Brief Report: Use of the Infant-Toddler Checklist in Infant Siblings of Children with Autism Spectrum Disorder

Chandni Parikh[1,*], **Ana-Maria Iosif**[2], **Sally Ozonoff**[1]

[1]Department of Psychiatry & Behavioral Sciences, MIND Institute, University of California-Davis Health, Sacramento, CA

[2]Department of Public Health Sciences, University of California-Davis, Davis, CA

Abstract

Few studies have explored autism spectrum disorder (ASD) screening in the first year of life. The current investigation examines the psychometric properties of the Infant-Toddler Checklist starting in the first year of life in a sample at elevated and average risk for ASD based on family history. 283 participants were followed from 6 to 36 months, when diagnostic outcome was determined. The results indicated low to moderate sensitivity, specificity, and positive predictive value across ages for broadly distinguishing any delays from typical development, as well as for more narrowly discriminating children with ASD from those who were typically developing. Implications for utilizing ASD screening tools in the first year of life with high risk samples are discussed.

Keywords

screening; infants; ASD; Infant-Toddler Checklist

Introduction

Although heterogeneous in the timing of symptom emergence, autism spectrum disorder (ASD) begins early in life, with signs first appearing between 9 and 24 months for most children (Ozonoff et al., 2010; Zwaigenbaum et al., 2007), whether they demonstrate a regressive or an early onset course. Parents often report concerns around their child's first birthday (Macari et al., 2018; Sacrey et al., 2016), but there is a significant delay until formal diagnosis for most children, which occurs at a mean age of four in the United States and has not changed appreciably in two decades (Baio et al., 2018). This "diagnostic odyssey" of 2-3 years indicates a profound need for effective methods of screening during infancy, the most

critical time period when significant gains from early intervention are observed (Koegel, Koegel, Ashbaugh, & Bradshaw, 2014).

More than a decade ago, the American Academy of Pediatrics recommended the use of an ASD-specific standardized screener at the 18- and 24-month well-child visits (Johnson & Myers, 2007). Since the publication of these guidelines, there has been an increase in screening practices but far from universal implementation (Arunyanart et al., 2012; Khowaja, Hazzard, & Robins, 2015). Currently, there are multiple screeners available for children 12 months and older; however, there are few choices for ASD screeners in the first year of life. This may be due, at least in part, to the protracted period of gradual symptom emergence and phenotypic variability in the first years of life (Towle & Patrick, 2016; Zwaigenbaum et al., 2015).

In evaluating the accuracy of a screening instrument, the most important psychometric features are sensitivity (SE), the percentage of affected children who screen positive for the disorder, specificity (SP), the percentage of unaffected individuals who screen negative for the disorder, and positive predictive value (PPV), the percentage of children who screen positive who are ultimately diagnosed with the disorder. In order to calculate these indices, it is necessary to follow both screen-positive and screen-negative cases to the age at which the disorder can conclusively be determined present or absent (Sheldrick et al., 2015).

Among the most widely used instruments with applicability to younger children is the Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001). The M-CHAT is a parent questionnaire developed for use in the general population that is normed from 16 to 30 months. A validation study with low-risk toddlers using the M-CHAT-Revised, with follow-up, (M-CHAT-R/F) indicated SE of 85% (CI: 79 – 92%) and SP of 99% (CI: 99 – 100%; Robins et al., 2014). A recent meta-analysis of 13 studies using the M-CHAT reported somewhat lower psychometric indices, including a pooled SE of 83% (95% CI: 75-90%), SP of 51% (95% CI: 41-61%), and PPV of 53% (95% CI: 43-63%), in high-risk children (i.e., those with identified developmental concerns prior to screening) but only 6% PPV in low-risk children (Yuen, Penner, Carter, Szatmari, & Ungar, 2018). Many studies included in the meta-analysis did not report SE and SP because screen-negative cases were not further evaluated to confirm diagnostic status, resulting in only partial calculation of psychometric properties (Miller et al., 2011; Toh, Tan, Lau, & Kiyu, 2018; Yuen et al., 2018).

