SciMedCentral

Annals of Pediatrics & Child Health

Research Article

The Reliability and Validity of a Clinical Observation Tool and Scoring Guide for the Diagnostic Evaluation of Autism Spectrum Disorder by Community Pediatricians

Elizabeth A. Jimenez[1]*, Virginia F. Wright[1], Jessica Brian[1], Michelle Shouldice[2], Michelle Gordon[3], Sarah Barker[3], Nicky Jones-Stokreef[3], Janine Flanagan[4], Lisa Feitelberg[2], Suzanne Stead[3], Pamela Green[1], Rachel Barber[2], and Wendy S. Roberts[5]

[1]*Holland Bloorview Kids Rehabilitation Hospital, Canada*

[2]*The Hospital for Sick Children, Canada*

[3]*Orillia Soldiers' Memorial Hospital, Canada*

[4]*St. Joseph's Health Centre, Canada*

[5]*ISAND, Integrated Services for Autism and Neurodevelopmental Disorders, Canada*

***Corresponding author**

Elizabeth Jimenez, Holland Bloor view Kids Rehabilitation Hospital, 50 Kilgour Road, Toronto Ontario, M1C 3E9, Canada, Tel: 416-753-6057; Fax: 416-753-6046; Email: ejimenez@hollandbloorview.ca

Submitted: 24 January 2016

Accepted: 19 August 2017

Published: 22 August 2017

ISSN: 2373-9312

Keywords

- Autism spectrum disorder
- Early intervention
- Early diagnosis
- Community pediatrician

Abstract

Existing direct diagnostic tools for ASD tend not to be practical for use by most community paediatricians. We examined the potential of using selected activities modified from the Autism Diagnostic Observation Schedule (ADOS) to inform diagnostic decision-making.

Objectives: We evaluated the use of a unique, specially designed observation form and scoring guide (Pediatric Autism Early Diagnosis Tool; PAED), and then compared the diagnostic results with the validated ADOS Module 1. Completing a diagnostic assessment using this new method does not involve the complex scoring normally required with the ADOS, thus making it much more accessible for pediatricians to learn.

Methods: Twenty-eight pre-verbal children (mean age 33.6 months; SD=9.5m) referred to community clinics were assessed on the abbreviated process/PAED Tool and on the standard ADOS on separate visits. Both diagnostic methods included a detailed developmental history for ASD. A DSM-IV-TR diagnostic category was assigned for each approach. Videotaped sessions were scored by independent raters for PAED Tool reliability.

Results: PAED Tool inter-rater reliability was excellent using a video-scoring approach (ICC=0.86). Classification agreement between evaluation methods was excellent (weighted kappa=0.83), and sensitivity and specificity were both high (91%-100%).

Conclusions: When used with a detailed history and physical examination, the abbreviated battery and PAED Tool yielded good diagnostic accuracy. Further trials will involve DSM-5 modifications.

ABBREVIATIONS

ASD: Autism Spectrum Disorder

INTRODUCTION

Autism Spectrum Disorder (ASD) is a developmental disorder characterized by impairments in communication, social interaction and restricted, repetitive, and/or stereotyped behaviors/interests [1]. Current estimated rates of ASD are 1 in 68 children [2], with prevalence rates increasing dramatically in recent decades. Given the increase in recognition/prevalence of ASD, pediatricians in the community are increasingly likely to encounter a child with suspected ASD.

Outcomes in ASD can be improved with early diagnosis and intervention. Speech and language and targeted behavioral interventions increase skill development and reduce unwanted behaviors [3]. Indeed, current standards of care for children with ASD include referral for speech and language therapy and intervention approaches such as Applied Behavior Analysis/

Intensive Behavioral Intervention [4]. Admission for many ASD-specific programs typically depends on an accurate diagnosis supported by appropriate documentation.

The diagnosis of ASD is made based on clinical criteria specified in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR - text revision) [5], and now DSM-5 [1], and the ruling out of other possible developmental diagnoses. In order to determine whether these criteria are met, a thorough and accurate developmental history and direct observation of specified skills are required. Published position statements and practice guidelines emphasize that a comprehensive evaluation should include determination of a categorical DSM based diagnosis, preferably with standardized tools that operationalize the DSM criteria [6]. The DSM-IV-TR criteria for autism have shown good sensitivity and specificity [7].

While it is difficult to sub-classify the various forms of ASD, in some jurisdictions funding sources try to differentiate the more severe forms of autistic disorder from the milder "higher functioning" forms (previously Pervasive Developmental Disorder-not otherwise specified, PDD-NOS and Asperger Syndrome) [5]. Although the term ASD is now used as the umbrella term to describe the entire population of individuals with the full range of autism spectrum conditions, we will adhere to the severity differentiation between 'Autism' and 'ASD' for the purposes of the current paper.

