Psychological Assessment

1040-3590/16/$12.00 http://dx.doi.org/10.1037/pas0000317

Investigating the Accuracy of a Novel Telehealth Diagnostic Approach for Autism Spectrum Disorder

Christopher J. Smith
Southwest Autism Research and Resource Center, Phoenix, Arizona

Agata Rozga
Georgia Institute of Technology

Nicole Matthews
Southwest Autism Research and Resource Center, Phoenix, Arizona

Ron Oberleitner
Behavior Imaging Solutions, Boise, Idaho

Nazneen Nazneen and Gregory Abowd
Georgia Institute of Technology

Research indicates that a substantial amount of time elapses between parents' first concerns about their child's development and a formal diagnosis of autism spectrum disorder (ASD). Telehealth presents an opportunity to expedite the diagnostic process. This project compared a novel telehealth diagnostic approach that utilizes clinically guided in-home video recordings to the gold standard in-person diagnostic assessment. Participants included 40 families seeking an ASD evaluation for their child and 11 families of typically developing children. Children were between the ages of 18 months and 6 years 11 months; mean adaptive behavior composite = 75.47 (*SD* = 15.94). All parent participants spoke English fluently. Families completed the Naturalistic Observation Diagnostic Assessment (NODA) for ASD, which was compared to an in-person assessment (IPA). Agreement between the 2 methods, as well as sensitivity, specificity, and interrater reliability, were calculated for the full sample and the subsample of families seeking an ASD evaluation. Diagnostic agreement between NODA and the IPA was 88.2% ($\kappa = 0.75$) in the full sample and 85% ($\kappa = 0.58$) in the subsample. Sensitivity was 84.9% in both, whereas specificity was 94.4% in the full sample and 85.7% in the subsample. Kappa coefficients for interrater reliability indicated 85% to 90% accuracy between raters. NODA utilizes telehealth technology for families to share information with professionals and provides a method to inform clinical judgment for a diagnosis of ASD. Due to the high level of agreement with the IPA in this sample, NODA has potential to improve the efficiency of the diagnostic process for ASD.

Keywords: autism, diagnosis, video, telehealth, remote assessment

Supplemental materials: http://dx.doi.org/10.1037/pas0000317.supp

There are substantial delays between parents' first concerns about their child's development and a diagnosis of autism spectrum disorder (ASD; Wiggins, Baio, & Rice, 2006). These delays will likely worsen, given that prevalence rates for the disorder continue to climb and access to qualified health care professionals is limited in many communities (Autism and Developmental Dis-

Christopher J. Smith, Southwest Autism Research and Resource Center, Phoenix, Arizona; Agata Rozga, School of Interactive Computing, Georgia Institute of Technology; Nicole Matthews, Southwest Autism Research and Resource Center; Ron Oberleitner, Behavior Imaging Solutions, Boise, Idaho; Nazneen Nazneen and Gregory Abowd, School of Interactive Computing, Georgia Institute of Technology.

This study was presented as a poster at the American Academy of Child and Adolescent Psychiatry's 62nd Annual Meeting, San Antonio, TX, October 2015, and the International Meeting for Autism Research, Salt Lake City, UT, May 2015.

All phases of this study were supported by National Institute of Mental Health's Small Business Innovation Research Grant 9 R44 MH099035 awarded to BIS. Subcontracts with SARRC and Georgia Institute of Technology support Christopher J. Smith, Nicole Matthews, and Agata Rozga and their work on this study. Christopher J. Smith's and Nicole Matthew's employer, Southwest Autism Research and Resource Center (SARRC), will be paid in the future by Behavior Imaging Solutions (BIS) to conduct the reviews of cases for people who pay them for the commercial version of the Naturalistic Observation Diagnostic Assessment (NODA). Ron Oberleitner is the chief executive officer of BIS, the company that will commercialize NODA. Gregory Abowd was co-advisor for Nazneen Nazneen during her graduate studies, which may present a conflict of interest that is registered with and managed by Georgia Institute of Technology. The remaining authors have no conflicts of interest to disclose.

The authors express gratitude to the families who participated in this research and to Raun Melmed of the Melmed Center and SARRC for providing valuable insight during this project.

Correspondence concerning this article should be addressed to Christopher J. Smith, SARRC, 300 North 18th Street, Phoenix, AZ 85006. E-mail: csmith@autismcenter.org

abilities Monitoring Network Surveillance Year 2000 Principal Investigators, 2007, 2014; Liptak et al., 2008; Mandell, Novak, & Zubritsky, 2005; Thomas, Ellis, McLaurin, Daniels, & Morrissey, 2007). Lengthy wait lists for diagnostic evaluations delay early intensive intervention, which is critical for optimal outcomes (Howlin, Magiati, & Charman, 2009). Telehealth approaches have been investigated as a means of treatment delivery in ASD, but few have explored the potential for such technologies to support diagnostic assessments (Baharav & Reiser, 2010; Parmanto, Pulantara, Schutte, Saptono, & McCue, 2013; Vismara, Young, & Rogers, 2012; Wainer & Ingersoll, 2015). The current project examined a method that guides families to collect clinically relevant videos in the home and share them with diagnostic professionals using telehealth technology. If validated, this approach may present one avenue for reducing the time between parent concerns and diagnosis.