The Infant-Toddler Checklist (ITC; Wetherby & Prizant, 2002) is a parent report instrument developed as a broadband screener for communication delays that has the advantage of screening cutoff scores and standardized norms for infants as young as 6 months. It was initially developed to screen for communication delays and demonstrated good psychometric properties for this purpose (SE = 88.9%, SP = 88.9%; Wetherby et al., 2004). A later validation study examined how well the ITC distinguished children with ASD from a general population sample (Wetherby, Brosnan-Maddox, Peace, & Newton, 2008). In a sample of 5,385 children aged 6-24 months, 56 out of 60 children who were eventually diagnosed with ASD had screened positive on their first and/or subsequent ITC, resulting in

a 93.3% SE pooled across ages, ranging from 20% at 6-8 months to 94.7% at 21-24 months. Estimates of SP and PPV were not reported.

Despite being one of the only ASD screening instruments with norms available for the first year of life, there are few studies of the ITC's usage in infancy. Pierce et al. (2011) screened low-risk infants in a pediatric office at 12 months and found a 75% PPV for detecting a range of developmental delays, including ASD, language delays, and other developmental concerns. Table 1 provides a list of the studies that have utilized the ITC and published on its psychometric properties. Other than the Wetherby et al. (2004) validation sample, no studies have implemented the ITC in the first year of life and conducted longitudinal follow up of the screen-negatives, permitting calculation of complete psychometric indices.

The objective of the current study was to address several evidence gaps in the literature: a) employing the ITC in the first year of life; b) following all cases to 36 months of age when diagnostic outcomes could be determined for the full sample, permitting calculation of SE, SP, PPV, and negative predictive value (NPV); and c) examining the ITC's utility in a high-risk cohort. In this study, we examined not only how well the ITC discriminated children with ASD from those who were typically developing, but also how it distinguished any delays (ASD plus other developmental problems) from typical development, since the ITC was originally developed as a broadband screener for communication delays (Wetherby et al., 2008).

Methods

Participants and Procedures

Using a prospective infant sibling design, $n = 324$ infants were enrolled in the study but 41 were excluded from the present analyses because they did not have outcome classification at 36 months. The final sample included $n = 283$ infants with $n = 97$ typically developing siblings (low-risk) and $n = 186$ infants with a family history of ASD (high-risk). High-risk infants had at least one older sibling with ASD (proband), confirmed using the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012) and the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003), a widely used parent report instrument with validated cutoffs for ASD and satisfactory psychometric properties (e.g., 60-92% sensitivity and 52-92% specificity for distinguishing ASD vs. non-ASD across studies; Norris & Lecavalier, 2010). Low-risk status was confirmed by an intake screener and proband SCQ scores below the ASD range. Exclusion criteria included birth before 32 weeks of gestation and a known genetic disorder in the proband. Parents provided informed consent and the study was approved by the university's Institutional Review Board.

All participants were enrolled by 6 months ($n = 210$) or 9 months ($n = 73$) of age. Data was collected at up to seven visits (6, 9, 12, 15, 18, 24, and 36 months). Three outcome groups were classified at the 36-month visit: an ASD group ($n = 46$; 45 from the high-risk group and 1 from the low-risk group) who met *Diagnostic and Statistical Manual of Mental Disorders* 5th ed. (DSM-5; APA, 2013) criteria for ASD and obtained ADOS-2 comparison scores of 4 or higher; a Non-Typically Developing group (Non-TD; $n = 38$) that

demonstrated Mullen scores over 1.5 standard deviations below the normative mean and/or ADOS-2 comparison scores   3, but did not meet criteria for ASD; and a Typically Developing group (TD; $n = 199$) whose scores on the Mullen were in the average range or above and ADOS-2 comparison scores were below 3.