There are several diagnostic and screening measures used by psychologists, developmental pediatricians and child psychiatrists to assist in the diagnostic assessment. In reviewing recent literature, we found that most of the published measures focus on parent-reported descriptions of a child's behavior. These include the CARS-2 (Childhood Autism Rating Scale-2), Gilliam Autism Rating Scale, and the M-CHAT (Modified Checklist for Autism in Toddlers). However, when completing a diagnostic assessment, it is necessary to include both observational measures and parent report. This allows the diagnostician to have all of the important information available, so the most accurate diagnosis is made. Currently in the literature, there are only a few measures published that involve observation of a child's skills and behaviors, and ones that have been published to date are described briefly below.

The Screening Tool for Autism in Two-year–olds (STAT) [8], is a specific ASD screening tool rather than a diagnostic measure. It is used in the clinical setting to identify children aged 24-35 months, at risk for ASD. The Communication and Symbolic Behavior Scales -Developmental Profile (CSBS-DP) [9], is designed to identify children between 12 and 24 months of age at risk for general developmental delay rather than ASD in particular. Again, the CSBS-DP is a screening tool and not a diagnostic measure. However, a specific scoring system (Systematic Observation of Red Flags) linked with the CSBS allows identification of those at risk for ASD [10].

There are a few additional ASD screening tools involving observational components that have been developed in the last few years. One is called the ADEC (Autism Detection in Early Childhood) [11]; it was found by one research group to have high internal consistency, and adequate reliability when analysed for its ability to screen for ASD. The RITA-T (Rapid Interactive Screening Test for Autism in Toddlers), a clinical process for diagnostic decision-making [12], was evaluated in an initial study. It had excellent sensitivity for identifying ASD, and moderate specificity/positive predictive value in a high risk sample. One additional interactive screen was described in 2014. The Three-item Direct Observation Screen (TIDOS) for autism spectrum disorder [13], is a screening measure that assesses joint attention, eye contact, and responsiveness to name. In an initial study to explore the usefulness of this screen, each of the observed screening items showed moderate sensitivity and specificity for identifying ASD, except for responsiveness to name, which had poor sensitivity.

The issue with all of the observational measures described above is that they are useful for screening for ASD, but a screen is only part of the process of identifying children with ASD early, with an accurate diagnostic process. What is needed is a brief standardized observational tool that can be used as part of a diagnostic assessment (in conjunction with an appropriate developmental history and physical exam). Having children screened accurately for ASD is important, but having children accurately diagnosed when they are young is in fact the most important goal of the clinical assessment process. In the end, if more children are diagnosed with ASD more efficiently and accurately, this will allow them to more quickly receive intervention. This is a particularly important issue being faced in underserved communities internationally.

The Autism Diagnostic Observation Schedule (ADOS) [14], and complementary Autism Diagnostic Interview (ADI) – Revised [15], are more comprehensive than all of the tools described above, and are considered the closest thing to "gold standard" measures in ASD diagnostic protocols. The ADOS takes the form of a semi-structured assessment of social interaction, communication, play, and imaginative use of play materials. The items operationalize the DSM-IV and ICD-10 criteria for diagnosis of Autism or ASD. The ADOS is available in four modules, and one is chosen based on the developmental and expressive language level of the child. The pre-verbal module (Module 1) has 10 activities and 29 rating categories. The scoring algorithm for social behaviors and communication skills assists in the differentiation between a diagnosis of Autism and ASD. These scores are then considered in light of the findings from the clinician's physical exam and history (sometimes supported by the ADI-R), observation and other information to complete the diagnosis. Each module of the ADOS takes at least 30 to 45 minutes to administer, and requires special testing equipment. Users also require specific training and reliability checks in order for the test results to be considered valid. A revised version of the ADOS (the ADOS-2) [16], corresponds to DSM-5 criteria.

Considerable work has been done to establish the psychometric properties of the ADOS; it has demonstrated excellent inter-rater reliability in live versus video scoring and live versus live scoring, excellent intra-rater reliability using live ratings with test repeated once during a 9-month interval, excellent internal consistency, strong discriminative validity, and good sensitivity and specificity when used to differentiate classic Autism from milder forms of ASD [14]. The issue of live versus

video scoring for reliability was specifically addressed by the ADOS' developers. Indeed, live versus video scoring needs to be evaluated whenever the options for either approach exist within a clinical test. In the case of the ADOS, live ratings were optimal from a reliability perspective.