Practice parameters from the American Academy of Child and Adolescent Psychiatry have recommended that professionals first determine a diagnosis and then conduct a multidisciplinary evaluation to identify factors that may have contributed to developmental delay (Volkmar et al., 2014). The recommended diagnostic process includes a parent interview to assess developmental history and direct observation of the child (Huerta & Lord, 2012; Volkmar & Klin, 2005), though these procedures should inform, not replace, clinical judgment. The use of recommended semistructured assessments to collect this information may be hampered by required training, cost and lengthy administration time. Ultimately, skilled professionals evaluate development through some method, but ultimately rely on clinical judgment to diagnose (Charman & Gotham, 2013). Despite consistent recommendations for two methods of assessment (interview and observation), most practitioners rely on only one method to diagnose ASD (Rice et al., 2014) which may affect the validity of the diagnostic outcome lengthy administration time.

Store-and-forward telehealth approaches to diagnosis may facilitate sharing of both current behavior examples and developmental history with diagnostic professionals. These systems support video recordings of live events, which are subsequently shared with a clinical expert for review and assessment. This approach may offer several key advantages particularly relevant to remote diagnosis of ASD (Oberleitner, Laxminarayan, Suri, Harrington, & Bradstreet, 2014). It enables families to record videos in their home, in the course of their day-to-day activities, which ensures the capture of natural expressions of child behavior that are widely acknowledged as crucial to an accurate and comprehensive assessment. Moreover, because home recordings can be carried out over the course of several days, they may mitigate some of the shortcomings associated with a single clinic-based or live telehealth assessment, such as the child's reactivity, their current mood or level of fatigue, or the likelihood that low-frequency behaviors may not be observed. Developmental history can also be shared through a parent survey within the telehealth system. From a practical standpoint, such an approach minimizes the need to coordinate schedules with a clinician and reduces the need for remotely located families to travel long distances to a clinic. Finally, beyond the opportunity to provide a timely diagnosis directly to the family, it may also enable clinical centers to more efficiently make use of their limited resources by triaging families on waiting lists for diagnostic assessments.

Pilot studies have demonstrated parents' ability to collect videos of child behavior in the home and share them with diagnosticians who, in turn, determined their relevance for ASD diagnosis (Nazneen et al., 2015; Smith, Oberleitner, Treulich, McIntosh, & Melmed, 2009). Still, comparison of the resulting diagnostic outcomes to a gold-standard, in-person assessment (IPA) has not yet been reported. The current report presents a comparison of the Naturalistic Observation Diagnostic Assessment (NODA), a store-and-forward telehealth approach to ASD diagnosis that relies on parent-collected videos, to an independently conducted IPA.

Method

Participants

Participants included 51 children in the southwestern United States and at least one parent of each child. The full sample included 11 children who were typically developing (TD) and 40 children whose parents were seeking an evaluation for ASD in response to advertisements for the study (EV subgroup). TD children were recruited from a database of children who were previously evaluated for a clinical program that included typically developing peers as part of the treatment model. Children were between the ages of 18 months and 6 years 11 months and had no known genetic condition. All parent participants spoke English fluently and were evaluated by English-speaking raters. See Table 1 for additional participant demographics. Study procedures were approved by the Western Institutional Review Board, and informed consent was obtained from at least one parent or guardian of each child. Evaluations were conducted after participants were provided informed consent, and there were no exclusions on the basis of results of the IPA.

The primary NODA rater had a master's degree in psychology and 10 years of experience conducting ASD assessments. To demonstrate usability of the NODA system and determine interrater reliability, 10 secondary raters (clinical or research professionals with a minimum of 10 years of experience conducting observational assessments for ASD) were recruited from different regions of North America, and each was assigned five cases. Informed consent was obtained from each secondary rater. The primary rater and secondary raters were blind to the child's group membership (EV or TD), the results of the IPA, and results from the other raters. Although the primary rater was employed by the research center, she worked remotely (i.e., off-site) and did not have direct contact with the staff members who conducted the IPAs. The principal investigator conducted a 30-min training on the web-based assessment portal and NODA procedures (described in the Method section) with each rater.

Procedure

In-person assessment (IPA). All participants completed the IPA during their first visit to the center. The IPA included the Autism Diagnostic Interview—Revised (ADI–R; Rutter, Le Couteur, & Lord, 2003); the Autism Diagnostic Observation Schedule—Second Edition (ADOS–2; Lord et al., 2012); either the Mullen Scales of Early Learning (MSEL; Mullen, 1995) for participants up to 68 months or the Kaufman Brief Intelligence Test—Second Edition (KBIT–2; Kaufman & Kaufman, 2004) for

Table 1
Sample Characteristics for Participants Who Were Either Seeking an Evaluation for ASD or Were Typically Developing

	ASD evaluation[a]			Typically developing			Full sample		
Variable	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Age in months	40	52.78	17.58	11	42.55	11.07	51	50.60	16.84
Males	30			6			36		
Ethnicity									
Caucasian	15			6			21		
Hispanic	19			3			22		
Black	3			1			4		
Other	3			1			4		
MSEL[b]	26	74.38	16.18	9	111.78	15.87	35	84.00	22.95
FSIQ[c]	6	91.17	16.65	0			6	91.17	16.65
ABC[d]	40	69.98	11.80	11	95.45	12.95	51	75.47	15.94
ADOS Comp[e]	34	6.53	2.45	2[f]	2.00	1.41	36	6.28	2.61

