, et al., 2012; Lord et al., 2012). A recent meta-analysis indicated pooled sensitivity ranging from .77 to .90 and specificity ranging from .62 to .90 for the *ADOS-2* (Dorlack, Myers, & Kodituwakku, 2018). Sensitivity and specificity were not published in the *ADI-R* manual; however, original research literature prior to measure publication indicates that sensitivity ranged from .48 to .88 and specificity was 1.00 (Cox et al., 1999; Gilchrist et al., 2001). More recent literature suggests that the sensitivity of this measure ranges from .53 to .92, and specificity ranges from .62 to .95 (Falkmer, Anderson, Falkmer, & Horlin, 2013; Risi et al., 2006). The use of these instruments together has been reported to increase sensitivity to between .70 and .98 and specificity to

.80 to .96 (Risi et al., 2006; Ventola et al., 2006), which improves diagnostic accuracy compared to each measure alone (Kim & Lord, 2012a).

Generalization of Results of Previous Research

Setting Type

Much of the literature that has been published on the accuracy of the *ADOS-2* and the *ADI-R* has been completed using evaluations from populations recruited specifically for research, and these are the studies on which the published accuracy psychometrics were based. However, it is important to understand the accuracy of these measures when used in actual clinical practice in comparison with evaluations conducted in research settings. Previous research has suggested that the sensitivity and specificity of the *ADOS-G* and *ADOS-2* are lower in clinical settings compared to the research context (de Bildt et al., 2009; Kamp-Becker et al., 2018; Langmann, Becker, Poustka, Becker, & Kamp-Becker, 2017).

Diverse Samples

When children with ASD are recruited for research studies, there are often strict inclusion criteria required for participation, including lack of comorbid diagnoses, that may not generalize to a clinical community sample (de Bildt et al., 2004; Neuhaus, Beauchaine, Bernier, & Webb, 2017; Tomanik, Pearson, Loveland, Lane, & Shaw, 2007). For example, the de Bildt et al. (2004) study only included individuals with intellectual disability. The Neuhaus et al. (2017) study was a subset of individuals in the Simons Simplex Collection (SSC), which only includes families with one child with ASD (i.e.,

no siblings with ASD), and no immediate or extended family members with a history of ASD. The Tomanik et al. (2007) study recruited participants from the community, but participants were excluded if they had a known genetic disorder, specific learning disability, or major sensory impairment. Excluding individuals based on these characteristics creates a more homogeneous sample for recruitment and analysis for research studies but does not reflect the true nature of the ASD population. Among people with ASD, there is a higher risk of ASD among siblings (Virkud, Todd, Abbacchi, Zhang, & Constantino, 2009), over 50% of the ASD population has intellectual disability (Charman et al., 2011), and there are high amounts of comorbidity within the ASD population with genetic disorders and other developmental disorder (Mannion & Leader, 2013). Results from studies which limit the recruitment sample based on participant characteristics may not be generalizable to the diagnostic clinical setting in which children with complex presentations are evaluated. Therefore, researchers have called for the investigation of *ADOS-2* and *ADI-R* agreement in community-based samples, indicating that a clinical sample may reveal additional moderators (Neuhaus et al., 2017; Tomanik et al., 2007).

Studies investigating the clinical utility of the *ADOS-2* and *ADI-R* for ASD diagnosis with diverse populations have yielded variable results. Differential diagnoses for ASD include ruling out specific language impairment, intellectual disability/global developmental delay, Attention-Deficit/Hyperactivity Disorder (ADHD), Oppositional Defiant Disorder (ODD), conduct disorder (CD), attachment disorders, social anxiety disorder, major depressive disorder, obsessive-compulsive disorder, psychosis, Tourette's disorder, and selective mutism as possible explanations for ASD symptoms (Green, Kidd,

& Accordino, 2016). Further complicating ASD assessment and diagnosis is that ASD commonly co-occurs with many of these diagnoses. Therefore, extensive training and expertise is required to administer these diagnostic instruments and accurately interpret results to clarify clinical diagnosis.

Agreement of Diagnostic Measures

Even though the *ADOS-2* and the *ADI-R* have high levels of accuracy when administered by clinicians with appropriate training, at times, these measures do not agree with one another and produce discrepant results. This can complicate the ASD diagnostic procedure, particularly for newer clinicians or those who may have less experience conducting ASD evaluations. It is important to determine the factors which predict agreement to assist clinicians in making a final diagnosis in situations when these diagnostic measures do not agree.

Child and family characteristics have been shown to be related to ASD diagnostic measures in general but have not been investigated in relation to the agreement of the *ADOS-2* and *ADI-R* specifically. Characteristics related to higher likelihood of ASD include male gender (Kreiser & White, 2014), higher ASD severity (Hill, Gray, Kamps, & Varela, 2015), increased behaviors during testing (e.g., anxiety, opposition, or inattention (Tomanik et al., 2007), lower gestational age (Joseph et al., 2017; Larsson et al., 2005), advanced birth order (Croen, Najjar, Fireman, & Grether, 2007; Glasson et al., 2004; Tomeny, Barry, & Bader, 2014), older parental age at birth (Croen et al., 2007; Hultman, Sandin, Levine, Lichtenstein, & Reichenberg, 2011; Larsson et al., 2005; Lauritsen, Pedersen, & Mortensen, 2005; Parner et al., 2012), lower socioeconomic status

(Larsson et al., 2005), and the presence of family psychiatric history (Larsson et al., 2005).

In situations in which the *ADOS-2* and *ADI-R* produce discrepant results, investigating which factors predict final diagnosis may provide additional data to multidisciplinary diagnostic teams or individual clinicians to consider in coming to an accurate diagnostic conclusion. Gaining clarity regarding these issues will help simplify and expedite the diagnostic process for ASD. Additionally, this may help inform which cases may prove to be more diagnostically complex and should be evaluated by clinicians with more expertise compared to those cases in which coming to a diagnostic conclusion may be more straightforward. Decreasing time between first concern of ASD and receiving an ASD diagnosis will give individuals with ASD access to services as early as possible, improving the likelihood of success in adulthood.

Screening for ASD

In addition to optimizing the diagnostic evaluation process for children with ASD, identifying which children would benefit most from an ASD evaluation is imperative to ensure that children with ASD receive ASD-specific early intervention services. Implementation of screening tools for ASD can help identify children for evaluation at earlier ages. Currently, ASD screening for all children is recommended at 18 and 24 months by the *American Academy of Pediatrics* (Johnson & Myers, 2007; Towle & Patrick, 2016). The *Modified Checklist for Autism in Toddlers, Revised* (*M-CHAT-R;* Robins, Fein, & Barton, 2009) and the *Communication and Symbolic Behavior Scales, Developmental Profile, Infant Toddler Checklist* (*CSBS-ITC;* Wetherby & Prizant, 2003)

are two screening tools, which are short, simple parent questionnaires and are available at no cost. The *M-CHAT-R* is an ASD-specific screening tool whereas the *CSBS-ITC* is a developmental screening tool focused on assessing language and social skills. These screening instruments have been utilized in primary care settings to identify children with developmental delays and ASD, but these measures may also have clinical utility in tertiary care settings to determine whether a child referred for evaluation should receive an ASD-specific evaluation or another type of developmental evaluation.

Statement of the Problem

It is imperative for clinicians and researchers working with individuals with ASD to continue to study and improve the ASD screening and diagnostic process. ASD evaluations are crucial to inform appropriate types of intervention services. The diagnostic process is currently lengthy, requires significant training, and resources are low. More information is needed to clarify and simplify the ASD diagnostic process to empower and enable more clinicians to accurately conduct ASD assessments. Optimization of this process will help more clinicians feel confident and qualified in conducting ASD evaluations. Increasing the number of available clinicians to conduct ASD clinical evaluations will benefit individuals with ASD and their families by increasing accessibility to ASD evaluations to qualify for ASD-specific early intervention services. Reviewing the use of these diagnostic measures in diverse, real-world samples, investigating the agreement between these diagnostic measures, and exploring the clinical utility of screening measures is a first step toward improving the diagnostic process and better serving the ASD community. The overall purpose of this project was to provide an

in-depth analysis of current ASD diagnostic procedures and to provide guidance to clinicians regarding appropriate clinical practice for ASD evaluations.

SYSTEMATIC REVIEW AND META-ANALYSIS OF THE CLINICAL UTILITY OF THE *ADOS-2* AND THE *ADI-R* IN DIAGNOSING AUTISM SPECTRUM DISORDERS IN CHILDREN

JENNA B. LEBERSFELD, MARISSA SWANSON, CHRISTIAN D. CLESI, SARAH E. O'KELLEY

In preparation for the *Journal of Autism and Developmental Disorders*

Format adapted for dissertation

ABSTRACT

Evaluations to determine the presence of autism spectrum disorder (ASD) typically utilize a behavioral observation and ASD-focused clinical interview. The *Autism Diagnostic Observation Schedule, Second Edition* (*ADOS-2*) and the *Autism Diagnostic Interview, Revised* (*ADI-R*) are two of the most well-researched and commonly used measures in ASD evaluations. This meta-analysis aimed to determine whether these measures performed better in research settings compared with clinical settings. Eligible peer-reviewed articles included children under 18 years old referred for an ASD evaluation using the *ADOS-2* and/or the *ADI-R* in the United States, Canada, or the United Kingdom. Twenty-two articles were included and analyzed using the hierarchical summary receiver operating characteristics (HSROC) model of Rutter and Gatsonis (2001). Overall, the *ADOS-2* performed better than the *ADI-R* across all settings. *ADOS-2* sensitivity ranged from .89 to .92 and specificity ranged from .81 to .85. Results regarding the accuracy of the *ADOS-2* in research compared with clinical settings were mixed. For the *ADI-R*, overall accuracy estimates indicated sensitivity of .75 and specificity of .82. The *ADI-R* was more accurate in research samples compared to clinical samples, with similar sensitivity (Research =.73, Clinical = .71) but higher specificity (Research = .85, Clinical = .72) in research samples. In general, only a small number of studies conducted in clinical settings were identified, and further research is needed to investigate how these measures operate outside research settings.

INTRODUCTION

A comprehensive psychological evaluation for children referred for possible autism spectrum disorder (ASD) is most accurately conducted by a multidisciplinary team through the use of information from multiple sources, including a clinical observation of the child and an ASD-focused clinical interview with parents or caregivers (Kim & Lord, 2012a; Risi et al., 2006; Stewart, Vigil, Ryst, & Yang, 2014). The measures which have shown the highest accuracy and are therefore most commonly used are the *Autism Diagnostic Observation Schedule, Second Edition* (*ADOS-2*; Lord et al., 2012) and the *Autism Diagnostic Interview, Revised* (*ADI-R*; Howes et al., 2017; Penner et al., 2017; Rutter et al., 2003). The *ADOS-2* is a 45- to 60-minute, semi-structured observation designed to elicit behaviors commonly demonstrated by individuals with ASD. The *Autism Diagnostic Observation Schedule, Generic* (*ADOS-G*; Lord et al., 2000), is the original version of the *ADOS-2*. Algorithms were updated for the *ADOS-G* for improved diagnostic validity; these became the basis for the *ADOS-2* (Gotham, Risi, Pickles, & Lord, 2007) and are more closely aligned with DSM-5 criteria for ASD diagnosis. The *ADI-R* is a clinician-administered interview with a parent or caregiver that consists of detailed questions about development and underlying behaviors associated with ASD (Rutter et al., 2003). This interview takes between 90 and 150 minutes. Both instruments require specialized training and experience to administer and score appropriately for diagnostic accuracy. The multidisciplinary team, often led by a licensed

clinical psychologist or physician (e.g., developmental/behavioral pediatrician), takes the results of these measures as well as other information gathered during the evaluation and uses clinical judgement to render a final diagnosis. The *ADOS-2* and *ADI-R* were initially developed as research tools and have been studied at length in the research literature, and these measures have also been published for use in clinical settings to aid diagnosis. The accuracy of the *ADOS-2* and *ADI-R* has been shown to be lower in clinical settings compared to the research context (de Bildt et al., 2009; Kamp-Becker et al., 2018; Langmann et al., 2017; Zander et al., 2016, 2017), but the majority of these studies were conducted in other countries outside of the United States where it was developed, including the Netherlands (de Bildt et al., 2009; Oosterling, Roos, et al., 2010), Greece (Papanikolaou et al., 2009), Australia (Dereu, Roeyers, Raymaekers, Meirsschaut, & Warreyn, 2012; Gray, Tonge, & Sweeney, 2008), Germany (Kamp-Becker et al., 2018) and Sweden (Zander et al., 2015; Zander et al., 2017). Differences in sociocultural norms as well as the use of language translations of the originally published measures may have increased the error associated with the use of these measures in clinical settings. Given that children referred for clinical evaluations in the community often have more complex presentations than samples recruited for research studies, it was hypothesized that these two diagnostic tools may be less accurate in clinical settings than the reported psychometrics from large scale studies conducted in the research setting. Therefore, the purpose of this systematic review and meta-analysis was to determine the accuracy and clinical utility of the *ADOS-2* and the *ADI-R* administered in English in community clinical settings in the United States, Canada, and the United Kingdom.

METHODS

This systematic review and meta-analysis utilized methods outlined in the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) of Diagnostic Test Accuracy studies guidelines (McInnes et al., 2018) and the Handbook for Diagnostic Test Accuracy Reviews (Deeks, 2013). This protocol was registered with PROSPERO 2018 (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018111589, Registration number: CRD42018111589). The research was approved by the Institutional Review Board (IRB) at the university at which this research was conducted.

Eligibility Criteria

Target Condition

The target condition included in this review is Autism Spectrum Disorder (ASD), as defined by any edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 2013) or International Classification of Diseases (ICD), contemporary to the year of article publication. Given that the definition of ASD has evolved over time, this paper uses the DSM-5 definition of ASD, which includes all previous definitions of any ASD diagnosis (e.g., Autistic Disorder, Asperger Syndrome, Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS)).

Index Tests

The index tests evaluated in this review were the *ADOS-2* and the *ADI-R*. Precursors which are considered comparable to the *ADOS-2* were also included as index tests (i.e., the *ADOS-Toddler* (Luyster et al., 2009)) and *ADOS-G* with revised algorithms (Gotham et al., 2007), as they formed the basis for the *WPS ADOS-2* publication. For ease of reference, any papers using the *ADOS-G* with revised algorithms, *ADOS-T*, or *ADOS-2* will be referred to collectively as the "*ADOS-2*" for the duration of this paper. Articles which included either one or both of the *ADI-R* and the *ADOS-2* were eligible. Older *ADOS* versions were not considered eligible index tests (i.e., *ADOS-G* without the revised algorithms (Lord et al., 2000)), *PL-ADOS* (DiLavore, Lord, & Rutter, 1995). Index tests administered for the purpose of assessing treatment outcome as opposed to initial ASD diagnosis were excluded. Papers which used the *ADI-R Toddler* diagnostic algorithms were excluded from this analysis given that these algorithms are not yet published for clinical use or widely used in the research literature.

Reference Standard

The reference standard for a final ASD or non-ASD diagnosis was the final consensus diagnosis of a comprehensive evaluation for ASD. Ideally, papers would have included both the *ADOS-2* and the *ADI-R* and involved a doctoral-level clinical psychologist in the diagnostic process, but given the limited number of available studies, requiring these criteria for inclusion proved overly restrictive for this review. Therefore, the comprehensive evaluation must have included any version of the *ADOS* and any ASD-focused clinical interview. Papers which used the *ADOS-2* as the index test were

required to include some type of ASD-focused clinical interview in the evaluation, but this interview did not necessarily have to be the *ADI-R*. For studies in which the *ADI-R* served as the index test, the evaluation did not need to include the *ADOS-2* specifically, but must have included the administration of some version of the *ADOS* (i.e., *PL-ADOS, ADOS-G, ADOS-G* with revised algorithms from Gotham 2007*, ADOS-T,* or *ADOS-2*). Papers were excluded in which the *ADI-R* was administered but for which no version of the *ADOS* was administered. Additionally, any studies which used another method for determining ASD or non-ASD diagnosis (e.g., pre-determined algorithm based on a combination of *ADOS* and *ADI-R* results) were not included, given that this type of methodology for determining ASD diagnosis does not reflect clinical practice. Studies which did not report a final consensus clinical diagnosis and included only diagnoses reported by a parent, pediatrician, or educator, and/or other forms of ASD diagnosis were not included. Thus, a conservative approach to the reference standard for diagnosis was taken.

Participants

Participants were children referred for an ASD evaluation under the age of 18 years. Studies which included mixed samples of children and adults were reviewed to determine whether data for children alone could be extracted. This was not the case for any of the included articles, therefore, samples for all included articles were comprised solely of children under the age of 18 years old. All children in included articles were administered the appropriate *ADOS-2* module based on chronological age and language level, as specified in the *ADOS-2* manual. Children with co-occurring disorders were

included, given that over 50% of children with ASD have a co-occurring disorder and excluding individuals with co-occurring disorders would limit the generalizability of results to the general population and would severely restrict the number of included articles (Mannion & Leader, 2013). Other disorders may have included, but were not limited to, the following: language disorder, intellectual disability (ID), developmental delay (DD), attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder (CD), anxiety, and depression. Children for whom a previous ASD diagnosis was rendered prior to recruitment and administration of the index tests were not included, as the focus of this review was on the utility of the *ADOS-2* and *ADI-R* in *initial* diagnostic evaluations. Additionally, studies were excluded if children were required to have certain results on the *ADOS-2* or *ADI-R* to participate in the study. For example, some studies only analyzed data and reported results from participants who received an *ADOS-2* classification of "autism spectrum" or "autism" following evaluation and excluded those who received a classification of "non-spectrum." Studies employing similar methods were not included.

Setting

Ideally, the samples of children in included articles would have received an initial clinical evaluation to evaluate the presence of an ASD within a community setting. A community setting was defined as children referred for a comprehensive clinical evaluation for ASD at a community-based clinic. Patients seen in community clinics may have been referred for a variety of reasons, including parent concern, provider concern (e.g., doctor, teacher, other clinician), and/or failing an ASD or developmental screener

(e.g., in a pediatrician's office). However, given the limited number of studies conducted in exclusively within community settings, children recruited for research evaluations for ASD were also included. This allowed for the comparison of the accuracy of the diagnostic measures between research and clinical settings. Studies which evaluated children for ASD who initially presented due to a specific genetic disorder were excluded from the meta-analysis; however, these articles are summarized separately (Table 7).

Data for all included articles were collected in the United States, Canada, or the United Kingdom (i.e., England, Scotland, Wales, Northern Ireland). Studies which utilized databases with children evaluated in one of these countries but were published outside of these countries were considered for inclusion. These geographic restrictions were chosen to minimize heterogeneity regarding language of test administration and sociocultural differences.

Language

Only articles for which an English full-text version was available were included in this study, as it was not feasible to provide translations of the non-English titles and abstracts due to limited resources. Articles for which English title and abstract review warranted full-text review, but for which English full-text articles were not initially available, were submitted through the UAB Inter Library Loan program to attempt to retrieve an English version of the article. However, for the articles submitted in this manner, no English full-text articles were available. Excluding non-English articles has not biased estimates of effectiveness in conventional medical literature (Egger et al., 2003; Jüni, Holenstein, Sterne, Bartlett, & Egger, 2002; David Moher et al., 2000), but

does have an effect on reviews of complementary and alternative medicine (Deeks, 2013; Klassen, Lawson, Moher, & others, 2005; Moher, Pham, Lawson, Klassen, & others, 2003). Given that the methodology presented in the articles reviewed is comparable to conventional medicine as opposed to complementary and alternative medicine, it is believed that excluding non-English articles did not affect the results of this paper.

Sufficient Statistics

Articles included sufficient information to calculate the number of true positives, false positives, true negatives, and false negatives for the *ADOS-2* and/or the *ADI-R* classifications compared to the diagnostic outcome of the reference standard. For some articles, these metrics were stated directly in the text or presented in supplementary materials. For articles in which these numbers were not directly stated, if the sensitivity and specificity of the measure as well as the total number of participants with an ASD diagnosis and a non-ASD diagnosis were provided, these statistics were calculated using the Review Manager software provided by the Cochrane Library (Review Manager, 2014).

Study Design

Due to the anticipated limited number of studies available for review, article eligibility included any articles which presented original research. Both prospective and retrospective study designs were included as well as cross-sectional or longitudinal designs. Participants could have been recruited through a variety of avenues, including, but not limited to, clinical referrals from caregivers, pediatricians, and other

professionals, clinical samples, community samples, and research samples. No minimum sample size was specified. Following the completion of initial article eligibility determination, case studies and case series (defined as $n \leq 10$) as well as papers which included samples with specific genetic conditions were separated from other types of articles, as it was determined that these types of selected samples were likely less representative of the population of children referred for an ASD evaluation. Including case studies and case series in this paper was deemed important to portray a comprehensive picture of the current literature. Case studies and case series papers were instead summarized in narrative form and presented below. Review articles and meta-analyses were not included, but citations within were reviewed.

Publication Status

Only articles which were published in peer-reviewed journals were considered for inclusion in this review, since the level of rigor of non-peer-reviewed literature cannot be accurately known. Theses, dissertations, conference abstracts, posters, oral presentations, editorials, interviews, replies to other articles, announcements, magazine articles, reports, books, conference proceedings, commentaries, and other "grey literature" were excluded. Although there is some evidence within intervention efficacy research showing that the exclusion of these types of products may bias the included sample due to an exclusion of negative findings, it is unclear whether this same type of bias exists for meta-analyses of diagnostic test accuracy studies (Deeks, 2013; Smart, 1964; Vogel & Windeler, 2000).

Search Strategy

Searches were conducted on 9/16/2018 from the following databases: PsycINFO (ProQuest), ERIC (ProQuest), PubMed/MEDLINE, and the Cochrane Database of Systematic Reviews (including Cochrane Central Register of Controlled Trials (CENTRAL)). These electronic databases have been widely used in other systematic literature reviews and comparison studies (Dorlack et al., 2018). On 9/18/2018, searches of the following individual journals were also conducted: Journal of Autism and Developmental Disorders, Research in Autism Spectrum Disorders, Autism Research, and Autism. These journals were selected due to their high level of rigor and relevance to the research topic. References of relevant articles were scanned individually for additional articles for inclusion. Google Scholar was used informally to identify keywords but was not included in the formal search strategy due to its lack of search reproducibility.

The search strategy was designed to maximize the number of possibly relevant studies identified with the initial search criteria and was formalized with the assistance of the UAB Libraries Reference Department. All articles published since the original publication date of each measure were considered. The *ADI-R* was published by WPS in 2003. Therefore, searches for the *ADI-R* were conducted from 2003 through the date of the search in September 2018. Given that the *ADOS* revised algorithms were published in 2007, searches to identify articles with the *ADOS* were conducted from 2007 to the present. Two searches were conducted for each database and journal. Search 1 was conducted to identify any article which included the *ADOS* (any version) and/or the *ADI-R* from 2007 to the present. Search 2 was created to identify articles which included the

ADI-R from 2003 to 2006. The PsychINFO and ERIC databases were both searched simultaneously through ProQuest. All other searches were conducted separately. For database searches, the "ADOS Search" and "ADI Search" both included keywords identified through common misspellings encountered organically and those listed in PsychINFO under "Tests." For ASD-specific journal searches, only commonly accepted test abbreviations were included in the searches, since it is unlikely that a misspelling of a commonly used test for ASD diagnosis would be overlooked in an ASD-specific journal. All searches also included a wildcard for ASD diagnosis (i.e., "autis*"), except for the Research in Autism Spectrum Disorders journal for which wildcards were not supported. Given the topic of this journal, this specifier was determined to be nonessential. Detailed search terms are included in Appendix B.

Study Selection

Citations from searches (n = 11,672) were exported into EndNote, and duplicate articles (n = 2,591) were eliminated automatically in EndNote. An additional 949 duplicate articles were identified manually. Therefore, 8,132 unique citations were reviewed. Following review of titles and abstracts, the full text of possibly included articles were obtained electronically and reviewed for eligibility. All titles, abstracts, and possibly relevant full-text articles were reviewed by two authors (JL and MS). Given differences in initial article eligibility identification between the two authors resulting in low initial agreement (i.e., 11 papers identified for inclusion by both authors, 37 papers identified by only one author but not the other, yielding 23% agreement), the inclusion criteria were clarified. All articles identified for possible inclusion by one or both authors

(all 48 articles) were then re-reviewed by the same two authors. Following the second review for eligibility, 16 papers were identified by both authors for inclusion and 10 articles were identified by only one author (yielding 62% agreement). Through discussion between these two authors, one of the 10 discrepant articles was agreed upon for inclusion in the review as a case series paper but not included in the meta-analysis. One of the 16 articles initially agreed upon for inclusion by both authors i.e., (Corsello et al., 2007) was excluded following discussion between JL and SO due to differences in diagnostic classification methodology. The remaining nine articles with discrepancy between JL and MS were reviewed by two clinical psychologists with research backgrounds working in an academic medical center who were also experts in autism spectrum disorder who were not involved with the study. These outside reviewers had 100% agreement with one another. One of the articles which both outside reviewers initially excluded due to suspected missing statistics was discussed among JL and the external reviewers and determined that appropriate statistics were present for inclusion in the meta-analysis. All other decisions by the outside reviewers for the remaining eight discrepant articles were implemented, with six articles included and two articles excluded. This resulted in a total of 22 articles for inclusion in the meta-analysis, with 14 articles included in the *ADOS-2* analyses and 13 papers included in the *ADI-R* analyses. Despite the complexity of the inclusion criteria, the additional steps taken to rectify initial low agreement likely resulted in the inclusion of all appropriate papers in the meta-analysis.

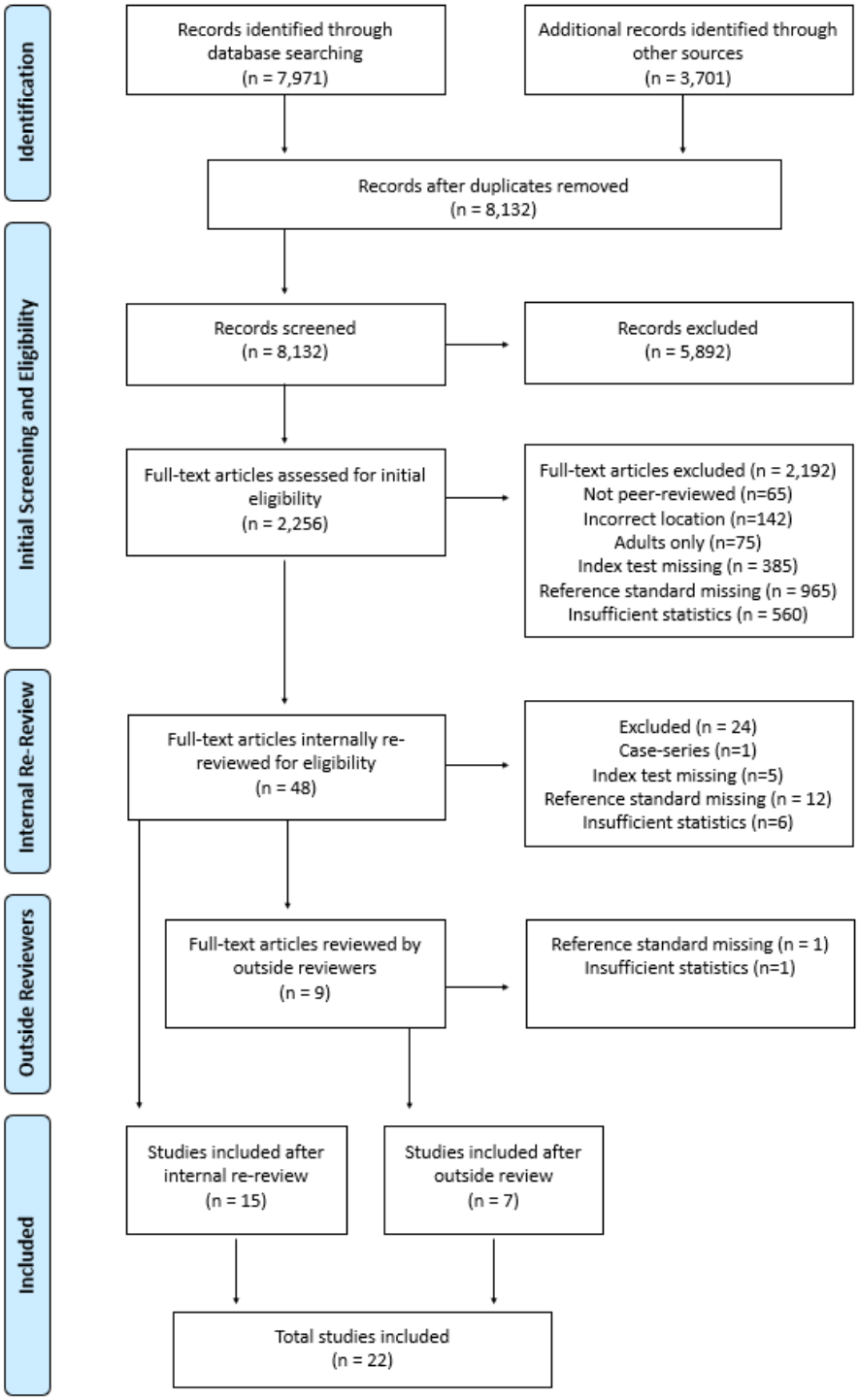

Figure 1. PRISMA flow diagram.

Data Collection Process

All data were initially extracted by JL including information regarding study characteristics, participant characteristics, measures administered, clinical evaluation characteristics, quality of study, and risk of bias. Data were organized in an online spreadsheet which was initially piloted by JL. Information relevant for statistical analyses was extracted by both JL and CC. Agreement was calculated, and any discrepancies were resolved via consensus discussion between JL and CC. The QUADAS-2 was used to assess risk of bias for each study. The QUADAS-2 was adapted and operationalized from Vllasaliu et al. (2016).

Table 1. Data Extraction

Study Characteristics
Location
Recruitment Strategy
Study Design
Participant Selection and Inclusion Criteria
Referral Question
Screening and Evaluation Methodology
Clinical Setting (Yes/No)
Participant Characteristics
Sample Size
Participant Age Range, Mean, and SD
Participant Gender
Socioeconomic Status
Participants Included in Study but Excluded from Analysis
Participant Diagnoses Prior to the Evaluation
Measures Administered
ADOS Version
Appropriate ADOS Module Administered (Yes/No)
ADI-R Version
Training and Experience of Index Test Administrators

Clinical Evaluation Characteristics
Reference Standard Administered (Yes/No)
Initial Diagnostic Evaluation (Yes/No)
DSM or ICD Version Used for Final Diagnosis
Clinical Psychologist Involved in Diagnosis (Yes/No)
Narrative Description of Diagnostic Team
ADOS-2
Module
Algorithm
True Positives
False Positives
True Negatives
False Negatives
Sensitivity
Specificity
PPV
NPV
ADI-R
True Positives
False Positives
True Negatives
False Negatives
Sensitivity
Specificity
PPV
NPV
Final Diagnosis
ASD Group N
Non-ASD Group N

A total of 116 data points was extracted by JL and CC, and 112 data points were agreed upon (97%) across the 22 articles. Two discrepancies were due to referring to the wrong text in the table, and two discrepancies were due to typographical errors or calculation errors.

