

Identifying Autism in a Brief Observation

Terisa P. Gabrielsen, PhD, NCSP^a, Megan Farley, PhD^a, Leslie Speer, PhD, NCSP^a, Michele Villalobos, PhD^a, Courtney N. Baker, PhD^b, Judith Miller, PhD^a

abstract

BACKGROUND: Pediatricians, neurologists, and geneticists are important sources for autism surveillance, screening, and referrals, but practical time constraints limit the clinical utility of behavioral observations. We analyzed behaviors under favorable conditions (ie, video of autism evaluations reviewed by experts) to determine what is optimally observable within 10-minute samples, asked for referral impressions, and compared these to formal screening and developmental testing results.

METHODS: Participants ($n = 42$, aged 15 to 33 months) were typically developing controls and children who screened positive during universal autism screening within a large community pediatric practice. Diagnostic evaluations were performed after screening to determine group status (autism, language delay, or typical). Licensed psychologists with toddler and autism expertise, unaware of diagnostic status, analyzed two 10-minute video samples of participants' autism evaluations, measuring 5 behaviors: Responding, Initiating, Vocalizing, Play, and Response to Name. Raters were asked for autism referral impressions based solely on individual 10-minute observations.

RESULTS: Children who had autism showed more typical behavior (89% of the time) than atypical behavior (11%) overall. Expert raters missed 39% of cases in the autism group as needing autism referrals based on brief but highly focused observations. Significant differences in cognitive and adaptive development existed among groups, with receptive language skills differentiating the 3 groups.

CONCLUSIONS: Brief clinical observations may not provide enough information about atypical behaviors to reliably detect autism risk. High prevalence of typical behaviors in brief samples may distort clinical impressions of atypical behaviors. Formal screening tools and general developmental testing provide critical data for accurate referrals.

FREE

WHAT'S KNOWN ON THIS SUBJECT: Behavioral observations influence a clinician's decision to diagnose or refer, and may even override formal screening results. In the case of autism spectrum disorder, an expected rate of atypical behavior during the span of a medical visit is unknown.

WHAT THIS STUDY ADDS: We are the first to quantify the high base rates of typical behavior in young children who have autism and language delay. When observation times are brief, the preponderance of typical behaviors may negatively impact referral decision accuracy.

^aDepartments of Psychiatry and Educational Psychology, University of Utah, Salt Lake City, Utah; and ^bDepartment of Child and Adolescent Psychiatry and Behavioral Sciences, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

Drs Gabrielsen and Miller conceptualized and designed the study, recruited and evaluated participants, trained raters to reliability, analyzed the data, and drafted the original manuscript; Drs Farley and Speer generated data for the study as blinded raters and critically reviewed the manuscript; Dr Villalobos participated in design and implementation of data collection, evaluated participants, and critically reviewed the manuscript; Dr Baker designed the analysis and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted.

Dr Gabrielsen's current affiliation is Department of Counseling, Psychology and Special Education, Brigham Young University, Provo, UT. Dr Farley's current affiliation is the Waisman Center, University of Wisconsin-Madison, Madison, WI. Dr Speer's current affiliation is Center for Autism, Cleveland Clinic Children's, Cleveland, OH. Dr Villalobos's current affiliation is Department of Psychiatry, University of North Carolina, Chapel Hill, Asheville, NC. Dr Baker's current affiliation is Department of Psychology, Tulane University, New Orleans, LA. Dr Miller's current affiliation is Center for Autism Research, Children's Hospital of Philadelphia and University of Pennsylvania, Philadelphia, PA.

www.pediatrics.org/cgi/doi/10.1542/peds.2014-1428

DOI: 10.1542/peds.2014-1428

Accepted for publication Nov 20, 2014

Symptoms of autism spectrum disorder (ie, social/communication deficits and restricted interests/repetitive behaviors)¹ become apparent over time, as gaps between typical and atypical development widen in childhood.²⁻⁷ Thus, both typical and atypical behaviors present simultaneously. Little is known empirically about ratios of typical to atypical behaviors in children who have typical or delayed development, leaving physicians to gather observational data during brief observations (10–20 minutes) with limited reference points. Surveillance and screening during primary care visits³ and clinical judgment of neurologists and geneticists (referrals for neurodevelopmental concerns) are key to early detection and referral. Although standardized parent report screening tools for autism are available,³ clinical impression is critical in decision-making, and often overrides information obtained from screening tools.⁸ Standardized observational screening measures are promising (ie, Systematic Observation of Red Flags,⁹ Screening Tool for Autism in Two-Year Olds¹⁰), but have not been adopted and may not be practical in primary care.

Little research exists to determine ratios of typical or atypical behaviors exhibited by children who have autism spectrum disorder during the time-span of an average medical visit.¹¹ During brief observations, low-frequency atypical behaviors may not stand out among high frequencies of typical behavior. Thus, we aimed to determine the ratio of signal (atypical behaviors) to noise (typical behaviors) in behaviors of young children who have autism in brief observations, and relate this information to clinical judgment and standardized test data.

METHODS

Participants

Institutional review boards of participating institutions approved all

methods and parents gave written consent for screening, evaluation, and video recording. Children aged 15 to 33 months were recruited through a 3-tiered autism screening process in a large suburban pediatric practice.¹² Screening had high participation rates (80% of families [$n = 796$] completed screening questionnaires, verified against clinic schedules). The sample was representative of many community clinics, comprising middle- to lower-socioeconomic status families from diverse racial and ethnic backgrounds, although African Americans were underrepresented. Participants were recruited, screened, and evaluated in English or Spanish.