Note. ASD = autism spectrum disorder; MSEL = Mullen Scales of Early Learning Composite Score; FSIQ = Full Scale IQ; ABC = adaptive behavior composite; ADOS Comp = Autism Diagnostic Observation Schedule Comparison score.
[a] Referred for ASD evaluation. [b] For participants ≤ 68 months of age; there were eight incomplete assessments in the *referred* group and two incomplete assessments in the *typically developing* group. [c] For participants older than 68 months from the Kaufmann Brief Intelligence Test. [d] From the Vineland Adaptive Behavior Scales. [e] Comparison score for Modules 1–3 ($n = 36$); toddler module ($n = 8$) does not have a comparison score. [f] Only two comparison scores are reported for the typically developing group because six participants were previously assessed with the first edition of the ADOS, which did not include a comparison score, and three participants were assessed with the second edition of the ADOS toddler module.

participants 69 months and older; and the Vineland Adaptive Behavior Scales—Second Edition (VABS–2; Sparrow, Cicchetti, & Balla, 2005). Six of the 11 TD children were previously evaluated with the first edition of the ADOS (ADOS; Lord, Rutter, Dilavore, & Risi, 1999), which did not include a comparison score. The rest of the IPA was completed during their participation in this study. Assessments were completed by experienced raters who were blind to the subject group (EV or TD) and to the information collected in NODA.

The principal investigator, a psychologist with 20 years of experience evaluating individuals with ASD for research purposes, completed a *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM–5*; American Psychiatric Association, 2013) diagnosis for each participant on the basis of the assessment results and clinical judgment. Results of the IPA were not provided to the family until after they completed the NODA procedures. Thus, families were not informed about the significance of their child's behavior before collecting videos for NODA.

Naturalistic Observation Diagnostic Assessment (NODA). NODA included collection of both developmental history and video data. First, caregivers completed a brief developmental history interview, and responses were stored in the family's online account. The NODA application, installed on a mobile device, guided parents to record their child in four 10-min scenarios: (1) family meal time, (2) playtime with others, (3) playtime alone, and (4) parent concerns. The first three scenarios provided opportunities for the child to demonstrate typical social-communication skills and play-based behaviors. Instructions to the parent to introduce specific social presses were included in the app (e.g., interact with your child playfully, say your child's name to get his attention, ask your child where something is in the room, give your child time to initiate or respond, point at something and direct your child's attention to it). Pilot studies demonstrated that these instructions improved the clinical utility of the videos (Nazneen et al., 2015). To avoid predisposing parents toward collecting examples of behaviors that indicate ASD (e.g., hand mannerisms, poor eye contact, odd behavior), NODA included instructions that created opportunities for demonstrating typical social communicative behavior. The fourth scenario was less structured and simply asked parents to record any behavior that caused them concern. Additional instructions for each scenario suggested that parents use a mounting device (i.e., tripod) to set up and frame the recording ahead of time and to ensure relevant people and objects (i.e., the child's face, any toys the child was playing with, the child's social partner if relevant) were clearly in view. Each recording stopped automatically after 10 min, at which time parents had the option to either upload or delete the video. Parents had the capability to view the video before uploading if desired. More details about the content of the app can be found in the online supplemental materials and were previously published (see Nazneen et al., 2015).

Raters logged in to a web-based assessment system that enabled them to review children's developmental histories and the videos uploaded by families, to complete a *DSM–5* checklist for ASD, and to render a diagnosis (ASD or not ASD). While reviewing videos, the raters "tagged" examples of atypical behavior by pausing the video and selecting a term from a predefined list of descriptors, or "tags" (e.g., no social response) that were built into the interface. Each tag was automatically mapped by the NODA system to a specific *DSM–5* criterion. The behaviors represented by tags and their mappings to *DSM* criteria were informed by the *DSM–5* and determined by a team of experienced diagnosticians involved in this project. After tagging the videos, the raters reviewed the developmental history and then completed a *DSM–5* checklist within NODA. To assist the raters in making the determination as to whether each *DSM–5* criterion was satisfied, tags that had been inserted in the videos during the review process were listed below each criterion. Each tag linked to a relevant moment in the video for the raters to review if needed. On the basis of clinical judgment, the raters determined whether there was enough evidence from the developmental history and the tagged behaviors to satisfy each *DSM–5* criterion for ASD and ultimately whether to assign a diagnosis. After determining the final diagnostic category

(ASD or not ASD), the raters scored their confidence in the diagnosis on a scale from 1 (*extremely low*) to 5 (*extremely high*). More details about the content of the assessment portal can be found in the online supplemental materials and were previously published (see Nazneen et al., 2015).