DATA ANALYSIS

Diagnostic Accuracy Measures

The principal diagnostic accuracy measures reported were the sensitivity and specificity of the *ADOS-2* and the *ADI-R* using the published algorithms for each measure. Sensitivity and specificity were calculated for each measure across all papers as well as calculated individually using setting type as a covariate (clinical, research, or both). For each measure, summary receiver-operating characteristic (SROC) curves of the sensitivity and specificity for each paper by setting were generated.

Synthesis of Results

Articles were organized using the Review Manager software (RevMan; Review Manager, 2014). Each article citation was imported, and the statistical information required to calculate the true positives, true negatives, false positives, and false negatives for each measure were entered into the program. Articles were classified by setting: clinical, research, or both. A clinical setting was defined as any article for which evaluations were conducted in a community setting and were not recruited specifically for research. A research setting included any studies which recruited participants for research. Some studies included participants from both community and research settings and were classified as such. For the *ADI-R* analysis, the model converged using these three groups and statistics are reported accordingly. For the *ADOS-2* analyses, having

three groups did not allow the model to converge. Therefore, this combined group was re-categorized as part of the "research" setting for the *ADOS-2* analyses. This was viewed as the more conservative approach because if, as hypothesized, the administration of this measure in research settings was more accurate than clinical settings, including clinical evaluations in the research group would hypothetically lower the accuracy and close the gap between the clinical and research results.

Meta-Analysis

The statistical methods used for the meta-analyses were the hierarchical summary receiver operating characteristic (HSROC) model of Rutter and Gatsonis (Rutter, 1995; Rutter & Gatsonis, 2001) and were computed using the MetaDAS SAS macro (Takwoingi & Deeks, 2010). This hierarchical model (Rutter, 1995; Rutter & Gatsonis, 2001) was chosen for this analysis as recommended in the Cochrane Handbook for Diagnostic Test Accuracy and has many strengths. The model produces the pooled estimated mean of the sensitivity and specificity and accounts for the correlation between sensitivity and specificity across studies. Separate pooling of sensitivity and specificity results in underestimation of these statistics, since it does not take into account the inherent trade-off between these statistics (Deeks, 2001). Positive and negative predictive values are influenced by prevalence in the sample, which introduce heterogeneity and uncertainty. The chosen method for statistical analysis uses a Bayesian model to determine random effects and was preferred to using fixed effects due to the large amount of heterogeneity commonly seen among diagnostic test accuracy studies. Additionally, the HSROC method was recommended when covariates will be included in the model.

This model also produces the Diagnostic Odds Ratio (DOR), a global estimate of overall test accuracy. The DOR is a summary of the diagnostic accuracy of a test and in this sample can be interpreted as how many times higher the odds are of a person with ASD to score in the ASD range on the diagnostic test compared to someone without ASD. Interpretation of DOR can be used to compare across tests and models.

The meta-analytical methods were calculated separately for the *ADOS-2* and the *ADI-R*. Data were exported from Review Manager. The HSROC model was computed for all studies included for each measure individually, and then setting was included as a covariate to determine whether setting had an effect on diagnostic test accuracy. For the *ADI-R*, the setting covariate included all three groups: clinical, research, and both. For the *ADOS-2* analysis, the covariates for setting were separated into clinical only and research and both combined.

Outliers and Sensitivity Analysis

Studies were plotted graphically on HSROC plots and were visually inspected for outliers. One article (Molloy, 2011) was identified as an outlier in the *ADOS-2* analysis via visual inspection, and no outliers were identified for the *ADI-R* analysis. Sensitivity analyses were conducted by removing the outlier article and repeating the analyses. Results were compared with and without the outlier to determine the effect of this specific study on the results.

The Gotham (2007) and Gotham (2008) papers provided separate sensitivity and specificity estimates based on differing DSM-IV criteria for two instances: Autism vs. Non-spectrum (NS) and ASD vs. NS. In the Autism vs. NS analysis, PDD-NOS and

Asperger Disorder cases were not included, and *ADOS-2* classifications of ASD were classified as non-spectrum. In the ASD vs NS condition, children with Autistic Disorder were not included, and *ADOS-2* classifications of "autism spectrum" and "autism" were both considered classifications of ASD. Including both estimates in a single analysis would result in the inclusion of the non-spectrum cases more than once. To avoid this, separate analyses were conducted for the Autism vs. NS and ASD vs. NS estimates for the Gotham 2007 and Gotham 2008 articles. Additionally, results were analyzed both including all the outliers and in separate analyses with the outliers excluded. For clarity, analytic approaches for the *ADOS-2* are defined in Table 2.

Table 2. *ADOS-2* Analytical Approaches

Approach	Molloy 2011	Gotham 2007 & 2008
1	Included	ASD vs. NS
2	Included	Autism vs. NS
3	Excluded	ASD vs. NS
4	Excluded	Autism vs. NS
5	Included	Excluded
6	Excluded	Excluded

RESULTS

Study characteristics for the 22 articles included in the meta-analysis are outlined in Table 3.

Quality of the Included Studies

Quality of the studies including risk of bias and applicability concerns are presented in Figure 2. Risk of bias was unclear or high risk for 12 of the 22 papers (54%), and there were concerns regarding the use of the reference standard (unclear or high risk for all articles). However, this was primarily due to the clinicians' knowledge of the results of the index tests prior to the implementation of the reference standard, as opposed to using blind raters to come to a diagnostic conclusion. This is common practice in clinical settings, as the index tests (i.e., the ASD diagnostic measures) are inextricably linked and used as a primary source of information in the reference standard (i.e., ASD diagnostic evaluation and final clinical diagnosis). Overall, there was low risk of bias from the index tests, flow and timing, and applicability of the findings to practice.

Table 3. Study Characteristics

Study	Test(s)		Total *N*	Sex		Age	Diagnosis	
	ADOS-2	ADI-R		Male *n*	Female *n*	*M* or Range	ASD *n*	Non-ASD *n*
1. Baird 2006		X	255	223	32	12 years	158	97
2. Bishop 2017	X	X	289	203	86	8 years	142	126
3. Camodeca 2018	X		483	355	128	10 years	127	356
4. Dykens 2017	X		146	72	74	11 years	32	114
5. Gillentine 2017	X	X	18	12	6	9 years	7	10
6. Gotham 2007	X		1,630	[1]	[1]	41 to 104 months	1,351	279
7. Gotham 2008	X		1,282	923	359	37 to 118 months	1,068	214
8. Grzadzinski 2016	X	X	212	176	36	9 years	164	48
9. Guthrie 2013	X		82[2]	64	18	19 months	56	12
10. Harris 2008		X	63	63	0	8 years	38	25
11. Havdahl 2016	X		389	288	101	[3]	255	163
12. Kim 2012		X	695[4]	353	160	33 months	491	203
13. LeCouteur 2008		X	101	81	20	36 months	77	24
14. Luyster 2009	X		206	158	48	15 to 26 months	59	147
15. Mazefsky 2006		X	78	56	22	4 years	59	19
16. Molloy 2011	X		584	507	77	3 to 9 years	329	255
17. Risi 2006		X	1,039	818	221	27 to 94 months	881	158
18. Ventola 2006		X	45	37	8	26 months	36	9
19. Wiggins 2008		X	142	112	30	26 months	73	69
20. Wiggins 2015	X	X	922	581	341	59 months	584	338
21. Ziats 2016	X	X	18	14	4	14 years	8	10
22. Zwaigenbaum 2016	X		381	215	166	39 months	103	278

[1]57% to 86% male... [2]Diagnosis deferred n = 14. [3]4 years for younger group, 9 years for older group. [4]One participant diagnosis not reported.

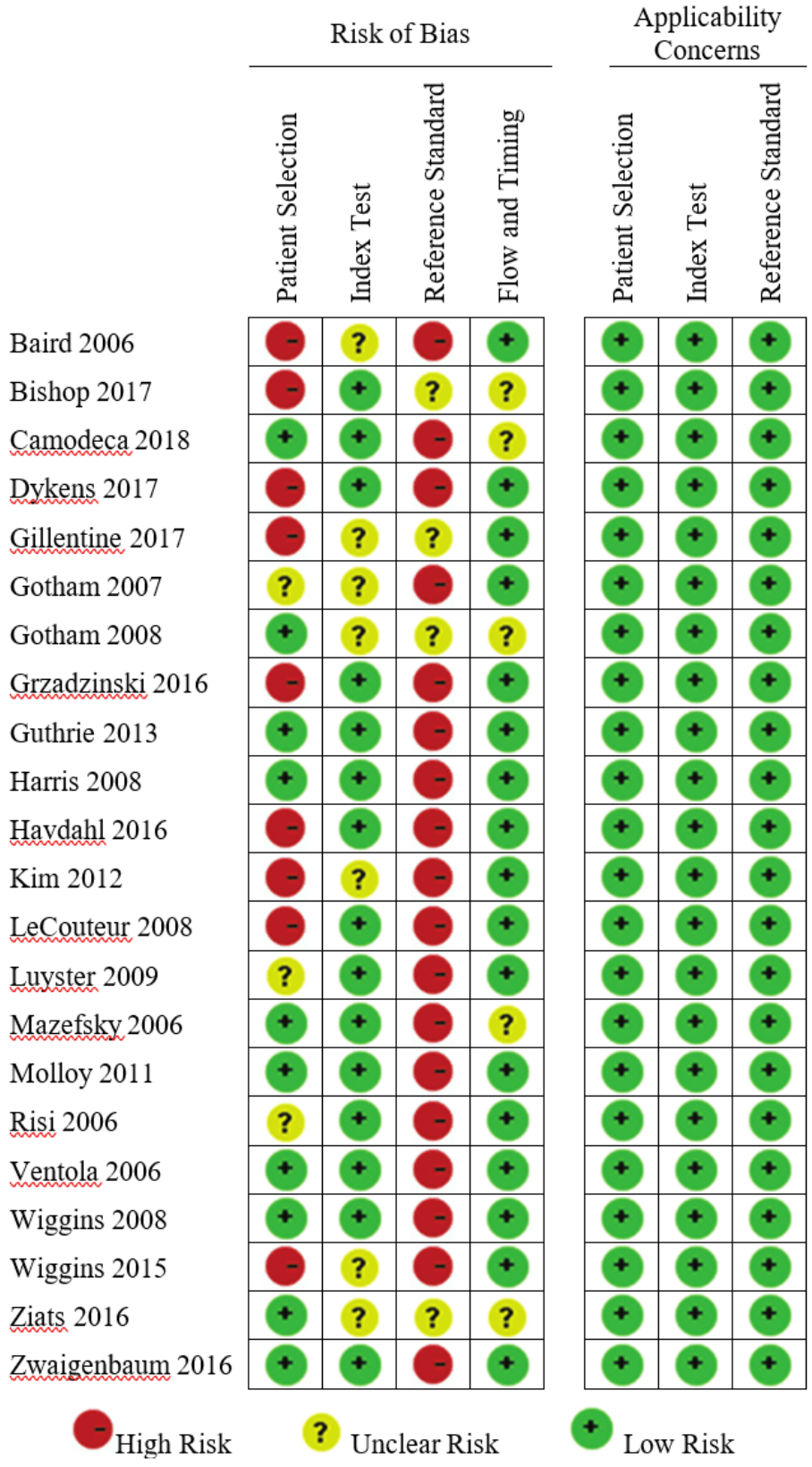

Figure 2. QUADAS-2 risk of bias and applicability concerns.

Table 4. Sensitivity and Specificity of ADOS-2 Overall and by Evaluation Setting

Approach	Overall				Research or Both				Clinical				-2LL	*p*
	n	Se	Sp	DOR	n	Se	Sp	DOR	n	Se	Sp	DOR		
1	14	.89	.81	36.5	11	.89	.80	34.8	3	.89	.80	31.0	7.02	.071
2	14	.92	.83	52.7	11	.92	.83	59.2	3	.89	.80	30.9	5.81	.120
3	13	.89	.83	42.3	11	.89	.81	36.3	2	.88	.90	71.1	3.23	.357
4	13	.92	.85	61.9	11	.92	.83	59.7	2	.88	.90	70.8	2.83	.418
5	12	.91	.81	47.0	9	.93	.81	53.8	3	.89	.80	31.0	7.87	.049*
6	11	.92	.84	55.7	--	--	--	--	--	--	--	--	7.53	.057

* p < .05; ASD = Autism Spectrum Disorder; NS = Non-Spectrum; AUT = Autism; Se = sensitivity; Sp = specificity; DOR = diagnostic odds ratio; -2LL = -2 Log Likelihood Difference; "--" = data not available

Table 5. Sensitivity and Specificity of Published ADOS Algorithms

Module and Algorithm	Gotham 2007				Gotham 2008			
	AUT vs. NS		ASD vs. NS		AUT vs. NS		ASD vs. NS	
	Se	Sp	Se	Sp	Se	Sp	Se	Sp
Module 1, No Words, NVMA > 15 mo.	.95	.94	.82	.79	.86	.80	--	--
Module 1, Some Words	.97	.91	.77	.82	.89	.91	.95	1.00
Module 2, Younger	.98	.93	.84	.77	.94	1.00	.65	.88
Module 2, Older	.98	.90	.83	.83	--	--	--	--
Module 3	.91	.84	.72	.76	.82	.92	.60	.75

NVMA = Nonverbal Mental Age; ASD = Autism Spectrum Disorder; NS = Non-Spectrum; AUT = Autism; Se = sensitivity; Sp = specificity; "--" = data not available

Estimates of overall sensitivity (Range = .89 – .92) and specificity (Range = .81 – .85) of the *ADOS-2* (see Table 4) as well as individual estimates for identified articles (Sensitivity Range = .85 – 1.00; Specificity Range = .44 – 1.00; see Figure 5) were generally comparable to published algorithms (see Table 5). Addition of the setting covariate was significant (*-2LL* = 7.87, *p* = .049) when the Gotham 2007 and 2008 papers were excluded (Table 4). The highest DOR was reported within clinical samples when the Molloy 2011 paper was excluded, and the Gotham 2007 and Gotham 2008 papers utilized the Autism vs. Non-Spectrum algorithms (Table 4). When all articles were included, the DOR was higher for research compared with clinical samples; however, inclusion of the setting covariate was not significant (p = .071). Exclusion of the Molloy 2011 article had little effect on sensitivity of the clinical sample but increased the specificity of the clinical sample from .80 to .90, which is higher than specificities in the research samples (.81 and .83; Table 5).

Interpretation of the SROC plot for all three setting types (clinical, research, and both) for Approach 1 appears to indicate that research samples have higher levels of accuracy compared with clinical samples and combined clinical and research samples. However, the Molloy 2011 paper was identified as an outlier with a specificity of .44 (Figure 5). When this paper was removed (Figure 4), visual inspection of the SROC curve for Approach 3 suggests that there was no longer a difference between accuracy of the *ADOS-2* in research and clinical settings. Using Approach 3, accuracy of the *ADOS-2* for studies including both research and clinical evaluations was lower than either research or clinical settings individually.

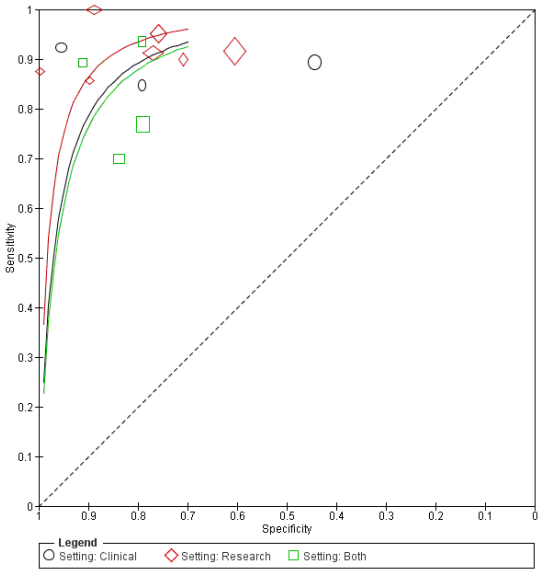

Figure 3. SROC plot of *ADOS-2* by setting for Approach 1 (Molloy, 2011 included).
Note: Size of shape indicates sample size.

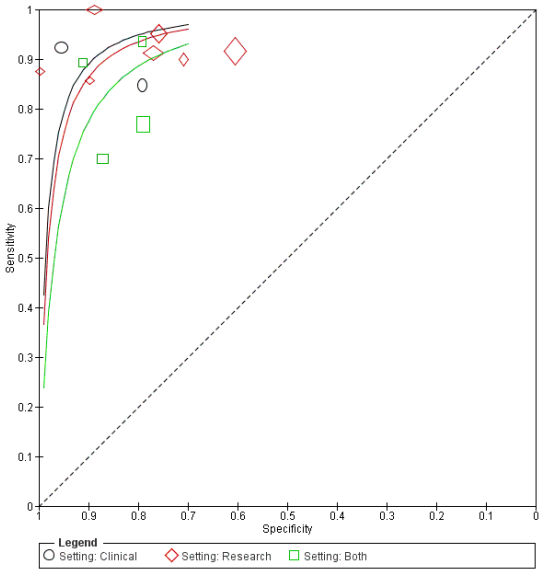

Figure 4. SROC plot of *ADOS-2* by setting for Approach 3 (Molloy, 2011 excluded).
Note: Size of shape indicates sample size.

Study	TP	FP	FN	TN	Setting	Sensitivity (95% CI)	Specificity (95% CI)
Bishop2017	116	10	8	38	Both	0.94 [0.88, 0.97]	0.79 [0.65, 0.90]
Gotham2007	783	27	37	225	Both	0.95 [0.94, 0.97]	0.89 [0.85, 0.93]
Gotham2008	639	19	109	164	Both	0.85 [0.83, 0.88]	0.90 [0.84, 0.94]
Luyster2009	50	4	6	41	Both	0.89 [0.78, 0.96]	0.91 [0.79, 0.98]
Molloy2011	294	142	35	113	Clinical	0.89 [0.86, 0.92]	0.44 [0.38, 0.51]
Grzadzinski2016	139	10	25	38	Clinical	0.85 [0.78, 0.90]	0.79 [0.65, 0.90]
Camodeca2018	84	8	7	169	Clinical	0.92 [0.85, 0.97]	0.95 [0.91, 0.98]
Wiggins2015	536	133	48	205	Research	0.92 [0.89, 0.94]	0.61 [0.55, 0.66]
Guthrie2013	46	7	5	17	Research	0.90 [0.79, 0.97]	0.71 [0.49, 0.87]
Havdahl2016	214	39	11	123	Research	0.95 [0.91, 0.98]	0.76 [0.69, 0.82]
Zwaigenbaum2016	94	64	9	214	Research	0.91 [0.84, 0.96]	0.77 [0.72, 0.82]
Dykens2017	18	14	0	114	Research	1.00 [0.81, 1.00]	0.89 [0.82, 0.94]
Gillentine2017	6	1	1	9	Research	0.86 [0.42, 1.00]	0.90 [0.55, 1.00]
Ziats2016	7	0	1	10	Research	0.88 [0.47, 1.00]	1.00 [0.69, 1.00]

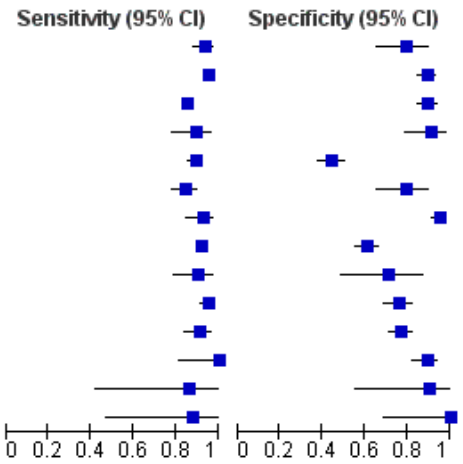

Figure 5. Forest plot of *ADOS-2* by setting using the Gotham ASD vs. NS estimates.

ADI-R

Study	TP	FP	FN	TN	Setting	Sensitivity (95% CI)	Specificity (95% CI)
Risi2006	614	202	88	326	Both	0.87 [0.85, 0.90]	0.62 [0.57, 0.66]
Bishop2017	108	16	34	110	Both	0.76 [0.68, 0.83]	0.87 [0.80, 0.93]
Grzadzinski2016	110	14	54	34	Clinical	0.67 [0.59, 0.74]	0.71 [0.56, 0.83]
Mazefsky2006	24	12	8	31	Clinical	0.75 [0.57, 0.89]	0.72 [0.56, 0.85]
Kim2012	552	60	56	94	Research	0.91 [0.88, 0.93]	0.61 [0.53, 0.69]
Ventola2006	19	3	17	6	Research	0.53 [0.35, 0.70]	0.67 [0.30, 0.93]
Gillentine2017	7	3	0	7	Research	1.00 [0.59, 1.00]	0.70 [0.35, 0.93]
Wiggins2015	450	90	134	248	Research	0.77 [0.73, 0.80]	0.73 [0.68, 0.78]
Ziats2016	6	2	1	7	Research	0.86 [0.42, 1.00]	0.78 [0.40, 0.97]
Harris2008	26	5	12	20	Research	0.68 [0.51, 0.82]	0.80 [0.59, 0.93]
Wiggins2008	24	4	49	65	Research	0.33 [0.22, 0.45]	0.94 [0.86, 0.98]
Baird2006	132	2	26	95	Research	0.84 [0.77, 0.89]	0.98 [0.93, 1.00]
LeCouteur2008	40	0	37	24	Research	0.52 [0.40, 0.63]	1.00 [0.86, 1.00]

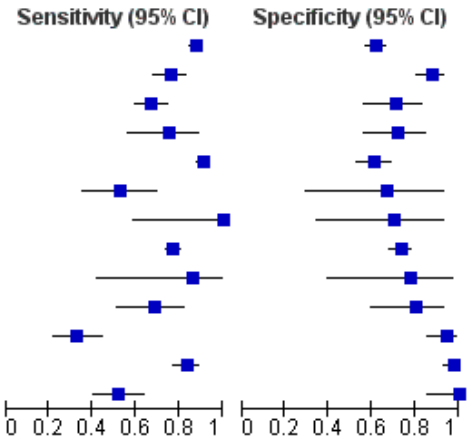

Figure 6. *ADI-R* forest plot by setting.

Table 6. Sensitivity and Specificity of *ADI-R* Overall and by Evaluation Setting

Setting	n	Sens	Spec	DOR
Overall	13	.75	.82	13.6
Research	9	.73	.85	15.8
Both	2	.82	.76	15.9
Clinical	2	.71	.72	6.2

Table 7. Sensitivity and Specificity of Published *ADI-R* Algorithms

Article	*n*	Se	Sp
Cox 1999			
ADI-R at 20 months, diagnosis at 42 months	45	.19	1.00
ADI-R at 42 months, diagnosis at 42 months	45	.48	1.00
Gilchrist 2001	53	.88	1.00

Se = Sensitivity, Sp = Specificity

Though the *ADI-R* manual does not directly report sensitivity and specificity, the articles referenced in the manual which utilized the published algorithms report sensitivities which varied widely, from .19 to .88 (see Table 7). The articles cited in the *ADI-R* manual reported perfect specificities of 1.00 (Cox et al., 1999; Gilchrist et al., 2001). In the present study, the articles identified during the meta-analysis showed overall sensitivity of .75 and specificity of .82 (Table 7). In the current meta-analysis, individual articles had a wide range of sensitivities (.33 – 1.00) and specificity (.61 – 1.00, Figure 6).

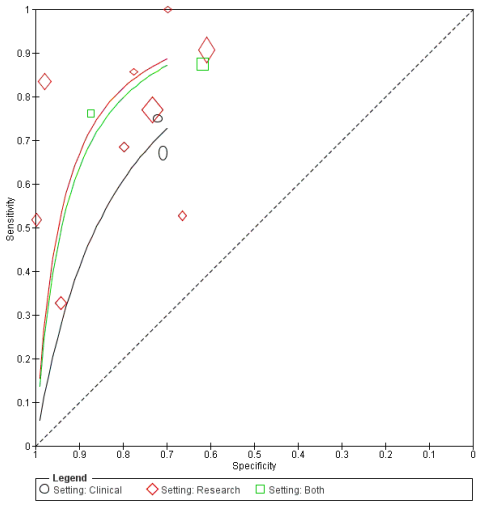

Figure 7. *ADI-R* SROC plot by setting. Note: size of shape indicates sample size.

Inclusion of the setting covariate in the model compared to the model without the covariate was trending toward significance (*-2LL* difference = 11.788, $p = .067$). Clinical and research samples had comparable sensitivity (Clinical = .71, Research = .73), but articles utilizing both research and clinical samples had higher sensitivity (.82). Specificity was higher for research samples (.85) compared to clinical samples (.72) and those including both research and clinical evaluations in the study (.76, see Table 6 and Figure 6).

Case Studies, Case Series, and Genetic Syndromes

Many rare diseases and genetic syndromes have significant overlap with ASD symptoms, and it is important to investigate the accuracy of the ASD diagnostic measures in these populations. Study characteristics are presented in Table 8.

Table 8. Characteristics of Case Studies, Case Series, and Genetic Syndromes Papers

Study	Test(s) ADOS-2	Test(s) ADI-R	*n*	Participant Characteristics	Findings
1. Baron-Cohen et al. (2006)		X	1	Child born to two parents with confirmed Asperger syndrome	ASD evaluation at 26 months of age. Child met criteria for *autism* on the *ADI-R* and received a diagnosis of Asperger syndrome
2. Bennett, Hodgetts, Mackenzie, Haqq, & Zwaigenbaum (2017)	X		10	Prader-Willi syndrome	Clinical best estimate diagnoses indicated that none of the children had ASD, although three children scored above the cut-off for ASD on the *ADOS-2*.
3. Bukelis, Porter, Zimmerman, & Tierney (2007)		X	1	Smith-Lemli-Opitz syndrome.	Co-occurring ASD and intellectual disability Scored in the *autism* range on the *ADI-R*.
4. Cooper & Hanstock (2009)		X	1	Differential diagnosis of ASD and depression	"Abnormal" scores on the *ADI-R*. Child received a diagnosis of "high-functioning autism."
5. Correia, Café, Almeida, Mouga, & Oliveira (2015)		X	1	FRAXE mutation	Exceeded the *autism* cut-off on the *ADI-R* and received a diagnosis of autistic disorder.
6. Merritt, Hart, & LeGrow (2017)	X		1	Say-Barber-Biesecker-Young-Simpson syndrome	Scores indicated ASD, although the classification was not specified, and the child received an ASD diagnosis following evaluation.
7. Reaven, Hepburn, & Ross (2008)	X	X	3	Psychosis	All children scored above published cut-offs on both the *ADOS-2* and the *ADI-R*, although none of the children received a final diagnosis of ASD.
8. Schaaf et al. (2011)		X	2	16p11.2 chromosomal rearrangements (one with a microdeletion and one with a duplication)	One child met criteria for *PDD-NOS* but did not meet cut-offs on the *ADI-R*. The other met criteria on the *ADI-R* and received a diagnosis of autistic disorder.
9. Suter, Treadwell-Deering, Zoghbi, Glaze, & Neul (2014)		X	4	MECP2 mutations without Rett Syndrome	Three participants met criteria for *autism* on the *ADI-R*. Classifications were consistent with final diagnosis in all four cases.
10. Thurm et al. (2016)	X	X	33	Smith-Lemli-Opitz syndrome.	Most individuals had elevated scores on the *ADOS-2* and the *ADI-R*, but only about half received a diagnosis of ASD
11. Treadwell-Deering, Powell, & Potocki (2010)		X	15	Potocki-Lupski Syndrome (Duplication 17p11.2),	Eight received an ASD evaluation. On the *ADI-R*, five met criteria for *autism* and received an ASD diagnosis. Two met criteria on the *ADI-R* but were not diagnosed with ASD.
12. Urraca et al. (2013)		X	13	Interstitial duplication 15q11.2-q13 syndrome	Eleven children received an ASD diagnosis, and all 11 had *ADI-R* classifications of *autism*
13. Zhang et al. (2009)		X	2	Neuroligin-4 missense mutation	Both had diagnoses of ASD and scored at or above the cut-off on the *ADI-R* for autism

DISCUSSION

This study utilized a systematic review to investigate the accuracy of the *ADOS-2* and the *ADI-R* in clinical settings compared to research settings, and it was hypothesized that these measures would perform better in research settings given the heterogeneity and complexity of children referred for an ASD evaluation in clinical samples. Overall, *ADOS-2* accuracy from the meta-analysis was generally comparable to accuracy reported in published literature. For the *ADI-R*, sensitivity and specificity in literature cited in the published manual reported variable sensitivity and perfect specificity. Accuracy estimates reported in articles identified during the current meta-analysis paint a more nuanced picture, with overall sensitivity at .75 and specificity at .82. For the *ADI-R*, accuracy in clinical studies was below that for research-only studies or those utilizing both research and clinical samples, indicating that the *ADI-R* performed less accurately in clinical settings.

Overall, the *ADOS-2* was more accurate than the *ADI-R* across measures of sensitivity and specificity in both research and clinical settings. For the *ADOS-2*, when comparing samples of children evaluated in clinical settings with those whose evaluations were completed in research settings (or which used a combination of clinical and research evaluations), analyses indicated that sensitivity was comparable or slightly lower for clinical samples compared to research samples. For specificity, results were mixed. Some analyses indicated comparable or slightly lower specificity in clinical compared to

research samples, whereas when an outlier was excluded, results showed that specificity in clinical samples was higher than research samples. This suggests that given the small number of studies identified that were conducted in solely clinical settings, a single article can have a large effect on results. Therefore, more research of evaluations initially performed for clinical purposes needs to be conducted to determine exactly how the *ADOS-2* performs in clinical settings. Given current findings, specificity of the *ADOS-2* may be more variable across clinical settings whereas sensitivity may remain relatively stable.

The genetics studies which utilized comprehensive ASD assessments for their participants suggest overall high accuracy of the ASD diagnostic measures. Many of these studies utilized the *ADI-R*, but many did not use the most recent *ADOS-2* and instead utilized an older version of this measure. This was the case even for many articles published following the 2007 release of the revised algorithms or the updated *ADOS-2* in 2012. It is unknown what factors impacted others using out of date materials, but this could be due to a lapse in publication date of the articles, or the studies may have been part of a research study in which the measures had been previously specified by the research protocol and could not be updated. Research on rare genetic disorders and other complex cases is encouraged with updated diagnostic measures to investigate how these measures perform for more complex and genetically diverse populations, especially given the significant overlap of ASD and many rare genetic disorders.