Screening Process and Group Assignments

Participants were screened with the Modified Checklist for Autism in Toddlers (M-CHAT)¹³ and the Infant Toddler Checklist.¹⁴ Children who screened positive on at least 1 questionnaire or whose parents or providers were concerned despite negative screening ($n = 192$) were contacted by phone for follow-up, then invited for in-person evaluation, if warranted, at no cost. In-person evaluations included the gold standard autism observational measure (Autism Diagnostic Observation Schedule [ADOS]^{15,16}), a developmental measure (Mullen Early Learning Scales),¹⁷ and a measure of adaptive functioning (Vineland Adaptive Behavior Scales, Second Edition, Survey Interview).¹⁸ After evaluation, 14 children were identified with early signs of autism spectrum disorder and 16 with suspected language delays, but not autism (14 were selected based on age match to the Autism group). One child was identified as typically developing. Thirteen additional age-matched typically developing children were recruited from the same neighborhoods using the same screening instruments and test battery. Table 1 shows screening results and group assignment.

Table 2 shows characteristics of the sample. A χ^2 analysis revealed no significant differences on demographic variables, but lower rates of subsidized insurance (proxy for socioeconomic status) in the Autism group are noted.

Cognitive testing results (Table 3) found the Typical group to have average abilities, Language group mean scores 1 SD below average, and Autism group mean scores 2 SDs below average. The mean Gross Motor score in the Autism group was 1 SD below average. Autism and Language groups showed lower abilities in Communication, Socialization, and Motor adaptive domains than the Typical group. Receptive Language on the Mullen Early Learning Scales was the only score among adaptive and cognitive measures with significant differences between each group, pairwise.

Study Procedures

Video Segments

Two samples (10-minute segments from clinical evaluations) were chosen for analysis to examine whether children behaved differently after becoming familiar with the examiner and room: (1) the first 10 minutes of an ADOS,^{15,16} and (2) 30 minutes after starting the ADOS. Each 10-minute video was divided into sixty 10-second clips, viewed consecutively with 4-second breaks for behavioral coding (5040 intervals across 42 children, 2 videos each).

Behaviors

Five behavioral categories were rated to reflect broad interactional behaviors that might be noted by providers familiar with autism, but not necessarily specialists. Behavioral categories were based on diagnostic criteria,¹⁹ ADOS scoring algorithms,^{15,16} and the Systematic Observation of Red Flags.⁹ These included social responding, vocalizations, play, social initiations, and a discrete behavior, response to name. Table 4 contains operational

TABLE 1 Screening Results of Children in Pediatric Primary Care^{12a}

				Totals, n (%)
Tier 1 universal screening: total screened				796 (100)
Lost to follow-up (incomplete screens)				60 (8)
Children who had negative screens				544 (68)
Children who had positive screens (advanced to tier 2)				192 (24)
	Lost to follow-up	Screening decision: autism evaluation	Screening decision: no autism evaluation	Totals ^b
Tier 2 at-risk follow-up				
M-CHAT with follow-up interview	38 (5)	19 (2)	68 (8)	125 (16)
ITC with follow-up to confirm	40 (5)	45 (6)	58 (7)	143 (18)
Negative initial screens, but parent or provider concern		5 (<1)		5 (<1)
Children who had positive screen or concern		69 (8)		
Less: children positive on both screeners		-17 (2)		
Children invited for in-person evaluation (advanced to tier 3)		52 (6)		
	Lost to follow-up	Evaluated	Excluded by age match	Totals
Tier 3: indicated evaluations	21 (2)	31 (4)		52 (6)
Group assignments: autism group ^a		14 (2)		14 (2)
Language group ^c		16 (2)	2 (<1)	14 (2)
Typical group		1 (<1)		
Plus: controls recruited (negative screening and evaluation results)		+13		14 (n/a)

ITC, Infant Toddler Checklist from the Communication and Symbolic Behavior Scales¹⁴; M-CHAT, Modified Checklist for Autism in Toddlers¹⁵; n/a, not applicable to the total of 796 originally screened owing to addition of recruited controls.

^a These data include children who completed evaluations after publication of Miller et al 2011.

^b Totals are separated by screener, and some children screened positive on both, so total exceeds 192.

^c A total of 16 children were in this group, with 2 excluded owing to lack of age match with Autism group.

definitions used to determine whether behaviors were typical or atypical.

Ratings

Behaviors were rated for each category in each interval by using

partial interval recording (a method of recording occurrence at any time during the interval).^{20,21} Each behavior category received only 1 rating per interval, even if multiple behaviors were observed. Atypical behaviors were prioritized to

maximize detection, even if typical behaviors were present in the same segment. The order of priority for ratings was “atypical,” “typical,” “unclear,” and “no opportunity.” “Unclear” was the code for behaviors not visible (eg, off camera or back turned) so “unclear” and “no opportunity” codes were not of interest for this study. Raters could review segments without restriction. After completing each 10-minute video, raters indicated whether they would refer the child for an autism evaluation based solely on the observation.