Analyses

NODA was compared to the IPA by calculating percentage of agreement, kappa, sensitivity, and specificity, first for the full sample ($N = 51$) and then for the EV group ($n = 40$). Additionally, agreement at the *DSM–5* symptom level (A1 to A3 and B1 to B4) was measured by summing the values (1 = present, 0 = absent) on the subcriteria and calculating a two-way random effects model intraclass correlation coefficient (ICC; Type II; Shrout & Fleiss, 1979). Variables derived from each assessment method were used to investigate differences between participants for whom NODA and IPA were discordant. Kappa and ICC were also used to determine interrater reliability between the primary NODA rater and the secondary raters.

Results

Within the full sample, the diagnostic procedures (NODA and IPA) agreed in 88.2% of cases ($\kappa = .75$, 95% confidence interval [CI: .56, .94]). The sensitivity of NODA for a diagnosis of ASD was .85 (95% CI [.67, .94]) and the specificity was .94 (95% CI [.71, 1.00]). As a measure of agreement among the *DSM–5* symptom criteria, ICC was .86 (95% CI [.73, .92]). For interrater reliability, the secondary raters agreed with the primary rater in 78% of cases, and kappa was 0.56 (95% CI [.53, .59]) and ICC was .85 (95% CI [.73, .91]). Seven of the 10 secondary NODA raters agreed with the primary rater on four of the five cases that were assigned to them; of the remaining raters, two agreed on three of their five cases and one agreed on all five cases.

In the EV subgroup, the two diagnostic procedures agreed in 85% of cases ($\kappa = 0.58$, 95% CI [.27, .89]), with a sensitivity of .85 (95% CI [.67, .94]) and a specificity of .86 (95% CI [.42, .99]). As a measure of agreement at the *DSM–5* symptom level, ICC was .60 (95% CI [.25, .79]). For interrater reliability, the secondary raters agreed with the primary rater in 72% of cases, kappa was .37 (95% CI [.15, .58]) and ICC was .72 (95% CI [.47, .85]). Of the 40 children in this group, 33 met criteria for ASD on the basis of the IPA, and 29 met criteria on the basis of NODA. Of the seven participants who did not meet criteria for ASD on the basis of the IPA, six also did not meet criteria on the basis of NODA (see Table 2).

Participants for whom NODA and IPA were concordant ($n = 34$) were compared to participants who were discordant ($n = 6$) across variables derived from each assessment method (see Table 3). From the IPA, we created a *developmental estimate* variable, consisting of the MSEL composite score ($n = 26$) or the KBIT–2 ($n = 6$). For participants missing the MSEL composite score because one or more subscales was incomplete ($n = 8$), we used the VABS–2 adaptive behavior composite (ABC), which was strongly and positively correlated with MSEL in the full sample ($n = 35$; $r = .75$, $p < .001$; 95% CI [.57, .86]). The groups did not differ significantly in age, $t(38) = 0.38$, $p = .70$; $d = 0.16$, or the VABS–2 ABC, $t(38) = 1.53$, $p = .13$; $d = 0.78$, but the discordant group had a significantly higher developmental estimate, $t(38) = 2.36$, $p = .02$; $d = 1.87$. Among the six discordant cases, the ADI–R and ADOS–2 disagreed on ASD or non-ASD in 66.7% of cases, compared to 27.5% among the 36 concordant cases. Fisher's exact test determined that group differences in disagreement on these instruments approached significance ($p = .08$).

Table 2
Characteristics and Category Agreement Between Diagnostic Methods Among Participants Seeking an ASD Evaluation ($n = 40$)

Variable	IPA category	
	Non-ASD ($n = 7$)	ASD ($n = 33$)
Males: *n* (%)	5 (71.43)	25 (75.75)
Age in months: *M* (*SD*)	53.14 (22.24)	52.70 (16.85)
Cognitive functioning: *M* (*SD*)	83.14 (6.18)	75.18 (17.94)
Primary NODA rater (%)		
ASD (%)	1 (14.29)	28 (84.85)
non-ASD (%)	6 (85.71)	5 (15.15)
Confidence: *M* (*SD*)	3.43 (.51)	3.76 (.83)
Secondary NODA raters		
ASD (%)	3 (42.86)	26 (78.7)
Non-ASD (%)	4 (57.14)	7 (21.21)
Confidence: *M* (*SD*)	3.14 (1.07)	3.87 (1.08)

Note. IPA = in-person assessment; NODA = Naturalistic Observation Diagnostic Assessment.

Five continuous variables were created to represent ASD global symptom categories by summing the number of tags assigned by a rater (see Table 3). The confidence scores from the raters and the repetitive behavior category were normally distributed and were analyzed with *t* tests. The distributions from the remaining categories were nonnormal and were analyzed with Mann–Whitney *U* tests. The concordant group had significantly higher confidence scores from the primary rater, $t(38) = -2.51$, $p = .02$, $d = 1.00$; more repetitive behavior tags, $t(38) = 2.52$, $p = .016$, $d = 1.35$; and significantly more tags overall ($Z = 2.54$, $p = .01$), compared to the discordant group; no other significant differences were observed.