Sources of Heterogeneity

One limitation of this meta-analysis is that additional sources of heterogeneity were not investigated due to the limited number of eligible articles identified for inclusion. One consideration is that the definition of autism spectrum disorder has shifted over time, with the previous version of the DSM separating out different autism spectrum disorders including Autistic Disorder, Asperger Syndrome, and Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS). As the field progressed and additional research was conducted, it became clear that differentiating among different types of ASD conditions was difficult, and the conceptualization of ASD became instead to be viewed as a "spectrum" of symptoms presenting along a continuum of severity. In addition, in the DSM-IV, three categories of autism spectrum symptoms were specified: communication, social abilities, and restricted and repetitive behaviors and interests. As more research was conducted, it became apparent that there was not specifically a core deficit in communication skills overall, but specifically difficulties with social aspects of communication. Therefore, the current conceptualization of ASD as specified in the DSM-5 indicates only two core categories of autism spectrum related symptoms: social communication and restricted and repetitive behaviors and interests. The *ADOS-2* revised algorithms reflect this change in conceptualization of ASD. However, the *ADI-R* has not yet been updated, and the *ADI-R* manual states that the measure only reliably differentiates between those with Autistic Disorder (DSM-IV) and other non-spectrum conditions, not those with milder ASD symptoms. These factors further complicate the already multifaceted ASD diagnostic process. The current meta-analysis did not differentiate between articles which used the DSM-IV and DSM-5 criteria for ASD

diagnosis. For these results to be directly applicable to current clinical practice, an investigation of the accuracy of these diagnostic measures should be conducted considering only the current conceptualization of ASD as specified in the DSM-5.

Risk and Sources of Bias

The QUADAS-2 indicated no studies with concerns of the applicability of the results to practice. This is likely due to the eligibility criteria of the studies included in the meta-analysis which specified the types of measures and evaluations which were considered acceptable based on current clinical practice. However, just under half of the articles had high risk of bias regarding patient selection. This was most often due to not enrolling a consecutive or random sample of participants in the study. Additionally, high or unclear levels regarding risk of bias of the reference standard were indicated for all studies. This is due to the nature of conducting ASD evaluations in that the reference standard was almost always interpreted with knowledge of the index tests, as is true in clinical practice. Clinicians making ASD diagnoses were therefore not blind to the results of the index tests; in fact, clinicians utilize the results of the index tests as part of the information used to make the final clinical diagnosis. Therefore, the reference standard is inherently influenced by the results of the index tests and cannot be interpreted separately. Although the index test results and reference standard are inextricably linked in this type of evaluation, this was not seen as a flaw of the study, given its reflection of true clinical practice.

Only peer-reviewed published articles were considered for inclusion in this meta-analysis, given that the reliability of the information presented from other types of

sources cannot be determined. However, this also means that many other sources of potentially useful information were excluded without consideration. There is a clear publication bias within the intervention literature wherein studies with negative findings are often not accepted for publication. However, given that this paper presents diagnostic accuracy results instead of intervention results, the effect of excluding non-peer-reviewed articles is not known. It is important to consider this as a possible source of bias, and current research is underway to determine the effect of excluding these types of publications from meta-analyses of diagnostic test accuracy studies.

Conclusion

ASD diagnostic measures may be less accurate in clinical compared to research settings, but more research utilizing solely clinical populations is needed. The *ADOS-2* indicated high levels of sensitivity and specificity across settings, and it is recommended that all ASD evaluations include this measure. Given the complexity of children referred in clinical practice, it is imperative that these ASD diagnostic measures are administered and scored accurately according to the published guidelines. Formal training and experience using the measures in clinical settings is necessary to maintain the integrity of the measures when implementing them as part of a diagnostic clinical evaluation for children referred for the question of ASD.

Acknowledgements

The authors would like to thank the UAB Libraries Reference Department for their support in formalizing and improving the search strategy for this project, Sarah

Ryan, Ph.D. and Cassandra Newsom, Ph.D. for serving as article reviewers, and Dustin Long, Ph.D. for providing assistance with the biostatistical analyses. This research was supported in part by the Health Resources and Services Administration (HRSA) Maternal and Child Health Bureau (MCH) Leadership Education in Neurodevelopmental and Related Disabilities (LEND; PI: Biasini), UAB Civitan International Science Center and Foundation for Children with Intellectual and Developmental Disabilities McNulty Scientist Award (O'Kelley), and the UAB Civitan-Sparks Clinics.

REFERENCES

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

Baron-Cohen, S., Scott, F., Wheelwright, S., Johnson, M., Bisarya, D., Desai, A. & Ahluwalia, J. (2006). Can Asperger syndrome be diagnosed at 26 months old? A genetic high-risk single-case study. *Journal of Child Neurology*, *21*(4), 351–6.

Bennett, J., Hodgetts, S., Mackenzie, M., Haqq, A. & Zwaigenbaum, L. (2017). Investigating Autism-Related Symptoms in Children with Prader-Willi Syndrome: A Case Study. *International Journal of Molecular Sciences*, *18*(3), 517.

Bukelis, I., Porter, F. D., Zimmerman, A. W. & Tierney, E. (2007). Smith-Lemli-Opitz syndrome and autism spectrum disorder. *American Journal of Psychiatry*, *164*(11), 1655–1661.

Cooper, K. L. & Hanstock, T. L. (2009). Confusion between depression and autism in a high functioning child. *Clinical Case Studies*, *8*(1), 59–71.

Correia, F., Café, C., Almeida, J., Mouga, S. & Oliveira, G. (2015). Autism spectrum disorder: FRAXE mutation, a rare etiology. *Journal of Autism and Developmental Disorders*, *45*(3), 888–892.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Leventhal, B. L. & Lord, C. (2007). Between a ROC and a hard place: decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *48*(9), 932–40.

Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J., … Wheelwright, S. (1999). Autism spectrum disorders at 20 and 42 months of age: stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *40*(5), 719–32.

De Bildt, A., Sytema, S., van Lang, N. D. J., Minderaa, R. B., van Engeland, H. & de Jonge, M. V. (2009). Evaluation of the ADOS revised algorithm: the applicability in 558 Dutch children and adolescents. *Journal of Autism and Developmental Disorders*, *39*(9), 1350–8. doi:10.1007/s10803-009-0749-9

Deeks, J. J. (2001). Systematic reviews of evaluations of diagnostic and screening tests.

British Medical Journal, *323*, 157–162.

Deeks J. J., Wisniewski S, & Davenport C. (2013). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2013. Available from: http://srdta.cochrane.org/.

Dereu, M., Roeyers, H., Raymaekers, R., Meirsschaut, M. & Warreyn, P. (2012). How useful are screening instruments for toddlers to predict outcome at age 4? General development, language skills, and symptom severity in children with a false positive screen for autism spectrum disorder. *European Child \& Adolescent Psychiatry*, *21*(10), 541–551.

DiLavore, P. C., Lord, C. & Rutter, M. (1995). The pre-linguistic autism diagnostic observation schedule. *Journal of Autism and Developmental Disorders*, *25*(4), 355–379.

Dorlack, T. P., Myers, O. B. & Kodituwakku, P. W. (2018). A comparative analysis of the ADOS-G and ADOS-2 algorithms: preliminary findings. *Journal of Autism and Developmental Disorders*, 1–12.

Egger, M., Juni, P., Bartlett, C., Holenstein, F., Sterne, J. & others. (2003). How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess*, *7*(1), 1–76.

Gilchrist, A., Green, J., Cox, A., Burton, D., Rutter, M. & Le Couteur, A. (2001). Development and current functioning in adolescents with Asperger syndrome: a comparative study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 227–40.

Gotham, K., Risi, S., Pickles, A. & Lord, C. (2007). The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, *37*(4), 613.

Gray, K. M., Tonge, B. J. & Sweeney, D. J. (2008). Using the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule with young children with developmental delay: evaluating diagnostic validity. *Journal of Autism and Developmental Disorders*, *38*(4), 657–667.

Howes, O. D., Rogdaki, M., Findon, J. L., Wichers, R. H., Charman, T., King, B. H., … others. (2017). Autism spectrum disorder: Consensus guidelines on assessment, treatment and research from the British Association for Psychopharmacology. *Journal of Psychopharmacology*, 0269881117741766.

Jüni, P., Holenstein, F., Sterne, J., Bartlett, C. & Egger, M. (2002). Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *International Journal of Epidemiology*, *31*(1), 115–123.

Kamp-Becker, I., Albertowski, K., Becker, J., Ghahreman, M., Langmann, A., Mingebach, T., … others. (2018). Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *European Child \& Adolescent Psychiatry*, 1–15.

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.

Klassen, T. P., Lawson, M. L., Moher, D. & others. (2005). Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. *Journal of Clinical Epidemiology*, *58*(8), 769–776.

Langmann, A., Becker, J., Poustka, L., Becker, K. & Kamp-Becker, I. (2017). Diagnostic utility of the autism diagnostic observation schedule in a clinical sample of adolescents and adults. *Research in Autism Spectrum Disorders*, *34*, 34–43.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., Bishop, S. & others. (2012). *Autism diagnostic observation schedule: ADOS-2*. Western Psychological Services Los Angeles, CA.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., … others. (2009). The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(9), 1305–1320.

Mannion, A. & Leader, G. (2013). Comorbidity in autism spectrum disorder: A literature review. *Research in Autism Spectrum Disorders*, *7*(12), 1595–1616.

McInnes, M. D., Moher, D., Thombs, B. D., McGrath, T. A., Bossuyt, P. M., Clifford, T., … others. (2018). Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*, *319*(4), 388–396.

Merritt, J., Hart, J. C. & LeGrow, T. L. (2017). Autism spectrum disorder in Say-Barber-Biesecker-Young-Simpson syndrome. *British Medical Journal Case Reports 2017*, 1-4.

Moher, D., Klassen, T. P., Schulz, K. F., Berlin, J. A., Jadad, A. R., Liberati, A. & others. (2000). What contributions do languages other than English make on the results of meta-analyses? *Journal of Clinical Epidemiology*, *53*(9), 964–972.

Moher, D., Pham, B., Lawson, M., Klassen, T. & others. (2003). The inclusion of reports

of randomised trials published in languages other than English in systematic reviews. *Health Technol Assess, 7*(41), 1–90.

Oosterling, I. J., Roos, S., de Bildt, A., Rommelse, N., de Jonge, M., Visser, J., … Buitelaar, J. (2010). Improved diagnostic validity of the ADOS revised algorithms: a replication study in an independent sample. *Journal of Autism and Developmental Disorders, 40*(6), 689–703. doi:10.1007/s10803-009-0915-0

Papanikolaou, K., Paliokosta, E., Houliaras, G., Vgenopoulou, S., Giouroukou, E., Pehlivanidis, A., … Tsiantis, I. (2009). Using the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule-Generic for the diagnosis of autism spectrum disorders in a Greek sample with a wide range of intellectual abilities. *Journal of Autism and Developmental Disorders, 39*(3), 414–420.

Penner, M., Anagnostou, E., Andoni, L. Y. & Ungar, W. J. (2017). Systematic review of clinical guidance documents for autism spectrum disorder diagnostic assessment in select regions. *Autism*, 1362361316685879.

Reaven, J. A., Hepburn, S. L. & Ross, R. G. (2008). Use of the ADOS and ADI-R in children with psychosis: Importance of clinical judgment. *Clinical Child Psychology and Psychiatry, 13*(1), 81–94.

Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., … Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(9), 1094–103.

Rutter, C. M. (1995). Regression methods for meta-analysis of diagnostic test data. *Acad Radiol, 2*, S48–S56.

Rutter, C. M. & Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine, 20*(19), 2865–2884.

Rutter, M., Le Couteur, A., Lord, C. & others. (2003). Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services, 29*, 30.

Schaaf, C. P., Goin-Kochel, R. P., Nowell, K. P., Hunter, J. V., Aleck, K. A., Cox, S., … Shinawi, M. (2011). Expanding the clinical spectrum of the 16p11. 2 chromosomal rearrangements: three patients with syringomyelia. *European Journal of Human Genetics, 19*(2), 152.

Smart, R. G. (1964). The importance of negative results in psychological research.

Canadian Psychologist/Psychologie Canadienne, *5*(4), 225.

Stewart, J. R., Vigil, D. C., Ryst, E. & Yang, W. (2014). Refining best practices for the diagnosis of autism: A comparison between individual healthcare practitioner diagnosis and transdisciplinary assessment. *Nevada Journal of Public Health*, *11*(1), 1.

Suter, B., Treadwell-Deering, D., Zoghbi, H. Y., Glaze, D. G. & Neul, J. L. (2014). Brief report: MECP2 mutations in people without Rett syndrome. *Journal of Autism and Developmental Disorders*, *44*(3), 703–711.

Takwoingi, Y. & Deeks, J. (2010). *MetaDAS: A SAS macro for meta-analysis of diagnostic accuracy studies. User Guide Version 1.3. 2010 July. Available from: http://srdta.cochrane.org/.*

Thurm, A., Tierney, E., Farmer, C., Albert, P., Joseph, L., Swedo, S., … others. (2016). Development, behavior, and biomarker characterization of Smith-Lemli-Opitz syndrome: an update. *Journal of Neurodevelopmental Disorders*, *8*(1), 12.

Treadwell-Deering, D. E., Powell, M. P. & Potocki, L. (2010). Cognitive and behavioral characterization of the Potocki-Lupski syndrome (duplication 17p11. 2). *Journal of Developmental \& Behavioral Pediatrics*, *31*(2), 137–143.

Urraca, N., Cleary, J., Brewer, V., Pivnick, E. K., McVicar, K., Thibert, R. L., … Reiter, L. T. (2013). The interstitial duplication 15q11. 2-q13 syndrome includes autism, mild facial anomalies and a characteristic EEG signature. *Autism Research*, *6*(4), 268–279.

Vllasaliu, L., Jensen, K., Hoss, S., Landenberger, M., Menze, M., Schütz, M., … Freitag, C. M. (2016). Diagnostic instruments for autism spectrum disorder (ASD). *The Cochrane Library*.

Vogel, U. & Windeler, J. (2000). Factors modifying frequency of publications of clinical research results exemplified by medical dissertations. *Deutsche Medizinische Wochenschrift (1946)*, *125*(5), 110–113.

Zander, E., Sturm, H. & Bölte, S. (2015). The added value of the combined use of the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule: Diagnostic validity in a clinical Swedish sample of toddlers and young preschoolers. *Autism*, *19*(2), 187–199.

Zander, E., Willfors, C., Berggren, S., Choque-Olsson, N., Coco, C., Elmund, A., … others. (2016). The objectivity of the Autism Diagnostic Observation Schedule (ADOS) in naturalistic clinical settings. *European Child \& Adolescent Psychiatry*, *25*(7), 769–780.

Zander, E., Willfors, C., Berggren, S., Coco, C., Holm, A., Jifält, I., … others. (2017). The Interrater Reliability of the Autism Diagnostic Interview-Revised (ADI-R) in

Clinical Settings. *Psychopathology*, *50*(3), 219–227.

Zhang, C., Milunsky, J. M., Newton, S., Ko, J., Zhao, G., Maher, T. A., … others. (2009). A neuroligin-4 missense mutation associated with autism impairs neuroligin-4 folding and endoplasmic reticulum export. *Journal of Neuroscience*, *29*(35), 10843–10854.

AGREEMENT OF THE *ADOS-2* AND *ADI-R* IN A CLINICAL SAMPLE OF CHILDREN REFERRED FOR AN AUTISM SPECTRUM DISORDER EVALUATION

JENNA B. LEBERSFELD, KRISTI C. GUEST, SARAH E. O'KELLEY

In preparation for the *Journal of Autism and Developmental Disorders*

Format adapted for dissertation

ABSTRACT

The aim of the present study was to determine which child and family factors predicted agreement between the *Autism Diagnostic Observation Schedule* (*ADOS-2*) classification and the *Autism Diagnostic Interview, Revised* (*ADI-R*) classification included in diagnostic evaluations for autism spectrum disorder (ASD). Participants were 356 children ages one to 18 years of age (n = 205 ASD, n = 151 non-ASD) who were referred for a psychological evaluation at a tertiary care clinic at an academic medical center. The accuracy of the *ADOS-2* alone was similar to the use of both the *ADOS-2* and the *ADI-R* combined, with the *ADOS-2* alone having higher sensitivity and the combination of measures having higher specificity. The *ADOS-2*, *ADI-R*, and final clinical diagnosis were in agreement 71% of the time. Child age was the only significant predictor of agreement in this sample, with younger children having higher levels of agreement between measures. Results suggest that the *ADOS-2* should be administered for all ASD diagnostic evaluations, but for younger children with more straightforward clinical presentations, the administration of the *ADI-R* may not be necessary to achieve diagnostic accuracy.

INTRODUCTION

Children for whom there is a question of autism spectrum disorder (ASD) are often referred for a comprehensive evaluation to determine whether a child meets criteria for a diagnosis of ASD. These evaluations are often conducted by a licensed clinical psychologist or pediatrician working in conjunction with a multidisciplinary team with expertise in children with ASD. The diagnostician conducts an observation and interaction session with the child as well as an interview with the caregivers to evaluate the child's symptoms of ASD. The members of the multidisciplinary team then consult with one another to determine whether the child's behavior can be explained by a diagnosis of ASD.

The clinical observation which is most commonly used is the *Autism Diagnostic Observation Schedule, Second Edition* (*ADOS-2*). During the *ADOS-2* administration, the clinician and the child participate in semi-structured, developmentally-appropriate activities which were designed to evaluate symptoms of ASD. These activities range from playing with toys to having conversations and discussing emotions. The *Autism Diagnostic Interview, Revised* (*ADI-R*) is a semi-structured clinical interview during which the interviewer asks the child's primary caregivers about the child's behaviors related to ASD symptom areas both currently and in the past.

The *ADOS-2* and the *ADI-R* have shown high levels of accuracy in research settings both separately and in combination when compared to a best estimate clinical

diagnosis (Kim & Lord, 2012a; Risi et al., 2006). When quantifying the accuracy of a diagnostic test, sensitivity and specificity are two types of psychometric properties that are often reported. Sensitivity measures the likelihood that an individual with a given disorder will have a positive test result (i.e., indicating the presence of the target diagnosis), and sensitivity indicates how likely it is that a person without the disorder will have a negative test result (i.e., indicating the absence of the target diagnosis). Published *ADOS-2* sensitivity ranges from .60 to .95 and specificity ranges from .75 to 1.00 (Lord, Luyster, Gotham, & Guthrie, 2012; Lord, Rutter, et al., 2012). These accuracy measures are not published in the *ADI-R* manual; however, original research literature prior to measure publication indicates that sensitivity ranged from .48 to .88 and specificity was 1.00 (Cox et al., 1999; Gilchrist et al., 2001).

Most often, the *ADOS-2* and the *ADI-R* results agree with one another; that is, both measures indicate that a child has ASD or both measures have non-ASD results. However, at times these two measures produce conflicting results, which poses a challenge for the diagnostician and the multidisciplinary team. In these situations, diagnostic teams must take into consideration all information gathered during the evaluation and use their clinical judgement to come to a final diagnostic conclusion. This takes considerable time, clinical expertise, and experience and may pose difficult for less-experienced diagnostic teams (Risi et al., 2006).

The relationship between some child and family factors and agreement of these diagnostic measures have been investigated with variable results. For example, lower IQ was identified as a moderator of *ADOS-2* and *ADI-R* agreement by Neuhaus et al. (2017) but not by others (Tomanik et al., 2007). Lower adaptive behavior, fewer behavioral

difficulties, higher household income, and non-African-American race were shown to be moderators of *ADOS-2* and *ADI-R* agreement in one study (Neuhaus et al., 2017). Younger age was related to higher agreement in one study (de Bildt et al., 2004), while another found no relationship (Neuhaus et al., 2017). Biological sex, birth order, parental age at birth, and parental education were not shown to moderate agreement (Neuhaus et al., 2017).

In an attempt clarify results from existing literature and to aid diagnosticians in conducting ASD evaluations and coming to diagnostic conclusions, this paper aimed to determine whether other types of information typically gathered or known during a diagnostic evaluation, such as child and family characteristics, predicted agreement of the *ADOS-2* and the *ADI-R*. Additionally, this study investigated whether these same child and family factors could be used to predict final diagnosis within cases for which the *ADOS-2* and the *ADI-R* disagreed. It was hypothesized that the following child characteristics would be predictive of agreement and final diagnosis within agreement cases: lower age at evaluation, male biological sex, lower cognitive ability, lower adaptive behavior, lower language skills, later birth order, higher birth weight, and higher gestational age at birth. Additionally, the following family factors were hypothesized to predict agreement between the diagnostic measures as well as final diagnosis: older maternal age at birth, higher levels of maternal education, biological parent marital status (married), and private insurance type.

METHODS

Participants

Participants were 360 children ages one through 18 years old who received a routine comprehensive clinical evaluation for ASD at a tertiary care clinic in Alabama from 2013 through 2018. Only children for whom a conclusive final clinical diagnosis was rendered were included. Four children received an inconclusive final diagnosis (i.e., rule-out ASD: n = 3, provisional ASD: n = 1) and were excluded from the sample. Of the remaining 356 children, 205 (58%) received a final diagnosis of ASD and 151 (42%) received a non-ASD final diagnosis (e.g., attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder or disruptive behavior disorder, developmental coordination disorder, stereotypic movement disorder, language or speech disorder, anxiety disorder, depressive disorder, adjustment disorder, specific learning disorder, intellectual disability, unspecified neurodevelopmental disorder, or global developmental delay). Additional demographics are presented in Table 1. Group differences between the ASD and non-ASD groups were calculated using χ^2 tests for categorical variables and *t*-tests for continuous variables. Given that this is a clinical sample and not a research sample, all measures were not administered to all participants, resulting in varied sample sizes for each measure. For further explanation, please refer to the Procedures section.

Table 1. Participant Demographics: Child Characteristics

Child Characteristics	Total		ASD		Non-ASD		p
	n	% or mean (SD)	n	% or mean (SD)	n	% or mean (SD)	
Age, years	356	6.3 (3.8)	205	6.1 (3.7)	151	6.6 (3.9)	
Gender	356	--	205	--	151	--	
Male	279	78	160	78	119	79	
Female	77	22	45	22	32	21	
Race	353	--	204	--	149	--	**
White	209	59	110	54	99	66	
Black	95	27	56	27	39	26	
Other	49	14	38	19	11	7	
Individualized Education Program (IEP) Special Education Eligibility Category	169	--	99	--	70	--	
Autism	54	32	25	25	29	41	
Developmental Delay	63	37	43	43	20	29	
Speech or Language Impairment	18	11	13	13	5	7	
Other	34	20	18	18	16	23	
Gestational Age (weeks)	338	37.9 (3.0)	197	38.3 (2.6)	141	37.3 (3.4)	**
Order of Child in Home	337	--	196	--	141	--	
First	181	54	113	58	68	48	
Second	104	31	53	27	51	36	
Third	10	11	23	12	14	10	
Fourth or More	4	5	7	4	8	6	

Note: Some percentages may not add up to 100 due to rounding. N's differ between groups due to unreported variables. IEP Eligibility Categories are not applicable for children under 3 years of age. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

Table 2. Participant Demographics: Family Characteristics

	Total		ASD		Non-ASD		p
	n	% or mean (SD)	n	% or mean (SD)	n	% or mean (SD)	
Maternal Age at Birth	345	27.2 (6.2)	200	28.5 (6.5)	145	25.5 (5.4)	***
Maternal Age at Eval	345	33.4 (7.2)	200	34.3 (7.1)	145	32.1 (7.1)	**
Maternal Education	324	--	184	--	140	--	
Less Than High School	45	14	20	11	25	18	
Graduated High School/GED	128	40	68	37	60	43	
Some College	91	28	54	29	37	26	
Graduated College	44	14	31	17	13	10	
Graduate or Medical School	16	5	11	6	5	4	
Biological Parent Marital Status	334	--	189	--	145	--	*
Married	135	40	86	46	49	34	
Not Married	199	60	103	55	96	66	
Number of Caregivers in Home	236	--	132	--	104	--	
1	51	22	24	18	27	26	
2	157	67	95	72	62	60	
3	18	8	9	7	9	9	
4+	10	4	4	3	6	6	
Number of Children in Home	350	--	203	--	147	--	
1	105	30	63	31	42	29	
2	119	34	73	36	46	31	
3	93	27	51	25	42	29	
4	24	7	11	5	13	9	
5+	9	3	5	3	4	3	

Note: Percentages may not add up to 100 due to rounding. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

The male to female gender ratio of children in this sample was 4:1 which is comparable to gender ratios that have been reported within the ASD population. However, in this sample, this 4:1 gender ratio was also seen in the non-ASD group, indicating that within the group of children for whom ASD was a concern but who ultimately did not receive an ASD diagnosis, this male to female gender ratio of 4:1 is also reflected. Race of the participants was generally representative of the state in which the evaluations were conducted (Kreiser & White, 2014). The groups differed on clinician-documented race [$X^2(1) = 10.18$, $p < .01$], with the ASD group having a lower percentage of White individuals (54%) and greater percentage of Other races (19%) compared to the non-ASD group (White = 66%, Other = 7%). The percentage of Black individuals did not differ between groups (ASD = 27%, Non-ASD = 26%).

Mean gestational age differed by one week between groups [$t(254) = 2.91$, $p < .005$], with those in the non-ASD group being born one week earlier on average than those in the ASD group [$t(254) = 2.91$, $p < .005$). On average, maternal age at birth of children in the ASD group was three years older than the non-ASD group [$t(338) = 4.71$, $p < .001$]. Maternal age at evaluation also differed between groups, with mothers in the ASD group being on average about two years older than mothers in the non-ASD group [$t(343) = 2.78$, $p < .01$]. Children in both groups were more likely to have biological parents who were not married than married; however, this effect was significantly more pronounced in the non-ASD group compared to the ASD group [$\chi^2(1) = 4.67$, $p < .05$].

Procedure

Clinical Evaluation

All children included in this sample received a comprehensive clinical evaluation for ASD at a tertiary care clinic located at an academic medical center in Alabama. About 1,200 referrals were made to this clinic per year for evaluation to address concerns regarding developmental delays, learning difficulties, and emotional or behavioral functioning, with about half having a parent or provider concern of ASD. However, only between 75 to 150 comprehensive ASD evaluation clinic slots (i.e., those that utilize the *ADOS-2* and *ADI-R*) were available each year from 2013 through 2017. In 2018, an additional psychologist conducted ASD evaluations at this clinic resulting in an increased number of ASD evaluations completed.

During the ASD evaluation, children were evaluated by a multidisciplinary clinical team which was led by a licensed clinical psychologist with ASD expertise and both clinical and research training on the *ADOS-2* and the *ADI-R*. The clinical team also usually included a developmental behavioral pediatrician or general pediatrician, and the child may have also received evaluations from other disciplines including optometry, audiology, nutrition, speech/language pathology, occupational therapy, physical therapy, and social work. Following the evaluation, all members of the multidisciplinary team gathered to discuss the case and finalize a consensus team decision regarding the child's diagnosis. Recommendations for services and intervention options were then delivered to the child's family verbally and through written reports. All procedures were conducted clinically, and retrospective use for research was approved by the Institutional Review Board (IRB) at the university at which the evaluations and research were conducted.

Chart Review and Data Extraction

Participant data were collected via retrospective chart review of these comprehensive ASD evaluations. Due to the clinical nature of these charts, the types of information available varied from child to child, and the n's per group are presented in Tables 1, 2, 3, and 5. Unless specified below, in general, the chart review and data extraction process was systematized by first prioritizing information gathered from the reports generated by diagnosticians from the university's tertiary care clinic evaluation, followed by information from medical records, caregiver report on intake paperwork completed prior to evaluation, school records, and reports from outside evaluations (e.g., psychologist, speech language pathologist). Given the inherent complexity of this chart review, data extraction, and data entry process, variables had consensus-driven operational definitions, and graduate and undergraduate students were trained to 95% reliability for extraction of data. Any discrepancies were resolved through discussion with lab members and faculty mentors.

Variable Definitions and Limitations

Child Characteristics

Race. Although participant race is presented in Table 1, race in this dataset is clinician-rated as opposed to self-identified by the participants' families. Given the inherent error associated with this method of data collection, race is presented for descriptive purposes only and will not be explored further.

Gestational age. Gestational age was extracted from medical birth records whenever possible. When medical records were not available, information was gleaned from the psychology report or intake paperwork completed by caregivers prior to evaluation. When a source of information indicated that the child was born "full-term," gestational age was estimated to be 40 weeks.

Maternal age at birth. Maternal age at birth was primarily obtained from medical birth records. When medical birth records were unavailable, maternal age at birth was estimated by subtracting child age at evaluation from the maternal age at evaluation.

Maternal age at evaluation. Parent age is reported on the intake paperwork completed by caregivers prior to the ASD evaluation. However, at times there was a significant delay between the completion of the paperwork and the ASD evaluation. Therefore, maternal age at evaluation was estimated by adding the maternal age reported on the intake paperwork to the number of years between intake and evaluation.

Final Clinical Diagnosis

Final clinical psychological diagnoses were made using the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* or *Fifth Edition* (*DSM-IV, DSM-5*). The *DSM-IV* defined multiple types of autism spectrum disorders including Asperger's Disorder; Autistic Disorder; and Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS). The *DSM-5* was updated to reflect the conceptualization of ASD as a single disorder with differing levels of severity. For this study, *DSM-IV* diagnoses of Asperger's Disorder; Autistic Disorder; or Pervasive

Developmental Disorder, Not Otherwise Specified were considered an ASD diagnosis. All other diagnoses were classified as non-ASD.

Measures

ASD Measures

Autism Diagnostic Observation Schedule, Second Edition (*ADOS-2*; Lord et al., 2012)

The *ADOS-2* is a semi-structured, 45- to 60-minute observation and interaction session with an evaluator and the child which is used to aid in the diagnosis of ASD. Five different modules are available, and the appropriate module is determined based on the child's age and language level. The child and the clinician engage in developmentally appropriate activities specifically developed to evaluate ASD-specific symptomology. Child behaviors are then scored using an algorithm. This measure yields a classification of *autism, autism spectrum,* or *non-spectrum*. A comparison score can also be calculated which ranges from 1 to 10, with higher scores indicating a greater level of ASD symptoms. The *ADOS-2* sensitivity estimates range from 0.61 to 0.97 and specificity ranges from 0.47 to 1.00 (Gotham et al., 2008; Kamp-Becker et al., 2011; Molloy, Murray, Akers, Mitchell, & Manning-Courtney, 2011; Oosterling, Rommelse, et al., 2010; Oosterling, Roos, et al., 2010; Zander et al., 2015). A recent meta-analysis indicated pooled sensitivity ranging from .77 to .90 and specificity ranging from .62 to .90 for the *ADOS-2* (Dorlack et al., 2018).