Raters and Reliability

Two licensed psychologists, expert in early childhood development and autism spectrum disorder (and ADOS¹⁵ research reliable), rated behaviors. Both were unaware of study hypotheses and child-specific information other than age. Raters achieved initial reliability through practice videos. Inter-rater reliability

TABLE 2 Demographic Characteristics

	Autism (n = 14)	Language (n = 14)	Typical (n = 14)	Total (n = 42)
Gender				
Male	12 (86)	9 (64)	9 (64)	30 (71)
Female	2 (14)	5 (36)	5 (36)	12 (29)
Race/ethnicity				
Caucasian, non-Hispanic	8 (57)	8 (57)	8 (57)	24 (57)
African American	1 (7)	—	—	1 (2)
Native American	—	—	1 (7)	1 (2)
Hispanic	4 (29)	4 (29)	4 (29)	12 (29)
Pacific Islander	1 (7)	2 (14)	—	3 (7)
Asian	—	—	1 (7)	1 (2)
Spanish-speaking only	3 (21)	—	3 (21)	6 (14)
Insurance				
Private insurance	12 (86)	9 (64)	9 (64)	30 (71)
CHIP/Medicaid	2 (14)	5 (36)	2 (14)	9 (21)
Self-insured	—	—	3 (21)	3 (7)
Age range, mo	16.3–33.0	16.0–33.2	15.4–32.5	15.4–33.2
Mean age, mo (SD)	22.7 (4.8)	23.0 (5.5)	24.5 (5.5)	23.4 (5.2)

All data are presented as n (%) unless otherwise indicated. CHIP, Children's Health Insurance Plan (subsidized insurance for family income above Medicaid qualification level, but below ability to pay for private insurance).

TABLE 3 Cognitive and Adaptive Development

Subdomains	Autism (n = 14)	Language (n = 14) ^a	Typical (n = 14)
	Mean (SD); 95% CI		
Mullen Scales of Early Learning (t scores, mean, 50; SD, 10)			
Visual Reception (VR)	30 (12) ^b ; 23–37	39 (11); 33–46	49 (13) ^b ; 42–56
Receptive Language (RL)	26 (7) ^{b,d} ; 22–30	39 (13) ^{c,d} ; 32–47	54 (10) ^{b,d} ; 48–60
Expressive Language (EL)	29 (9) ^b ; 23–34	33 (9) ^d ; 28–38	53 (11) ^{b,d} ; 47–59
Gross Motor (GM)	42 (12) ^b ; 30–48	51 (14); 43–59	59 (10) ^b ; 53–64
Fine Motor (FM)	36 (13); 28–43	44 (14); 36–52	46 (12); 39–52
Vineland Adaptive Behavior Scales (standard scores, mean, 100; SD, 15)			
Communication (Com)	79 (13) ^e ; 71–87	87 (11) ^f ; 62–94	108 (10) ^{e,f} ; 94–116
Daily Living (DL) ^f	92 (11) ^g ; 86–98	98 (11); 75–103	113 (14) ^g ; 91–123
Socialization (Soc)	89 (9) ^h ; 82–92	93 (10) ⁱ ; 71–100	109 (11) ^{h,i} ; 96–117
Motor (Mot) ^j	88 (9) ^j ; 83–94	93 (11) ^k ; 55–97	103 (7) ^{j,k} ; 93–109

No significant differences in scores by age were found. Significant differences between genders were found only for Expressive Language on the Mullen Scales of Early Learning within the Language group $F(1,12) 16.7; P = .002$. Mean score for girls in the Language group was 41.2 (SD, 6.7; median, 39) and for boys it was 28.1 (SD, 5.2; median, 30). CI, confidence interval.

^a One Vineland in the Language group was not completed.

^b Autism < Typical, $P \leq .001$; Cohen's d Effect Size (ES)_{VR} = 1.53, ES_{RL} = 3.31, ES_{EL} = 2.43, ES_{GM} = 1.508.

^c Autism < Language, $P = .003$; ES_{RL} = 1.324.

^d Language < Typical, $P \leq .001$; ES_{RL} = 1.28, ES_{EL} = 2.07.

^e Autism < Typical, $P = .005$; ES_{Com} = 1.54.

^f Language < Typical, $P = .004$; ES_{Com} = 1.169.

^g Distribution of Vineland DL scores was significantly skewed (-1.184) $se_{skew} = 0.369$. Kruskal-Wallis $\chi^2(2) = 10.369$; $P = .006$, ASD (mean rank, 16.04) < Typical (29.32), $P = .01$.

^h Autism < Typical, $P = .02$; ES_{Soc} = 1.35.

ⁱ Language < Typical, $P = .01$; ES_{Soc} = 0.99.

^j Distribution of Vineland Mot was significantly skewed (-1.941) $se_{skew} = 0.374$. Kruskal-Wallis $\chi^2(2) = 12.644$; $P = .002$, ASD (mean rank, 15.07) < Typical (29.39), $P = .003$.

^k Language (16.46) < Typical, $P = .01$.

was calculated by exact agreement (number of agreements/total observations) on 20% of study videos (5040 of 25 200 individual codes). Inter-rater reliability was 82% overall. Agreement was 84% on presence of behavior, 87% on absence of behavior. Agreement on typical behaviors was 97%, and on atypical behaviors it was 35%, which is discussed in more detail later. The κ between raters was 0.67.

Analytic Approach

Exclusion

One video of a child (age 33 months) in the Autism group was excluded from analyses other than Vocalizations because 46% of intervals were off camera. The child's second video was included.

Methods

We analyzed rates of typical and atypical behaviors using non-parametric Kruskal-Wallis H and Mann-Whitney U to allow for

significantly skewed distributions and occasional outliers. Demographic data and referral impressions were analyzed using χ^2 tests. ANOVA was used to analyze normally distributed developmental testing standard scores and relationships between referral impressions and age. Correlations were calculated using Spearman's ρ . In all analyses P values $\leq .01$ were considered significant. Non-significant results are not reported in the text.