Characteristics for the six discordant cases are presented in Table 3. One participant did not meet *DSM–5* criteria for ASD on the basis of the IPA, but the primary NODA rater endorsed ASD with high confidence. The second rater did not endorse ASD but with low confidence (rating of 1). The MSEL was completed even though the participant was older than the 68-month ceiling (rater error). He was 82 months old and had an MSEL composite score of 80. The ADI–R endorsed autism, but the ADOS–2 did not; appropriate social initiations were frequently noted throughout the ADOS–2 despite a prominent expressive language impairment (MSEL expressive language score of 22). The five remaining discordant cases met criteria for ASD only on the basis of the IPA; three did not meet criteria on the ADI–R but met ADOS–2 criteria for autism; the remaining two met criteria on both the ADI–R and the ADOS–2. Although the primary rater tagged behaviors across categories for these five cases, there was insufficient evidence to endorse *DSM–5* criteria. As indicated previously, the primary rater's confidence scores were significantly lower for the discordant cases compared to the concordant cases. For two of these five cases, the secondary rater was in agreement with the IPA results and endorsed full *DSM–5* criteria for ASD.

Table 3
Demographics, IPA Assessment, and Total NODA Tags in Symptom Categories for IPA and NODA Concordant and Discordant Groups and Discordant Participants

	Group comparison					Discordant participants					
	Concordant (n = 34)		Discordant (n = 6)								
Variable	%	*M* (*SD*)	%	*M* (*SD*)	*p*	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6
Gender	76 (M)		67 (M)		.63[a]	M	M	F	M	F	M
Age in months		53.32 (17.52)		55.33 (19.39)	.78[b]	38	47	65	82	68	32
IPA											
ASD	82		83			1	1	1	0	1	1
ADI–R/ADOS–2 agreement[c]	74		33		.08[a]	0	1	0	0	0	1
Developmental estimate[d]		74.67 (16.93)		93.00 (10.37)	.03[e]	79[f]	91[g]	92[g]	80[g]	109[h]	93[g]
VABS ABC		68.79 (12.09)		76.67 (7.69)	.09[e]	79	76	76	64	88	77
NODA tag categories[i]											
Social impairment		6.09 (5.29)		2.50 (2.88)	.06[e]	1	3	1	0	1	5
Verbal impairment		6.47 (4.80)		2.83 (2.48)	.06[e]	7	4	3	0	1	2
Nonverbal impairment		3.68 (3.64)		1.67 (1.75)	.12[e]	1	2	1	0	1	5
Repetitive behaviors		4.50 (2.83)		1.50 (1.38)	.02[f]	0	3	0	1	3	2
Sensory component		1.50 (1.99)		.50 (.58)	.40[e]	1	0	0	0	1	1
Stereotyped mannerisms		1.97 (3.05)		.67 (1.21)	.17[e]	3	0	0	0	1	0
Total tags		24.21 (15.84)		9.67 (5.43)	.01[e]	13	12	5	1	15	12
NODA *DSM–5* criteria[j]											
ASD	82		17			0	0	0	1	0	0
A1 Social reciprocity	88		33			0	0	0	1	0	1
A2 Nonverbal communication	88		33			0	0	0	1	1	0
A3 Relationships	85		33			0	0	0	1	0	1
B1 Repetitive behavior	88		83			0	1	1	1	1	1
B2 Rituals and routines	71		17			0	0	0	0	0	1
B3 Preoccupations	44		0			0	0	0	0	0	0
B4 Sensory component	47		33			0	0	0	1	0	1
NODA rater confidence											
Primary		3.82 (.72)		3.00 (.89)	.02[b]	4	2	3	4	3	2
Secondary		3.79 (.98)		3.40 (1.52)	.44[b]	5	4	*	1	3	4

Note. IPA = in-person assessment; NODA = Naturalistic Observation Diagnostic Assessment; Concordant = agreement between IPA and NODA; Discordant = disagreement between IPA and NODA; sub = subject; M/F = male/female; ADI–R = Autism Diagnostic Interview—Revised; ADOS–2 = Autism Diagnostic Observation Schedule—Second Edition; VABS ABC = Vineland Adaptive Behavior Scales adaptive behavior composite; *DSM–5* = *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.); MSEL = Mullen Scales of Early Learning; KBIT–2 = Kaufman Brief Intelligence Test—Second Edition.
[a] Fisher's exact. [b] *t* test. [c] ADI–R/ADOS agreement: 1 = scales agree on diagnostic category, 0 = scales disagree on diagnostic category. [d] Developmental estimate includes MSEL composite score (n = 26), KBIT–2 (n = 6), or VABS ABC (n = 8). [e] Mann Whitney U. [f] VABS ABC. [g] Mullen Scales of Early Learning developmental composite. [h] Kaufmann Brief Intelligence Test Full Scale IQ. [i] NODA tag categories = total number of tags from the primary NODA rater in each category. [j] *DSM–5* criteria endorsed by the primary NODA rater: 1 = criterion endorsed, 0 = criterion not endorsed.

Discussion

This report focuses on an initial validation of NODA, a telehealth diagnostic system that guides parents to collect short videos of child behavior and remotely share them with a clinician who conducts a diagnostic assessment for ASD. Although all analyses were conducted on both the full sample (including TD children) and the subgroup of families seeking an ASD evaluation for their child (EV subgroup), the results from the subgroup present the most pertinent evidence regarding the accuracy of NODA. However, because NODA is a novel approach to diagnosis for ASD, it important to demonstrate that it does not yield false positives among typically developing children.