Autism Diagnostic Interview, Revised (*ADI-R;* Rutter et al., 2003)

The *ADI-R* is a semi-structured diagnostic interview given to a parent or caregiver by a trained clinician. The time of administration of the comprehensive *ADI-R* interview ranges from 90 to 150 minutes, and the interview consists of detailed questions about development and underlying behaviors associated with ASD and focuses on whether the child currently demonstrates ASD-specific behaviors and whether they exhibited certain ASD symptoms in the past (Rutter et al., 2003). This measure yields a classification of *autism* or *not autism*.

The utility of the *ADI-R* diagnostic algorithm was found to be most appropriate for children with a non-verbal mental age over two-years-old (Lord et al., 1994), and many younger children referred for evaluation at this clinic did not yet meet these criteria. A version of the *ADI-R* for toddlers was developed (Kim & Lord, 2012b) for research purposes for use with children under four years old; however, it is not widely available and has not yet been published for clinical use.

Sensitivity and specificity were not published in the *ADI-R* manual; however, original research literature prior to measure publication indicates that sensitivity ranged from .48 to .88 and specificity was 1.00 (Cox et al., 1999; Gilchrist et al., 2001). More recent literature suggests that the sensitivity of this measure ranges from .53 to .92, and specificity ranges from .62 to .95 (Falkmer et al., 2013; Risi et al., 2006).

Table 3. Constructs and Measures

Construct	Measure	n
ASD	*ADOS-2* Toddler Module	20
	ADOS-2 Module 1	152
	ADOS-2 Module 2	67
	ADOS-2 Module 3	96
	ADOS-2 Module 4	4
	ADI-R	309
Cognitive	*Bayley-III*	21
	DAS-II	130
	Leiter-3	10
	RIAS	1
	SB-5	7
	WASI	1
	WASI-II	8
	WISC-IV	39
	WISC-V	19
Adaptive	*ABAS-Parent*	92
	ABAS-Teacher	2
	AGS	26
	Vineland	107
Language	*CELF-4*	8
	CELF-5	7
	CELF-P2	36
	OWLS	7
	OWLS-II	48
	PLS-4	4
	PLS-5	216
	REEL	2
Sensory	*SPM-Home*	1
	SPM-School	2
	SPM-Preschool	26
	SPM-Version Not Specified	8
	SSP	31
	SSP2	44

Cognitive Measures

During clinical assessments, the most appropriate cognitive measure for each child is chosen based on a variety of factors including child age, language level, and developmental level. The following measures were administered and included in this dataset: *Bayley Scales of Infant and Toddler Development, Third Edition* (*Bayley-III*); *Differential Ability Scales, Second Edition (DAS-II); Leiter International Performance Scale, Third Edition (Leiter-3); Reynolds Intellectual Assessment Scales (RIAS); Stanford-Binet Intelligence Scales, Fifth Edition (SB-5); Wechsler Abbreviated Scale of Intelligence, First Edition (WASI) and Second Edition (WASI-II);* and *Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) and Fifth Edition (WISC-V)*. The majority of these tests were administered within the clinic setting, but other scores were retrieved from previous evaluations through the school system or other outpatient clinics. All of these tests measure cognitive ability and produce an overall standard score with a mean of 100 and a standard deviation of 15, such as the Full Scale IQ (FSIQ). This accurately reflects the variation in clinical practice and individualized cognitive assessment; therefore, these measures were collapsed across participants to create the cognitive ability variable. It is acknowledged that the *Bayley-III* is a developmental measure, and the cognitive scale may not demonstrate the same stability over time as other measures of cognitive ability. Nevertheless, this was the best estimate of cognitive ability available for 21 children, and rather than exclude these children from the cognitive ability analyses, the *Bayley-III* estimates were included. Please refer to Table 3 for the number of administrations of each measure in the sample.

Adaptive Measures

A variety of measures were also used to estimate adaptive skills in this sample. Tests were selected based on the age of the child as well as whether an interview or questionnaire was more clinically appropriate. The following tests were used to measure adaptive behavior in this sample: *Adaptive Behavior Assessment System, Parent (ABAS-Parent) and Teacher (ABAS-Teacher) Rating Forms; American Guidance Service Early Screening Profiles, Self-Help/Social Profile (AGS);* and *Vineland Adaptive Behavior Scales (Vineland).* Similar to the cognitive tests, results from the majority of these measures were obtained through evaluations at this clinic; however, results may have been obtained from other reports and records provided from outside agencies. These tests also produce a standard score of global adaptive functioning, such as the General Adaptive Composite (GAC) or the Adaptive Behavior Composite (ABC). These standard scores were collapsed across measures to create the adaptive behavior variable.

Language Measures

At this multidisciplinary clinic, children are referred for a speech-language evaluation if language is an area of concern for the family or provider; this is also seen as an important component of the multidisciplinary assessment. Language measures are selected based on the clinical need of the child. Tests were most often administered by speech-language pathologists as part of a formal speech-language evaluation at this multidisciplinary clinic; however, results of many of these tests were also obtained through school records from evaluations conducted by the public school system. Measures administered were the following: *Clinical Evaluation of Language*

Fundamentals, Fourth Edition (CELF-4), Fifth Edition (CELF-5), and *Preschool-2 (CELF-P2); Oral and Written Language Scales, First Edition (OWLS)* and *Second Edition (OWLS-II); Preschool Language Scales, Fourth Edition (PLS-4)* and *Fifth Edition (PLS-5);* and the *Receptive-Expressive Emergent Language Test (REEL).* These tests measure the language construct and produce a standard score estimating overall language ability. This score was collapsed across measures to create the overall language variable for this sample.

Sensory Measures

A child receives a measure evaluating sensory differences when there is a reported concern with this area of functioning, which occurs often in children referred for ASD evaluation. This is most often determined and administered by an occupational therapist during the multidisciplinary clinic evaluation for ASD; however, the results of some measures were also retrieved through school records. The tests used to measure sensory characteristics were the *Sensory Processing Measure, Home Version (SPM-Home), School Version (SPM-School), Preschool Version (SPM-Preschool) and the Short Sensory Profile, First Edition (SSP) and Second Edition (SSP-2).* Although these measures produce different types of results, with the *SPM* specifying the type of sensory differences (e.g., sensory seeking behaviors, sensory sensitivities, sensory avoidance, etc.,) research shows that when comparing across the *SPM* and *SSP*, results are most accurate when dichotomizing the variable (Dugas et al., 2018; Hansen & Jirikowic, 2013). Therefore, for this study, the sensory variable was dichotomized for each child to indicate “typical” or “atypical” sensory characteristics.

Characteristics of Clinical Evaluations

Due to the clinical nature of these evaluations, the same measures were not administered to all children. Of the 356 children, 11 received the *ADOS-G* at the beginning of 2013 prior to transitioning to use of the *ADOS-2* at this clinic, and six children did not receive the administration of any *ADOS* version. These 17 children are not included in analyses of the *ADOS-2* but are included in analyses of the *ADI-R*. Twenty children were administered the *ADOS-2* Toddler Module which does not render a classification of *autism*, *autism spectrum*, or *non-spectrum*. Instead, levels of concern are given (i.e., *little-to-no, mild-to-moderate*, or *moderate-to-severe*). Of note, only one child who received the *ADOS-2* Toddler Module also received the *ADI-R*. For the other 19 children, an ASD-focused clinical interview was administered which addressed the child's development and symptoms of ASD but with questions more appropriate for toddler-aged children.

For 22 evaluations, Spanish-language interpreters were used to aid in completion of the evaluation for clinical interview and *ADI-R*. The *ADOS-2* was administered in English in all cases based on the primary language of the child as well as caregiver report of child's preferred language. Results from these evaluations are included in the analyses to reflect the diversity of the sample accurately.

DATA ANALYSIS

Statistical analyses were conducted using IBM SPSS Statistics software version 25. Based on the *ADOS-2* classification, the *ADI-R* classification, and the final diagnosis, eight distinct categories of "agreement" were defined, as indicated in Table 4. Agreement categories 1 and 2 show that the *ADOS-2*, *ADI-R*, and final diagnosis were all in agreement, with both measure classifications and final diagnosis indicating ASD for category 1 and non-ASD for category 2. When dichotomizing the agreement variable, these two categories were combined and labeled as the "Agree" group. Categories 3 through 6 indicate four different ways in which the measures and final diagnosis can disagree, reflecting a clinical assessment in which the classifications derived following administration of the *ADOS-2* and *ADI-R* conflicted with one another, and the multidisciplinary team used all information gathered during the evaluation as well as clinical judgement to come to a final diagnostic conclusion. For the dichotomous agreement variable, these four categories were combined together and labeled as the "Disagree" group. The final two categories 7 and 8 represent rare cases (n = 8) in which the *ADOS-2* and *ADI-R* classifications were in agreement with one another, but the clinical team made a final clinical diagnosis that was in contrast to the classifications derived from the ASD measures. These cases were excluded from dichotomization of the agreement variable. Other missing cases in the investigations of agreement reflect cases for which the *ADOS-2* or the *ADI-R* were not administered (n = 43).

Table 4. Agreement Categories

	ADOS-2	*ADI-R*	Final Diagnosis
Agree			
1	ASD	ASD	ASD
2	Non-ASD	Non-ASD	Non-ASD
Disagree			
3	ASD	Non-ASD	ASD
4	Non-ASD	ASD	ASD
5	Non-ASD	ASD	Non-ASD
6	ASD	Non-ASD	Non-ASD
Not Included			
7	ASD	ASD	Non-ASD
8	Non-ASD	Non-ASD	ASD

For this study, *ADOS-2* analyses were conducted in two ways: by excluding children who received the *ADOS-2* Toddler Module and by re-coding the levels of concern such that *little-to-no* concern was coded as *non-ASD,* and *mild-to-moderate* and *moderate-to-severe* levels of concern were coded as *ASD.* For all analyses which involved both the *ADOS-2* classification and the *ADI-R* classification, the *ADOS-2* Toddler Module classifications adapted for this study were not included. This is because only one child who received the *ADOS-2* Toddler Module also received an *ADI-R,* for reasons described previously regarding clinic procedures. For all other analyses, the *ADOS-2* classifications both with and without the Toddler Module classifications were used. Additionally, for eight cases, the *ADOS-2* classification and *ADI-R* classification were in agreement, but this classification disagreed with the final clinical diagnosis (see Table 5). These cases were excluded from analyses of agreement between the *ADOS-2* and the *ADI-R* but were included for analyses of agreement between the ASD measure and the final diagnosis as well as to determine the accuracy of each instrument compared to final diagnosis.

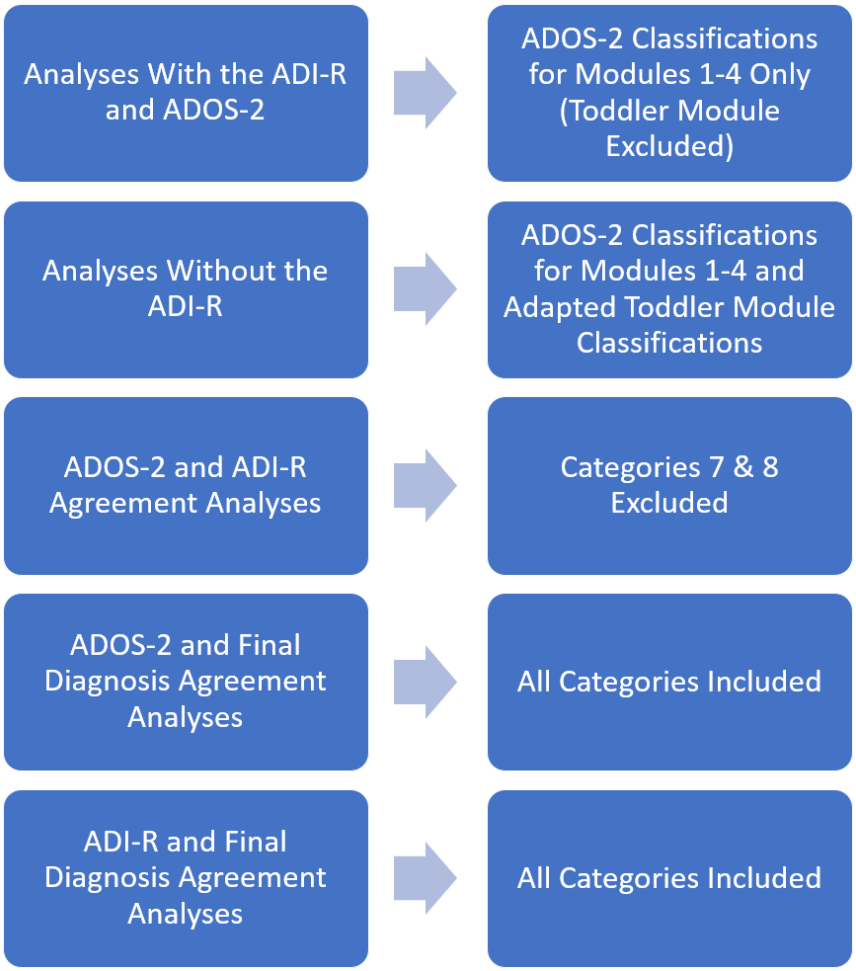

Figure 1. Inclusion and exclusion criteria for agreement analyses.

The sensitivity, specificity, positive predictive value, and negative predictive value for the *ADOS-2* and the *ADI-R* were calculated as individual measures of test accuracy in predicting the final ASD or non-ASD diagnosis. Youden's statistic was used to calculate a singular measure maximizing the sensitivity and specificity. Although this statistic does not have clinical significance in and of itself, it can be used to compare the overall accuracy of different measures or combinations of measures (Youden, 1950). Youden's statistic is defined as sensitivity + specificity – 1 (Youden, 1950). To determine the statistical agreement between pairs of variables (i.e., the *ADOS-2* classification, the *ADI-R* classification, and final diagnosis), Cohen's kappa (κ) (Cohen, 1960) was

calculated and interpreted using published guidelines (Landis & Koch, 1977). Although many guidelines are available for use in interpretation of the Cohen's κ statistics, guidelines adapted from Landis & Koch (1977) were used wherein values <.20 are considered poor agreement, .21 to .40 indicate fair agreement, .41 to .60 show moderate levels of agreement, .61 to .80 is good agreement, and .81 to 1.00 is considered very good. Binomial logistic regressions were conducted to determine which child and family factors predicted agreement. Point-biserial correlations were conducted between the dichotomous Agreement variable and continuous child and family factors, and tests were calculated between the dichotomous Agreement variable and dichotomous or categorical child and family factors to determine significant associations between child and family factors and agreement. Only variables with significant ($p < .05$) or trending ($p < .10$) correlations or χ^2tests were included in the regressions.

RESULTS

Child Characteristics Between Groups

Table 5. Results of Measures Administered During the ASD Evaluation

	Total		ASD		Non-ASD		p
	n	% or mean (SD)	n	% or mean (SD)	n	% or mean (SD)	
ADOS-2 Classification	319		184		135		***
Autism	173	54	165	90	8	6	
Autism Spectrum	33	10	13	7	20	15	
Non-Spectrum	113	35	6	3	107	80	
ADOS-2 Toddler Module Level of Concern	20		12		8		***
Moderate-to-Severe	11	55	11	92	0	0	
Mild-to-Moderate	2	10	1	8	1	13	
Little-to-No	7	35	0	0	7	88	
ADOS-2 Comparison Score	319	5.4 (3.2)	183	7.6 (1.9)	136	2.5 (1.9)	***
ADI-R Classification	309		185		124		***
Autism	180	58	150	81	30	24	
Not Autism	129	42	35	19	94	76	
Cognitive Score	236	76.3 (18.8)	109	75.2 (20.3)	127	77.3 (17.5)	
Adaptive Score	227	66.4 (14.2)	128	64.1 (14.8)	99	69.5 (12.9)	**
Language Score	328	67.7 (18.6)	195	62.2 (17.9)	133	75.9 (16.6)	***
Sensory Classification	112		67		45		
Typical	15	13	8	12	7	16	
Atypical	97	87	59	88	38	84	

Note: Percentages may not add up to 100 due to rounding. Cognitive, adaptive, and language scores are presented as standard scores with a mean of 100 and a standard deviation of 15.

As expected, diagnostic groups differed on measures of ASD symptomology including *ADOS-2* classification [$X^2(2) = 232.19$, $p < .001$], *ADOS-2* comparison score [$t(317) = 23.8$, $p < .001$], and *ADI-R* classification [$\chi^2(1) = 98.7$, $p < .001$], with the ASD group having a greater number of *autism* or *autism spectrum* classifications for both the *ADOS*-2 and the *ADI-R* and higher *ADOS-2* comparison scores than the non-ASD group. Groups also differed on adaptive behavior skills, with the ASD group having poorer overall adaptive functioning than the non-ASD group [$t(225) = -2.87$, $p < .005$], although both group means were in the well below average range. Language abilities differed significantly between groups, with the non-ASD group scoring higher and in the below average range compared with the ASD group whose mean language score was in the well below average range [$t(326) = -7.00$, $p < .001$].

ASD Diagnostic Measure Accuracy

Table 6. Accuracy of ASD Diagnostic Measures Compared to Final Diagnosis

	Se	Sp	Youden's	PPV	NPV
ADOS-2 with Toddler	.97	.80	.77	.87	.95
ADOS-2 without Toddler	.97	.79	.76	.86	.95
ADI-R	.81	.76	.57	.83	.73
ADOS-2 and *ADI-R*[1]	.79	.94	.73	.95	.75
ADOS-2 or *ADI-R*[2]	.99	.60	.59	.79	.99

Note: Se = sensitivity; Sp = specificity; PPV = positive predictive value; NPV = negative predictive value; [1]Child met ASD criteria on both the *ADOS-2* and the *ADI-R*; [2]Child met ASD criteria on either or both of the *ADOS-2* and the *ADI-R*.

The *ADOS-2* both with and without the Toddler Module was highly accurate, and sensitivity and specificity are comparable to the published values. The *ADI-R* accuracy was lower than the *ADOS-2* but was still relatively high compared to levels reported in the literature. Requiring a child to meet ASD classification on both the *ADOS-2* and the *ADI-R* reduced sensitivity compared to the *ADOS-2* alone but increased specificity.

Relaxing the criteria such that children who met ASD classification on either or both of the *ADOS-2* and the *ADI-R* increased sensitivity but significantly decreased specificity to unacceptable levels.

Table 7. Agreement Categories

1. Agree (ASD)
2. Agree (Non-ASD)
3. Disagree (*ADOS-2* & DX = ASD)
4. Disagree (*ADI-R* & DX = ASD)
5. Disagree (*ADOS-2* & DX = Non-ASD)
6. Disagree (*ADI-R* & DX = Non-ASD)
7. Not Included (*ADOS-2* & *ADI-R* = ASD, DX = Non-ASD)
8. Not Included (*ADOS-2* & *ADI-R* = Non-ASD, DX = ASD)

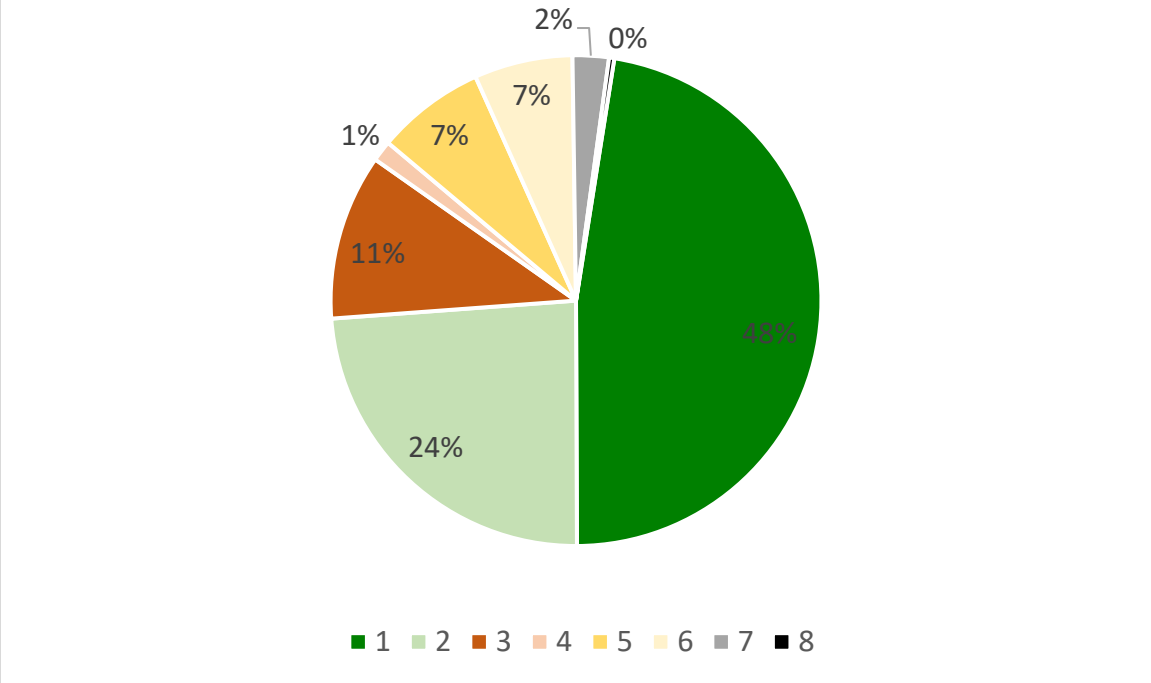

Figure 2. Frequencies of agreement categories.

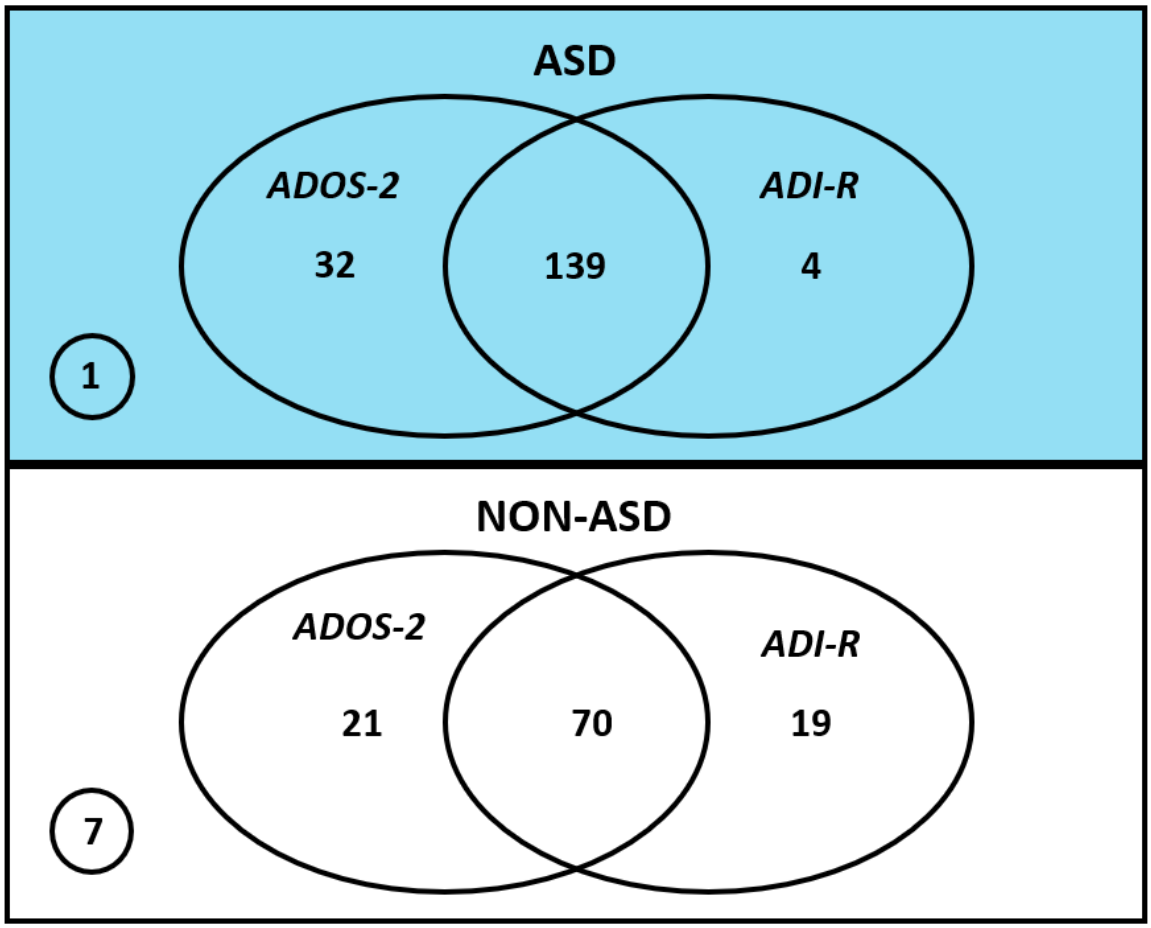

Figure 3. Agreement of the *ADOS-2*, *ADI-R*, and final clinical diagnosis.

Overall, the *ADOS-2*, *ADI-R*, and final clinical diagnosis agreed with one another most of the time (72%). For just under half of the cases (48%) both ASD measures and final diagnosis indicated ASD. For 24% of the evaluations, both ASD measures and final diagnosis resulted in a non-ASD diagnosis (Figure 2). For instances in which both diagnostic measures and final diagnosis disagreed, 18% of the time, the diagnosis was consistent with the *ADOS-2* result (ASD: 11%, non-ASD: 7%) whereas the diagnosis was consistent with the *ADI-R* classification in only 8% of cases (ASD: 1%, non-ASD: 7%). Among the eight cases in which the *ADOS-2* and *ADI-R* classifications were in agreement but disagreed with the final diagnosis, for seven of these cases the tests resulted in classifications of ASD but the diagnostician ultimately made a non-ASD

diagnosis (Figure 3, bottom-left corner of Non-ASD box). There was only one instance in which the *ADOS-2* and *ADI-R* results indicated non-ASD but the clinical team made an ASD diagnosis (Figure 3, bottom-left corner of ASD box).

Using the Landis & Koch (1977) guidelines specified above (<.20 = poor, .21 to .40 = fair, .41 to .60 = moderate, .61 to .80 = good, and .81 to 1.00 = very good), agreement between the *ADOS-2* and the *ADI-R* was moderate. The *ADOS-2* classification both with and without the Toddler Module classifications included resulted in good agreement with final diagnosis, whereas the *ADI-R* showed only moderate agreement with final diagnosis. All agreement analyses showed high levels of significance; however, it has been noted that it is possible to achieve significance using Cohen's κ with low levels of agreement, and therefore it is recommended that the strength of the κ value be interpreted outside of significance values (Bakeman & Gottman, 1997).

Table 8. Agreement between ASD Diagnostic Measures and Final Diagnosis

	ADI-R				Final Diagnosis			
	n	*κ*	*SE*	*p*	*n*	*κ*	*SE*	*p*
ADOS-2 with Toddler	285	.438	.054	***	339	.784	.034	***
ADOS-2 without Toddler	--	--	--	--	319	.777	.036	***
ADI-R	--	--	--	--	309	.565	.048	***

* = p < .05; ** = p < .01; *** = p < .001; SE = standard error

Child age was significantly negatively correlated with agreement (r = -.191, p < .005), indicating that for younger children the *ADOS-2* and the *ADI-R* were more likely to agree with one another, and likelihood of agreement decreased for older children. No other variables were significantly correlated or associated with agreement. Maternal age at birth trended toward significance (r = .103, p < .10), with older maternal age associated with higher likelihood of agreement between measures. Therefore, these two variables were included as predictors in a binomial logistic regression predicting the dichotomous

agreement between the *ADOS-2* and the *ADI-R*. This resulted in a significant model [$\chi^2(2) = 10.527$, $p < .005$, $n = 275$] which correctly classified 73.1% of cases. When both predictors were included in the model, only child age was significant (see Table 11), indicating that for every one-year decrease in child age at evaluation, the odds of agreement increases by a factor of 1.10.

Table 9. Correlations Between Agreement and Child and Family Factors

Variable Correlated with Agreement	r_{pb}	*n*	*p*
Age	-.191	285	.001**
Cognitive	.004	191	.877
Adaptive	-.088	173	.953
Language	-.072	266	.250
Gestational Age	.050	270	.409
Order of Child in Home	.031	272	.607
Number of Children in Home	-.059	280	.322
Number of Caregivers in Home	.124	176	.102
Maternal Age at Birth	.103	275	.088
Maternal Age at Evaluation	-.021	268	.728

* = p < .05, ** = p < .01, *** = p < .001

Table 10. Associations Between Child and Family Factors and Agreement

Variable Associated with Agreement	χ^2	*df*	*n*	*p*
Gender	.024	1	285	.877
Maternal Education (High School or Less / Some College or More)	.430	1	257	.512
Biological Parent Marital Status	1.175	1	268	.278
Insurance Type (Public/Private)	.296	1	284	.568
IEP (Yes/No)	1.037	1	227	.309
Sensory	.08	1	88	.777

Table 11. Logistic Regression Predicting Agreement Based on Child Age and Maternal Age at Birth

	B	SE	Wald	*df*	*p*	Odds Ratio
Age	-.100	0.36	7.593	1	.006**	.905
Maternal Age at Birth	.038	.023	2.575	1	.109	1.038
Constant	.717	.694	1.066	1	.302	2.048

** = p < .01

DISCUSSION

The purpose of this study was to investigate the child and family factors that predicted the agreement of the *ADOS-2* and the *ADI-R* in an attempt to streamline and simplify the diagnostic process in a university-based clinical sample. It was believed that further investigation of complex cases in which the *ADOS-2* and the *ADI-R* disagreed would help to simplify the ASD diagnostic process and provide insight to less experienced diagnosticians conducting ASD evaluations. A simplification of this process would help increase the number of ASD evaluations that could be conducted, which would reduce wait-times and contribute to a lower age of diagnosis for these children. This would allow them to be identified and receive ASD-specific early intervention services, which have been shown to produce the best outcomes in adolescence and adulthood for people with ASD.

Accuracy of ASD Diagnostic Measures

Overall, in this sample, the *ADOS-2* and *ADI-R* demonstrated high levels of accuracy in predicting whether or not a child received a final clinical diagnosis of ASD, and the *ADI-R* accuracy was lower than the *ADOS-2* overall. These are similar to the results reported in the literature (Lord, Rutter, et al., 2012). Lower accuracy of the *ADI-R* compared with the *ADOS-2* may be due to the strict cut-off criteria in the *ADI-R* and the reflection of DSM-IV criteria for Autistic Disorder which has not yet been updated for

DSM-5. This *ADI-R* may not account for milder presentations of ASD symptoms that are understood to represent the broader spectrum of skills and behaviors. Within the ASD diagnostic clinic at this tertiary care center, the *ADI-R* diagnostic classification was used as only one piece of information and considered amongst all other information gathered during the evaluation to come to a final diagnostic conclusion.

The highest levels of overall accuracy were achieved by using the *ADOS-2* alone or when combining the *ADOS-2* and the *ADI-R* and requiring a classification of *autism* or *autism spectrum* on both measures for an ASD diagnosis. When using the *ADOS-2* alone to predict ASD final clinical diagnosis in a high-risk sample, sensitivity is high but specificity is lower, although in an acceptable range diagnostically. When using both the *ADOS-2* and the *ADI-R* together, results are reversed, with high levels of specificity and lower, but still acceptable, levels of sensitivity.