Measures

Infant-Toddler Checklist (Wetherby et al., 2008).—The ITC is a parent report questionnaire about social-communication, language, and symbolic development. It yields three composite scores (social, speech, and symbolic) as well as a total score. It is normed for use with children 6 to 24 months of age. A positive screen for communication delays, including ASD, is defined as a score in the bottom 10th percentile on the social, symbolic, or total composites.

Outcome Measures: Autism Diagnostic Observation Schedule-2 (Lord et al., 2012), Mullen Scales of Early Learning (Mullen, 1995).—The ADOS-2 is a semi-structured play-based interaction and observation that provides a cutoff validated to distinguish ASD from non-ASD cases; it has been widely used in prospective studies of high-risk samples (Zwaigenbaum et al., 2009). It provides a comparison score that ranges from 1 to 10 and can be used to classify participants into non-spectrum (below 4) and ASD (4 and above) groups (Gotham, Pickles, & Lord, 2009). The ADOS-2 was used for the determination of outcome at 36 months of age. Five modules are available for children of different ages and verbal levels; at 36 months, modules 1 and 2 were employed. These modules have excellent psychometric properties, with reported sensitivities of 77-84% and specificities of 77-82% for distinguishing ASD from non-spectrum (Lord et al., 2012). The Mullen Scales of Early Learning (MSEL) is a standardized developmental test for children birth to 68 months that measures motor, cognitive, and language skills. The MSEL manual reports good internal, test-retest, and interrater reliability and convergent validity (Mullen, 1995). Both outcome measures were administered by examiners who had completed rigorous research training and achieved 80% or higher reliability with a trainer prior to administration and checked periodically throughout the study. All examiners were kept unaware of child risk group, as well as scores and diagnoses from previous visits.

Statistical Analysis

Following Wetherby et al. (2008), a positive screen was defined as a score in the bottom 10th percentile on the social, symbolic, or total scales. We examined statistical measures of the performance of this binary classification, including sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV), at each visit. Following the approach of Wetherby et al. (2008), two sets of analyses were conducted. First, the ASD and Non-TD groups were collapsed into an "any delays" group and compared to the TD group. Second, we compared the ASD group alone to TD.

Since sample size varied by visit, we used multiple imputations for incomplete longitudinal binary data, so that estimates would be based on the same sample across ages. We generated 100 data sets using multilevel multiple imputations with fully conditional specification with the logistic regression (Yamaguchi, Misumi, & Maruo, 2018) and combined the results

according to Rubin's rules (Rubin, 1987). All analyses were implemented in SAS 9.4 (SAS Institute, Cary, North Carolina).

Results

Table 2 summarizes the performance of the ITC screener cutoff for distinguishing participants with any delays (e.g., ASD plus Non-TD) from TD from 6 to 24 months. The psychometric properties were below generally accepted levels of 80% or higher sensitivity and specificity (Clark & Harrington, 1999), with indices improving over time and the highest values obtained at the 24-month visit.

Next, following Wetherby et al. (2008), we examined the ITC's psychometric properties as an ASD-specific screener by comparing the ASD and TD groups alone, excluding the Non-TD children. The psychometric values were moderate, with strongest indices at 24 months: SE = 77% (95% CI: 64– 90%), SP = 85% (95% CI: 80– 91%), PPV = 55% (95% CI: 42– 68%), and NPV = 94% (95% CI: 90– 98%). Sensitivity indices were higher at all ages for specifically detecting ASD than for more broadly detecting any delays, but PPV was substantially reduced. Values for all ages are provided in Table 3.

Discussion

The current study examined the psychometric properties of the ITC in a sample of infants at high- and low-risk for ASD. This is the first study to provide longitudinal ITC data from a high-risk sample, with all children tested before the age of 12 months and followed through to outcome at 36 months, allowing for a more detailed look at psychometric properties of the ITC in the first year of life.