There was substantial agreement between NODA and IPA for diagnostic categories (ASD, non-ASD) on the basis of the *DSM–5*. Confidence intervals were quite large for the statistics measuring agreement, which may be due to the relatively small sample size in this initial validation study. Sensitivity was the same in the analyses of the full sample and the EV subgroup, but specificity dropped from 94.4% to 85.7% because fewer true negative cases were included once TD children were removed. Kappa coefficients were 0.75 (full sample) and 0.58 (EV subgroup) for comparing diagnostic outcomes between NODA and IPA and 0.56 (full sample) and 0.37 (EV subgroup) for interrater reliability. To evaluate kappa, one must consider the number of codes to be assigned in the comparison when determining the level of accuracy represented by kappa (Bakeman & Quera, 2011). As the number of codes increases, so does the magnitude of kappa for an associated level of accuracy (e.g., a kappa of 0.30 represents 85% accuracy when there are two codes, but to achieve 85% with five codes, a kappa of 0.64 is required). Because there were only two codes in this study (ASD, not ASD), the kappa coefficients indicate 85% to 90% accuracy between IPA and the primary NODA rater, as well as between the primary and secondary NODA raters.

In the full sample, ICCs indicated moderate to high agreement between IPA and NODA, and between raters regarding specific *DSM–5* symptom criteria. These results were inflated due to the inclusion of typically developing children. In the EV subgroup, the ICC between IPA and NODA was .60. Inspection of the data revealed the greatest number of disagreements in three criteria pertaining to restricted, repetitive patterns of behaviors and interests (i.e., B2 to B4). The number of disagreements on each of these items was nearly double the number of disagreements on A1 to A3 and B1 (e.g., seven for A2, A3, and B1, and 14 for B2). The lower ICC may also be due to the fact that ratings were made on different information. That is, the IPA ratings were based on information collected with assessments during the IPA, and the NODA ratings were based on behaviors captured on video in the home setting. Agreement between the NODA raters was higher, and although ratings were based on the same information (behaviors captured on video at home), the greatest number of disagreements were observed on the same three criteria. These analyses suggest that behaviors related to rigidity (B2), fixated interested (B3), and hyper- or hyporeactivity to sensory input (B4) may be the most difficult symptoms to detect with NODA. More specific questions on the developmental history questionnaire may help to compensate for this difficulty.

Due to the heterogeneous presentation of ASD, any one assessment method and clinical judgment is likely associated with some level of outcome variability. In this project, NODA disagreed with the IPA in six cases. These participants had higher cognitive abilities according to the IPA, fewer tagged behaviors in NODA, and significantly lower confidence scores from the primary rater in comparison to the confidence scores from concordant cases. Although the sample of discordant cases was small and results must be interpreted with caution, they suggest that children with higher cognitive ability and fewer observable behaviors may require additional assessment to determine the appropriate diagnosis. Notably, of the six discordant cases, the ADI–R and ADOS–2 disagreed in four cases (66.7%) compared to only nine disagreements among the 34 concordant cases (27.5%). This lack of consensus on standardized, gold-standard assessments is illustrative of the diversity of clinical presentation and the likelihood that IPA results may also vary among different diagnosticians depending on which methods of assessment they employ. In practice, a lower confidence score by the NODA rater could serve as a decision point for bringing the child in for an IPA or perhaps sharing the information with a second or even third NODA rater.

The identification and recruitment procedure for secondary raters emphasized NODA's ability to connect families to clinical professionals regardless of location. Secondary raters were located in different regions of North America and were able to complete NODA assessments on their own schedule (e.g., evenings and weekends) with relative ease after just 30 min of training on using the system. Most reported completing a single diagnostic assessment in less than an hour. Thus, NODA has potential to improve efficiency of the diagnostic process by creating easy access to professionals regardless of location.

Clinical judgment is a vital component in the IPA, and it plays a prominent role in NODA as well. NODA informs clinical judgment with data collected by families in their home and provides the clinician with a systematic and structured way to annotate behavior examples to support diagnostic determinations. With NODA, diagnosis is not based solely on observed behaviors present in one or two short video segments, and methods that attempt to do so have been observed to be less reliable (Gabrielsen et al., 2015). Instead, parents are guided to record specific scenarios that occur naturally in most homes and are given simple instructions to create opportunities for the child to express typical social communication. Still, clinical judgment is often based on a two-way exchange of information between patients and clinicians rather than a single opportunity to share information. Although not utilized in this initial validation study, the NODA system also includes a feature to allow raters to request additional information from families (e.g., rerecording a scenario with additional social presses from the parent), which shows up in the form of an alert within the family's NODA application (see Nazneen et al., 2015, for more details). This feature provides an additional opportunity for the rater to solicit clinically relevant information to clarify the nature of the child's behavior and perhaps improve the accuracy and confidence of clinical judgment.