Overall, this indicates that the administration of the *ADOS-2* is certainly a crucial piece of an ASD evaluation, and this is generally accepted best practice in the field (Penner et al., 2017). However, given that overall accuracy was similar with the *ADOS-2* alone compared with the *ADOS-2* and the *ADI-R* combined, this introduces the question of whether the administration of the *ADI-R* is a necessary piece of an ASD evaluation and whether or not it is additive to the clinical diagnostic process within a high-risk sample such as the sample in this study. There is often a long delay between time of referral and the completion of a diagnostic evaluation for ASD due to limited resources; therefore, it is imperative to maximize clinician resources and use time wisely. The results of this study indicate that the *ADOS-2* should likely be administered to all children referred to tertiary care settings for a concern of possible ASD, but that the *ADI-R* may be used at

the clinician's discretion, such as when presented with a particularly complex case. In other instances, an ASD-focused clinical interview can likely be substituted for the *ADI-R* without loss of overall diagnostic accuracy if the *ADOS-2* is administered. However, given that this study did not include a direct comparison of the accuracy of the *ADI-R* compared to unstructured ASD-focused clinical interview, further research should investigate this question.

Agreement of ASD Diagnostic Measures

Overall agreement was moderate between the *ADOS-2* and the *ADI-R*, and these measures agreed with the final diagnosis 71% of the time. High levels of agreement between these measures have been reported in the literature (de Bildt et al., 2004; Neuhaus et al., 2017). When the *ADOS-2* and the *ADI-R* produced conflicting classifications, the diagnostic team usually sided with the *ADOS-2* when making an ASD diagnosis, but diagnosticians were split between siding with the *ADOS-2* or the *ADI-R* when making a non-ASD diagnosis. Child age was the only significant child factor in predicting agreement; with younger children having higher likelihood of agreement, and there were no significant family factors. This is contrary to Neuhaus et al. (2017), in which child age did not predict agreement between these measures. In the current study, no other child factors predicted agreement, whereas Neuhaus et al. (2017) showed greater agreement between measures for children with lower IQ, poorer adaptive behavior, and fewer behavioral difficulties. These differences may be explained by differing samples between these studies. The Neuhaus et al. (2017) study utilized data from the Simons Simplex Collection (SSC) which is a large dataset consisting of a consortium of families

with only one child with ASD. All children in the SSC sample were diagnosed with ASD by research-reliable clinicians, whereas the current sample included children with ASD as well as other non-ASD diagnoses. Exclusionary criteria in the SSC included children with a mental age below 18 months, known genetic conditions, pregnancy or birth complications, premature birth, and low birth weight, and in the current clinical sample, many children with these characteristics were included. Finally, overall mean age of the sample was about three years older in the Neuhaus et al. (2017) study compared to the current study [Neuhaus et al. (2017) *M* = 108.3 months (9.0 years), *SD* = 42.8 months (3.6 years); Current study *M* = 6.3 years, *SD* = 3.8 years]. These differences may account for the varied results in factors predicting agreement between this study and Neuhaus et al. (2017).

The ASD diagnostic measures were more likely to disagree with each other for older children compared with younger children. This may be because younger children who were referred for an ASD evaluation were more likely to have clearer symptoms, whereas children who received an ASD evaluation at a later age may have had a more nuanced and complex presentation. Additionally, the *ADI-R* diagnostic algorithms are based on observations from when the child was four years old and its clinical utility may be more affected by age and language levels for younger children (Kim, Thurm, Shumway, & Lord, 2013). Diagnostic teams evaluating younger children may be able to streamline the diagnostic process by administering the *ADOS-2* alone and feel confident that the *ADI-R* would agree with the *ADOS-2* classification if it were administered, making this measure a less critical component of the evaluation. However, for children for whom diagnostic presentation is unclear and could benefit from further clarification,

the *ADI-R* could be administered to serve as another piece of data to add to the diagnostic process and assess some of the more nuanced presentations of the older children referred for diagnosis. Given that no other child or family factors were predictive of agreement, it may be difficult for a diagnostician to predict in advance whether a child could benefit from a shorter ASD evaluation with a less experienced diagnostician or whether a full comprehensive ASD evaluation would be required. Further research should investigate the accuracy of these diagnostic measures across age ranges of the referred child as well as how clinician training and experience affects the accuracy and agreement of these measures compared to final clinical diagnosis.

Generalizability of Sample and Diagnostic Process

Results indicated that the ASD and non-ASD groups differed on gestational age by an average of one week (ASD: M = 38.3, SD = 2.6, Non-ASD: M = 37.3, SD = 3.4). According to the National Institute of Child Health and Human Development, children born following 37 weeks through 38 weeks and 6 days gestation are considered "early term" and are 5% more likely to have an intellectual or developmental disability compared to those born after 39 weeks gestation (Spong, 2013). However, given that the non-ASD children in this sample were not typically developing and instead represent a wide range of non-ASD disorders, this observed group difference in the current study may not be clinically relevant for guiding diagnosis.

Although previous research has indicated that the *ADOS-2* and *ADI-R* are less accurate in clinical settings compared to research settings (de Bildt et al., 2015; Oosterling, Roos, et al., 2010), possibly due to the heterogenous samples referred for

evaluation in the community compared to more homogenous research samples, all psychologists at this tertiary care clinic who were scoring and interpreting these measures were highly trained and achieved research reliability on these measures. In addition, this site is a member of the University Centers for Excellence in Developmental Disabilities Education, Research, and Service (UCEDD) and Leadership Education in Neurodevelopmental and Related Disabilities (LEND) programs, providing training at the graduate level in ASD assessment. Therefore, although this study serves as a methodological improvement using retrospective review of evaluations initially conducted solely for clinical purposes compared to evaluations from samples recruited specifically for research, the highly trained clinicians in this clinic may contribute to the high levels of accuracy of these measures, especially since the study was unable to account for the intangible effects of clinical expertise on the accuracy of these measures. This may not reflect other community clinics at which diagnosticians have been trained clinically but not at the level of research reliability or those without significant ASD expertise. This suggests that diagnosticians may need significant training, possibly at the level of research reliability, to make adequately accurate clinical judgements when using these measures. The field would benefit from future research using similar methodology investigating the accuracy and agreement of the *ADOS-2* and the *ADI-R* when these measures were administered by diagnosticians who may be more representative of general community practitioners who have significant clinical expertise and experience working with children with ASD but have not been trained to the level of research reliability (Molloy et al., 2011).

Of note, the final clinical diagnosis rendered by this multidisciplinary team was not validated by another outside institution, which is a limitation. Furthermore, although many different clinical psychologists administered the *ADOS-2* (*n* = 6) and the *ADI-R* (*n* = 5), one lead clinical psychologist (SO) conducted or supervised administration of 41% of all *ADOS-2* evaluations and 61% of all *ADI-R* administrations at this clinic. When evaluating the accuracy of diagnostic measures, it may be prudent to confirm the clinical diagnosis for at least a subset of children to measure agreement between the diagnostic outcomes in different clinics. This would include the involvement of an independent outside institution to confirm the result of the evaluation for reliability purposes to improve confidence in the accuracy of the final clinical diagnosis.

Many of the variables utilized in this study were measures regarding the biological mother. However, oftentimes the person accompanying the child to the clinical evaluation and participating in the clinical interview was not the biological mother. Information was not available in the database regarding the relationship of the person who completed the clinical interview to the child. Therefore, there is an inherent limitation in relating maternal characteristics to accuracy of the *ADI-R* given that the reporter was not always the biological mother. ASD and Non-ASD groups differed significantly on maternal age at birth and maternal age at evaluation, with the ASD group having higher maternal age at birth and maternal age at evaluation compared to the Non-ASD group. This is consistent with the literature indicating that children born to mothers of higher age are at greater risk for the development of ASD (Croen et al., 2007; Lauritsen et al., 2005; Parner et al., 2012). Additionally, previous research has indicated the relationship between advanced paternal age and increased risk for ASD (Hultman et

al., 2011); however, paternal characteristics were not evaluated for this study. Future research should investigate whether paternal characteristics are related to ASD evaluation and diagnosis, including the agreement accuracy of the diagnostic measures.

Regarding biological parent marital status, children in the ASD group were more likely to have parents that were married compared with the non-ASD group. Although the literature shows that the stress of parenting a child with ASD often leads to stress on the parental relationship and may lead to greater rates of separation and divorce for these couples (Hartley et al., 2010), similar findings have been reported for families of children with other developmental disorders including ADHD (Wymbs et al., 2008), and the literature is not clear regarding whether parental marital status may be directly related to ASD diagnosis. The overall rate of biological parents who were not married in our sample was higher than reported in the general population (U.S. Census Bureau, 2018). More specific information regarding whether the biological parents were divorced, separated, or never married was not investigated, and may be additive in future research.

An inherent flaw with this dataset is that race and ethnicity were not directly reported by the child or family and were instead rated by the clinician based on their best estimate. The authors are aware that this method of collecting this variable would result in an inaccurate measure of race and ethnicity, and therefore further analysis of this variable was not completed. However, race and ethnicity as well as other cultural and identity characteristics have routinely been excluded from the research literature regarding ASD and other developmental disorders. Consideration of these variables is of the utmost importance to better understand these populations, particularly those who identify as minorities. It is recommended that all clinics regularly collect patient-reported

information regarding race, ethnicity, and identity, whether as part of the initial intake paperwork or asked by clinicians during the clinical interview. This information can then be entered into a clinical database for use in future research.

Additionally, given the nature of retrospective chart review of clinical evaluations, children were administered different measures according to clinical need resulting in missingness that is not random. Different sample sizes per group and within measures must be taken into account when interpreting the results of this study, particularly when evaluating results regarding the effects of child and family factors and non-ASD measures administered including cognitive, adaptive, language, and sensory measures.

Conclusion

Overall, this study indicates that the *ADOS-2* alone as well as in combination with the *ADI-R* produced the highest levels of accuracy in predicting final clinical diagnosis. It may be suggested that diagnostic teams conducting ASD diagnostic evaluations should always administer the *ADOS-2* and can use clinical judgment to determine whether the administration of the *ADI-R* would be beneficial to the evaluation. In particular, evaluations with children with more complex or nuanced presentations could benefit from the administration of the *ADI-R*.

REFERENCES

Bakeman, R. & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. (2nd ed.). New York, NY, US: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511527685

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J., … Wheelwright, S. (1999). Autism spectrum disorders at 20 and 42 months of age: stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *40*(5), 719–32.

Croen, L. A., Najjar, D. V., Fireman, B. & Grether, J. K. (2007). Maternal and paternal age and risk of autism spectrum disorders. *Archives of Pediatrics & Adolescent Medicine*, *161*(4), 334–40.

De Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F. & Minderaa, R. (2004). Interrelationship between autism diagnostic observation schedule-generic (ADOS-G), autism diagnostic interview-revised (ADI-R), and the diagnostic and statistical manual of mental disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders*, *34*(2), 129–137.

De Bildt, A., Sytema, S., Zander, E., Bölte, S., Sturm, H., Yirmiya, N., … others. (2015). Autism diagnostic interview-revised (ADI-R) algorithms for toddlers and young preschoolers: application in a non-US sample of 1,104 children. *Journal of Autism and Developmental Disorders*, *45*(7), 2076–2091.

Dorlack, T. P., Myers, O. B. & Kodituwakku, P. W. (2018). A comparative analysis of the ADOS-G and ADOS-2 algorithms: preliminary findings. *Journal of Autism and Developmental Disorders*, 1–12.

Dugas, C., Simard, M. N., Fombonne, E., and Couture, M. (2018). Comparison of two tools to assess sensory features in children with autism spectrum disorder. *The American Journal of Occupational Therapy, 72*(1), 1-9.

Falkmer, T., Anderson, K., Falkmer, M. & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: a systematic literature review. *European Child & Adolescent Psychiatry*, *22*(6), 329–40. doi:10.1007/s00787-013-0375-0

Gilchrist, A., Green, J., Cox, A., Burton, D., Rutter, M. & Le Couteur, A. (2001). Development and current functioning in adolescents with Asperger syndrome: a comparative study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 227–40.

Gotham, K., Risi, S., Dawson, G., Tager-Flusberg, H., Joseph, R., Carter, A., … Lord, C. (2008). A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(6), 642–51. doi:10.1097/CHI.0b013e31816bffb7

Hansen, K. D. & Jirikowic, T. (2013). A comparison of the Sensory Profile and Sensory Processing Measure Home Form for children with fetal alcohol spectrum disorders. *Physical & Occupational Therapy in Pediatrics*, *33*(4), 440-452.

Hartley, S. L., Barker, E. T., Seltzer, M. M., Floyd, F., Greenberg, J., Orsmond, G., & Bolt, D. (2010). The relative risk of timing of divorced in families of children with an autism spectrum disorder. *Journal of Family Psychology, 24*(4), 449-457. doi:10.1037/a0019847

Hultman, C. M., Sandin, S., Levine, S. Z., Lichtenstein, P., & Reichenberg, A. (2011). Advancing paternal age and risk of autism: new evidence from a population-based study and meta-analysis of epidemiological studies. *Molecular Psychiatry, 16,* 1203-1212.

Kamp-Becker, I., Ghahreman, M., Heinzel-Gutenbrunner, M., Peters, M., Remschmidt, H. & Becker, K. (2011). Evaluation of the revised algorithm of Autism Diagnostic Observation Schedule (ADOS) in the diagnostic investigation of high-functioning children and adolescents with autism spectrum disorders. *Autism*, *17*(1), 87–102.

Kim, S. H. & Lord, C. (2012a). 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.

Kim, S. H. & Lord, C. (2012b). New autism diagnostic interview-revised algorithms for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Autism and Developmental Disorders*, *42*(1), 82–93.

Kim, S. H., Thurm, A., Shumway, S. & Lord, C. (2013). Multisite study of new autism diagnostic interview-revised (ADI-R) algorithms for toddlers and young preschoolers. *Journal of Autism and Developmental Disorders*, *43*(7), 1527–38. doi:10.1007/s10803-012-1696-4

Kreiser, N. L. & White, S. W. (2014). ASD in females: are we overstating the gender difference in diagnosis? *Clinical Child and Family Psychology Review*, *17*(1), 67–84.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Lauritsen, M. B., Pedersen, C. B. & Mortensen, P. B. (2005). Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *46*(9), 963–71.

Lord, C., Luyster, R. J., Gotham, K. & Guthrie, W. (2012). *ADOS-2. Autism Diagnostic Observation Schedule.* (2nd ed.). Torrance, CA: Western Psychological Services.

Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T. & Manning-Courtney, P. (2011). Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism: The International Journal of Research and Practice*, *15*(2), 143–62. doi:10.1177/1362361310379241

Neuhaus, E., Beauchaine, T. P., Bernier, R. A. & Webb, S. J. (2017). Child and family characteristics moderate agreement between caregiver and clinician report of autism symptoms. *Autism Research, 11(3),* 476-487.

Oosterling, I. J., Rommelse, N., De Jonge, M., Van Der Gaag, R. J., Swinkels, S., Roos, S., … Buitelaar, J. (2010). How useful is the Social Communication Questionnaire in toddlers at risk of autism spectrum disorder? *Journal of Child Psychology and Psychiatry*, *51*(11), 1260–1268.

Oosterling, I. J., Roos, S., de Bildt, A., Rommelse, N., de Jonge, M., Visser, J., … Buitelaar, J. (2010). Improved diagnostic validity of the ADOS revised algorithms: a replication study in an independent sample. *Journal of Autism and Developmental Disorders*, *40*(6), 689–703. doi:10.1007/s10803-009-0915-0

Parner, E. T., Baron-Cohen, S., Lauritsen, M. B., Jørgensen, M., Schieve, L. A., Yeargin-Allsopp, M. & Obel, C. (2012). Parental age and autism spectrum disorders. *Annals of Epidemiology*, *22*(3), 143–150.

Penner, M., Anagnostou, E., Andoni, L. Y. & Ungar, W. J. (2017). Systematic review of clinical guidance documents for autism spectrum disorder diagnostic assessment in select regions. *Autism*, 1362361316685879.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., … Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*(9), 1094–103.

Rutter, M., Le Couteur, A., Lord, C. & others. (2003). Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services*, *29*, 30.

Spong, C. Y. (2013). Defining "term" pregnancy: recommendations from the Defining "Term" Pregnancy Workgroup. *Jama*, *309*(23), 2445–2446.

Tomanik, S. S., Pearson, D. A., Loveland, K. A., Lane, D. M. & Shaw, J. B. (2007). Improving the reliability of autism diagnoses: Examining the utility of adaptive behavior. *Journal of Autism and Developmental Disorders*, *37*(5), 921–928.

U.S. Census Bureau. (2018, November). Current Population Survey. Retrieved from https://www.census.gov/programs-surveys/cps.html

Wymbs, B. T., Pelham, W. E., Molina, B. S. G., Gnagy, E. M., Wilson, T. K., & Greenhouse, J. B. (2008). Rate and predictors of divorce among parents of youth with ADHD. *Journal of Consulting and Clinical Psychology, 76*(5), 735–744. doi:10.1037/a0012719

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35.

Zander, E., Sturm, H. & Bölte, S. (2015). The added value of the combined use of the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule: Diagnostic validity in a clinical Swedish sample of toddlers and young preschoolers. *Autism*, *19*(2), 187–199.

CLINICAL UTILITY OF TWO SCREENING MEASURES FOR AUTISM SPECTRUM DISORDER: *M-CHAT-R* AND *CSBS-DP-ITC*

JENNA B. LEBERSFELD, SARAH E. O'KELLEY, KIRSTIN J. BAILEY, ELIZABETH M. GRIFFITH, FRED J. BIASINI, KRISTI C. GUEST

In preparation for the *Journal of Autism and Developmental Disorders*

Format adapted for dissertation

ABSTRACT

Young children with ASD often require a comprehensive clinical evaluation and diagnosis to access ASD-specific early intervention services. Wait-times from initial family concern to diagnosis and intervention are long due, in part, to the high numbers of children referred for a concern of ASD and low availability of diagnostic providers. A systematic triage process in a tertiary care center utilizing the *M-CHAT-R* and *CSBS-ITC* screening tools was implemented to help determine which children would be best served through a comprehensive ASD evaluation in place of a broad developmental evaluation. Caregivers completed the *M-CHAT-R* (n = 75) and the *CSBS-ITC* (n = 216) prior to children receiving a comprehensive ASD diagnostic evaluation utilizing the *ADOS-2* and the *ADI-R*. Analyses indicated that the *CSBS-ITC* predicted the *ADOS-2* and final clinical diagnosis whereas the *M-CHAT-R* only predicted the *ADOS-2* classification. Neither screening tool predicted the *ADI-R*. Both measures had high levels of sensitivity but low specificity at the published cut-off scores. Overall, the *CSBS-ITC* showed evidence of clinical utility in the tertiary care triage process, whereas the *M-CHAT-R* showed some possible benefit, but more research is needed due to mixed results. Additionally, more conservative cut-off values for the screening measures in at-risk samples may help improve the triage process.

INTRODUCTION

Enrollment in early intervention services which utilize strategies specific for children with autism spectrum disorder (ASD) is crucial for children with ASD to acquire the necessary skills to achieve the best possible quality of life. However, a comprehensive and/or specialized ASD evaluation is often required for children to access services, and there is a shortage of opportunities across the country to receive these types of evaluations (Oswald et al., 2017). A multisite study of children across the United States conducted by the Centers for Disease Control and Prevention (CDC) showed that 85% of families whose children later went on to be diagnosed with ASD had developmental concerns for their child by 36 months of age. However, only 42% of these children received an evaluation by 36 months of age, and 39% of these children did not receive a comprehensive evaluation before four years old (Baio, 2018). Furthermore, the rate of ASD diagnosis is increasing and is currently reported to affect one in 59 children (Baio, 2018). The increasing rate of ASD in children and the insufficient number of providers available to conduct evaluations for these families creates a significant delay between when a family first expresses concern regarding their child's development and when a comprehensive ASD evaluation can be completed. It is imperative to determine which high-risk children referred for evaluation are best served by receiving a comprehensive, ASD-focused evaluation as opposed to another type of evaluation (e.g., broad developmental, psychoeducational, behavioral, etc.).

A variety of screening tools have been created to foster earlier identification of children at risk for ASD. Screening for autism spectrum disorder (ASD) in primary care settings is recommended by the American Academy of Pediatrics for all children at ages 18 months and 24 months (Johnson & Myers, 2007; Towle & Patrick, 2016). Level 1 screening tools have been created for primary care providers for use with the general population and consist of short checklists or questionnaires. These types of measures require minimal time and effort on behalf of the primary care provider and caregiver and require little-to-no training to use and interpret. The primary purpose of Level 1 of screening tools is to identify toddlers at-risk for a developmental disorder and differentiate them from typically developing children (Johnson, Myers, & others, 2007). Implementation of screening tools has been shown to reduce the delay between caregiver or provider concern and ASD diagnosis by two years (Robins, 2014). Multiple screening tools have been developed for use in primary care but are not often used in more specialized settings.

Secondary care settings include schools, community mental health clinics, and other private providers who may work with children with certain types of developmental disorders but are not highly specialized. Level 2 screening tools were designed for implementation in these types of settings to differentiate children at-risk for ASD from children at-risk for other developmental disorders. These types of screening measures are more time consuming to administer and often require a trained clinician to directly observe child behaviors (Johnson et al., 2007). In the United States, two-tiered screening approaches are not routinely used due to time constraints in primary care settings (Khowaja, Robins, & Adamson, 2018). Therefore, it is recommended that a child who

fails a Level 1 or Level 2 screening tool be referred for an ASD diagnostic evaluation, often conducted at a tertiary care center.

Tertiary care centers provide highly specialized diagnosis and treatment of complex cases, and no screening measures have been specifically designed for use in tertiary care settings providing ASD-focused clinical evaluations. Level 1 screening tools originally developed for use in primary care settings may be useful in tertiary care clinics to determine which children are at increased risk for ASD and require an ASD-focused comprehensive evaluation compared to those for whom a general developmental evaluation would be appropriate. Level 2 screening tools are not feasible for use in the initial triage and clinic assignment process due to the time needed to administer and score these measures. Using Level 1 screening tools to help identify which children would benefit most from an ASD evaluation would allow children with ASD to access ASD-focused early intervention services as early as possible. However, the clinical utility of Level 1 screening measures in this type of tertiary care setting is unknown.

To optimize the intake and triage process for a high-risk, clinically referred sample, a tertiary care clinic in the southeastern United States implemented two Level 1 screening tools in children ages four years old and younger: the *Modified Checklist for Autism in Toddlers, Revised* (*M-CHAT-R*; Robins, 2009) and the *Communication and Symbolic Behavior Scales Developmental Profile Infant Toddler Checklist* (*CSBS-ITC*; Wetherby, 2002). The purpose of this study was to determine the clinical utility of these screening tools, both individually and combined, in predicting the results of measures commonly used during a comprehensive ASD psychological evaluation as well as the final clinical diagnosis rendered by the multidisciplinary team who conducted the

evaluation. It was predicted that scores from these screening measures, the *M-CHAT-R* and the *CSBS-ITC*, would predict scores and classifications on the ASD diagnostic measures (i.e., the *Autism Diagnostic Observation Schedule, Second Edition* (*ADOS-2*) and the *Autism Diagnostic Interview, Revised* (*ADI-R*)) as well as final clinical diagnosis.

METHODS

Participants

Participants were 217 children referred for a comprehensive psychological evaluation including a question of ASD at a tertiary care center in the southeastern United States from the years 2013 to 2018. Demographics are presented in Tables 1 and 2. Group differences were calculated using *t*-tests for continuous variables and χ^2 tests for categorical variables. When initial χ^2 tests resulted in cells with expected counts less than five, categories were collapsed. If the test continued to have expected cell counts less than five, categories were dichotomized, and the result for the Fisher's Exact Test was interpreted.

Table 1. Participant Demographics: Child Characteristics

	Total		ASD				Non-ASD			
	n	*M* (*SD*)	*n*	*M* (*SD*)	Min.	Max.	*n*	*M* (*SD*)	Min.	Max.
Age at Screening (years)	217	2.64 (.79)	129	2.63 (.78)	1.00	4.42	88	2.66 (.81)	1.33	5.25
Delay Between Screening and ASD Evaluation (years)	217	1.55 (1.24)	129	1.51 (1.16)	.16	6.36	88	1.60 (1.36)	0.12	7.44
Age at Evaluation (years)	217	4.18 (1.78)	129	4.13 (1.70)	1.33	10.00	88	4.26 (1.89)	1.92	11.33
Gestational Age (weeks)	213	37.96 (3.08)	128	38.37 (2.78)	23	45	85	37.35 (3.41)	23	41
	n	%	n	%			n	%		
Gender	217	--	129	--			88	--		
Male	170	78	103	80			67	76		
Female	47	28	26	20			21	24		
Race	214	--	128	--			86	--		
White	126	59	71	55			55	64		
Black	60	28	37	29			23	27		
Other	28	13	20	16			8	9		
Individualized Education Program (IEP)	168	--	98	--			70	--		
Yes	131	78	77	79			54	77		
No	37	22	21	21			16	23		
Special Education Eligibility Category	112	--	67	--			45	--		
Autism	35	31	17	25			18	40		
Developmental Delay	43	38	31	46			12	27		
Speech or Language Impairment	14	13	11	16			3	7		
Other	20	18	8	12			12	27		
Order of Child in Home	204	--	124	--			80	--		
First	105	52	69	56			36	45		
Second	64	31	36	29			28	35		
Third	25	12	16	13			9	11		
Fourth or More	10	5	3	2			7	9		

Note: Some percentages may not add up to 100 due to rounding. N's <217 indicate unknown or missing data. Special education eligibility categories are missing or unknown for many young children in this sample due to ineligibility for an IEP prior to age 3 years old.

Table 2. Participant Demographics: Family Characteristics

	Total		ASD				Non-ASD			
	n	*M* (SD)	*n*	*M* (SD)	Min.	Max	*n*	*M* (SD)	Min.	Max
Maternal Age at Birth	212	27.52 (6.35)	127	29.10 (6.60)	18	46	85	25.16 (5.17)	17	41
Maternal Age at Evaluation	207	31.67 (6.74)	127	32.98 (6.79)	21	52	85	29.68 (6.22)	20	54
	n	%	*n*	%			*n*	%		
Maternal Education	197	--	114	--			83	--		
Less Than High School	24	12	7	5			17	20		
Graduated High School/GED	79	40	42	33			37	45		
Some College	58	29	38	29			20	24		
Graduated College	28	14	20	16			8	10		
Graduate or Medical School	8	4	7	5			1	1		
Insurance Type	217	--	129	--			88	--		
Public	172	79	94	73			78	89		
Private	45	21	35	27			10	11		
Biological Parent Marital Status	206	--	121	--			85	--		
Married	90	44	58	48			32	38		
Not Married	116	56	63	52			53	62		
Number of Caregivers in Home	143	--	83	--			60	--		
1	31	22	18	27			13	22		
2	92	64	57	69			35	58		
3	12	8	6	7			6	10		
4+	8	6	2	2			6	10		
Number of Children in Home	212	--	128	--			84	--		
1	65	30	41	32			24	28		
2	73	34	49	38			24	28		
3	55	25	31	24			24	28		
4	12	6	4	3			8	10		
5+	7	3	3	2			4	5		

Note: Percentages may not add up to 100 due to rounding. Different n's per group are a result of unknown or missing data which could not be determined following chart review and data extraction.

Groups differed significantly on gestational age, with the ASD group being born one week later on average compared to the non-ASD group [ASD = 38.37, Non-ASD = 37.35, $t(211) = 2.41$, $p = .017$]. Groups differed on special education eligibility category [$\chi^2(3) = 9.85$, $p = .020$], with the ASD group having fewer classifications of ASD compared to the Non-ASD group, based on the child's special education category at the time of referral and clinic visit. Groups also differed on maternal age at birth [ASD M = 29.10, Non-ASD $M = 25.16$, $t(205) = 4.85$, $p < .001$] and maternal age at evaluation [ASD $M = 32.98$, Non-ASD $M = 29.68$, $t(210) = 3.58$, $p = < .001$], with mothers in the ASD group being older compared to those in the Non-ASD group. Maternal education differed between groups, with mothers in the ASD group having attended more years of school and having higher degrees compared to those in the Non-ASD group [$\chi^2(4) = 15.21$, $p = .004$]. Both groups had a higher percentage of public insurance compared to private insurance, but this percentage was larger for the Non-ASD group compared to the ASD group. Groups did not differ on other child or family factors.

Procedures

Clinic Assignment

At the tertiary care center, caregivers of children ages four years of age or younger at the time of referral are asked to complete two screening questionnaires, the *M-CHAT-R* and the *CSBS-ITC*, as part of intake materials prior to being considered for evaluation. Children referred to this center are assigned to one of the following types of psychological evaluation clinics based on the presenting concern(s): general developmental clinic, psychoeducational clinic, behavioral assessment clinic, or ASD

clinic. The results of the screening tools as well as parent or provider concern are used to determine the clinic assignment. Children for whom a caregiver or provider expressed concern regarding possible ASD receive a comprehensive clinical evaluation focusing on the question of ASD, regardless of the results of the screening tools.

Regarding terminology, literature published on screening tools commonly use the terms “pass” to refer to scores which do not indicate concern and “fail” for scores which indicate that some level of concern is present but do not necessarily indicate a negative outcome. “Pass/fail” terminology is routinely used in the medical field and has been utilized consistently in the social/behavioral research with screening instruments as well. To remain consistent with the existing literature, these “pass” and “fail” labels for the screening tools will be used throughout this paper.

For clinic assignment at the tertiary care center, children who “fail” (i.e., score in a range which indicates concern) one or both of the screening instruments are considered for assignment to the ASD clinic slot. In rare cases, when screening tools suggest possible ASD concerns but no parent or provider concern is noted, a clinician may review the chart and decide that an ASD evaluation is not necessary (e.g., if the child appeared to fail the screening measures due to significant motor delays and ASD seemed unlikely). The majority of children assigned to the ASD clinic receive the *ADOS-2* and the *ADI-R* as ASD diagnostic measures. Children who “pass” (i.e., score in a range indicating no concern) both screening tools and do not have a parent or provider concern of ASD are referred to a different type of evaluation and do not receive the ASD diagnostic assessments. Children who did not receive an ASD evaluation are not included in this sample.