One finding of this study is that the ITC did not perform as well in a sample of infants at elevated (familial) risk for ASD as it did in community validation studies (Wetherby et al., 2004, 2008), whether it was used to identify ASD alone or any delays (ASD + Non-TD). Specifically, when comparing children with any delays (i.e. ASD + Non-TD) to TD children, PPV ranged from a low of 31% at 6 months to a high of 62% at 24 months in our study. Longitudinally, SE ranged from 51% to 62% and SP from 42% to 85% in our study. SE was improved, but PPV worsened, in analyses focusing on the detection of ASD alone from TD. The psychometric indices we obtained are all below recommended values for screening instruments (Clark & Harrington, 1999) and below the indices reported by Wetherby et al. (2004, 2008). Thus, we suggest caution when using the ITC in the context of a high-risk sample.

The lower PPV found in the current sample, relative to Wetherby et al. (2008), is somewhat surprising, given the higher prevalence of ASD in high-risk groups. The PPV of any screening instrument depends on the base rate of the condition in the population within which it is being tested. Clark and Harrington (1999) showed that even when screening tools have acceptable SE and SP (i.e., 80% and above), they demonstrate low PPV indices when used to detect low prevalence disorders such as ASD. This was demonstrated in the meta-analysis of the M-CHAT as well (Yuen et al., 2018). In the current sample, 46 of 283 participants (16.3%) had ASD, which is a much higher base rate than in the general

population, suggesting we might have expected more optimal (rather than worse) performance of the ITC relative to the Wetherby et al. (2008) validation study.

In both the current investigation and previous studies (Wetherby et al., 2008), screening accuracy is weakest in the first year of life and improves over time, with the strongest results obtained at the latest ages. This is not surprising, given the heterogeneity of symptom onset, which may begin in the first year of life in some children but unfolds along different timelines and at different rates across affected children. This may raise the question of whether it is worth screening for ASD in the first year of life. Since early treatment, which is known to optimize outcomes, depends upon early detection, we believe that striving for the goal of first year screening is not only worthwhile, but critical. It is possible that combining measures or conducting two-stage screenings (e.g., following a positive ITC with the M-CHAT) would bolster the performance of screening tools in the first year of life, which may be a fruitful approach for future investigation.

One of the strengths of this study is that it followed all participants, including screen-negative cases, to outcome at 36 months, permitting calculation of false negative rates, SE, and SP, which not all prior studies have been able to do. Another strength is the extensive longitudinal data, with all participants tested in the first year of life, providing psychometric properties at earlier ages for the ITC. Some of the limitations of the study include reduced generalizability of the results to a general community sample and small sample sizes relative to most screening studies.

ASD screening practices have shown exciting advances in recent years. However, the field still needs a systematic approach for identifying ASD risk that addresses the challenges of phenotypic variability and differing patterns of ASD symptom onset early in life. The abundance of research demonstrating that signs of ASD emerge in the latter half of the first year of life (Landa, Gross, Stuart, & Faherty, 2013; Ozonoff et al., 2010) calls for a closer look at the use of screening tools under the age of 18 months. Early recognition of ASD-related behaviors affords the opportunity for more timely provision of interventions that can have positive cascading effects on adaptive, cognitive, and social development.

Acknowledgments:

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®).

Arunyanart W, Fenick A, Ukritchon S, Imjaijitt W, Northrup V, & Weitzman C (2012). Developmental and autism screening: A survey across six states. Infants & Young Children, 25(3).

Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, … Dowling NF (2018). Prevalence of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveillance Summaries, 67(6), 1–23. 10.15585/mmwr.ss6706a1

Clark A, & Harrington R (1999). On diagnosing rare disorders rarely: Appropriate use of screening instruments. The Journal of Child Psychology and Psychiatry and Allied Disciplines, 40(2), 287–290. https://doi.org/ DOI: undefined

Dudova I, Markova D, Kasparova M, Zemankova J, Beranova S, Urbanek T, & Hrdlicka M (2014). Comparison of three screening tests for autism in preterm children with birth weights less than 1,500 grams. Neuropsychiatric Disease and Treatment, 10, 2201–2208. 10.2147/NDT.S72921 [PubMed: 25484588]

Gotham K, Pickles A, & Lord C (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. Journal of Autism and Developmental Disorders, 39(5), 693–705. [PubMed: 19082876]

Johnson CP, & Myers SM (2007). Identification and evaluation of children with autism spectrum disorders. Pediatrics, 120(5), 1183–1215. 10.1542/peds.2007-2361 [PubMed: 17967920]

Khowaja MK, Hazzard AP, & Robins DL (2015). Sociodemographic barriers to early detection of autism: Screening and evaluation using the M-CHAT, M-CHAT-R, and Follow-Up. Journal of Autism and Developmental Disorders, 45(6), 1797–1808. 10.1007/s10803-014-2339-8 [PubMed: 25488122]

Koegel LK, Koegel RL, Ashbaugh K, & Bradshaw J (2014). The importance of early identification and intervention for children with or at risk for autism spectrum disorders. International Journal of Speech-Language Pathology, 16(1), 50–56. 10.3109/17549507.2013.861511 [PubMed: 24328352]

Landa RJ, Gross AL, Stuart EA, & Faherty A (2013). Developmental trajectories in children with and without autism spectrum disorders: The first 3 years. Child Development, 84(2), 429–442. 10.1111/j.1467-8624.2012.01870.x [PubMed: 23110514]

Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, & Bishop SL (2012). Autism Diagnostic Observation Schedule Manual (2nd ed.). Torrance, CA: Western Psychological Services.

Macari SL, Wu GC, Powell KK, Fontenelle S, Macris DM, & Chawarska K (2018). Do parents and clinicians agree on ratings of autism-related behaviors at 12 months of age? A study of infants at high and low risk for ASD. Journal of Autism and Developmental Disorders, 48(4), 1069–1080. [PubMed: 29181689]

Miller JS, Gabrielsen T, Villalobos M, Alleman R, Wahmhoff N, Carbone PS, & Segura B (2011). The Each Child Study: Systematic screening for autism spectrum disorders in a pediatric setting. Pediatrics, 127(5), 866 LP–871. 10.1542/peds.2010-0136 [PubMed: 21482605]

Mullen EM (1995). Mullen Scales of Early Learning. MN: Circle Pines, AGS.

Norris M, & Lecavalier L (2010). Screening Accuracy of Level 2 Autism Spectrum Disorder Rating Scales: A Review of Selected Instruments. Autism, 14(4), 263–284. 10.1177/1362361309348071 [PubMed: 20591956]

Oosterling IJ, Swinkels SH, van der Gaag RJ, Visser JC, Dietz C, & Buitelaar JK (2009). Comparative analysis of three screening instruments for autism spectrum disorder in toddlers at high risk. Journal of Autism and Developmental Disorders, 39(6), 897–909. 10.1007/s10803-009-0692-9 [PubMed: 19205862]

Ozonoff S, Iosif A-M, Baguio F, Cook IC, Hill MM, Hutman T, … Young GS (2010). A prospective study of the emergence of early behavioral signs of autism. Journal of the American Academy of Child & Adolescent Psychiatry, 49(3), 256–266.e2. 10.1016/j.jaac.2009.11.009 [PubMed: 20410715]

Pierce K, Carter C, Weinfeld M, Desmond J, Hazin R, Bjork R, & Gallagher N (2011). Detecting, studying, and treating autism early: The one-year well-baby check-up approach. The Journal of Pediatrics, 159(3), 458–465.e6. [PubMed: 21524759]

Robins DL, Casagrande K, Barton M, Chen C-MA, Dumont-Mathieu T, & Fein D (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). Pediatrics, 133(1), 37–45. [PubMed: 24366990]

Robins DL, Fein D, Barton ML, & Green JA (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. Journal of Autism & Developmental Disorders, 31(2), 131. [PubMed: 11450812]

Rubin DB (1987). The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for

creating a few imputations when fractions of missing information are modest: The SIR. Journal of the American Statistical Association, 82(398), 543–546. 10.2307/2289460

Rutter M, Bailey A, & Lord C (2003). The social communication questionnaire: Manual. Western Psychological Services.