NODA conveys the information needed for an initial diagnosis of ASD for most children. It is not intended to eliminate the need for future evaluations but to accelerate the pathway to treatment. Practice parameters indicate the need for additional evaluations to identify potential factors responsible for the developmental delay and for treatment planning (Volkmar et al., 2014); neither is necessary for the initial diagnosis. Likewise, the *DSM–5* includes several terms to specify severity of the disorder that may vary by context and fluctuate over time (American Psychiatric Association, 2013). Thus, these features are to provide additional information to help further characterize the individual's presentation once the diagnostic criteria are satisfied and are not a necessary component of the initial diagnosis. NODA is intended only to accelerate the diagnostic process by improving access to professionals who can provide information to parents about their child's development. The sooner parents get this information the sooner they can pursue a behavioral intervention program, the recommended treatment for developmental delays (Howard, Sparkman, Cohen, Green, & Stanislaw, 2005).

Although there are many potential benefits of a store-and-forward telehealth approach to diagnosis, this study focused on only the initial validation of NODA in making a diagnostic determination of ASD. Results indicate this approach can yield diagnostic information comparable to that of an IPA for most children. Other benefits should be carefully investigated. One goal of telehealth is to decrease the time between parent concerns and diagnosis. Randomized controlled trials in active diagnostic centers can determine whether NODA can actually decrease time from parents' concerns to receiving a diagnosis of ASD and also decrease time until they access intervention. An additional potential use of this approach is to triage cases on waiting lists for diagnostic assessments to separate clear-cut cases from children who will require an IPA to make the initial diagnostic determination. Also, NODA may be used to supplement an IPA for more complex cases where the clinician wishes to observe how the child behaves at home. Finally, the social validity of the procedure should also be investigated to better understand parent impressions for collecting videos on their child and sharing them remotely with a clinician they never met who, in turn, is evaluating their child's behavior.

In practice, NODA is designed to generate a detailed report that describes the specific behavioral examples (tagged in the videos)

that support each *DSM–5* criterion, a clinician summary, and recommendations for next steps. Possible modes of delivery include electronic delivery of the report alone or along with an opportunity to consult remotely with the NODA clinician. Alternatively, the report can be released to the referring diagnostic professional, who can meet with the parent in person, explain the results, and offer their own clinical interpretation. The optimal delivery of the final report generated from NODA needs to be investigated.

Limitations and Future Directions

This study demonstrated accuracy of a novel telehealth approach that may improve the diagnostic process for ASD; however, some limitations exist that should be considered when interpreting the findings. For one, the IPA was conducted before families completed NODA. Thus, parents may have learned about their child's behavior and development during the process, which may have influenced the type of behavior they captured on video for NODA. To minimize the possible order effect, we did not discuss results of the IPA with parents until after NODA videos were collected. Parents were not given information about the diagnostic relevance of their child's specific behaviors until after the NODA videos were obtained. Additionally, video collection was semistructured (i.e., uniform duration of 10 continuous minutes across four specific scenarios, and instructions for parents to shape the interaction), which makes it unlikely that parents would be able to selectively capture behavior that supports or does not support a diagnosis of ASD. By design, NODA does not allow families to pause and restart videos, which should reduce the possibility of families' submitting biased video footage. Future research may examine whether NODA's accuracy differs as a function of the order of IPA and NODA. Further, sampling bias may have inflated the rate of ASD cases (33 of 40 = 82.5%) among families seeking an evaluation. Some participants may have been previously identified with developmental delays but were never evaluated for ASD and their parents may have participated in this study for the free evaluation. This possible bias should be considered when interpreting the effect of the high rate of ASD diagnosis.

This study included only two participant groups (TD and EV) and two outcome categories (ASD and non-ASD). The utility of NODA may be improved by including a third category to classify children as non-ASD but developmentally delayed. For some children the primary evidence for delays is the absence of typical behavior, and a comparison to the rates of typical behavior expressed by TD children may be helpful in determining a diagnostic category. Pilot data were collected from TD children in this project to quantify rates of typical behavior, but this topic needs to be explored in focused investigation in a much larger sample. The resulting normative standards from future efforts may help to support a diagnosis of ASD or developmental delays for some children. Differential diagnosis for developmental disorders is a key area for future inquiry with NODA.

Determining reliability and validity of a new diagnostic method for a disorder as complex as ASD requires a series of studies conducted over time. Although the results of this project provide strong preliminary evidence for NODA, data were collected in a relatively small sample of participants ages 18 to 71 months. The broad age range may have limited the applicability of NODA to a more specific age group (i.e., early childhood). Further, NODA was designed to improve efficiency of the diagnostic process for ASD, but the present study addressed diagnostic accuracy in comparison to only the IPA and interrater reliability. Thus, reliability, validity, and efficiency of NODA need to be further investigated in future studies with larger samples.

References

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

Autism and Developmental Disabilities Monitoring Network Surveillance Year 2000 Principal Investigators. (2007). Prevalence of autism spectrum disorders—Autism and developmental disabilities monitoring network, six sites, United States, 2000. *MMWR, 56(SS01),* 1–11. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5601a1.htm

Autism and Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators. (2014). Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR, 63*(SS02), 1–21. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6302a1.htm

Baharav, E., & Reiser, C. (2010). Using telepractice in parent training in early autism. *Telemedicine and e-Health, 16,* 727–731. http://dx.doi.org/10.1089/tmj.2010.0029

Bakeman, R., & Quera, V. (Eds.). (2011). Observer agreement and Cohen's kappa. *Sequential analysis and observational methods for the behavioral sciences* (pp. 57–71). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139017343.006