ASD Clinical Evaluation

All children included in this sample received a comprehensive clinical evaluation for ASD at a tertiary care clinic in Alabama. During the ASD evaluation, children were evaluated by a multidisciplinary clinical team which was led by a licensed clinical psychologist with ASD expertise and both clinical and research training on the *ADOS-2* and the *ADI-R*. The clinical team also usually included a developmental behavioral pediatrician or general pediatrician, and the child may have also received evaluations from other disciplines including social work, optometry, audiology, nutrition, speech/language pathology, occupational therapy, and physical therapy. Following the evaluation, all members of the multidisciplinary team gathered to discuss the case and finalize a consensus multidisciplinary team decision regarding the child's diagnosis. Recommendations for services and intervention options were then delivered to the child's family verbally and through written reports. All procedures were conducted clinically, and retrospective use for research was approved by the Institutional Review Board (IRB) at the university at which the evaluations and research were conducted.

Chart Review and Data Extraction

Participant data were collected via retrospective chart review of these comprehensive ASD evaluations. Due to the clinical nature of these charts, the same measures were not administered to all children and the types of information available varied from child to child, and n's per group are presented in Tables 1, 2, 3, and 5. Unless specified below, the chart review and data extraction process was systematized by first

prioritizing information gathered from the reports generated by diagnosticians from the university's tertiary care clinic evaluation, followed by information from medical records, caregiver report on intake paperwork completed prior to evaluation, school records, and reports from outside evaluations (e.g., psychologist, speech language pathologist). Given the inherent complexity of this chart review, data extraction, and data entry process, variables had consensus-driven operational definitions, and graduate and undergraduate students were trained to 95% reliability for extraction of data. Any discrepancies were resolved through discussion with lab members and faculty mentors who also work as part of the ASD diagnostic team.

Variable Definitions and Limitations

Child Characteristics

Race. Although participant race is presented in Table 1, race in this dataset is clinician-rated as opposed to self-identified by the participants' families. Given the inherent error associated with this method of data collection, race is presented for descriptive purposes only and will not be explored further.

Gestational age. Gestational age was extracted from medical birth records whenever possible. When medical records were not available, information was gleaned from the psychology report or intake paperwork completed by caregivers prior to evaluation. When a source of information indicated that the child was born "full-term," gestational age was estimated to be 40 weeks.

Maternal age at birth. Maternal age at birth was primarily obtained from medical birth records. When medical birth records were unavailable, maternal age at birth was estimated by subtracting the child's age at evaluation from the mother's age at evaluation.

Maternal age at evaluation. Parent age is reported on the intake paperwork completed by caregivers prior to the ASD evaluation. However, at times there was a significant delay between the completion of the paperwork and the ASD evaluation. Therefore, maternal age at evaluation was estimated by adding the maternal age reported on the intake paperwork to the number of years from intake to evaluation.

Final Clinical Diagnosis

Final clinical psychological diagnoses were made using the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* or *Fifth Edition* (*DSM-IV*, *DSM-5*). The *DSM-IV* defined multiple types of autism spectrum disorders including Asperger's Disorder; Autistic Disorder; and Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS). The *DSM-5* was updated to reflect the conceptualization of ASD as a single disorder with differing levels of severity. For this study, *DSM-IV* diagnoses of Asperger's Disorder; Autistic Disorder; or Pervasive Developmental Disorder, Not Otherwise Specified were considered an ASD diagnosis. All other diagnoses were classified as non-ASD.

Measures

Screening Tools

Modified Checklist for Autism in Toddlers, Revised (*M-CHAT-R*; Robins et al., 2009)

The *M-CHAT-R* is a 20-item, ASD-focused yes/no questionnaire completed by a parent or caregiver which takes less than five minutes to complete, is free to use, and is available for download online. The items presented on the *M-CHAT-R* focus on concerns commonly seen in toddlers with ASD (e.g., "If you point at something across the room, does your child look at it?", "Does your child play pretend or make-believe?", "Does your child make unusual finger movements near his or her eyes?"). The *M-CHAT-R* is scored by adding up the number of items "failed" to receive a Total Raw Score ("*M-CHAT-R* Total"). Higher scores indicate greater levels of concern. The cut-off indicating concern on this measure is a Total Raw Score of three ("Total3 Score"). This means that any child who fails zero, one, or two items is considered to have "passed" the screening instrument, indicating no concern of ASD symptoms. Any score of three or greater indicates that the child has "failed" the screening tool. Whether a child "passed" or "failed" the measure was called the *M-CHAT-R* Decision for this study. A follow-up interview is available (*M-CHAT-R/F*; Robins, 2009), which is administered to clarify whether failed items on screening questionnaires are truly failed; however, the follow-up interview is not completed as part of this clinic's intake and triage process. Given that these children have already been referred for an in-person evaluation and the evaluation clarifies any reported symptoms in more depth, it was believed that the follow-up interview would not be additive to the triage, clinic assignment, and evaluation process nor save clinical time. Previous research has indicated that in a primary care sample,

without the follow-up interview, the sensitivity and specificity for the *M-CHAT-R* Total3 Score exceeded .90 (sensitivity = .91; specificity = .95; Robins et al., 2014). However, NPV was .999, and PPV was only .138 (Robins et al., 2014). This indicates that almost all children who passed the screening tool truly did not have ASD, but that many individuals scoring in the at-risk range for ASD on this measure did not receive a final clinical diagnosis of ASD (Robins et al., 2014).

Communication and Symbolic Behavior Scales, Developmental Profile (*CSBS-DP*; Wetherby & Prizant, 2003)

The *CSBS Developmental Profile* (*CSBS-DP*) is a developmental screening tool that was not developed specifically to focus on questions related to ASD symptomology, but instead was targeted toward identifying children demonstrating early delays in communication and social skills for the purpose of early identification and early intervention. In this clinic, one portion of the *CSBS-DP*, the *CSBS Infant-Toddler Checklist* (*CSBS-ITC*), is used, which is a 24-item parent questionnaire that is available at no cost. Each item has three to five answer choices and the total possible raw score is 47. Higher scores indicate better development. Additional *CSBS* follow-up parent interview questions and a brief behavioral observation are also available in the commercially published measure but are not used in this clinic for the purposes of triage and clinic determination.

The *CSBS-ITC* was initially developed for children ages two years old or younger, and normed composite scores are available for Social, Speech, and Symbolic scales as well as a Total Composite (Wetherby, 2002) for children ages 6 to 24 months. Per the

manual, a child was considered to have "failed" this screening tool if they performed less than 1.25 standard deviations below the mean of the normative sample based on any single composite score or the Total Composite score. The normative sample for the *CSBS-ITC* consisted of 2,188 children ages 6 to 24 months from eight sites across the United States and Canada. Sources for recruiting including doctor offices, childcare centers, and health fairs or other community events. Whether a child "passed" or "failed" the measure was called the *CSBS* Decision for this study. For a 24-month-old child, the following raw-score cut-offs indicate concern: Social ≤18, Speech ≤9, Symbolic ≤12, Total ≤42. The published means and standard deviations for the 24-month-old age group raw scores are as follows: Social 22.39 (3.02), Speech 11.94 (2.12), Symbolic 15.03 (1.99), Total 49.36 (5.60). The *CSBS-DP* manual indicates that this measure can be used for children up to age six years old if the developmental level is estimated to be less than two years old (Wetherby, 2002). Even for children outside the normed age range whose developmental level is unknown, as was the case with many referred families in this study, children older than 24 months of age can be compared with the 24-month age group, and norms for 24-month-olds were used in a previous study for children ages 25 to 44 months when comparing ASD classifications across screening instruments (Oosterling et al., 2009). However, Wetherby (2002) cautions the evaluator from interpreting and reporting standard scores and percentiles when using the measure outside of the normed age range. Therefore, for this study, raw scores were analyzed and interpreted. Specifically, the *CSBS-ITC* Social raw scores (*CSBS-ITC* Social) and *CSBS-ITC* Total raw scores (*CSBS-ITC* Total) were used for this study. The *CSBS-ITC* Social was used because social skills are an area of weakness for individuals with ASD and Wetherby

(2008) indicated that the Social Composite score can be used to distinguish children with ASD from those with other developmental disorders (DD). The *CSBS-ITC* Total was included to capture a broader measure of social communication and symbolic behaviors. For children at-risk for ASD ages 8 to 44 months, psychometrics for the *CSBS-ITC* Decision using the published cut-off value have been reported as follows: sensitivity .71, specificity .59, PPV .78, and NPV .50 (Oosterling et al., 2009).

The *M-CHAT-R* and *CSBS-ITC* were determined to be the best available measures for triage purposes based on the highest psychometric properties available for screening measures that were brief, easy to use, and free of charge. It was acknowledged that the accuracy of these measures was slightly lower than levels suggested for screening in ASD populations (sensitivity and specificity above .70; Dumont-Mathieu & Fein, 2005). However, it was predicted that the use of these two measures together would improve the clinical utility for triage of children referred for ASD assessment.

Table 3. Constructs and Measures

Construct	Measure	n
ASD	*ADOS-2* Toddler Module	19
	ADOS-2 Module 1	125
	ADOS-2 Module 2	47
	ADOS-2 Module 3	16
	ADI-R	178
Cognitive	*Bayley-III*	20
	DAS-II	83
	Leiter-3	7
	SB-5	2
	WASI-II	1
	WISC-V	4
Adaptive	*ABAS-Parent*	43
	AGS	15
	Vineland	64
Language	*CELF-5*	1
	CELF-P2	16
	OWLS-II	3
	PLS-4	5
	PLS-5	176
	REEL	2
Sensory	*SPM-Preschool*	21
	SPM-Version Not Specified	4
	SSP	15
	SSP2	28

ASD Diagnostic Measures

Autism Diagnostic Observation Schedule, Second Edition (*ADOS-2*; Lord et al., 2012)

The *ADOS-2* is a semi-structured, 45- to 60-minute observation and interaction session with an evaluator and the child used to aid in the diagnosis of ASD. Age- and language-appropriate activities, interview questions, and conversation topics with specific presses are administered by the clinician to yield different levels of symptoms associated with ASD. For Modules 1 through 4, total raw scores are used to determine diagnostic

classifications of *autism*, *autism spectrum disorder*, or *non-spectrum*. Children under 30 months old without phrase speech were administered the Toddler Module, and the total raw score is used to determine a level of concern (i.e., *little-to-no, mild-to-moderate,* or *moderate-to-severe*) is specified instead of a classification. For Modules 1, 2, and 3, a comparison score ranging from one to 10 can be calculated which indicates the level of ASD severity, with higher scores indicating greater levels of ASD symptomatology in comparison to children of the same age and language level. Comparison scores cannot be calculated for the Toddler Module or Module 4 based on the published clinical algorithms. The *ADOS-2* sensitivity estimates range from 0.61 to 0.97 and specificity ranges from 0.47 to 1 (Gotham et al., 2008; Kamp-Becker et al., 2011; Molloy et al., 2011; Oosterling, Roos, et al., 2010; Zander et al., 2015). A recent meta-analysis indicated pooled sensitivity ranging from .77 to .90 and specificity ranging from .62 to .90 for the *ADOS-2* (Dorlack et al., 2018).

Autism Diagnostic Interview, Revised (*ADI-R;* Rutter et al., 2003)

The *ADI-R* is a semi-structured diagnostic interview given to a parent or caregiver by a trained clinician. This measure takes about 90 to 150 minutes to administer and yields a classification of *autism* or *not autism*. Sensitivity ranges from .53 to .92, and specificity ranges from .62 to .95 (Falkmer et al., 2013; Risi et al., 2006). Of note, only one child who received the *ADOS-2* Toddler Module also received the *ADI-R*. For the other children, an ASD-focused clinical interview was administered which addressed the child's development and symptoms of ASD but with questions more appropriate for toddler-aged children. This is because the utility of the *ADI-R* diagnostic algorithm was

found to be most appropriate for children with a non-verbal mental age over two-years-old (Lord et al., 1994), and many younger children referred for evaluation at this clinic did not yet meet these criteria. A version of the *ADI-R* for toddlers was developed (Kim & Lord, 2012b) for research purposes for use with children under four years old; however, it is not widely available and has not yet been published for clinical use.

Cognitive Measures

During clinical assessments, the most appropriate cognitive measure for each child is chosen based on a variety of factors including child age, language level, and developmental level. The following measures were administered and included in this dataset: *Bayley Scales of Infant and Toddler Development, Third Edition* (*Bayley-III*); *Differential Ability Scales, Second Edition (DAS-II); Leiter International Performance Scale, Third Edition (Leiter-3); Stanford-Binet Intelligence Scales, Fifth Edition (SB-5); Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II);* and *Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V)*. Most of these tests were administered within the clinic setting, but other scores were retrieved from previous evaluations through the school system or other outpatient clinics. All of these tests measure cognitive ability and produce an overall standard score with a mean of 100 and a standard deviation of 15, such as the Full Scale IQ (FSIQ). This reflects variation in clinical practice and individualized cognitive assessment, therefore, these measures were collapsed across participants to create the cognitive skills variable. It is acknowledged that the *Bayley-III* is a developmental measure, and the cognitive scale may not demonstrate the same stability over time as other measures of cognitive ability.

Nevertheless, this was the best estimate of cognitive ability available for 20 children, and rather than exclude these children from the cognitive ability analyses, the *Bayley-III* estimates were included. Please refer to Table 3 for the number of administrations of each measure in the sample.

Adaptive Measures

A variety of measures were also used to measure adaptive skills in this sample. Tests were selected based on the age of the child as well as whether an interview or questionnaire was more clinically appropriate. The following tests were used to measure adaptive behavior in this sample: *Adaptive Behavior Assessment System, Parent Rating Forms (ABAS-Parent); American Guidance Service Early Screening Profiles, Self-Help/Social Profile (AGS);* and *Vineland Adaptive Behavior Scales (Vineland).* Similar to the cognitive tests, results from the majority of these measures were obtained through evaluations at this clinic; however, results may have been obtained from other reports and records provided from outside agencies. These tests also produce a standard score of global adaptive functioning, such as the General Adaptive Composite (GAC) or Adaptive Behavior Composite (ABC). These standard scores were collapsed across measures to create the adaptive behavior variable.

Language Measures

At this multidisciplinary clinic, children are referred for a speech-language evaluation if language is an area of concern for the family or provider; this is also seen as an important component of the multidisciplinary assessment for ASD and DD. Language

measures are selected based on the clinical need of the child. Tests were most often administered by speech-language pathologists as part of a formal speech-language evaluation at this multidisciplinary clinic; however, results of many of these tests were also obtained through school records from evaluations conducted by the public school system. Measures administered were the following: *Clinical Evaluation of Language Fundamentals, Fifth Edition (CELF-5) and Preschool-2 (CELF-P2); Oral and Written Language Scales, Second Edition (OWLS-II); Preschool Language Scales, Fourth Edition (PLS-4) and Fifth Edition (PLS-5);* and the *Receptive-Expressive Emergent Language Test (REEL).* These tests measure the language construct and produce a standard score estimating overall language ability. This score was collapsed across measures to create the overall language variable for this sample.

Sensory Measures

A child receives a measure evaluating sensory differences when there is a reported concern with this area of functioning, which occurs often in children referred for ASD evaluation. This is most often determined and administered by an occupational therapist during the multidisciplinary clinic evaluation for ASD; however, the results of some measures were also retrieved through school records. The tests used to measure sensory characteristics were the *Sensory Processing Measure, Home Version (SPM-Home), Preschool Version (SPM-Preschool) and the Short Sensory Profile, First Edition (SSP) and Second Edition (SSP-2).* Although these measures produce different types of results, with the *SPM* specifying the type of sensory differences (e.g., sensory seeking behaviors, sensory sensitivities, sensory avoidance, etc.), research shows that when

comparing across the *SPM* and *SSP*, results are most accurate when dichotomizing the variable. Therefore, for this study, the sensory variable was dichotomized for each child to indicate “typical” or “atypical” sensory characteristics.

Clinical Evaluation Considerations

Given the clinical nature of this sample, the types of measures administered as well as availability of information collected via chart review differ among children, resulting in varied n’s per group for different variables. Additionally, the routine administration of the *M-CHAT-R* at this clinic began in 2016, replacing the use of the *M-CHAT*. Therefore, whereas the *CSBS-ITC* was administered to 216 children between 2013 to 2018, the *M-CHAT-R* was administered to only 76 children from 2016 to 2018. Only 75 children received both the *M-CHAT-R* and the *CSBS-ITC*.

DATA ANALYSIS

Statistical analyses were conducted using IBM SPSS Statistics software version 25. To determine the relationship between the screening tools and child and family factors, Pearson correlations (r) were calculated between two continuous variables, point-biserial correlations (r_{pb}) were calculated for one dichotomous and one continuous variable, and χ^2 tests were conducted between two categorical variables. For χ^2 calculations in which one or more expected cell counts were fewer than five children, categories were combined to increase degrees of freedom and the analysis was repeated. If expected cell counts continued to be fewer than five, categorical variables were dichotomized. If one or more expected cell counts contained fewer than five children following dichotomization, Fisher's Exact Test was interpreted, which provides a *p*-value but does not provide the value of a statistic.

T-tests and χ^2 tests were conducted between the ASD and Non-ASD groups to determine whether groups differed on the ASD screening measures, ASD diagnostic measures, and child cognitive, adaptive, language, and sensory characteristics. Levels of agreement between the screening tools, diagnostic measures, and final diagnosis were calculated using Cohen's kappa (κ).

When including results from both the *M-CHAT-R* and the *CSBS-ITC* in the same model, it was necessary to select whether the *CSBS-ITC* Total or *CSBS-ITC* Social score would be more appropriate. The *CSBS-ITC* Social and *CSBS-ITC* Total could not both be

included in the model due to the high correlation and clear multicollinearity between these two variables due to the fact that the *CSBS-ITC* Social is incorporated as part of the *CSBS-ITC* Total score for each child. The *CSBS-ITC* Social was chosen for use over the *CSBS-ITC* Total in models combined with the *M-CHAT-R* Total given that Wetherby (2008) suggested that the *CSBS-ITC* Social would be the best differentiator of children with ASD compared to children with other developmental delays.

To investigate how well the screening instruments predicted the results of the ASD measures as well as final diagnosis when the *M-CHAT-R* and *CSBS-ITC* were considered as independent variables together in the same model, three separate binomial logistic regressions were conducted with the *M-CHAT-R* Total and the *CSBS-ITC* Social as independent variables predicting (1) *ADOS-2* classification, (2) *ADI-R* classification, and (3) final clinical diagnosis. A linear regression was also conducted for the *M-CHAT-R* Total and the *CSBS-ITC* Social as independent variables in the same model with the *ADOS-2* comparison score as the dependent variable. Follow-up, exploratory binomial logistic regressions were completed separately with the *M-CHAT-R* Total, *M-CHAT-R* Decision, *CSBS-ITC* Social, *CSBS-ITC* Total, and *CSBS-ITC* Decision as independent variables in different models predicting each of the dichotomous dependent variables individually: *ADOS-2* classification, *ADI-R* classification, and final clinical diagnosis. Follow-up linear regressions were completed with each independent variable separately (*M-CHAT-R* Total, *M-CHAT-R* Decision, *CSBS-ITC* Social, *CSBS-ITC* Total, and *CSBS-ITC* Decision) to determine how well each variable predicted the continuous *ADOS-2* comparison score. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were also calculated to quantify the accuracy of each of the

screening measures compared to *ADOS-2* classification, *ADI-R* classification and final clinical diagnosis.

To determine the cut-off value which optimized both sensitivity and specificity for the *M-CHAT-R* Total, *CSBS-ITC* Social, and the *CSBS-ITC* Total in predicting the *ASD* diagnostic measure classification as well as final diagnosis, the Youden index was calculated. The Youden index maximizes the distance between the point on the ROC curve for a given cut-off value and the diagonal chance line (Hajian-Tilaki, 2013). Other methods for determining optimal cut-off values were considered; however, the cost of a false positive and a false negative in the use of screening tools in this tertiary care center triage process were determined to be equal, and therefore the use of the Youden Index was most appropriate. In the situation of a false positive, a child receives an ASD-focused evaluation when they do not have ASD, and another child on the waiting list has a longer delay to receive appropriate diagnosis and intervention. The cost of a false negative, in which a child receives a non-ASD evaluation when they do in fact have ASD, means that this child will most likely be referred to a future ASD evaluation by the treating psychologist and will not be completely missed. In each case, appropriate diagnosis and intervention are delayed for a single child. Therefore, sensitivity and specificity were deemed to be equally important for using these screening tools as part of the triage process. ROC curves were produced and presented as visual representations of overall test accuracy and AUC was calculated as an exploratory analysis to compare the average accuracy of the screening tools across all values of the continuous variable.

For all analyses, the *ADOS-2* classification variable includes all *ADOS-2* modules available, including the Toddler Module. When used clinically, a classification for the

ADOS-2 Toddler Module cannot be calculated; instead, levels of concern are given. Therefore, these levels of concern were adapted to parallel the classifications produced in other *ADOS-2* modules such that *little-to-no* concern was coded as *non-ASD*, and *mild-to-moderate* and *moderate-to-severe* levels of concern were coded as *ASD*.

Results were interpreted both with and without the Holm-Bonferroni correction for the inflation of Type 1 error. Extreme outliers were defined as any data point ±3 standard deviations from the mean. Four outliers were identified. Outlier #1 was identified due to the screening tools having been completed when the child was 5.25 years old (M = 2.64, SD = .79, +3SD = 5.01). This child failed both the *M-CHAT-R* and the *CSBS-ITC* screening tools and received a general developmental evaluation at age six years old. The child then completed an ASD evaluation at age 9 and received a non-ASD final diagnosis. For outliers #2 and #3, screening tools were initially completed within the typical age range (#2 = 3.92 years, #3 = 3.67 years); however, there was a significant delay between screening completion and ASD evaluation for both children (#2 = 7.44 years, #3 = 6.36 years respectively), resulting in ASD evaluations completed at much later ages than others (#2 = 11 years old, #3 = 10 years old). Outlier #2 initially passed both screening tools. This child attended a screening clinic evaluation at four years old and received a psychoeducational evaluation at nine years old. Only then was the child referred for an ASD evaluation at 11 years old and received a final non-ASD diagnosis. Outlier #3 initially failed both screening tools, and there was a delay in scheduling this child for an evaluation due to apparent difficulty coordinating interpretation services for this family. The child received a general developmental evaluation at seven years old and was referred for an ASD evaluation at 10 years old. This child ultimately received a final

clinical diagnosis of ASD. Screening tools for outlier #4 were completed at age 2.83 years and the child failed both screening measures, but there was a 6.25-year delay between screening and ASD evaluation. Following initial screening, the clinic had difficulty contacting the family to schedule their appointment. Two years later, another intake was completed with the family although updated screening tools were not completed. The child then received a general developmental evaluation at six years old and completed the ASD evaluation at nine years old. The final diagnosis for this child was a non-ASD diagnosis.

For three of these outliers, the child failed both screening tools but completed a non-ASD evaluation prior to receiving a comprehensive ASD evaluation. The reasons specific to each child are unknown; however, in general, this situation arises when the clinician making clinic assignments reviews the chart and determines that the child failed these screening tools due to a reason other than possible autism spectrum disorder such as significant motor delays, global delays, or complex medical history. All outliers were investigated and determined to be accurate data points; therefore, statistical analyses were completed both with and without the outliers. Results and interpretation did not change following the removal of the outliers; therefore, results are presented with the outliers included in the analyses.

RESULTS

Child and Family Factors

As shown in Tables 4a and 4b, the *M-CHAT-R* (Total and Decision) and *CSBS-ITC* (Social, Total, and Decision) showed strong, significant negative correlations and associations with one another as expected, given that lower *M-CHAT-R* scores suggest fewer ASD symptoms and higher *CSBS-ITC* scores suggest better-developed skills. Raw scores and decisions for both screening tools showed small-to-moderate significant correlations with the *ADOS-2* classification and comparison score. The *ADI-R* classification also had a small significant correlation with the *CSBS-ITC* Social and Total scores but was not associated with the *CSBS-ITC* Decision or *M-CHAT-R* Total or Decision.

Child age at screening had a small significant negative correlation with the *M-CHAT-R* Total score, and age at evaluation was negatively correlated with the *M-CHAT-R* Total and Decision. Child age at screening and evaluation were moderately positively correlated with the *CSBS-ITC* Social and Total raw scores and negatively correlated with the *CSBS-ITC* Decision. These results indicate that parents of younger children indicated higher levels of concern. Cognitive, adaptive, and language scores were moderately positively significantly correlated with *CSBS-ITC* Social and Total raw scores and showed small to moderate negative correlations with the *CSBS-ITC* Decision, indicating that parents and caregivers of children with better global development indicated better

development on the *CSBS-ITC*. Only adaptive scores showed moderate negative correlations with the *M-CHAT-R* whereas there was no association with cognitive or language scores and this measure, suggesting that the *M-CHAT-R* may be less influenced by child cognitive or language skills than the *CSBS-ITC*. Sensory classification was not significantly associated with results of either screening tool.

Children living in households with a higher number of caregivers had moderately lower total scores on the *M-CHAT-R* and a small positive correlation with *CSBS-ITC* Social raw score, indicating that parents and caregivers with more adult support at home rated their children as having fewer ASD symptoms and better social development compared to caregivers in households with fewer caregivers. Other child and family factors were not associated with the results of the screening tools.

Table 4a. Correlations between Screening Tools and Child and Family Factors for Combined Sample

	M-CHAT-R						*CSBS-ITC*								
	Total			Decision			Social			Total			Decision		
	n	r/r_{pb}	*p*	*n*	r/r_{pb}	*p*	*n*	r/r_{pb}	*p*	*n*	r/r_{pb}	*p*	*n*	r/r_{pb}	*p*
M-CHAT-R Total	--	--	--	76	.684	***	75	-.763	***	75	-.734	***	75	.504	***
M-CHAT-R Decision	76	.684	***	--	--	--	75	-.548	***	75	-.502	***	--	--	--
CSBS-ITC Social	75	-.763	***	75	-.548	***	--	--	--	216	.915	***	216	-.573	***
CSBS-ITC Total	75	-.734	***	75	-.502	***	216	.915	***	--	--	--	216	-.607	***
CSBS-ITC Decision	75	.504	***	--	--	--	216	-.573	***	216	.607	***	--	--	--
ADOS-2 Classification	75	.284	*	--	--	--	206	-.318	***	206	-.326	***	--	--	--
ADOS-2 Comp. Score	61	.371	**	61	.355	**	189	-.416	***	189	-.411	***	189	.254	***
ADI-R Classification	53	.250		--	--	--	178	-.150	*	178	-.152	*	--	--	--
Age at Screening (years)	75	-.244	*	76	-.207		216	.388	***	216	.464	***	216	-.264	***
Age at Eval (years)	76	-.271	*	76	-.255	*	216	.309	***	216	.374	***	216	-.264	***
Gender	76	-.145		--	--	--	216	.029		216	.035		--	--	--
IEP (Yes/No)	63	-.119		--	--	--	167	-.014		167	-.021		--	--	--
Gestational Age	76	.144		76	.137		212	.057		212	.073		212	-.017	
Order of Child in Home	74	.205		74	.176		203	-.123		203	-.144	*	203	.098	
Maternal Age at Birth	74	.205		74	.167		211	-.045		211	-.031		211	-.003	
Cognitive Score	39	-.176		39	.016		116	.356	***	116	.379	***	116	-.211	*
Adaptive Score	49	-.386	**	49	-.344	*	121	.377	***	121	.404	***	121	-.327	***
Language Score	74	-.133		74	-.140		203	.333	***	203	.448	***	203	-.282	***
Sensory Classification	30	-.159		--	--	--	68	-.144		68	-.113		--	--	--
Maternal Age at Evaluation	74	.028		74	-.032		211	.063		211	.101		211	-.111	
Maternal Education[ǂ]	70	.067		--	--	--	196	-.059		196	.010		--	--	--
Insurance Type	76	-.080		--	--	--	216	-.022		216	.046		--	--	--
Parent Marital Status	71	-.163		--	--	--	206	.114		206	.087		--	--	--
# of Caregivers in Home	70	-.314	**	70	-.125		142	.186	*	142	.145		142	-.051	
# of Children in Home	76	.070		76	.124		211	-.107		211	-.123		211	.077	

Note: Pearson correlations were calculated for two continuous variables and point-biserial correlations were calculated for one dichotomous and one continuous variable. * = p < .05, ** = p < .01, *** = p < .001. [ǂ]Maternal Education was dichotomized as follows: High School or Less, Some College or More.

Table 4b. Associations between Screening Tools and Child and Family Factors for Combined Sample

	M-CHAT-R Decision			*CSBS-ITC* Decision		
	n	$\chi^2(df)$	*p*	*n*	$\chi^2(df)$	*p*
M-CHAT-R Decision	--	--	--	75	┼	***
CSBS-ITC Decision	75	┼	***	--	--	--
ADOS-2 Classification	75	4.72(1)	*	206	6.44(1)	*
ADI-R Classification	53	.799(1)		178	1.63(1)	
Gender	76	┼		216	4.84(1)	*
IEP (Yes/No)	63	┼		167	.003(1)	
Sensory Classification	30	┼		68	┼	
Maternal Education╪	70	.302(1)		196	.003(1)	
Insurance Type	76	┼		216	.024(1)	
Parent Marital Status	71	1.579 (1)		206	.996(1)	

Note: * = p < .05, ** = p < .01, *** = p < .001. ┼The p-value for Fisher's Exact Test was interpreted. Fisher's Exact Test does not produce a test statistic. ╪Maternal Education was dichotomized as follows: High School or Less, Some College or More

ASD and Non-ASD Group Differences

Mean raw scores on the *CSBS-ITC* for the combined sample indicated Social scores comparable to those of an eight- to nine-month-old child and Total scores similar to those of a ten-month-old child. Interestingly, the ASD and Non-ASD group did not significantly differ on *M-CHAT-R* Total score or *M-CHAT-R* Decision but did show significant differences on the *CSBS-ITC* Social, Total, and Decision, with individuals in the non-ASD group obtaining scores on this measure which suggest better functioning (see Tables 5a and 5b). As predicted, the *ADOS-2* classification and comparison score and the *ADI-R* classification differed between groups with the ASD group having higher levels of ASD symptoms. Cognitive, adaptive, and language scores were better for the Non-ASD group compared to the ASD group, whereas sensory classification did not differ between groups.