RESULTS

Behavioral Category Results

Differences in Opportunities

Kruskal-Wallis and Mann-Whitney tests determined that opportunities for social interactions were different among groups in Responding and Initiating, with fewer opportunities in the Autism group. Opportunities for Response to Name were greater in the Autism group (shown in Table 4). Because of these differences, coded

behaviors were converted to percentage scores (eg, Atypical Percentage score = Atypical codes/[Atypical + Typical codes]) to standardize comparisons. No significant effects of time were found between samples at 10 minutes and 30 minutes, so the 2 observations were combined and analyzed on a per-child basis, limiting the effects of anomalous data in single samples.

Atypical behaviors occurred in 11% of intervals within the Autism group, compared with 2% of intervals in both Language and Typical groups. Group differences were significant for total number of atypical behaviors, $\chi^2(2) = 12.602$; $P = .002$ (Autism > Language, $P = .005$; and Typical, $P = .008$). Because typical and atypical percentage scores are complementary, differences in total typical behaviors in all analyses reflect significance for Atypical behaviors, but in the opposite direction (eg, total typical behaviors were significant for group, but Autism < Language or Typical). Results by behavioral category are shown in Fig 1 and Table 4.

Social Responding

Groups differed on typical and atypical responding overall, $\chi^2(2) = 9.899$; $P = .01$, Autism_{atypical} > Language_{atypical}; $P = .01$.

Vocalizations

The quality and repetitive nature of sounds was significantly different among groups, $\chi^2(2) = 13.624$; $P = .001$ (Autism_{atypical} > Language_{atypical}; $P = .01$).

Play

Although raw count ratios of play behaviors (Table 4) differ, comparison of percentages of typical to atypical play behaviors within each child were not significantly different. Mean ranks for atypical play were 27 in the Autism group, and 17.3 and 20.2 in the Language and Typical groups, respectively. Mean percentage scores of atypical play in all groups were <10%.

TABLE 4 Ratios of Typical to Atypical Behaviors

	Operational Definitions ^a		Typical:Atypical Occurrence Ratio		
	Typical Behavior	Atypical Behavior	Autism	Language	Typical
Responding	Taking offered toys, eye contact or directed facial expressions in response to social overtures, laughing or vocalizing in response to social games, catching a ball, or following verbal or gestured directions	Absence of any typical behaviors within 3 s of a clear opportunity presented by the adult	8:1 ^{b,c}	51:1 ^b	63:1 ^c
Vocalization	Babbling, crying, laughing, talking, with typical prosody	Stereotyped or repetitive (3 times in succession) vocalizations, or vocalizations with odd prosody	6:1	31:1	13:1
Play	Appropriate moving or holding of toys, objects, or body parts	Unusual examination, stereotyped or repetitive use of toys, objects, or the body (eg, identical action, such as banging, in more than 3 time intervals)	13:1	659:1 ^d	88:1
Initiating	Sharing attention or excitement, engaging others in play, requesting something, or initiating a turn in a back-and-forth game	No sharing, requesting, or initiating back and forth games despite a clear opportunity; odd overtures, such as using another's hand as a tool	18:1 ^{e,f}	67:1 ^e	136:1 ^f
RTN	Orienting or vocalizing to the speaker within 3 s of name being called	No response within 3 s of name being called	1:1 ^{g,h}	5:1 ^g	4:1 ^h

^a Raters were allowed to consider the subsequent interval to determine response to a bid that was presented at the end of the 10-s interval. The response was coded in the interval in which it was observed.

^b Kruskal-Wallis of total Responding behaviors (opportunities): $\chi^2(2) = 14.39; P = .001$; Mann-Whitney *U* Autism (70) < Language (88); *U* = 161.5; *z* = 2.921; *P* = .002.

^c Autism < Typical (92); *U* = 174.5; *z* = 3.518; *P* < .001.

^d Ratios are calculated on aggregate behaviors within the group. In the Language group, only 2 intervals were coded with atypical play, compared with 1319 intervals with typical play behavior.

^e Kruskal-Wallis of total Initiating behaviors (opportunities): $\chi^2(2) = 12.447; P = .002$ (Autism [10] < Language [19]; *U* = 161; *z* = 2.987; *P* = .003).

^f Autism < Typical (19); *U* = 167; *z* = 3.172; *P* = .001.

^g Kruskal-Wallis of total RTN behaviors (opportunities): $\chi^2(2) = 10.720; P = .005$ (Autism [19] > Language [10]; *U* = 36; *z* = -2.857; *P* = .004).

^h Autism > Typical (11); *U* = 40.5; *z* = -2.654; *P* = .007.

Social Initiation

Opportunities to initiate to the examiner (eg, to request desired toys or initiate another turn in a social game) were rated. The mean percentage scores of atypical initiating behavior were <10% and not significantly different between groups.

Response to Name

Differences between groups in the percentage of typical and atypical responses to name (RTN) were significant, $\chi^2(2) = 9.899; P = .007$ (Autism [mean rank, 29.68] > Language [16.64]; *P* = .01). Opportunities for RTN occurred at a low base rate (6% of intervals). All children in the Autism group responded to their names at least once (typical). Many children in the Language (57%) and Typical (50%) groups failed to respond to their names at least once (atypical).