A significant drawback of the ADOS for community practice is that it requires specific training and demonstration of rater reliability in administration and scoring prior to use in clinical practice. The ADOS is currently the recommended standardized tool for young children undergoing diagnostic assessment for ASD [6]. Despite this, most general pediatricians find it difficult to use the ADOS on a routine basis as it is time consuming to learn how to administer and score. This is a concern, since access to specialists is limited in many communities, and general pediatricians are increasingly required to confirm or rule out diagnoses of ASD in order to ensure timely referral to intervention programs. While their history taking skills are likely appropriate for this task, general pediatricians would be aided by a practical standardized measure to document their observations (analogous to a physical examination to elicit physical signs of a disorder).

In response to these needs,, our research team developed this abbreviated assessment battery (WR, PG, JB, MS) to provide a structured process for pediatricians to use in the diagnosis of ASD in community practice. We focused on developing a series of activities appropriate for use with young children who are non-verbal or using single words. The assessment battery's content consists of six significantly modified activities from the ten activities that make up the ADOS Module 1 (i.e., unstructured play, responding to name, bubbles, snack, pretend play, and social games). These items were chosen for their expected ability to inform diagnosis, together with potential ease and speed of administration, and minimal requirement for special equipment. Very simple rules were then constructed around administering these six items, replacing the ADOS' more rigorous specifications around toys and scripts used. Training on this abbreviated tool for community pediatricians has been possible in a half day workshop using demonstrations and videos.

The abbreviated battery testing process begins with observation of the child engaged in unstructured play. The test's six activities are then introduced into the flow of play. The pediatrician records on an assessment sheet (i.e., the Observation Record) descriptive comments about the child's behavior during each of these activities. There are written explanations on this Record to give cues as to what behaviors to focus on while observing the child's performance of each item.

Immediately following the assessment, information from the written notes is transferred from the Observation Record to a summary score form, the Pediatric Autism Early Diagnosis (PAED) Tool (see Appendix 1). On the PAED summary form, the pediatrician writes a short descriptive summary of 16 behavior items observed during the assessment battery's six activities; these behavior items (e.g. pointing to show, gestures) map directly onto DSM criteria. Next, the pediatrician fills out a DSM-IV-TR checklist for autistic disorder, informed directly by the PAED. The complex scoring that is part of completing the ADOS, is not necessary using this new tool.

Completing the abbreviated assessment allows the pediatrician to more accurately map the child's behaviors onto the three categories of impairment in ASD (communication, social interaction, and restricted and repetitive patterns of behavior). Overall the administration of the abbreviated test, together with completion of the PAED Tool and review of DSM criteria, was designed to take no more than 15 to 20 minutes.

We report here on our pilot findings using this abbreviated test with non-verbal or minimally verbal children. We had three research objectives: (1) to determine the inter- and intra-rater reliability of the PAED Tool scores that are derived from administration of the abbreviated test; (2) to evaluate whether the PAED Tool item scores accurately assist with diagnostic assignment in preschool children; and (3) to examine the DSM-IV-TR criteria identified by use of the abbreviated test and resulting PAED Tool scores to see if the same DSM-IV-TR criteria were identified using the ADOS with each child. To address our aims, we evaluated the consistency of scoring between and within trained raters, and compared the abbreviated test's PAED Tool scores with the full ADOS in terms of diagnostic agreement (i.e., sensitivity and specificity). We hypothesized that the selected items used for the abbreviated test would reliably allow operationalization of the DSM-IV-TR criteria, yielding similar diagnostic decision-making to that based on the full ADOS.

MATERIALS AND METHODS

This was a prospective measurement study with a convenience sample of children aged 2 to 6 years inclusive, with preverbal language level of 'less than phrase' speech. They were referred by community pediatricians or family physicians with suspected ASD, language and social concerns, and/or repetitive behaviors. The referrals were received at an academic children's rehabilitation/child development center (Holland Bloorview Kids Rehabilitation Hospital) in a large urban center, or one of its three satellite clinics (one was in a city 100 km away). The assessment process leading up to the diagnosis of Autism, PDD-NOS or Not Autism is typically done at these clinics using a series of two to three visits to permit completion of all components of the assessment, and to give the team time for sufficient consideration of findings before the final diagnosis is discussed with the family. The division of the various assessment components into separate test days is done to avoid child fatigue from prolonged or repeated testing of similar behaviors/abilities. It is essential for these tests that the child is functioning at his/her highest level of ability and interest, and that the tasks presented on a single testing day are novel for that session.

The study received approval from the lead center's (Holland Bloorview Kids Rehabilitation Hospital) research ethics board, and subsequent administrative approval at each of the satellite clinics, and each parent provided signed informed consent for participation in the study. Each parent also provided written consent for their child's assessments to be videotaped.