Table 5a. Results of Measures Administered During the ASD Evaluation: Continuous Variables

	Total		Final Clinical Diagnosis						
			ASD		Non-ASD				
	n[F]	mean (SD)	n	mean (SD)	n	mean (SD)	*t*	df	p
M-CHAT-R Total	76	7.30 (4.42)	45	7.96 (4.13)	31	6.35 (4.72)	-1.57	74	
CSBS-ITC Social	216	13.57 (5.78)	128	12.48 (5.56)	88	15.17 (5.75)	3.45	214	**
CSBS-ITC Total	216	29.99 (11.97)	128	27.72 (11.20)	88	33.30 (12.33)	3.45	214	**
ADOS-2 Comparison Score	189	5.54 (3.24)	112	7.86 (1.71)	77	2.17 (1.51)	23.53	187	***
Cognitive Score	117	76.92 (18.03)	50	71.32 (17.81)	67	81.10 (17.16)	-3.00	115	**
Adaptive Score	122	65.61 (13.39)	71	62.34 (13.69)	51	70.18 (11.61)	-3.32	120	**
Language Score	203	65.96 (17.44)	123	57.82 (12.67)	80	78.46 (16.39)	-9.56	139	***

Note: Cognitive, adaptive, and language scores are presented as standard scores with a mean of 100 and a standard deviation of 15. [F] N varies between groups due to unknown and missing data which was not available via chart review and data extraction. * = p < .05, ** = p < .01, *** = p < .001

Table 5b. Results of Measures Administered During the ASD Evaluation: Categorical Variables

	Total		Final Clinical Diagnosis						
			ASD		Non-ASD				
	n[╞]	%	n	%	n	%	χ^2	df	p
Final Clinical Diagnosis	217	--	129	59	88	41	--	--	--
M-CHAT-R Decision	76		45		31		2.80	1	
Pass	13	17	5	11	8	26			
Fail	63	83	40	89	23	74			
CSBS-ITC Decision	216		128		88		5.30	1	*
Pass	26	12	10	8	16	18			
Fail	190	88	118	92	72	82			
ADOS-2 Classification	188		112		76		165.29	2	***
Autism	105	56	105	94	0	0			
Autism Spectrum	17	9	5	4	12	16			
Non-Spectrum	66	35	2	2	64	84			
ADOS-2 Toddler Module Level of Concern	19		11		8				
Moderate-to-Severe	10	53	10	91	0	0	†	†	†
Mild-to-Moderate	2	11	1	9	1	13			
Little-to-No	7	37	0	0	7	88			
ADOS-2 Classification (dichotomized)	207		123		84		150.25	1	***
ASD	134	65	121	98	13	15			
Non-ASD	73	35	2	2	71	85			
ADI-R Classification	178		113		65		67.31	1	***
Autism	104	58	92	81	12	18			
Not Autism	74	42	21	19	53	82			
Sensory Classification	68		40		28				
Typical	12	18	6	15	6	21	┼	┼	
Atypical	56	82	34	85	22	79			

Note: Percentages may not add up to 100 due to rounding. [╞] N varies between groups due to unknown and missing data which was not available via chart review and data extraction. † χ^2 could not be calculated due to cells with predicted counts <5. [┼]The p-value for Fisher's Exact Test was interpreted. Fisher's Exact Test does not produce a test statistic. * = p < .05, ** = p < .01, *** = p < .001.

Agreement

As shown in Tables 6 and 7, the *M-CHAT-R* Decision and the *CSBS-ITC* Decision were in agreement for 87% of children and prior to correction had moderate agreement with each other. The *M-CHAT-R* Decision had fair agreement with the *ADOS-2* classification. Low but statistically significant agreement was shown between the *CSBS-ITC* Decision and the *ADOS-2* classification as well as between the *CSBS-ITC* Decision and final diagnosis. The *M-CHAT-R* Decision and *CSBS-ITC* Decision did not agree with the *ADI-R* classification over and above chance levels. Following correction for Type 1 error, agreement between the *M-CHAT-R* Decision and *CSBS-ITC* Decision as well as the *CSBS-ITC* Decision and the *ADOS-2* Classification continued to be significant; however, other results did not survive correction.

Table 6. Screening Tool Decisions

	M-CHAT-R Decision		
	Pass	Fail	Total
	n	n	n
CSBS-ITC Decision			
Pass	7	4	11
Fail	6	58	64
Total	13	62	75

Table 7. Agreement Between Measures

	M-CHAT-R Decision		*CSBS-ITC* Decision	
	n	κ	n	κ
CSBS-ITC Decision	75	.505***	--	--
ADOS-2 Classification	75	.212*	206	.144*
ADI-R Classification	53	.094	178	.069
Final Diagnosis	76	.162	216	.117*

Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$ prior to correction for Type 1 error

Predicting ASD Diagnostic Measures and Final Diagnosis

Logistic and linear regressions with the *M-CHAT-R* Total and *CSBS-ITC* Social both included as independent variables in the same model resulted in significant models when predicting the *ADOS-2* classification [$\chi^2(2) = 7.559$, $p < .05$], *ADOS-2* comparison score [$F(2,58) = 6.515$, $p < .005$], and *ADI-R* classification [$\chi^2(2) = 6.566$, $p < .05$] but not the final clinical diagnosis ($p > .05$). However, there were no significant individual variables in these models (all p's > .05). Although no multicollinearity was found between the *M-CHAT-R* Total and the *CSBS-ITC* Social (all variance inflation factors (VIFs) < 2.5 whereas a value of five to 10 indicates multicollinearity), a significant overall model with no significant individual predictors can occur when the predictors are correlated ($r = -.763$, $p < .001$). This does not necessarily indicate that there are no significant predictors, but instead indicates that one predictor does not show a significant effect over and above the effect of the other predictor. Therefore, to evaluate the effect of the *M-CHAT-R* and the *CSBS-ITC* individually, regressions with only one independent variable were interpreted.

Prior to correction, the *M-CHAT-R* Total and *M-CHAT-R* Decision significantly predicted the *ADOS-2* classification and *ADOS-2* comparison score but not the *ADI-R* classification or the final clinical diagnosis. The *CSBS-ITC* Social and *CSBS-ITC* Total significantly predicted the *ADOS-2* classification, *ADOS-2* comparison score, *ADI-R* classification, and final diagnosis. The *CSBS-ITC* Decision was a significant predictor of the *ADOS-2* classification, comparison score, and final diagnosis, but not the *ADI-R* classification.

Table 8a. Screening Instruments Predicting ASD Measure Classifications and Final Diagnosis

	ADOS-2 Classification				
	n	χ^2	*df*	*p*	R^2
M-CHAT-R Total	75	6.242	1	.012	.108
M-CHAT-R Decision	75	4.608	1	.032	.081
CSBS-ITC Social	206	21.378	1	.000	.135
CSBS-ITC Total	206	22.600	1	.000	.143
CSBS-ITC Decision	206	6.147	1	.013	.040
	ADI-R Classification				
M-CHAT-R Total	53	3.401	1	.065	.084
M-CHAT-R Decision	53	.799	1	.371	.020
CSBS-ITC Social	178	4.035	1	.045	.030
CSBS-ITC Total	178	4.159	1	.041	.031
CSBS-ITC Decision	178	1.646	1	.200	.012
	Final Diagnosis				
M-CHAT-R Total	76	2.469	1	.116	.043
M-CHAT-R Decision	76	2.750	1	.097	.048
CSBS-ITC Social	216	11.566	1	.001	.070
CSBS-ITC Total	216	11.565	1	.001	.070
CSBS-ITC Decision	216	5.195	1	.023	.032

Note: *p*-values prior to correction for Type 1 error. R^2 = adjusted R^2

Table 8b. Screening Instruments Predicting *ADOS-2* Comparison Score

	ADOS-2 Comparison Score				
	n	*F*	*df*	*p*	R^2
M-CHAT-R Total	61	9.410	1,59	.003	.123
M-CHAT-R Decision	61	8.510	1,59	.005	.111
CSBS-ITC Social	189	39.036	1,187	.000	.168
CSBS-ITC Total	189	38.107	1,187	.000	.165
CSBS-ITC Decision	189	12.853	1	.000	.059

Note: *p*-values prior to correction for Type 1 error. R^2 = Nagelkerke's R^2.

Table 9a. Strength of Screening Instruments Predicting ASD Measure Classifications and Final Diagnosis

	B	SE	Wald	*df*	*p*	Exp(B)
	ADOS-2 Classification					
M-CHAT-R Total	.143	.060	5.677	1	.017	1.153
M-CHAT-R Decision	1.386	.668	4.305	1	.038	4.000
CSBS-ITC Social	-.123	.028	19.111	1	.000	.884
CSBS-ITC Total	-.061	.014	20.025	1	.000	.941
CSBS-ITC Decision	1.054	.428	6.067	1	.014	2.868
	ADI-R Classification					
M-CHAT-R Total	.119	.067	3.172	1	.075	1.126
M-CHAT-R Decision	.730	.821	.789	1	.374	2.074
CSBS-ITC Social	-.027	0.13	4.060	1	.044	.973
CSBS-ITC Total	-.027	0.13	4.060	1	.044	.973
CSBS-ITC Decision	.611	.478	1.637	1	.201	1.843
	Final Diagnosis					
M-CHAT-R Total	.085	.055	2.382	1	.123	1.089
M-CHAT-R Decision	1.023	.627	2.662	1	.103	2.783
CSBS-ITC Social	-.084	.025	10.937	1	.001	.920
CSBS-ITC Total	-.041	.012	10.923	1	.001	.960
CSBS-ITC Decision	.964	.430	5.027	1	.025	2.622

Note: *p* values prior to correction for Type 1 error, B = beta, SE = standard error.

Table 9b. Strength of Screening Instruments Predicting *ADOS-2* Comparison Score

	B	SE	*t*	*p*
	ADOS-2 Comparison Score			
M-CHAT-R Total	.281	.092	3.068	.003
M-CHAT-R Decision	3.104	1.064	2.917	.005
CSBS-ITC Social	-.234	.037	-6.248	.000
CSBS-ITC Total	-.111	.018	-6.173	.000
CSBS-ITC Decision	2.420	.675	3.585	.000

Note: *p* values prior to correction for Type 1 error, SE = standard error.

Following correction, the *M-CHAT-R* Total and *M-CHAT-R* Decision continued to significantly predict the *ADOS-2* comparison score but not the *ADOS-2* classification. The *CSBS-ITC* Social and *CSBS-ITC* Total results remained significant when predicting the *ADOS-2* classification, *ADOS-2* comparison score, and final diagnosis, but neither measure was significant following correction as a predictor of the *ADI-R* classification.

The *CSBS-ITC* Decision no longer predicted the *ADOS-2* classification or final diagnosis after correction but continued to significantly predict the *ADOS-2* comparison score.

Table 10. Accuracy Statistics of Screening Tools Compared with ASD Measures Classification and Final Diagnosis

	M-CHAT-R Decision				*CSBS-ITC* Decision			
	Se	Sp	PPV	NPV	Se	Sp	PPV	NPV
ADOS-2 Classification	.91	.28	.67	.67	.92	.21	.68	.58
ADI-R Classification	.90	.18	.61	.57	.91	.15	.60	.55
Final Diagnosis	.89	.26	.63	.62	.92	.18	.62	.62
	Fail Both[1]				Fail One or Both[2]			
ADOS-2 Classification	.87	.36	.67	.63	.96	.17	.64	.71
ADI-R Classification	.87	.27	.63	.60	1.00	.00	.58	[3]
Final Diagnosis	.84	.32	.64	.59	.96	.16	.62	.71

Se = sensitivity, Sp = specificity, PPV = positive predictive value, NPV = negative predictive value. [1]Failing both the *M-CHAT-R* and the *CSBS-ITC* was considered a "Fail." Passing one or both was considered a "Pass." [2]Failing the *M-CHAT-R* and/or the *CSBS-ITC* was considered a "Fail." Passing both measures was considered a "Pass." [3]Could not be calculated due to denominator of zero.

As shown in Table 10, in general, the accuracy of screening measures at their published cut-off values in predicting results of more comprehensive diagnostic assessments as well as final clinical diagnosis were comparable across measures, with both the *M-CHAT-R* Decision and the *CSBS-ITC* Decision displaying high sensitivity (*M-CHAT-R* Decision: .89 to .91, *CSBS-ITC* Decision: .91 to .92) and low specificity (*M-CHAT-R* Decision: .18 to .28, *CSBS-ITC* Decision: .15 to .21). If failing both measures was required, specificity increased (.27 to .36) but at the cost of sensitivity (.84 to .87). Failing one or both measures improved sensitivity (.96 to 1.00) but decreased specificity (.00 to .17).

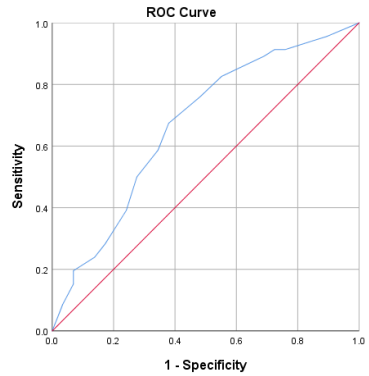

Figure 1. ROC curve of the *M-CHAT-R* Total and *ADOS-2* classification.

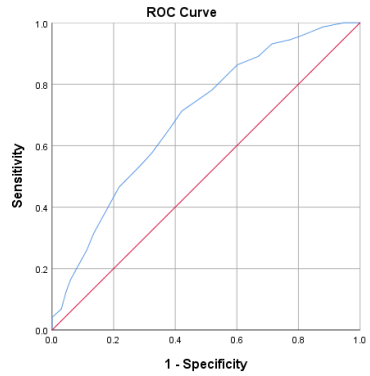

Figure 2. ROC curve of the *CSBS-ITC* Social and *ADOS-2* classification.

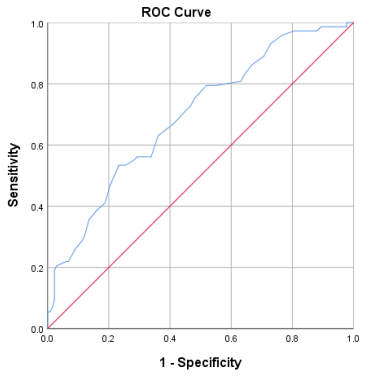

Figure 3. ROC curve of the *CSBS-ITC* Total and *ADOS-2* classification.

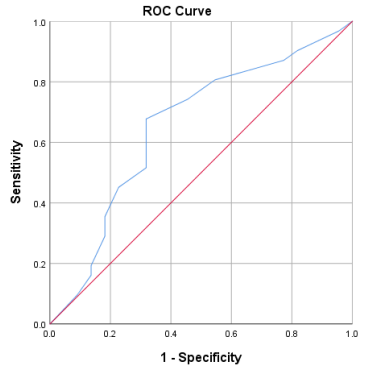

Figure 4. ROC curve of the *M-CHAT-R* Total and *ADI-R* classification.

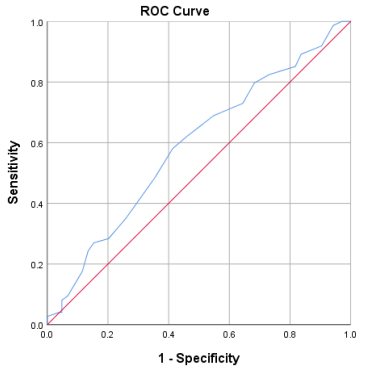

Figure 5. ROC curve of the *CSBS-ITC* Social and *ADI-R* classification.

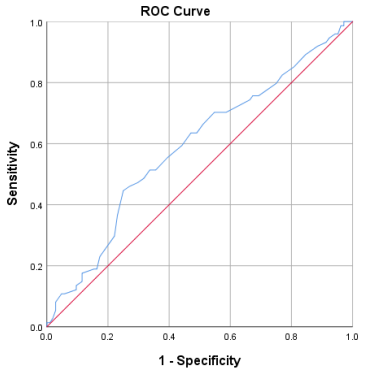

Figure 6. ROC curve of the *CSBS-ITC* Total and *ADI-R* classification.

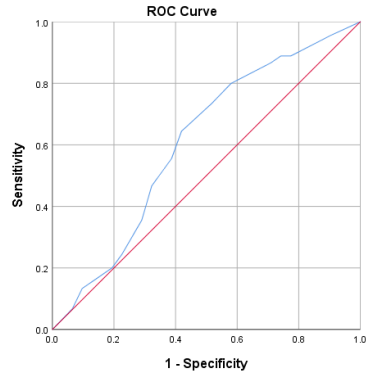

Figure 7. ROC curve of the *M-CHAT-R* Total and final diagnosis.

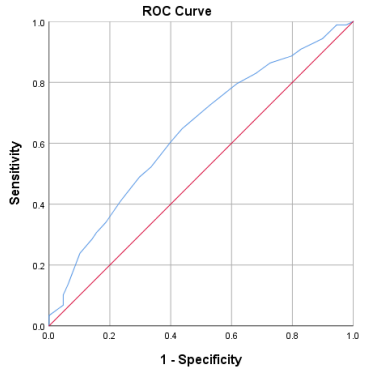

Figure 8. ROC curve of the *CSBS-ITC* Social and final diagnosis.

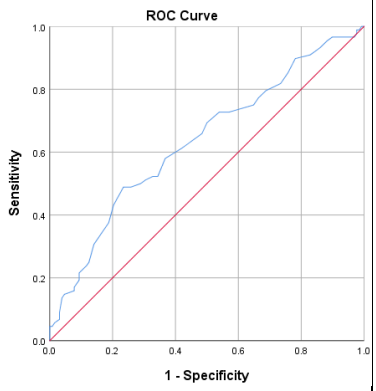

Figure 9. ROC curve of the *CSBS-ITC* Total and final diagnosis.

Table 11. Area Under the Curve for Screening Measures Predicting ASD Measures and Final Diagnosis

	M-CHAT-R Total		*CSBS-ITC* Social		*CSBS-ITC* Total	
	AUC	*p*	AUC	*p*	AUC	*p*
ADOS-2 Classification	.668	*	.692	***	.693	***
ADI-R Classification	.654		.588	*	.589	*
Final Diagnosis	.608		.636	**	.634	**

Note: AUC = Area Under the ROC Curve. * $p < .05$, ** $p < .01$, *** $p < .001$.

When considering the average accuracy all continuous values of the screening tools in predicting the ASD diagnostic measures and the final diagnosis, the *CSBS-ITC* Total and *CSBS-ITC* Social were comparable in predicting the *ADOS-2* classification. The *M-CHAT-R* Total was the best predictor of the *ADI-R* across values, and the *CSBS-ITC* Social and *CSBS-ITC* Total predicted the final diagnosis more accurately than the *M-CHAT-R* Total score across all values on average.

Table 12. Cut-Off Values Maximizing Accuracy of Screening Tools Predicting ASD Diagnostic Measures and Final Diagnosis

	M-CHAT-R Total				*CSBS-ITC* Social				*CSBS-ITC* Total			
	Cut-off	Se	Sp	Youden	Cut-off	Se	Sp	Youden	Cut-off	Se	Sp	Youden
ADOS-2 Classification	7	.67	.62	.295	13	.71	.58	.291	37	.53	.78	.301
ADI-R Classification	7	.68	.68	.359	14	.58	.58	.168	37	.45	.75	.196
Final Diagnosis	7	.64	.58	.225	13	.65	.56	.210	37	.49	.77	.255

Note: Se = sensitivity; Sp = specificity.

Based on the calculation of Youden's Index at all possible cut-off values for each measure, the values that maximize sensitivity and specificity in predicting the outcomes of the *ASD* diagnostic tools as well as the final clinical diagnosis were a cut-off score of 7 for the *M-CHAT-R*, 13 or 14 for the *CSBS-ITC* Social, and 37 for the *CSBS-ITC* Total (Table 12).

DISCUSSION

The purpose of this study was to investigate the clinical utility of the *M-CHAT-R* and *CSBS-ITC* screening tools in predicting ASD diagnosis as well as the impact of child and family factors on these measures to inform the triage and clinic assignment process.

Child Factors

Children with younger age at screening and evaluation were likely to have higher levels of reported concern on the screening measures. It may be the case that children with more significant symptoms were referred for an evaluation at a younger age, whereas children with milder symptoms may not be identified until later. Higher cognitive, adaptive, and language skills were associated with better development reported on all *CSBS-ITC* results, which is expected given that this is a measure of language and social abilities and is not ASD-specific. The *M-CHAT-R* was only significantly associated with adaptive skills and not with cognitive and language abilities. This may be due to the language scores demonstrating a floor effect, as 35% of the children with available language testing results received a score of 50, the lowest standard score possible on the Preschool Language Scales. Cognitive scores were normally distributed and may not have had an impact on *M-CHAT-R* scores. Ideally, screening measures for ASD would be unaffected by child cognitive, adaptive, and language abilities and would capture solely those children who demonstrate ASD symptoms regardless of other skill deficits.

However, given that the *M-CHAT-R* is an ASD-specific screening measure and children with ASD have been reported to have lower adaptive behavior skills compared with children with other developmental disorders, this association is expected (Hill et al., 2015). The cognitive, adaptive, and language scores were better for the Non-ASD group compared to the ASD group, which reflects differences reported in the literature (Hill et al., 2015; Paul, Chawarska, Cicchetti, & Volkmar, 2008). This suggests that when utilizing these screening measures as part of a triage process in a tertiary care setting to determine whether an ASD evaluation is appropriate, when screening measures indicate possible ASD and if results from prior testing are available, lower skill development in other areas may be taken into account as indicative of possible concern for ASD and would warrant an ASD evaluation.

The ASD and Non-ASD groups did not differ on whether parents reported typical or atypical sensory sensitivities. Specifically, in this clinic, sensory measures are not routinely given, and this measure is only used for children with reported sensory differences. However, sensory scores were available for a similar proportion of each diagnostic group with 31% of children in the ASD group and 32% of the Non-ASD group receiving a sensory measure. This may indicate that overall about one-third of children referred for an ASD evaluation at this clinic demonstrate sensory sensitivities regardless of final clinical diagnosis. However, due to this clinic process, this result is influenced by low sample size and a bias in reporting on this measure, given that there was an overall low level of “Typical” compared with “Atypical” sensory classifications.

Family Factors

Regarding family factors, households with a greater number of caregivers in the home reported lower levels of concern on the screening measures. This may be due to single-parent households experiencing higher levels of stress or greater difficulty with behavior management at home (Rezendes & Scarpa, 2011), resulting in a greater level of concern on the screening tools. Clinicians should be aware of the possible impact of family functioning on the results of these screening tools and future research should investigate the intersection and associations among family functioning, caregiver stress, and child behavior and development, specifically in regards to the effects on the results of screening and evaluation for their children.

ASD and Non-ASD groups differed significantly on maternal age at birth and maternal age at evaluation, with the ASD group having higher maternal age at birth and maternal age at evaluation compared to the Non-ASD group. This is consistent with the literature indicating that children born to mothers of higher age are at greater risk for the development of ASD (Croen et al., 2007; Lauritsen et al., 2005; Parner et al., 2012).

Group Differences

Contrary to expectations, ASD and Non-ASD groups did not differ on the *M-CHAT-R* Total score or Decision. This may be due to selection bias in the sample included, with the majority of children who received an ASD evaluation having moderate-to-high levels of concern on the *ADOS-2* and failing the *M-CHAT-R*. This may also be due to low power due to sample size, as the mean Total score was lower for the Non-ASD group than the ASD group. However, the effect size for the *M-CHAT-R* Total

Raw score difference between groups (Cohen's $d = 0.36$) suggests that the group difference between the ASD and non-ASD groups was small within this sample. The *CSBS-ITC* Social, Total, and Decision differed between the ASD and Non-ASD groups with the Non-ASD group having higher scores than the ASD group. This indicates that the Non-ASD group had better development in the areas of communication, language, and social abilities as indicated on this screening measure, and provides further evidence for the clinical utility of this measure within this high-risk referred sample.

Agreement and Predictors

The screening measures were associated with one another, as predicted, and the *M-CHAT-R* and *CSBS-ITC* Decisions were in agreement 87% of the time. Both the *M-CHAT-R* Decision and the *CSBS-ITC* Decision showed low-to-fair significant agreement with the *ADOS-2* Classification, and the *CSBS-ITC* Decision had low but significant agreement with the final diagnosis, but neither measure showed agreement with the *ADI-R* Classification. Similar results were found with the screening tools in predicting the ASD measures and final diagnosis. Following correction, the *ADOS-2* comparison score was significantly predicted by all screening variables for both the *M-CHAT-R* and the *CSBS-ITC*. Additionally, the *CSBS-ITC* Social and Total continued to predict the *ADOS-2* classification and final diagnosis. Although the *CSBS-ITC* Social and Total showed small significant negative correlations with the *ADI-R* classification, in general the screening measures did not agree with or predict the *ADI-R* classification. Overall, both screening tools were good predictors of the *ADOS-2* but only the *CSBS-ITC* could be used to predict final diagnosis. This suggests that although the *M-CHAT-R* mapped on

well to the *ADOS-2*, the *CSBS-ITC* performed better in predicting the ASD diagnostic measures and the final clinical diagnosis compared to the *M-CHAT-R* and therefore the *CSBS-ITC* may have better clinical utility for the purposes of use in the clinic assignment process in tertiary care centers conducting ASD evaluations.

Accuracy of Screening Tools

When using the clinical published cut-off scores, both screening tools showed high sensitivity and low specificity in predicting the results of the *ASD* diagnostic measures as well as the overall final clinical diagnosis, which is consistent with the original purpose of screening tools, as so many were failed. However, these cut-off values may not be the ideal values for use in a tertiary care setting. For example, when utilizing the *M-CHAT-R* in a primary care setting, it is recommended that the *M-CHAT-R* Follow-Up Interview is conducted for a score between three and seven, indicating that a follow-up interview should be conducted to clarify item responses. However, if a child scores an eight or above, they should be automatically referred for an ASD evaluation. For the *CSBS-ITC*, it is recommended that any child scoring at least one standard deviation below the mean be administered the follow-up caregiver questionnaire and/or the behavioral observation for further symptom clarification. Results from the accuracy of the *M-CHAT-R* in this clinical sample indicate that a value of seven would maximize accuracy for ASD diagnostic measures and final diagnosis. Results indicated that a cut-off score of 13 or 14 for the *CSBS-ITC* Social and 37 for the *CSBS-ITC* Total would maximize accuracy of these measures in this sample at risk for ASD. This may indicate that for periods of time during which the wait-list for an ASD evaluation is becoming

increasingly large, these more conservative cut-off scores could be used to prioritize evaluating children who likely have higher levels of ASD symptoms and therefore would benefit more from early access to interventions, services, and supports. Additionally, exceptionally high or low scores on these measures may indicate that the child demonstrates a clearer pattern of symptoms which may be able to be diagnosed from a shorter evaluation, which would allow a greater number of evaluations to be conducted and more children would be able to access services.

Generalizability of Sample and Diagnostic Process

Results indicated that the ASD and non-ASD groups differed on gestational age by an average of one week (ASD: M = 38.37, SD = 2.78 Non-ASD: M = 37.35, SD = 3.41). According to the National Institute of Child Health and Human Development, children born following 37 weeks through 38 weeks and 6 days gestation are considered "early term" and are 5% more likely to have an intellectual or developmental disability compared to those born after 39 weeks gestation (Spong, 2013). However, given that the non-ASD children in this sample were not typically developing and instead represent a wide range of non-ASD disorders, this observed group difference in the current study may not be clinically relevant for guiding screening and diagnosis.

Additionally, all psychologists at this tertiary care clinic who scored and interpreted the ASD diagnostic measures were highly trained and achieved research reliability on these measures. In addition, this site is a member of the University Centers for Excellence in Developmental Disabilities Education, Research, and Service (UCEDD) and Leadership Education in Neurodevelopmental and Related Disabilities

(LEND) programs, providing training at the graduate level in ASD assessment. Therefore, the highly trained clinicians in this clinic may contribute to the high levels of accuracy of these diagnostic measures and may influence the way the screening tools relate to these measures. This may not reflect other community clinics at which diagnosticians have been trained clinically but not at the level of research reliability or those without significant ASD expertise. The field would benefit from future research using similar methodology investigating the accuracy of these screening measures in clinical settings which may be more representative of general community, such as with practitioners who have significant clinical expertise and experience working with children with ASD but have not been trained to the level of research reliability (Molloy et al., 2011).

Of note, the final clinical diagnosis rendered by this multidisciplinary team was not validated by another outside institution, which is a limitation. Furthermore, although many different clinical psychologists administered the *ADOS-2* (n = 6) and the *ADI-R* (n = 5), one lead clinical psychologist (SO) conducted or supervised administration of 41% of all *ADOS-2* evaluations and 61% of all *ADI-R* administrations at this clinic. When evaluating the accuracy of diagnostic measures, it may be prudent to confirm the clinical diagnosis for at least a subset of children to measure agreement between the diagnostic outcomes in different clinics. This would include the involvement of an independent outside institution to confirm the result of the evaluation for reliability purposes to improve confidence in the accuracy of the final clinical diagnosis.

Limitations

Many of the variables utilized in this study were measures regarding the biological mother. However, oftentimes the person accompanying the child to the clinical evaluation, completing the screening questionnaires, and participating in the clinical interview was not the biological mother. Information was not available in the database regarding the relationship of the person who completed the clinical interview to the child, and the relationship of the person who completed the screening tools was not investigated. Therefore, there is an inherent limitation in relating maternal characteristics to the results of the screening measures given that the reporter was not always the biological mother.

Additionally, previous research has indicated the relationship between advanced paternal age and increased risk for ASD (Hultman et al., 2011); however, paternal characteristics were not evaluated for this study. Future research should investigate whether paternal characteristics are related to ASD screening, evaluation, and diagnosis.

An inherent flaw with this dataset is that race and ethnicity were not directly reported by the child or family and were instead rated by the clinician based on their best estimate. The authors are aware that this method of collecting this variable would result in an inaccurate measure of race and ethnicity, and therefore further analysis of this variable was not completed. However, race and ethnicity as well as other cultural and identity characteristics have routinely been excluded from the research literature regarding ASD and other developmental disorders. Consideration of these variables is of the utmost importance to better understand these populations, particularly those who identify as minorities. It is recommended that all clinics regularly collect patient-reported

information regarding race, ethnicity, and identity, whether as part of the initial intake paperwork or asked by clinicians during the clinical interview. This information can then be entered into a clinical database for use in future research.

Additionally, given the nature of retrospective chart review of clinical evaluations, children were administered different measures according to clinical need resulting in missingness that is not random. Different sample sizes per group and within measures must be taken into account when interpreting the results of this study, particularly when evaluating results regarding the effects of child and family factors and non-ASD measures administered including cognitive, adaptive, language, and sensory measures.

Many children whose caregivers completed these screening measures prior to evaluation were not included in this sample. This includes children who passed one or both screeners and for whom there was not a caregiver or provider concern. Conversely, most of the children who were included in this sample and ultimately received a comprehensive ASD evaluation failed one or both of the screening measures. Therefore, all results must be interpreted within the broader context of assessment and diagnosis of ASD and may not be representative of a general referred sample with other developmental delays.

Future Directions and Clinical Recommendations

Overall, results provide evidence for the ability of the *CSBS-ITC* to predict ASD diagnosis, and therefore suggests consideration of clinical utility in the triage process for tertiary care clinics. However, interestingly, results for the ASD-focused screener, the *M-*

CHAT-R were mixed. Although sensitivity of this measure was high, specificity was low, and many children in both the ASD and Non-ASD groups failed the *M-CHAT-R.* Item analysis of the *M-CHAT-R* would be useful to determine whether certain items are failed in this high-risk sample that do not necessarily indicate a diagnosis of ASD but instead are related to delays in other areas such as global developmental delays, language delays, or motor delays. If so, a higher threshold for the *M-CHAT-R* may be more useful in this referred sample for the purposes of informing clinical procedures. Additionally, some analyses may have had low power due to small sample size, since the *M-CHAT-R* is a newer measure and has only been in use at this clinic for a few years. Further investigation into the clinical utility of the *M-CHAT-R* should be conducted when a greater number of children who received the *M-CHAT-R* screening measure have been evaluated.