Sacrey L-AR, Bryson S, Zwaigenbaum L, Brian J, Smith IM, Roberts W, … Garon N (2016). The Autism Parent Screen for Infants: Predicting risk of autism spectrum disorder based on parent-reported behavior observed at 6–24 months of age. Autism, 22(3), 322–334. 10.1177/1362361316675120 [PubMed: 29671640]

Sheldrick RC, Benneyan JC, Kiss IG, Briggs-Gowan MJ, Copeland W, & Carter AS (2015). Thresholds and accuracy in screening tools for early detection of psychopathology. Journal of Child Psychology and Psychiatry, 56(9), 936–948. 10.1111/jcpp.12442 [PubMed: 26096036]

Toh T-H, Tan VW-Y, Lau PS-T, & Kiyu A (2018). Accuracy of Modified Checklist for Autism in Toddlers (M-CHAT) in detecting autism and other developmental disorders in community clinics. Journal of Autism and Developmental Disorders, 48(1), 28–35. 10.1007/s10803-017-3287-x [PubMed: 28866856]

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

Wetherby AM, Brosnan-Maddox S, Peace V, & Newton L (2008). Validation of the Infant-Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. Autism: The International Journal of Research and Practice, 12(5), 487–511. 10.1177/1362361308094501 [PubMed: 18805944]

Wetherby AM, Woods J, Allen L, Cleary J, Dickinson H, & Lord C (2004). Early indicators of autism spectrum disorders in the second year of life. Journal of Autism and Developmental Disorders, 34(5), 473–493. [PubMed: 15628603]

Wetherby A, & Prizant B (2002). Communication and Symbolic Behavior Scales Developmental Profile: First Normed Edition. Baltimore, MD: Brookes Publishing.

Yamaguchi Y, Misumi T, & Maruo K (2018). A comparison of multiple imputation methods for incomplete longitudinal binary data. Journal of Biopharmaceutical Statistics, 28(4), 645–667. [PubMed: 28886277]

Yuen T, Penner M, Carter MT, Szatmari P, & Ungar WJ (2018). Assessing the accuracy of the Modified Checklist for Autism in Toddlers: A systematic review and meta-analysis. Developmental Medicine & Child Neurology, 60(11), 1093–1100. [PubMed: 29992541]

Zwaigenbaum L, Bauman ML, Choueiri R, Fein D, Kasari C, Pierce K, … Smith Roley S (2015). Early identification and interventions for autism spectrum disorder: Executive summary. Pediatrics, 136, S1–S9. [PubMed: 26430167]

Zwaigenbaum L, Bryson S, Lord C, Rogers S, Carter A, Carver L, … Yirmiya N (2009). Clinical Assessment and Management of Toddlers With Suspected Autism Spectrum Disorder: Insights From Studies of High-Risk Infants. Pediatrics, 123(5), 1383–1391. Retrieved from http://10.0.6.6/peds.2008-1606 [PubMed: 19403506]

Zwaigenbaum L, Thurm A, Stone W, Baranek G, Bryson S, Iverson J, … Sigman M (2007). Studying the emergence of autism spectrum disorders in high-risk infants: Methodological and practical issues. Journal of Autism & Developmental Disorders, 37(3), 466–480. Retrieved from http://10.0.3.239/s10803-006-0179-x [PubMed: 16897376]

Table 1.