Abbreviated test and use of the PAED Tool

In this study, the usual clinical assessment (i.e. developmental history and physical examination) was completed at the first visit before administration of the abbreviated test or the ADOS Module

1 (Table 1). This information always precedes the administration of a structured assessment in our center's clinical context.

The abbreviated test was then administered and scored by the study's trained clinical PAED Tool examiner for that site. Each of the four examiners (two were board certified pediatricians and two worked in intake/family support roles in the clinics) had undergone a three hour standardized training by the authors (EJ, JF, NJ, WR) on the abbreviated test battery to learn how to administer it and score using the PAED Tool, and then assign DSM codes. Very importantly, none were ADOS trained, so there was no risk that skill levels related to ADOS administration and scoring would influence the use of the abbreviated test. One was considered a clinical expert as she had considerable experience using the PAED in clinic prior to participation in this study, while the others had no more than minimal experience with PAED Tool use in the community clinic and thus were more typical of community practitioners for whom the abbreviated test was designed.

The child's parents were present during the abbreviated test (as they would be in a clinical context), and the assessment was videotaped for use in the inter- and intra-rater evaluations (see the reliability study scoring schedule in Table 2). The videotaping approach was one typically used in our clinics, with the video camera positioned in the corner of the room, capturing all of the assessment action without being a source of distraction or anxiety to the child. During the abbreviated test's administration, the examiner used the Observation Record to note descriptive comments on the child's performance in each of the six activities. At the completion of live testing, the examiner filled out the abbreviated test's PAED Tool score form and then reviewed the findings from the developmental history and physical examination. She then selected the DSM-IV-TR criteria which were met, assigning a diagnostic category of Autism, PDD-NOS or Not Autism/ASD based on the DSM-IV-TR criteria.

Standard ADOS assessment

The ADOS Module 1 for preverbal children was administered and scored at a subsequent visit within 2 weeks (Table 1) at the same center by a different trained examiner (i.e., one of three ADOS examiners), blinded to the results of the abbreviated test. These examiners were designated users of the ADOS in their clinic setting before participating in the study. The ADOS examiner also was given the written results from the developmental history, other history and physical examination (i.e., the same information exactly as the other examiner had). After completing the ADOS scoring module, the examiner also selected the criteria met on the DSM score sheet based on all of the information available including the ADOS scores, and assigned a diagnostic category of Autism, PDD-NOS or Not Autism/ASD. The ADOS was also videotaped so that any scoring queries that arose could be checked by the research assistant and ADOS examiner. For the child's third visit, the family met with the clinic team and discussed the child's diagnosis and recommendations (Table 1), and while this is an integral part of the usual clinical care cycle, it was not evaluated for this study.

Video-scoring of the abbreviated test using the PAED Tool

Videotape scoring of this test was done by the examiner who had conducted the test with the child (designated generically as rater A), and occurred at least two weeks after the live assessment (intra-rater reliability) so that impact of memory on scores was minimized. This examiner was also blinded to the score from the live assessment session. It is possible that important nuances of the child's behavior obtained in the course of live assessment may be missed in the live scoring process while administering the test and simultaneously rating. Thus, we also had the live examiner (rater A) rate the videos a second time (at least two weeks later) to permit comparative evaluation of reliability of video coding. In both cases, the PAED Tool descriptive information was used by the examiner to fill out the DSM-IV-TR score sheets, and assign a diagnostic category based on DSM cut-off scores noted above.

Two evaluations of inter-rater reliability were possible. At all three sites, the child's video was scored by a separate abbreviated test trained examiner (rater B) who had not been present at the child's live assessment. The video-ratings of rater B were compared with those of rater A. Since there was the possibility that rater A's experience in rating the child live would have provided different insights into the child's abilities, we decided to also do an inter-rater evaluation with two examiners who had not seen the live assessment. This second more stringent look at reliability was possible for a subsample of the children (i.e., at two of the centers) where there was a second independent examiner available (rater C) who had not been present at the child's live assessment. Prior to scoring, the independent examiners separately reviewed the summary of the child's history and physical examination as had been done in the live-rating session, and then independently viewed and scored the videotaped abbreviated assessment using the Observation Record and PAED Tool. The DSM-IV-TR checklist for autistic disorder was then filled out using the same process as with the live assessor, allowing for determination of a diagnostic category based on DSM cut-off scores

Sample size

The target was to include 25 to 30 children for the reliability and sensitivity/specificity evaluation aspect of this study to ensure a sufficient number of children. A sample size of 20 to 30 subjects is typically used in rehabilitation–based reliability studies [17], and was sufficient to support the planned analyses [18]. Based on current experience in these clinics, enrolment of children sequentially from the referral list was expected to result in 60 to 80% of children screened having Autism or ASD, with the remaining 20 to 40% being children with communication, language and other non-autism developmental disorders.