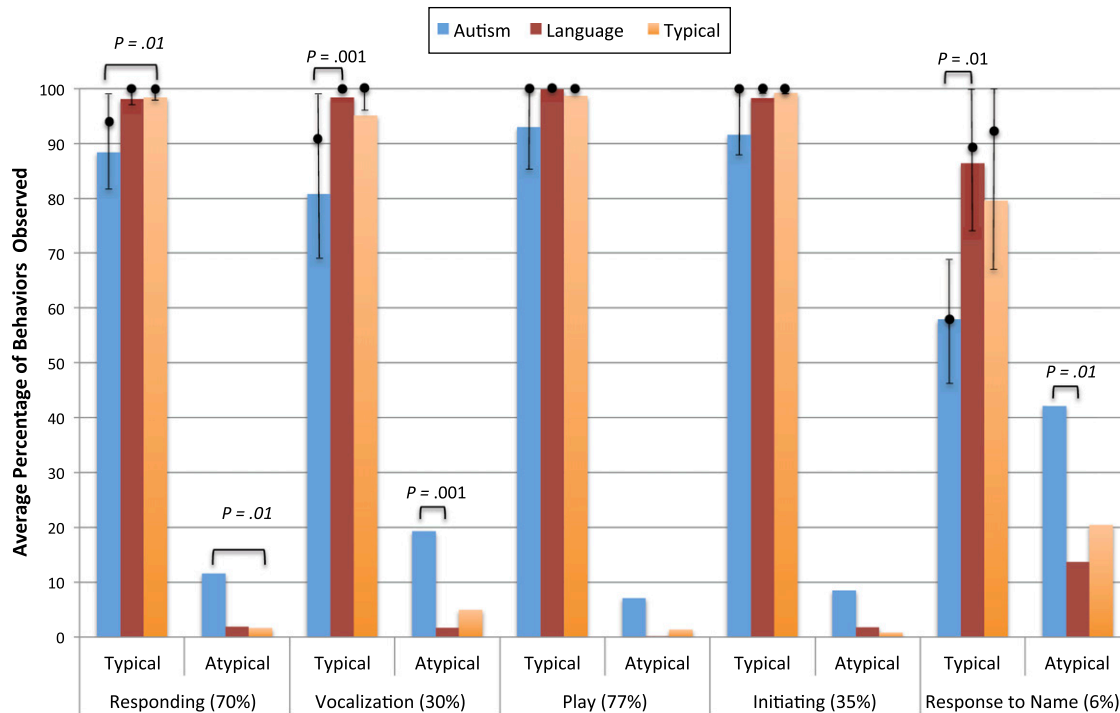
In 10-minute videos, average opportunities for RTN were 5 (SD,

3.6) in the Autism group, mean, 2.4 (SD, 2) for the Language group, and mean, 2.9 (SD, 2) for the Typical group. Typical responses to the first opportunity for RTN in the original codes were calculated as a more standardized measure of RTN. Typical overall RTN percentage scores in the Autism group (mean, 0.58; SD, 0.26) are similar to the overall percentage RTN on the first opportunity (0.56). In the Typical group, the percentage of typical responses (mean, 0.80; SD, 0.30) compares with overall response to the first opportunity (0.75). In the Language group, the percentage score for typical responses (mean, 0.86; SD, 0.15) compares to the percentage of responses to first RTN bids (0.64). Consistent with lower opportunities in the Language group, 6 out of 28 Language group videos contained no RTN opportunities, compared with 2 out of 28 in both the Autism and Typical groups. RTN opportunities on video did not necessarily include administration of the RTN item on the

ADOS, but correlations of the Atypical Percentage scores with ADOS scores were moderate for RTN and algorithm scores as shown in Table 5.

Referral Decision

At the end of each coding session, raters were asked, "Based on this 10-minute observation alone, would you refer this child for autism evaluation?" "Yes" or "No" responses were converted to "Correct" or "Incorrect" according to diagnostic group. Figure 2 shows rater judgments by group. Two referral decisions were made per child (2 videos). Rater judgment was most inaccurate for the Autism group (11 out of 28 videos [39%] incorrect). In the Language group, 7 out of 28 videos were incorrect (25%), and in the Typical group, 3 out of 28 were incorrect (11%). Accuracy was not attributable to rater or time (0–10 minutes vs 30–40 minutes into the evaluation). Within this small sample, sensitivity of the referral impression (ASD or not ASD) was 0.61, specificity



Behavior Categories (with Base Rates of Occurrence) in Samples

FIGURE 1

Differences in behavior occurrence between groups. Bars on Typical behavior columns illustrate the interquartile range, with dots indicating medians. In the case of an interquartile range where both 25th and 75th percentile scores were 100, only the dot is shown. Atypical score ranges are reflective of Typical score ranges and are not shown. Behaviors were rated as typical or atypical, then divided by the total behaviors (eg, typical/typical + atypical) to calculate percentage of behavior. Analysis comprised 3 groups, 2 time points (collapsed into 1 with no significant differences), and 12 dependent variables, including Total Atypical and Typical behaviors reported in the text. The majority of intervals included Play and/or Responding behaviors, whereas few intervals contained RTN behaviors. Significant differences shown for Atypical behaviors (shown) are reflected in Typical behaviors, but in the opposite direction.

was 0.82, and positive and negative predictive values were 0.63 and 0.81, respectively. There was a significant interaction between age and time, ($F [1,40] = 7.22; P = .01; \text{Effect Size} = 0.965$). Incorrect decisions were more likely to be made for younger children (mean, 20.2 months; SD, 4) in the

first 10 minutes, compared with correct decisions for later observations (at 30 minutes) or older children (mean, 24.7 months; SD, 5.1). All videos were included in analyses. For the video excluded from atypical/typical analysis, referral impression was Correct.