Studies using the Infant-Toddler Checklist as a screener in young children

Study	Age (months)	Sample Size	Recruitment Groups	Outcome Groups	Specificity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
				Autism spectrum disorder + communication delays vs. typical development				
Pierce et al. (2011)	12	10479	General	ASD, LD, DD, other	-	-	75	-
Wetherby et al. (2004)	< 24	3026	General	ASD, DD, TD	89	89	94	80
Wetherby et al. (2008)	18	5385	General	ASD, LD, TD	-	-	77	90
	24						79	88
				Autism spectrum disorder only vs. typical development				
Wetherby et al. (2004)	< 24	3026	General	ASD, TD	94	89	90	94
Wetherby et al. (2008)	18-24	5385	General	ASD, TD	93	-	-	-
Oosterling et al. (2009)	8-44	238	High-risk [1]	ASD, Non-ASD	71	59	78	50
Dudova et al. (2014)	24	155	<1500g birth weight	ASD, TD	85	85	38	98

Note. ASD = Autism spectrum disorder. TD = typical development. DD = developmental delay. LD = language delay.

[1] High-risk if scored positive on the Early Screening of Autistic Traits (ESAT) or due to clinical concern.

Table 2.

Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value of the Infant-Toddler Checklist for distinguishing any delays (ASD + Non-TD) from TD in a high-risk sample.

| | ASD + Non-TD outcome at 36 months | | TD outcome at 36 months | | | | | |
	NON-TD (True Positives)	TD (False Negatives)	NON-TD (False Positives)	TD (True Negatives)	Sensitivity[1] (95% CI)	Specificity[1] (95% CI)	PPV[1] (95% CI)	NPV[1] (95% CI)
6 months	38	21	73	56	62% (50% - 74%)	42% (33% - 50%)	31% (23% - 39%)	72% (63% - 82%)
9 months	37	34	70	106	52% (41% - 63%)	60% (53% - 67%)	35% (27% - 44%)	75% (68% - 82%)
12 months	47	29	70	109	62% (51% - 73%)	61% (54% - 68%)	40% (31% - 49%)	79% (73% - 86%)
18 months	35	37	36	141	51% (40% - 62%)	77% (71% - 84%)	49% (38% - 60%)	79% (73% - 85%)
24 months	38	32	23	150	57% (46% - 68%)	85% (80% - 91%)	62% (51% - 74%)	83% (77% - 88%)

[1]
Since sample size varied by visit, multiple imputations for incomplete longitudinal binary data were conducted. Estimates and CI were calculated after generating 100 data sets using multiple imputations and pooling the results.

Table 3.

Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value of the Infant-Toddler Checklist for distinguishing ASD from TD in a high-risk sample.

	ASD outcome at 36 months		Non-ASD outcome at 36 months					
	ASD (True Positives)	NON-ASD (False Negatives)	ASD (False Positives)	NON-ASD (True Negatives)	Sensitivity[1] (95% CI)	Specificity[1] (95% CI)	PPV[1] (95% CI)	NPV[1] (95% CI)
6 months	22	11	73	56	64 (48% – 79%)	42% (33% – 50%)	20% (13% – 27%)	83% (75% – 91%)
9 months	22	17	70	106	55% (40% – 70%)	60% (53% – 67%)	24% (16% – 32%)	85% (79% – 91%)
12 months	31	11	70	109	74% (60% – 87%)	61% (54% – 68%)	30% (22% – 39%)	91% (86% – 96%)
18 months	26	13	36	141	68% (54% – 82%)	77% (71% – 84%)	41% (30% – 52%)	91% (87% – 96%)
24 months	28	9	23	150	77% (64% – 90%)	85% (80% – 91%)	55% (42% – 68%)	94% (91% – 98%)

[1] Since sample size varied by visit, multiple imputations for incomplete longitudinal binary data were conducted. Estimates and CI were calculated after generating 100 data sets using multiple imputations and pooling the results.