Analyses

All statistical analyses were conducted using the statistical software package MedCalc.

Descriptive statistics [19] were applied to summarize participant characteristics and test results.

Reliability analyses: The DSM-IV-TR summary scores as

derived from the PAED Tool (out of 12 points reflecting the 12 DSM-IV-TR linked criteria) were calculated for each assessment, and the reliability of these scores was determined using Intraclass Correlation Coefficients (ICCs) and 95% CI's [20]. For both inter- and intra-rater reliability, the data of all examiners across participants from the four participating sites were pooled. Bland-Altman plots [20] (also known as Tukey Mean-Difference plots [21]) were constructed to determine if there were any systematic scoring biases. A sub-analysis of intra-rater reliability for pediatricians (n=2) and clinicians (n=2) was also conducted. This could not be done for inter-rater reliability as the pairings were sometimes between novice and expert raters (as would reflect clinical practice in which a diversity of experience is expected).

5.5.2. Diagnostic accuracy of the abbreviated battery (validity evaluation): The diagnostic category determined by the PAED Tool from the live administration of the test was compared with that obtained using the standard ADOS. An agreement analysis assessed the level of concordance for the three rating categories (Not Autism, PDD-NOS and Autism) using weighted kappa [20]. The primary sensitivity and specificity analysis was also conducted for Not Autism vs. Autism/ASD differentiation. In each case, two by two tables were set up to allow us to identify true positive (sensitivity) and true negative (specificity) rates [20]. Published work [14] with the ADOS Module 1 indicated greater than 80% sensitivity and specificity rates for Autism vs. non-ASD. Since the standard ADOS and abbreviated battery are used in conjunction (i.e., counter-checked) with a family interview and physical assessment before a diagnosis is given, this target level of accuracy was considered by our team to be sufficient.

RESULTS

The final sample consisted of 28 children (17 male); mean age was 33.3 months (SD =9.5; range: 22-59 months). For the inter-rater reliability subsample (n=17; 12 male), the mean age was 33.2 months (SD=9.6, range: 22-55 months). The PAED Tool mean scores (/12) are shown in Table 3.

Reliability

Intra-rater reliability (n=28) of the abbreviated test and PAED form was high when scored from video-viewing (A_{V1} versus $A_{V2,}$ n =27) (ICC = 0.92; 95%CI =0.83-0.96), and moderate for live versus video-viewing (A_{L1} versus A_{V1}) (ICC=0.72, 95%CI=0.48-0.86). For the intra-rater subgroup evaluation with pediatrician raters and clinical raters (i.e., A_{V1} versus A_{V2}), reliability was excellent in both groups, ICC=0.87 and 0.91 respectively).

Inter-rater reliability was fair (ICC=0.66, 95%CI=0.39-0.82) for assessing versus an independent rater (A_{V1} versus B_{V1}). However, when scoring from video, inter-rater reliability was excellent (ICC=0.86, 95%CI=0.66-0.95) when two raters (B_{V1} vs. C_{V1}) who had not been at the live session scored from video (n=17). Bland-Altman plots (Figure 1) did not reveal any scoring bias with the exception of the (A_{L1} versus A_{V1}) intra-rater comparison, in which there was a greater tendency to give higher video rating scores than live scores for children with higher PAED Tool scores (e.g., overestimating issues when able to review the video).

Classification: Based on the abbreviated battery and using live assessor PAED scores, 7/28 children were classified as Not Autism, 4 with PDD-NOS, and 17 children with Autism, while on the basis of the ADOS assessment, five children were classified as Not Autism, three children with PDD-NOS, and 20 children with Autism. The end result was that 4 of 28 children who were not classified in the same category by the PAED and ADOS (Table 4). However, overall classification agreement was excellent: weighted kappa (K) = 0.83, with category agreement being lowest for PDD-NOS (K=0.51). Sensitivity and specificity varied from 91%-100% for Autism/PDD-NOS versus Not Autism and also for Autism versus PDD-NOS.