In 1 instance, 1 rater noted a basis for an autism referral decision. For 1 video of a child in the Typical group, the rater recommended autism referral based solely on the child's atypical RTN in the 10-minute sample. The clinical impression of this child would otherwise not have indicated an autism referral.

TABLE 5 Correlations Between Atypical Behavior Percentages and ADOS Algorithm Scores

ADOS	Atypical Behaviors					
	Social Responding	Social Initiating	Vocalizations	Play	RTN	Total
Social affect (SA subtotal)	0.36	0.26	0.52	—	0.43	0.45
Restricted repetitive behavior (RRB subtotal)	—	—	0.60	0.50	—	0.54
Total score	0.46	0.28	0.54	0.48	0.41	0.52
RTN (item)	—	—	—	—	0.53	—

Spearman's ρ correlations between Atypical Percentage scores (atypical behaviors/[atypical + typical behaviors]) and ADOS algorithm scores. Overall (total score) correlations were moderate, as were RRB and RTN correlations, but differences in the nature of the 2 scores (Atypical Percentage scores as a ratio versus ADOS scores as a degree of atypicality) produced a wide range of individual category correlations. Atypical Percentage scores on Vocalization showed the highest correlation to both SA and RRB subtotal algorithm scores and the total algorithm score.

DISCUSSION

Our results suggest that, during brief observations, typical behaviors in some children who have autism can exceed atypical behaviors in frequency, to such a degree that it was often difficult even for clinicians experienced in autism spectrum disorder assessment to correctly determine if enough atypical behavior existed to merit a referral for autism evaluation. With low agreement

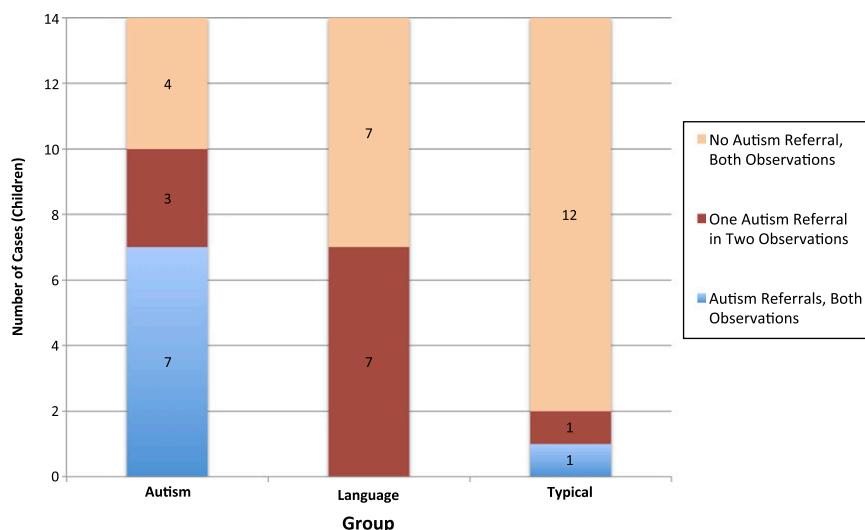


FIGURE 2

Accuracy of referral decisions. Children were observed twice (0 to 10 minutes; 30 to 40 minutes), each with a separate referral decision. Raters disagreed on autism referrals for all of the split decisions (1 view generated an autism referral, the other did not) except 1. In 1 Language case, the same rater gave different decisions for each view.

between raters on atypical behaviors (35%), it is possible that some atypical or typical behaviors were misidentified, but given the high rate of agreement on typical behaviors (97%), and low base rates of atypical behaviors (11% in the Autism group, 2% in non-Autism groups), it is unlikely any misidentified behaviors would overshadow the finding that typical behaviors were predominant, even in the Autism group. Rather, the low agreement on atypical behaviors highlights the difficulty of detecting atypical behaviors in brief observations, even for experts.

Although the Autism group demonstrated statistically higher rates of atypical behaviors and lower rates of typical behaviors compared with Language and Typical children overall, the ratio of behaviors in young children was such that typical behaviors in the Autism group still far exceeded atypical behaviors. Much attention has been drawn to atypical behaviors associated with autism.^{6,7,22-31} However, children who have autism do not engage in unusual behavior exclusively, and at the individual level, even experienced clinicians did not always agree on atypicality of behavior. Likewise, typical

development is characterized by occasional periods of repetitive play, rigidity, and failure to respond.³² Normative data regarding ratios of typical to atypical behaviors have been absent, leaving clinicians to rely on their own judgment about whether a child's behavior is excessively atypical.

We found children in the Autism group responded to their name over half the time (58%). Although this was significantly less than the Language (86%), or the Typical group (80%), the response rate in the Autism group is consistent with research establishing RTN as a highly specific but insensitive predictor of an autism diagnosis.²² Clinically, a single example of typical or atypical behavior (eg, RTN or making eye contact) may exert undue influence in referral decisions. Our results suggest that a positive RTN during a brief observation would be a poor single method of ruling out autism risk.

Less overt behaviors (ie, lower rates of social initiations, lower responsiveness, and repetitive behaviors) may be more difficult to detect than failure to respond to one's name. In these behaviors, ratios of atypical:typical in the Autism group

ranged from 1:6 (Vocalization) to 1:18 (Initiating). Atypical behaviors occur so rarely they may easily escape clinical detection during a brief observation. When weighing evidence in a decision-making process, atypical behaviors noted may be overshadowed by the many typical behaviors also likely to be observed during the same period of time.