It may be prudent to introduce a two-step triage and clinic assignment process for those children whose screening measures and initial intake paperwork indicate unclear ASD concern. This may include conducting the follow-up interview for the *M-CHAT-R* to clarify questions of development and ASD concern prior to clinic assignment. Clinicians may also consider using the *CSBS Developmental Profile*, including the additional caregiver questionnaire and/or the behavioral observation to gain more information about the child's development prior to completing the full evaluation.

Conclusion

Results indicate that the *CSBS-ITC* and the *M-CHAT-R* likely have clinical utility in the triage process in tertiary care settings. Given the time-consuming nature of ASD

assessments and the limited number of professionals qualified to conduct these types of evaluations, it is practical to utilize these ASD screening measures as part of a triage process to reserve ASD evaluations only for those children for whom it would be truly beneficial. Implementing the routine use of screening tools in clinical practice for young children will help reduce the delay between referral, evaluation, and diagnosis, and will help children receive the ASD-specific services they need.

REFERENCES

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., … others. (2018). Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveillance Summaries*, *67*(6), 1–23.

Dorlack, T. P., Myers, O. B. & Kodituwakku, P. W. (2018). A comparative analysis of the ADOS-G and ADOS-2 algorithms: preliminary findings. *Journal of Autism and Developmental Disorders*, 1–12.

Dumont-Mathieu, T. & Fein, D. (2005). Screening for autism in young children: The Modified Checklist for Autism in Toddlers (M-CHAT) and other measures. *Mental Retardation and Developmental Disabilities Research Reviews*, *11*(3), 253–262.

Falkmer, T., Anderson, K., Falkmer, M. & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: a systematic literature review. *European Child & Adolescent Psychiatry*, *22*(6), 329–40. doi:10.1007/s00787-013-0375-0

Gotham, K., Risi, S., Dawson, G., Tager-Flusberg, H., Joseph, R., Carter, A., … Lord, C. (2008). A replication of the Autism Diagnostic Observation Schedule (ADOS) revised algorithms. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(6), 642–51. doi:10.1097/CHI.0b013e31816bffb7

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine, 4*(2), 627-635.

Hill, T. L., Gray, S. A., Kamps, J. L. & Varela, R. E. (2015). Age and adaptive functioning in children and adolescents with ASD: The effects of intellectual functioning and ASD symptom severity. *Journal of Autism and Developmental Disorders, 45*(12), 4074–4083.

Johnson, C. P., Myers, S. M. & others. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics*, *120*(5), 1183–1215.

Kamp-Becker, I., Ghahreman, M., Heinzel-Gutenbrunner, M., Peters, M., Remschmidt, H. & Becker, K. (2011). Evaluation of the revised algorithm of Autism Diagnostic Observation Schedule (ADOS) in the diagnostic investigation of high-functioning children and adolescents with autism spectrum disorders. *Autism*, *17*(1), 87–102.

Khowaja, M., Robins, D. L. & Adamson, L. B. (2018). Utilizing two-tiered screening for early detection of autism spectrum disorder. *Autism, 22*(7), 881–890.

Kim, S. H. & Lord, C. (2012). New autism diagnostic interview-revised algorithms for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Autism and Developmental Disorders, 42*(1), 82–93.

Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T. & Manning-Courtney, P. (2011). Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism : The International Journal of Research and Practice, 15*(2), 143–62. doi:10.1177/1362361310379241

Oosterling, I. J., Roos, S., de Bildt, A., Rommelse, N., de Jonge, M., Visser, J., … Buitelaar, J. (2010). Improved diagnostic validity of the ADOS revised algorithms: a replication study in an independent sample. *Journal of Autism and Developmental Disorders, 40*(6), 689–703. doi:10.1007/s10803-009-0915-0

Oosterling, I. J., Swinkels, S. H., van der Gaag, R. J., Visser, J. C., Dietz, C. & Buitelaar, J. K. (2009). Comparative analysis of three screening instruments for autism spectrum disorder in toddlers at high risk. *Journal of Autism and Developmental Disorders*, *39*(6), 897–909.

Oswald, D. P., Haworth, S. M., Mackenzie, B. K. & Willis, J. H. (2017). Parental report of the diagnostic process and outcome: ASD compared with other developmental disabilities. *Focus on Autism and Other Developmental Disabilities, 32*(2), 152–160.

Paul, R., Chawarska, K., Cicchetti, D. & Volkmar, F. (2008). Language outcomes of toddlers with autism spectrum disorders: A two year follow-up. *Autism Research, 1*(2), 97–107.

Rezendes, D. L. & Scarpa, A. (2011). Associations between parent anxiety/depression and child behavior problems related to autism spectrum disorders: the roles of parenting stress and parenting self-efficacy. *Autism Research and Treatment, 2011,* 1-10. doi:10.1155/2011/395190

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., … Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*(9), 1094–103.

Robins, D. L., Casagrande, K., Barton, M., Chen, C.-M. A., Dumont-Mathieu, T. & Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*, *133*(1), 37–45.

Robins, D. L., Fein, D. & Barton, M. (2009). Modified checklist for autism in toddlers, revised, with follow-up (M-CHAT-R/F) TM. Self-published. www.mchatscreen.com.

Rutter, M., Le Couteur, A., Lord, C. & others. (2003). *Autism diagnostic interview-revised.* Los Angeles, CA: Western Psychological Services.

Towle, P. O. & Patrick, P. A. (2016). Autism spectrum disorder screening instruments for very young children: a systematic review. *Autism Research and Treatment*, 1–29.

Wetherby, A. M. & Prizant, B. M. (2003). *Communication and symbolic behavior scales (CSBS)*. Baltimore, MD: Brookes Publishing Company.

Zander, E., Sturm, H. & Bölte, S. (2015). The added value of the combined use of the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule: Diagnostic validity in a clinical Swedish sample of toddlers and young preschoolers. *Autism*, *19*(2), 187–199.

CONCLUSION

Overall, these studies investigated the measures used for ASD diagnosis and screening. When conducting ASD diagnostic evaluations, the *ADOS-2* performs well in clinical settings and should be considered a critical part of any *ASD* evaluation. Including the *ADI-R* in an ASD evaluation should be considered in more complex cases. These measures performed well in a clinic setting when administered by highly trained clinicians, and it is recommended that clinicians interested in evaluating children with a question of possible ASD seek out adequate training and experience prior to conducting ASD diagnostic evaluations independently. When considering implementing the use of screening tools to optimize clinic assignment in tertiary care settings, the *CSBS-ITC* is likely to be a useful addition to a screening and triage process for young children prior to conducting a comprehensive evaluation. Continued investigations into best practices for optimizing and streamlining the screening and diagnostic process for children with ASD will continue to improve access to early ASD interventions allow these children the best chance to reach their greatest potential.

REFERENCES

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

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., … others. (2018). Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveillance Summaries*, *67*(6), 1–23.

Beaudoin, A. J., Sébire, G. & Couture, M. (2014). Parent training interventions for toddlers with autism spectrum disorder. *Autism Research and Treatment, 2014,* 1-15.

Bryson, S. E., Rogers, S. J. & Fombonne, E. (2003). Autism spectrum disorders: early detection, intervention, education, and psychopharmacological management. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *48*(8), 506–16.

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T. & Baird, G. (2011). IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological Medicine*, *41*(3), 619–627.

Clark, M. L. E., Vinen, Z., Barbaro, J. & Dissanayake, C. (2018). School age outcomes of children diagnosed early and later with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *48*(1), 92–102.

Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J., … Wheelwright, S. (1999). Autism spectrum disorders at 20 and 42 months of age: stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *40*(5), 719–32.

Croen, L. A., Najjar, D. V., Fireman, B. & Grether, J. K. (2007). Maternal and paternal age and risk of autism spectrum disorders. *Archives of Pediatrics & Adolescent Medicine*, *161*(4), 334–40.

De Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F. & Minderaa, R. (2004). Interrelationship between autism diagnostic observation schedule-generic (ADOS-G), autism diagnostic interview-revised (ADI-R), and the diagnostic and statistical manual of mental disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders*, *34*(2), 129–137.

De Bildt, A., Sytema, S., van Lang, N. D. J., Minderaa, R. B., van Engeland, H. & de Jonge, M. V. (2009). Evaluation of the ADOS revised algorithm: the applicability in 558 Dutch children and adolescents. *Journal of Autism and Developmental Disorders*, *39*(9), 1350–8. doi:10.1007/s10803-009-0749-9

Dorlack, T. P., Myers, O. B. & Kodituwakku, P. W. (2018). A comparative analysis of the ADOS-G and ADOS-2 algorithms: preliminary findings. *Journal of Autism and Developmental Disorders*, 1–12.

Esler, A. N., Bal, V. H., Guthrie, W., Wetherby, A., Weismer, S. E. & Lord, C. (2015). The autism diagnostic observation schedule, toddler module: standardized severity scores. *Journal of Autism and Developmental Disorders*, *45*(9), 2704–2720.

Falkmer, T., Anderson, K., Falkmer, M. & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: a systematic literature review. *European Child & Adolescent Psychiatry*, *22*(6), 329–40. doi:10.1007/s00787-013-0375-0

Fein, D., Barton, M., Eigsti, I.-M., Kelley, E., Naigles, L., Schultz, R. T., … Tyson, K. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(2), 195–205. doi:10.1111/jcpp.12037

Gilchrist, A., Green, J., Cox, A., Burton, D., Rutter, M. & Le Couteur, A. (2001). Development and current functioning in adolescents with Asperger syndrome: a comparative study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 227–40.

Glasson, E. J., Bower, C., Petterson, B., de Klerk, N., Chaney, G. & Hallmayer, J. F. (2004). Perinatal factors and the development of autism: a population study. *Archives of General Psychiatry*, *61*(6), 618–27.

Gray, C. A. & Garand, J. D. (1993). Social stories: Improving responses of students with autism with accurate social information. *Focus on Autistic Behavior*, *8*(1), 1–10.

Gray, K. M., Keating, C. M., Taffe, J. R., Brereton, A. V., Einfeld, S. L., Reardon, T. C. & Tonge, B. J. (2014). Adult outcomes in autism: Community inclusion and living skills. *Journal of Autism and Developmental Disorders*, *44*(12), 3006–3015.

Green, I. W., Kidd, C. L. & Accordino, R. E. (2016). Autism Spectrum Disorder. In McDougle, Christopher J (Ed.), . Oxford University Press.

Hedley, D., Nevill, R., Uljarevi'c, 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.

Helt, M., Kelley, E., Kinsbourne, M., Pandey, J., Boorstein, H., Herbert, M. & Fein, D. (2008). Can children with autism recover? If so, how? *Neuropsychology Review*, *18*(4), 339–366.

Hill, T. L., Gray, S. A., Kamps, J. L. & Varela, R. E. (2015). Age and adaptive functioning in children and adolescents with ASD: The effects of intellectual functioning and ASD symptom severity. *Journal of Autism and Developmental Disorders*, *45*(12), 4074–4083.

Hultman, C., Sandin, S., Levine, S., Lichtenstein, P. & Reichenberg, A. (2011). Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. *Molecular Psychiatry*, *16*(12), 1203.

Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Johnson, C. P. & Myers, S. M. (2007). Identification and Evaluation of Children With Autism Spectrum Disorders. *Pediatrics*, *120*(5), 1183–1215. doi:10.1542/peds.2007-2361

Joseph, R. M., O'Shea, T. M., Allred, E. N., Heeren, T., Hirtz, D., Paneth, N., … Kuban, K. C. (2017). Prevalence and associated features of autism spectrum disorder in extremely low gestational age newborns at age 10 years. *Autism Research*, *10*(2), 224–232.

Kamp-Becker, I., Albertowski, K., Becker, J., Ghahreman, M., Langmann, A., Mingebach, T., … others. (2018). Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *European Child \& Adolescent Psychiatry*, 1–15.

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.

Kokina, A. & Kern, L. (2010). Social Story^{TM} interventions for students with autism spectrum disorders: A meta-analysis. *Journal of Autism and Developmental Disorders*, *40*(7), 812–826.

Kreiser, N. L. & White, S. W. (2014). ASD in females: are we overstating the gender difference in diagnosis? *Clinical Child and Family Psychology Review*, *17*(1), 67–84.

Langmann, A., Becker, J., Poustka, L., Becker, K. & Kamp-Becker, I. (2017). Diagnostic utility of the autism diagnostic observation schedule in a clinical sample of adolescents and adults. *Research in Autism Spectrum Disorders*, *34*, 34–43.

Larsson, H. J., Eaton, W. W., Madsen, K. M., Vestergaard, M., Olesen, A. V., Agerbo, E., … Mortensen, P. B. (2005). Risk factors for autism: perinatal factors, parental psychiatric history, and socioeconomic status. *American Journal of Epidemiology*, *161*(10), 916–925.

Lauritsen, M. B., Pedersen, C. B. & Mortensen, P. B. (2005). Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 46*(9), 963–71.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K. & Bishop, S. L. (2012). ADOS-2. Autism Diagnostic Observation Schedule. Manual (Part I): Modules 1-4. Western Psychological Services Los Angeles, CA.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*(1), 3.

Mannion, A. & Leader, G. (2013). Comorbidity in autism spectrum disorder: A literature review. *Research in Autism Spectrum Disorders, 7*(12), 1595–1616.

Murray, S. & Noland, B. (2012). *Video modeling for young children with autism spectrum disorders: A practical guide for parents and professionals*. Jessica Kingsley Publishers.

National Autism Center. (2015). *Findings and conclusions: National standards project, phase 2.* Randolph, MA: Author

National Collaborating Center for Mental Health (2013). *Autism: the NICE guideline on management and support of children and young people on the autism spectrum.* The British Psychological Society & The Royal College of Psychiatrists.

Neuhaus, E., Beauchaine, T. P., Bernier, R. A. & Webb, S. J. (2017). Child and family characteristics moderate agreement between caregiver and clinician report of autism symptoms. *Autism Research, 11(3),* 476-487.

Oswald, D. P., Haworth, S. M., Mackenzie, B. K. & Willis, J. H. (2017). Parental report of the diagnostic process and outcome: ASD compared with other developmental disabilities. *Focus on Autism and Other Developmental Disabilities, 32*(2), 152–160.

Parner, E. T., Baron-Cohen, S., Lauritsen, M. B., Jørgensen, M., Schieve, L. A., Yeargin-Allsopp, M. & Obel, C. (2012). Parental age and autism spectrum disorders. *Annals of Epidemiology, 22*(3), 143–150.

Penner, M., Anagnostou, E., Andoni, L. Y. & Ungar, W. J. (2017). Systematic review of clinical guidance documents for autism spectrum disorder diagnostic assessment in select regions. *Autism, 22(5),* 517-527.

Reichow, B., Barton, E. E., Boyd, B. A. & Hume, K. (2012). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews,* (10).

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., … Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*(9), 1094–103.

Robins, D. L., Fein, D. & Barton, M. (2009). Modified checklist for autism in toddlers, revised, with follow-up (M-CHAT-R/F) TM. Self-published. www.mchatscreen.com.

Rutter, M., Le Couteur, A., Lord, C. & others. (2003). Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services*.

Tomanik, S. S., Pearson, D. A., Loveland, K. A., Lane, D. M. & Shaw, J. B. (2007). Improving the reliability of autism diagnoses: Examining the utility of adaptive behavior. *Journal of Autism and Developmental Disorders*, *37*(5), 921–928.

Tomeny, T. S., Barry, T. D. & Bader, S. H. (2014). Birth order rank as a moderator of the relation between behavior problems among children with an autism spectrum disorder and their siblings. *Autism*, *18*(2), 199–202.

Towle, P. O. & Patrick, P. A. (2016). Autism spectrum disorder screening instruments for very young children: a systematic review. *Autism Research and Treatment*, 1–29.

Ventola, P. E., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J., … Fein, D. (2006). Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*, *36*(7), 839–47.

Virkud, Y. V., Todd, R. D., Abbacchi, A. M., Zhang, Y. & Constantino, J. N. (2009). Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *150*(3), 328–334.

Watkins, L., O'Reilly, M., Kuhn, M., Gevarter, C., Lancioni, G. E., Sigafoos, J. & Lang, R. (2015). A review of peer-mediated social interaction interventions for students with autism in inclusive settings. *Journal of Autism and Developmental Disorders*, *45*(4), 1070–1083.

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., … Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders*, *45*(7), 1951–1966.

Zander, E., Sturm, H. & Bölte, S. (2015). The added value of the combined use of the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule: Diagnostic validity in a clinical Swedish sample of toddlers and young preschoolers. *Autism*, *19*(2), 187–199.

APPENDIX A

INSTITUTIONAL REVIEW BOARD APPROVAL

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM
Office of the Institutional Review Board for Human Use

470 Administration Building
701 20th Street South
Birmingham, AL 35294-0104
205.934.3789 | Fax 205.934.1301 |
irb@uab.edu

APPROVAL LETTER

TO: Lebersfeld, Jenna B

FROM: University of Alabama at Birmingham Institutional Review Board
Federalwide Assurance # FWA00005960
IORG Registration # IRB00000196 (IRB 01)
IORG Registration # IRB00000726 (IRB 02)

DATE: 20-Jun-2019

RE: IRB-300001729
Screening and Diagnosis of Autism Spectrum Disorder

The IRB reviewed and approved the Continuing Review submitted on 10-Jun-2019 for the above referenced project. The review was conducted in accordance with UAB's Assurance of Compliance approved by the Department of Health and Human Services.

Type of Review:	Expedited
Expedited Categories:	5
Determination:	Approved
Approval Date:	20-Jun-2019
Approval Period:	One Year
Expiration Date:	19-Jun-2020

The following populations are approved for inclusion in this project:

- Children – CRL 1

The following apply to this project related to informed consent and/or assent:

- Waiver of HIPAA
- Waiver of Informed Consent

Documents Included in Review:

- waiverauthconsent.180428
- ipr.190610
- hsptrackedchanges.190610
- children.180428
- hspupdatedclean.190610

APPENDIX B

SEARCH STRATEGY

ADOS Search: “ADOS” OR “Autism Diagnostic Observation Schedule” OR “Autism Diagnostic Observation Scale” OR “Autism Diagnostic Observation Scales” OR “Autism Interview Schedule” OR "Autism Diagnostic and Observational Schedule" OR “Autism Observation Schedule” OR “Autism Diagnostic Observation” OR “Autism Diagnostic Scale” OR “Autism Diagnostic Observational Schedule” OR “Autism Diagnostic Observation Schedules” OR “Autism Diagnostic Observations Schedule” OR "Autism Diagnostic Observation Interview" OR “Autism Diagnosis Observation Schedule” OR “Autism Daignostic Observation Schedule” OR “Autism Diagnostic Schedule” OR "Autism Behavior Diagnostic Observation Schedule" OR "Autism Diagnostic Observation Autism Schedule" OR "Autism Diagnostic Observations Schedule” OR "Autism Diagnoctic Observation Scale" OR "Autism Spectrum Disorders Observation for Children" OR "Autism Diagnostic Observational System" OR "Autism Diagnostic Interview Schedule" OR "Autism Diagnostic Observation Schedul" OR “Autism Diagnosis Observation Schedule” OR "Autism Disagnostic Interview Revised" OR “Autism Diagnostic Behavior Scales" OR “Autism Diagnostics Observation Schedule" OR “Autism Diagnosis and Observation Scale” OR “Autism Diagnostic Observation System” OR "Autism Diagnostic Observation Schedules" OR “Autism Diagnostic Observational Scales” OR "Autism Diagnostic Obsevation Scale” OR "Autism Diagnostic Assessment Scale" OR "Autism Diagnostic and Observation Schedule" OR "Autisms Diagnostic Observation Schedule" OR "Autism Spectrum Disorder Observation for Children" OR "Autism Diagnositc Observation Schedule" OR "Autism Diagnostic Observation Schema"

ADI Search: “ADI” OR “Autism Diagnostic Interview” OR “Autism Diagnostic Instrument” OR “Autism Interview Schedule” OR "Autism Diagnostic and Observational Schedule" OR “Autism Diagnosis Interviewed” OR “Autism Diagnosis Interview” OR “Autism Diagnostic Interviewed” OR “Autism Diagnostic Inventory” OR "Autism Diagnostic Observation Interview" OR "Autism Diagnosic Interview Revised" OR "AutAutism Diagnostic Interview Revised" OR "Autism Diagnostic Interview Schedule" OR "Autism Diagnostics Interview" OR “Autism Diagnosis Interview” OR “Autism Disorders Interview” OR "Autism Diagnotic Interview" OR "Autism Diagnsostic Interview" OR "Autism Diagnostic Interviews" OR "Autism Interview Revised" OR "Autism Diagnostic Inventory"

Search 1: autis* AND [ADOS Search] OR [ADI Search] 2007 to present

Search 2: autis* AND [ADI Search] 2003 to 2006

PROQUEST (PsycInfo and ERIC)

[Autism Diagnosis]
autis*

AND

[ADOS Search]
(Exact("ADOS" OR "Autism Diagnostic Observation Schedule" OR "Autism Diagnostic Observation Scale" OR "Autism Diagnostic Observation Scales" OR "Autism Interview Schedule" OR "Autism Diagnostic and Observational Schedule" OR "Autism Observation Schedule" OR "Autism Diagnostic Observation" OR "Autism Diagnostic Scale" OR "Autism Diagnostic Observational Schedule" OR "Autism Diagnostic Observation Schedules" OR "Autism Diagnostic Observations Schedule" OR "Autism Diagnostic Observation Interview" OR "Autism Diagnosis Observation Schedule" OR "Autism Daignostic Observation Schedule" OR "Autism Diagnostic Schedule" OR "Autism Behavior Diagnostic Observation Schedule" OR "Autism Diagnostic Observation Autism Schedule" OR "Autism Diagnostic Observations Schedule" OR "Autism Diagnoctic Observation Scale" OR "Autism Spectrum Disorders Observation for Children" OR "Autism Diagnostic Observational System" OR "Autism Diagnostic Interview Schedule" OR "Autism Diagnostic Observation Schedul" OR "Autism Diagnosis Observation Schedule" OR "Autism Disagnostic Interview Revised" OR "Autism Diagnostic Behavior Scales" OR "Autism Diagnostics Observation Schedule" OR "Autism Diagnosis and Observation Scale" OR "Autism Diagnostic Observation System" OR "Autism Diagnostic Observation Schedules" OR "Autism Diagnostic Observational Scales" OR "Autism Diagnostic Obsevation Scale" OR "Autism Diagnostic Assessment Scale" OR "Autism Diagnostic and Observation Schedule" OR "Autisms Diagnostic Observation Schedule" OR "Autism Spectrum Disorder Observation for Children" OR "Autism Diagnositc Observation Schedule" OR "Autism Diagnostic Observation Schema"))

OR

[ADI Search]
(Exact("ADI" OR "Autism Diagnostic Interview" OR "Autism Diagnostic Instrument" OR "Autism Interview Schedule" OR "Autism Diagnostic and Observational Schedule" OR "Autism Diagnosis Interviewed" OR "Autism Diagnosis Interview" OR "Autism Diagnostic Interviewed" OR "Autism Diagnostic Inventory" OR "Autism Diagnostic Observation Interview" OR "Autism Diagnosic Interview Revised" OR "AutAutism Diagnostic Interview Revised" OR "Autism Diagnostic Interview Schedule" OR "Autism Diagnostics Interview" OR "Autism Diagnosis Interview" OR "Autism Disorders Interview" OR "Autism Diagnotic Interview" OR "Autism Diagnsostic Interview" OR

"Autism Diagnostic Interviews" OR "Autism Interview Revised" OR "Autism Diagnostic Inventory"))

PubMed/MEDLINE Search Strategy

[Autism Diagnosis]

Autis*

[ADOS]

"ADOS" OR "Autism Diagnostic Observation Schedule" OR "Autism Diagnostic Observation Scale" OR "Autism Diagnostic Observation Scales" OR "Autism Interview Schedule" OR "Autism Diagnostic and Observational Schedule" OR "Autism Observation Schedule" OR "Autism Diagnostic Observation" OR "Autism Diagnostic Scale" OR "Autism Diagnostic Observational Schedule" OR "Autism Diagnostic Observation Schedules" OR "Autism Diagnostic Observations Schedule" OR "Autism Diagnostic Observation Interview" OR "Autism Diagnosis Observation Schedule" OR "Autism Daignostic Observation Schedule" OR "Autism Diagnostic Schedule" OR "Autism Behavior Diagnostic Observation Schedule" OR "Autism Diagnostic Observation Autism Schedule" OR "Autism Diagnostic Observations Schedule" OR "Autism Diagnoctic Observation Scale" OR "Autism Spectrum Disorders Observation for Children" OR "Autism Diagnostic Observational System" OR "Autism Diagnostic Interview Schedule" OR "Autism Diagnostic Observation Schedul" OR "Autism Diagnosis Observation Schedule" OR "Autism Disagnostic Interview Revised" OR "Autism Diagnostic Behavior Scales" OR "Autism Diagnostics Observation Schedule" OR "Autism Diagnosis and Observation Scale" OR "Autism Diagnostic Observation System" OR "Autism Diagnostic Observation Schedules" OR "Autism Diagnostic Observational Scales" OR "Autism Diagnostic Obsevation Scale" OR "Autism Diagnostic Assessment Scale" OR "Autism Diagnostic and Observation Schedule" OR "Autisms Diagnostic Observation Schedule" OR "Autism Spectrum Disorder Observation for Children" OR "Autism Diagnositc Observation Schedule" OR "Autism Diagnostic Observation Schema"

[ADI]

"ADI" OR "Autism Diagnostic Interview" OR "Autism Diagnostic Instrument" OR "Autism Interview Schedule" OR "Autism Diagnostic and Observational Schedule" OR "Autism Diagnosis Interviewed" OR "Autism Diagnosis Interview" OR "Autism Diagnostic Interviewed" OR "Autism Diagnostic Inventory" OR "Autism Diagnostic Observation Interview" OR "Autism Diagnosic Interview Revised" OR "AutAutism Diagnostic Interview Revised" OR "Autism Diagnostic Interview Schedule" OR "Autism Diagnostics Interview" OR "Autism Diagnosis Interview" OR "Autism Disorders Interview" OR "Autism Diagnotic Interview" OR "Autism Diagnsostic Interview" OR "Autism Diagnostic Interviews" OR "Autism Interview Revised" OR "Autism Diagnostic Inventory"

Cochrane Database Search Strategy

[Autism Diagnosis]

#1 Autis*

[ADOS]

#2 “ADOS” OR “Autism Diagnostic Observation Schedule” OR “Autism Diagnostic Observation Scale” OR “Autism Diagnostic Observation Scales” OR “Autism Interview Schedule” OR "Autism Diagnostic and Observational Schedule" OR “Autism Observation Schedule” OR “Autism Diagnostic Observation” OR “Autism Diagnostic Scale” OR “Autism Diagnostic Observational Schedule” OR “Autism Diagnostic Observation Schedules” OR “Autism Diagnostic Observations Schedule” OR "Autism Diagnostic Observation Interview" OR “Autism Diagnosis Observation Schedule” OR “Autism Daignostic Observation Schedule” OR “Autism Diagnostic Schedule” OR "Autism Behavior Diagnostic Observation Schedule" OR "Autism Diagnostic Observation Autism Schedule" OR "Autism Diagnostic Observations Schedule” OR "Autism Diagnoctic Observation Scale" OR "Autism Spectrum Disorders Observation for Children" OR "Autism Diagnostic Observational System" OR "Autism Diagnostic Interview Schedule" OR "Autism Diagnostic Observation Schedul" OR “Autism Diagnosis Observation Schedule” OR "Autism Disagnostic Interview Revised" OR “Autism Diagnostic Behavior Scales" OR “Autism Diagnostics Observation Schedule" OR “Autism Diagnosis and Observation Scale” OR “Autism Diagnostic Observation System” OR "Autism Diagnostic Observation Schedules" OR “Autism Diagnostic Observational Scales” OR "Autism Diagnostic Obsevation Scale” OR "Autism Diagnostic Assessment Scale" OR "Autism Diagnostic and Observation Schedule" OR "Autisms Diagnostic Observation Schedule" OR "Autism Spectrum Disorder Observation for Children" OR "Autism Diagnositc Observation Schedule" OR "Autism Diagnostic Observation Schema"

[ADI]

#3 “ADI” OR “Autism Diagnostic Interview” OR “Autism Diagnostic Instrument” OR “Autism Interview Schedule” OR "Autism Diagnostic and Observational Schedule" OR “Autism Diagnosis Interviewed” OR “Autism Diagnosis Interview” OR “Autism Diagnostic Interviewed” OR “Autism Diagnostic Inventory” OR "Autism Diagnostic Observation Interview" OR "Autism Diagnosic Interview Revised" OR "AutAutism Diagnostic Interview Revised" OR "Autism Diagnostic Interview Schedule" OR "Autism Diagnostics Interview" OR “Autism Diagnosis Interview” OR “Autism Disorders Interview” OR "Autism Diagnotic Interview" OR "Autism Diagnsostic Interview" OR "Autism Diagnostic Interviews" OR "Autism Interview Revised" OR "Autism Diagnostic Inventory"

[ADOS AND ADI]

#4

#1 AND (#2 OR #3)

[ADI only]
#5
#1 AND #3

JADD Search Strategy

EBSCO database

[JADD only]
"Journal of Autism & Developmental Disorders" (SO Journal Name)

[Autism Diagnosis]
Autis* (Entire Document)

[ADOS]
"ADOS" OR "ADOS-2" OR "Autism Diagnostic Observation Schedule" (Entire Document)

[ADI]

"ADI" OR "ADI-R" OR "Autism Diagnostic Interview" (Entire Document)

Checked: Also search within the full text of the articles

Research in Autism Spectrum Disorders (published in 2007)

ScienceDirect

[Research in Autism Spectrum Disorders only]
In this journal or book title: Research in Autism Spectrum Disorders

[2007-Present]
Year(s): 2007-2018 (Search 2 with 2003-2006 not conducted because the journal was not published until 2007)

Find articles with these terms: (250 character limit)

[ASD diagnosis] (wildcards not supported)
Not included due to lack of wildcard support and likelihood that most articles would be about ASD given the topic of the journal

[ADOS]
"ADOS" OR "ADOS-2" OR "Autism Diagnostic Observation Schedule"

OR

[ADI]
"ADI" OR "ADI-R" OR "Autism Diagnostic Interview"

Autism Research (published in 2008)

Wiley Online Library

[ASD diagnosis]
Anywhere: autis*

[ADOS]
Anywhere: "ADOS" OR "ADOS-2" OR "Autism Diagnostic Observation Schedule"

[ADI]
Anywhere: "ADI" OR "ADI-R" OR "Autism Diagnostic Interview"