Charman, T., & Gotham, K. (2013). Measurement issues: Screening and diagnostic instruments for autism spectrum disorders—Lessons from research and practice. *Child and Adolescent Mental Health, 18,* 52–63. http://dx.doi.org/10.1111/j.1475-3588.2012.00664.x

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

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities, 26,* 359–383. http://dx.doi.org/10.1016/j.ridd.2004.09.005

Howlin, P., Magiati, I., & Charman, T. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities, 114,* 23–41. http://dx.doi.org/10.1352/2009.114:23-41

Huerta, M., & Lord, C. (2012). Diagnostic evaluation of autism spectrum disorders. *Pediatric Clinics of North America, 59,* 103–111. http://dx.doi.org/10.1016/j.pcl.2011.10.018

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test* (2nd ed.). Bloomington, MN: Pearson.

Liptak, G. S., Benzoni, L. B., Mruzek, D. W., Nolan, K. W., Thingvoll, M. A., Wade, C. M., & Fryer, G. E. (2008). Disparities in diagnosis and access to health services for children with autism: Data from the National Survey of Children's Health. *Journal of Developmental and Behavioral Pediatrics, 29,* 152–160. http://dx.doi.org/10.1097/DBP.0b013e318165c7a0

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule: ADOS*. Los Angeles, CA: Western Psychological Services.

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

Mandell, D. S., Novak, M. M., & Zubritsky, C. D. (2005). Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics, 116,* 1480–1486. http://dx.doi.org/10.1542/peds.2005-0185

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

Nazneen, N., Rozga, A., Smith, C. J., Oberleitner, R., Abowd, G. D., & Arriaga, R. I. (2015). A novel system for supporting autism diagnosis using home videos: Iterative development and evaluation of system design. *JMIR Mhealth Uhealth, 3,* e68. http://dx.doi.org/10.2196/mhealth.4393

Oberleitner, R., Laxminarayan, S., Suri, J., Harrington, J., & Bradstreet, J. (2014). The potential of a store and forward tele-behavioral platform for effective treatment and research of autism. In *Engineering in Medicine and Biology Society, 2004: 26th Annual International Conference of the IEEE* (Vol. 2, pp. 3294–3296). Piscataway, NJ: IEEE Service Center.

Parmanto, B., Pulantara, I. W., Schutte, J. L., Saptono, A., & McCue, M. P. (2013). An integrated telehealth system for remote administration of an adult autism assessment. *Telemedicine and e-Health, 19,* 88–94. http://dx.doi.org/10.1089/tmj.2012.0104

Rice, C., Carpenter, L., Bradley, C., Lee, L.-C., Pettygrove, S., Morrier, M., . . . Baio, J. (2014, May). *Diagnostic testing practices for autism spectrum disorder (ASD) in four U.S. communities*. Paper presented at the 2014 International Meeting for Autism Research (IMFAR), Atlanta, GA. Abstract retrieved from https://imfar.confex.com/imfar/2014/webprogram/start.html

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

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. http://dx.doi.org/10.1037/0033-2909.86.2.420

Smith, C., Oberleitner, R. S., Treulich, K., McIntosh, R., & Melmed, R. (2009, May). *Naturalistic Observation Diagnostic Assessment—The "NODA" pilot project*. Paper presented at the 2009 International Meeting for Autism Research (IMFAR), Chicago, IL. Abstract retrieved from https://imfar.confex.com/imfar/2009/webprogram/start.html

Sparrow, S. S., Cicchetti, V. D., & Balla, A. D. (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). Circle Pines, MN: American Guidance Service.

Thomas, K. C., Ellis, A. R., McLaurin, C., Daniels, J., & Morrissey, J. P. (2007). Access to care for autism-related services. *Journal of Autism and Developmental Disorders, 37,* 1902–1912. http://dx.doi.org/10.1007/s10803-006-0323-7

Vismara, L. A., Young, G. S., & Rogers, S. J. (2012). Telehealth for expanding the reach of early autism training to parents. *Autism Research and Treatment, 2012: Article ID 121878*. http://dx.doi.org/10.1155/2012/121878

Volkmar, F. R., & Klin, A. (2005). Issues in the classification of autism and related conditions. In F. R. Volkmar, A. Klin, R. Paul, & D. J. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (Vol. 1, 3rd ed., pp. 5–41). New York, NY: Wiley. http://dx.doi.org/10.1002/9780470939345.ch1

Volkmar, F., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., State, M., & the American Academy of Child and Adolescent Psychiatry (AACAP) Committee on Quality Issues (CQI). (2014). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 53,* 237–257. http://dx.doi.org/10.1016/j.jaac.2013.10.013

Wainer, A. L., & Ingersoll, B. R. (2015). Increasing access to an ASD imitation intervention via a telehealth parent training program. *Journal of Autism and Developmental Disorders, 45,* 3877–3890. http://dx.doi.org/10.1007/s10803-014-2186-7

Wiggins, L. D., Baio, J., & Rice, C. (2006). Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *Journal of Developmental and Behavioral Pediatrics, 27*(Suppl), S79–S87. http://dx.doi.org/10.1097/00004703-200604002-00005

Received November 19, 2015
Revision received March 3, 2016
Accepted March 4, 2016 ■