Expert clinicians, with advantages of focused and repeated observation conditions, identified only 61% of the Autism group observations as indicating need for autism referrals based on 10-minute behavioral samples. Identification of autism risk was more difficult with younger children during the first 10 minutes of the evaluation compared with older children or to the time period 30 to 40 minutes into the evaluation. Missed referral decisions may be related to inherent difficulties in characterizing atypical behaviors (agreement between expert raters was lowest for atypical behavior), low frequencies of atypical behavior relative to typical behavior, or the high degree of behavioral variability seen in the Autism group. If atypical behavior is difficult to pinpoint (even for experts), too infrequent to stand out in a brief visit, or occurring at different rates across children who have autism, there may not be enough reliable observational data available on a consistent basis for clinicians to develop their own threshold of concern. The fact that typical behaviors, in contrast, were so frequent in all children, and so salient (agreement on typical behavior was very high), may mean that typical behaviors create significant noise around which it is difficult to weigh the importance of the signal of infrequent atypical behavior.

Receptive language scores were the underlying characteristic differentiating the groups. In toddlers, receptive language assessment includes asking children to respond to familiar commands, give something, or

point to something. These tasks have distinctly social components to them (ie, social reciprocity, sharing, and joint attention), and may provide opportunities to observe social communication before expressive language delays become evident. The finding that receptive language ability is an important distinguishing characteristic suggests the impact of receptive language deficits on interactions between examiner and child may have been more clinically meaningful than other atypical behaviors. This may also explain false-positive autism referrals in the Language group.

Study limitations include the absolute number of children observed and the fact that referral decisions were artificially based on observation alone, rather than observation in combination with developmental history, parent concerns, direct interaction, and formal screening results. Sensitivity and specificity may not be replicable, given the small sample and exposure to focused coding of behaviors before the referral decision. Raters were autism experts, with opportunity for focused, detailed, and repeated observations (300 discrete ratings per observation), a much more detailed

experience than is available during busy medical visits. The final sample size was small, but derived from a community population, selected through universal screening. The sample is more likely to be representative of the range of clinical presentations facing general pediatricians and first-line specialty referrals (eg, neurology, Early Intervention, audiology) than other samples of referred populations with higher incidence rates.

CONCLUSIONS

Children who have autism display high rates of typical behavior alongside atypical behavior. Children who do not have autism show atypical behavior at times, albeit at a statistically smaller ratio than children who have autism. Even clinicians who have experience and expertise in autism may not detect differences in the atypical:typical behavior ratios in a 10- to 20-minute observation. Receptive language abilities are an important area of focus in early diagnosis of developmental delays.

These data suggest that decision-making processes for possible autism symptoms should include

consideration of all available data. In addition to behavioral observation, autism screening tools, parent observations, developmental testing, and a detailed history should be considered when making referral decisions. The high rate of false-negative referral impressions in the study is consistent with findings indicating that in developmental disabilities generally, clinical judgment does not improve on screening test results.⁸ Even a small number of clear examples of unusual behaviors might be sufficient to prompt further questions, or to begin the process of gathering additional developmental testing data (evaluation). The presence of typical behavior, or absence of clearly atypical behavior, during a brief period of observation is not sufficient or reliable enough to override other data indicating concern for autism.

ACKNOWLEDGMENTS

We acknowledge the valuable contributions of the families in the study, Granger Medical Pediatrics, Wee Care Pediatrics, Developmental Disabilities Inc, Paul Carbone, MD, and Lane Fischer, PhD, to data collection and the manuscript.

Address correspondence to Terisa Gabrielsen, PhD, 340-A MCKB, Brigham Young University, Provo, UT 84602. E-mail: terisa_gabrielsen@byu.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2015 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Portions of this study were funded by Centers for Disease Control and Prevention U01DD000068-01, The EACH CHILD Study, Judith Miller, Principal Investigator; University of Utah School of Medicine, Department of Pediatrics; University of Utah Graduate School, Steffensen Cannon Fellowship.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Association; 2013
2. Chawarska K, Paul R, Klin A, Hannigen S, Dichtel LE, Volkmar F. Parental recognition of developmental problems in toddlers with autism spectrum disorders. *J Autism Dev Disord*. 2007;37(1):62–72
3. Johnson CP, Myers SM; American Academy of Pediatrics Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics*. 2007; 120(5):1183–1215
4. Ozonoff S, Heung K, Byrd R, Hansen R, Hertz-Picciotto I. The onset of autism: patterns of symptom emergence in the first years of life. *Autism Res*. 2008;1(6): 320–328
5. Ozonoff S, Iosif AM, Baguio F, et al. A prospective study of the emergence of early behavioral signs of autism. *J Am Acad Child Adolesc Psychiatry*. 2010;49(3):256–266, e1–e2
6. Yirmiya N, Ozonoff S. The very early autism phenotype. *J Autism Dev Disord*. 2007;37:1–11