For a classification of Autism, the child must meet six total of the DSM-IV-TR criteria, including two criteria in DSM-IV-TR section 1, and one each in sections 2 and 3. The agreement between DSM-IV-TR ratings from the PAED Tool versus those from the ADOS (both live performance ratings as shown in Table 2) demonstrated that PAED section 1 (social interaction) was the strongest as far as individual item agreement (K statistics for its 4 items varied from 0.42 to 0.73 with the mean K = 0.57 indicating moderate agreement). In each of PAED Tool sections 2 (communication) and 3 (patterns of behaviors), there were two items (2a/2c and 3c/3d – see DSM-IV-TR item labels in Appendix 1) with K of 0.41 – 0.51, and two items with a K value <0.20. Specifically, for item 2d, 11 of 29 ratings were discordant, and for 8 of these, the use of the ADOS led to identification of more atypical behaviors. With regard to item 3a there were 14 ratings that were discordant and for 10 of them, only use of the ADOS identified an issue. The other two items with poor agreement between the ADOS and PAED Tool (2b and 3b) were affected by no scores reflecting presence of an atypical behavior (i.e., scoring cell 1,1), but there was perfect agreement for 25 and 27 of 'absence of characteristic' ratings (scoring cell 0,0) respectively.

DISCUSSION

This study was a pilot investigation, comparing the use of the full ADOS (already published and widely used in specialized clinical practice and research) to an abbreviated battery, with a specially designed PAED Tool, for its ability to diagnose ASD in the community setting. Reliability was excellent (ICC >0.90) for intra-rater viewing of the same video at two different times and for inter-rater scoring from video. Reliability was compromised in a few specific situations (please refer to the results section). The abbreviated test in combination with the PAED Tool showed excellent accuracy in diagnosis for live score rating, with sensitivity and specificity ranging from 91-100% in comparison with the full ADOS. Given these initial findings, together with reduced training needs, the abbreviated battery can be recommended for use by community practitioners who can benefit from its diagnostic accuracy and relative testing simplicity. Further evaluation with a larger sample and different evaluators will be required in order to confirm the results.

It should be kept in mind that the full ADOS is more comprehensive, involves more activities, and evaluates more specific skills and behaviors than the abbreviated version. Thus, some children with milder presentations may be picked up by the full ADOS when the abbreviated tool may miss them; indeed our results showed that when disagreement occurred between raters

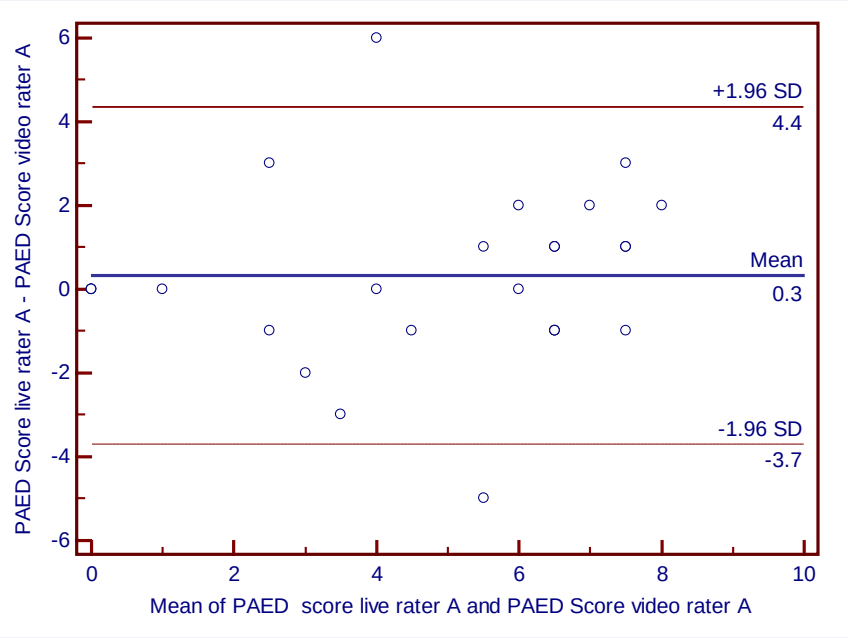

Figure 1 Bland-Altman Plot Intra-rater Reliability.

Table 1: Child's Clinic Visit Schedule.

Visit 1	Visit 2 (~ 2 weeks after visit 1)	Visit 3 (any time after visit 2)
History, physical	Review of history	Review of assessment findings with family and team recommendations
Abbreviated battery administration and PAED Tool scoring	ADOS administration and scoring	
Scoring of DSM-IV-TR	Scoring of DSM-IV-TR	

Table 2: Reliability Assessment Scoring Schedule for the PAED.