7. Zwaigenbaum L, Bryson S, Garon N. Early identification of autism spectrum disorders. *Behav Brain Res.* 2013;251: 133–146
8. Morelli DL, Pati S, Butler A, et al. Challenges to implementation of developmental screening in urban primary care: a mixed methods study. *BMC Pediatr.* 2014;14:16
9. Wetherby AM, Woods J, Allen L, Cleary J, Dickinson H, Lord C. Early indicators of autism spectrum disorders in the second year of life. *J Autism Dev Disord.* 2004;34(5):473–493
10. Stone WL, McMahon CR, Henderson LM. Use of the Screening Tool for Autism in Two-Year-Olds (STAT) for children under 24 months: an exploratory study. *Autism.* 2008;12(5):557–573
11. Halfon N, Stevens GD, Larson K, Olson LM. Duration of a well-child visit: association with content, family-centeredness, and satisfaction. *Pediatrics.* 2011;128(4): 657–664
12. Miller JS, Gabrielsen T, Villalobos M, et al. The each child study: Systematic screening for autism spectrum disorders in a pediatric setting. *Pediatrics.* 2011;127(5):866–871
13. Robins D, Fein D, Barton M. *The Modified Checklist for Autism in Toddlers (M-CHAT)*. Self-published; 1999
14. Wetherby A, Prizant B. *Communication and Symbolic Behavior Scales Developmental Profile—first normed edition*. Baltimore, MD: Brookes; 2002
15. Lord C, Rutter M, DiLavore P, Risi S. *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles, CA: Western Psychological Services; 1999
16. Lord C, Rutter M, DiLavore P, Risi S. *Autism Diagnostic Observation Schedule: Toddler Module*. Los Angeles, CA: Western Psychological Services; 2006
17. Mullen E. *Mullen Scales of Early Learning*. Bloomington, MN: Pearson, Inc; 1995
18. Sparrow S, Balla D, Cicchetti D. *Vineland Adaptive Behavior Scales*. 2nd ed. Survey Interview. Minneapolis, MN: Pearson, Inc; 2005
19. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Text Revision. Washington, DC: American Psychiatric Association. 2000
20. Harrop A, Daniels M. Methods of time sampling: a reappraisal of momentary time sampling and partial interval recording. *J Appl Behav Anal.* 1986;19(1): 73–77
21. Harrop A, Daniels M, Foulkes C. The use of momentary time sampling and partial interval recording in behavioral research. *Behav Psychother.* 1990;18: 121–127
22. Nadić AS, Ozonoff S, Young GS, Rozga A, Sigman M, Rogers SJ. A prospective study of response to name in infants at risk for autism. *Arch Pediatr Adolesc Med.* 2007;161(4):378–383
23. Pierce K, Carter C, Weinfeld M, et al. Detecting, studying, and treating autism early: the one-year well-baby check-up approach. *Pediatrics.* 2011;159:458–465
24. Chlebowski C, Robins DL, Barton ML, Fein D. Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics.* 2013;131(4). Available at: www.pediatrics.org/cgi/content/full/131/4/e1121
25. Young GS, Merin N, Rogers SJ, Ozonoff S. Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Dev Sci.* 2009;12(5):798–814
26. Wall DP, Kosmicki J, Deluca TF, Harstad E, Fusaro VA. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl Psychiatr.* 2012;2:e100
27. Allison C, Auyeung B, Baron-Cohen S. Toward brief “red flags” for autism screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist in 1,000 cases and 3,000 controls. *J Am Acad Child Adolesc Psychiatry.* 2012;51:202–212
28. Nadel S, Poss JE. Early detection of autism spectrum disorders: screening between 12 and 24 months of age. *J Am Acad Nurse Pract.* 2007;19(8): 408–417
29. Vanvuchelen M, Roeyers H, De Weerd W. Imitation assessment and its utility to the diagnosis of autism: evidence from consecutive clinical preschool referrals for suspected autism. *J Autism Dev Disord.* 2011;41(4): 484–496
30. Stone WL, Ousley OY, Hepburn SL, Hogan KL, Brown CS. Patterns of adaptive behavior in very young children with autism. *Am J Ment Retard.* 1999;104(2): 187–199
31. Landry R, Bryson SE. Impaired disengagement of attention in young children with autism. *J Child Psychol Psychiatry.* 2004;45(6):1115–1122
32. Evans DW, Leckman JF, Carter A, et al. Ritual, habit, and perfectionism: the prevalence and development of compulsive-like behavior in normal young children. *Child Dev.* 1997;68(1): 58–68

Identifying Autism in a Brief Observation

Terisa P. Gabrielsen, Megan Farley, Leslie Speer, Michele Villalobos, Courtney N. Baker and Judith Miller

Pediatrics 2015;135:e330; originally published online January 12, 2015;

DOI: 10.1542/peds.2014-1428

Updated Information & Services	including high resolution figures, can be found at: http://pediatrics.aappublications.org/content/135/2/e330.full.html
References	This article cites 23 articles, 5 of which can be accessed free at: http://pediatrics.aappublications.org/content/135/2/e330.full.html#ref-list-1
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Community Pediatrics http://pediatrics.aappublications.org/cgi/collection/community_pediatrics_sub Developmental/Behavioral Issues http://pediatrics.aappublications.org/cgi/collection/development:behavioral_issues_sub Autism/ASD http://pediatrics.aappublications.org/cgi/collection/autism:asd_sub
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://pediatrics.aappublications.org/site/misc/Permissions.xhtml
Reprints	Information about ordering reprints can be found online: http://pediatrics.aappublications.org/site/misc/reprints.xhtml

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2015 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™



PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Identifying Autism in a Brief Observation

Terisa P. Gabrielsen, Megan Farley, Leslie Speer, Michele Villalobos, Courtney N. Baker and Judith Miller

Pediatrics 2015;135:e330; originally published online January 12, 2015;
DOI: 10.1542/peds.2014-1428

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://pediatrics.aappublications.org/content/135/2/e330.full.html>

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2015 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