*Rater	Session data evaluated	Type of reliability evaluated	Time frame
Assessor 1	Live assessment		
Assessor 1	Video from live assessment	Intra-rater (live versus video) (n=27)	> 2 weeks after live assessment
Assessor 1	Video from live assessment	Intra-rater (video versus video) (n=28)	> 2 weeks after video scoring
Assessor 2	Video from live assessment	Inter-rater assessor 1 vs 2 (video versus video) (n=28)	No time requirement as was not present at live assessment – e.g. first time seeing the child via video
Assessor 3	Video from live assessment	Inter-rater assessor 2 vs 3 (video versus video) (n=17)	No time requirement as was not present at live assessment – e.g. first time seeing the child via video

*Raters who were paired worked at the same site – all trained on the PAED Tool. Only assessor 1 was present at the live assessment

using the DSM-IV-TR form, the ADOS measure ratings consistently identified more atypical behaviors. When considering this diagnostic limitation, it is important to keep in mind that the use of the abbreviated battery should be seen as the first step in the evaluation of a child with a possible ASD, who presents to his or her pediatrician. If, after having completed this evaluation, the pediatrician still feels uncertain about the diagnosis, the child should be referred to a tertiary center for evaluation by a developmental specialist, or specialized team, ideally using the full ADOS. However, in cases where the presentation is more pronounced, a diagnosis should be conferred by the community pediatrician using the abbreviated tool.

The diagnostic criteria published in the DSM-IV-TR were used in the present study, as these were the most up to date criteria available at the time. The unified ASD category proposed by the DSM-5 (published in 2013) aims to produce a clear diagnostic system that will identify the common characteristics of ASD across ages and ability levels [22]. The proposed diagnostic criteria for ASD became available to the public approximately

Table 3: PAED Tool Scores (/12).

Examiner A_L (live)			Examiner A_{V1} (Video 1st review)			Examiner A_{V2} (Video 2nd review)		
n	Mean (SD)	Min – max	n	Mean (SD)	Min - max	n	Mean (SD)	Min - max
28	5.1 (2.9)	0-9.0	28	4.8 (2.6)	0 – 8.0	28	5.4 (2.4)	1.0 -8.0
Examiner A_{V1} (Video 1st review)			Examiner B_{V1} (Video 1st review)			Examiner C_{V1} (Video 1st review)		
17	5.6 (2.0)	0-8.0	17	5.8 (2.7)	0 – 9.0	17	6.0 (2.8)	

Table 4: Scoring of Diagnostic Criteria for 299.00 Autistic Disorder (see from ADOS and PAED tool in Appendix 1).

Child	Characteristic observed* in ADOS assessment	Diagnosis from ADOS assessment	Characteristic observed during PAED Tool assessment	Diagnosis from PAED assessment
A	1b 1d 2d 3a 3b 3c 3 d	PDD	2a 3d	Speech delay
B	1a 1b 2a 2b 2c 3c 3d	ASD	1a 1b 1c 2a 2c 3a 3c 3d	PDD**
C	1d 2b 2c 3a 3c	ASD	1a 1c 1d 2d	PDD
D	1a 1b 2a 2c 2d 3a 3b 3d	ASD	1a 2a	Speech delay

*See Appendix 1 for item definitions
** This child was given a clinical diagnosis of PDD based on the diagnostic information and PAED results

two years before the official publication of the DSM-5, allowing several research groups to compare the accuracy of the two methods. Mahjouri and Lord's review [23], reveals specificity from 90 to 99%, and sensitivity from 10-93% in the larger studies, when looking at agreement between the DSM-5 and DSM-IV-TR diagnostic criteria.

In our recent review of the literature we also looked at research comparing the use of the proposed DSM-5 and DSM-IV-TR criteria, between 2013 and present. Looking at the highest quality studies, we found that most showed excellent agreement between the two diagnostic methods. Thus, most children undergoing diagnostic assessment for ASD using both DSM-IV-TR and DSM-5 will be identified equally well. Therefore, for clinical use, we have now changed the Observation Record and PAED Tool described in this paper to be linked with the use of a DSM-5 checklist (see Appendix 1). However, when using this DSM-5 checklist, clinicians should keep in mind that while the specificity of the DSM-5 is quite good (meaning that there will be fewer false positives that with the DSM-IV-TR), the sensitivity is not quite as good. Therefore there may be children who have a fair degree of clinical impairment who would benefit from intervention, but who might not be identified using the DSM-5 criteria.

LIMITATIONS

Study limitations include the small numbers in the inter-rater sample (n=17), and that there were not as many children in the non-ASD group as hoped for when designing the study; thus the abbreviated test and PAED Tool were not put to their fullest test.

With respect to future use of the PAED Tool, two items with rating issues from DSM-IV-TR sections 2 (communication – item d: lack of varied, spontaneous make-believe play or social imitative play…) and 3 (patterns of behaviors – item a: one or more stereotyped and restricted patterns of interest…) require review to determine whether modifications can be made to the abbreviated tool's scoring details (such as the Observation Record or the PAED Tool) to permit more accurate ratings on the DSM form.

Finally, a live inter-rater component was not included in this study, as in the initial evaluation of the abbreviated test we did not want to introduce another person into the live test situation. In future, this should be evaluated since it cannot be assumed that live and video ratings would necessarily be equivalent.

CONCLUSION

In conclusion, when used alongside the history and physical examination, the abbreviated battery demonstrated acceptable ability to classify children correctly, and holds promise as a shorter, but still reliable observational tool for community pediatricians, who would not be able to complete full ADOS evaluations. Further testing is needed to confirm the results with a larger sample and determine if there are any diagnostic changes given the new DSM-5 criteria.

ACKNOWLEDGEMENTS

Pui Ying Wong, librarian at Holland Bloorview Kids Rehabilitation Hospital assisted with searching of the literature. Joanne Jenah and Rohan Mahabir, administrative assistants at Holland Bloorview Kids Rehabilitation Hospital, assisted with typing and formatting of the manuscript.

REFERENCES

1. Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition (DSM-5). American Psychiatric Association Arlington VA. 2013.
2. Christensen DL, Baio J, Braun KVN, Bilder D, Charles J, Constantino JN, et al. Prevalence and characteristics of autism spectrum disorders among children aged 8 years – Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. Surveill Summar. 2016; 65: 1-23.

3. Anagnostou E, Zwaigenbaum L, Szatmari P, Fombonne E, Fernandez BA, Woodbury-Smith M, et al. Autism spectrum disorder: advances in evidence-based practice. CMAJ. 2014.

4. National standards report: the National Standards Project – addressing the need for evidence –based practice guidelines for autism spectrum disorders. Randolph MA. National Autism Center. 2009.

5. Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition text revision (DSM-IV-TR). American Psychiatric Association. Washington DC. 2000.

6. Plauche Johnson C, Myers SM. Identification and evaluation of children with autism spectrum disorders. Pediatrics. 2007; 120: 1183-1215.

7. Mahoney WJ, Szatmari P, MacLean,J, Bryson SE, Bartolucci B, Walter SD, et al. Reliability and accuracy of differentiating pervasive developmental disorder subtypes. J Am Acad Child Adolesc Psychiatry. 1998; 37: 278-285.

8. Stone W, Coonrod EE, Turner LM, Pozdol SL. Psychometric properties of the STAT for early autism screening. J Autism Dev Disord. 2004; 34: 691-701.

9. Wetherby A. Communication and symbolic behavior scales developmental profile, Preliminary Normed Edition. Paul H. Brookes Publishing Co. 2001.

10. Wetherby A, Woods J, Allen L, Cleary J, Dickinson H, Lord CE. Early indicators of autism spectrum disorders in the second year of life. J Autism Dev Disord. 2004; 34: 473-493.

11. Nah Y, Young RL, Brewer N, Berlingeri, G. Autism detection in early childhood: reliability and validity data for a level 2 screening tool for autistic disorder. Psychol Assess. 2014; 26: 251-226.

12. Choueiri R, Wagner S. A new interactive screening test for autism spectrum disorders in toddlers. J Pediatr. 2015; 167: 460-466.

13. Oner P, Oner O, Munir K. Three-item direct observation screen (TIDOS) for autism spectrum disorder. Autism. 2014; 18: 733-742.

14. Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, et al. The Autism Diagnostic Observation Schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord. 2000; 30: 205-223.

15. Lord C, Rutter M, LeCouteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994; 24: 659-685.

16. Lord CE, Rutter M. The Autism Diagnostic Observation Schedule, Second Edition (ADOS-2). WPS. Torrance CA. 2011.

17. Portnoy LG, Watkins MP. Foundations of clinical research. Applications to practice. East Norwalk, CT: Appleton and Lange. 1993.

18. Walter SD, Elizaziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998; 17: 101-110.

19. MedCalc Software 13.2.2 edition. Ostend Belgium. 2014.

20. Fleiss JL. Statistical methods for rates and proportions. Second Edition. New York: John Wiley and Sons Inc. 1981.

21. Kozak M, Wnuk A. Including the Tukey Mean-Difference (Bland–Altman) Plot in a Statistics Course, Teaching Stats. 2014; 36: 83-87.

22. Happe F. Why fold asperger syndrome into autism spectrum disorder in the DSM-5: Simons Foundation Autism Research Initiative News and Opinion. 2011.

23. Mahjouri S, Lord CE. What the DSM-5 portends for research, diagnosis and treatment of autism spectrum disorders. Curr Psychiatry Rep. 2012; 14: 739-747.

Cite this article

Jimenez EA, Wright VF, Brian J, Shouldice M, Gordon M, et al. (2017) The Reliability and Validity of a Clinical Observation Tool and Scoring Guide for the Diagnostic Evaluation of Autism Spectrum Disorder by Community Pediatricians. Ann Pediatr Child Health 5(4): 1136.