Title: Beyond Pass-Fail: Examining the Potential Utility of Two Thresholds in the Autism Screening Process

Running Title: Two Thresholds in Autism Screening

Megan Y. Roberts,[acd] Yael Stern,[ad] Lauren H. Hampton,[a] Jeffrey M. Grauzer,[a] Amanda Miller,[ac] Amy Levin,[ac] Benjamin Kornfeld,[abc] Matthew M. Davis,[abd] Aaron Kaat,[ad] & Ryne Estabrook[a]

Affiliations: [a]Northwestern University, 2240 Campus Drive, Evanston IL, 60208; [b]Ann & Robert H. Lurie Children's Hospital of Chicago, 225 East Chicago Ave, Chicago IL, 60611; [c]Northwestern University Center for Audiology, Speech, Language and Learning, 2315 Campus Dr., Evanston IL, 60208; [d]Northwestern University Institute for Innovations in Developmental Sciences, 633 North St. Clair, Chicago, IL 60611.

Address correspondence to: Megan Y. Roberts, Department of Communication Sciences and Disorders, 2240 Campus Drive, Frances Searle, Room 3-346, Evanston, IL 60208, [megan.y.roberts@northwestern.edu] 847-491-2416 (phone), 847-467-7141 (Fax)

Acknowledgments: We would like to thank Dean Barbara O'Keefe, Dr. Sumit Dhar, Dr. Dana Brazdziunas, Dr. Sarah Bauer, and Denise Eisenhauer for supporting Northwestern University's Developmental Diagnostics Program, without which this research would not have been possible.

Grant Funding: None

Number of Text Pages: 16

Number of Tables: 7

Number of Figures: 1

Lay Summary: This study examined the benefits of using two versus one cutoff score when screening for autism. Results indicate that having two scores and weighting test items based on predictive association with an autism diagnosis is better than using a single score and weighting each item equally. Using such an approach may reduce the wait time for specialty autism diagnostic evaluations, such that specialty evaluations may be reserved for those cases that are more ambiguous or more complex.

Abstract

Access to early intervention as early in development as possible is critical to maximizing long-term outcomes for children with autism spectrum disorders (ASD). However, despite the fact that ASD can be reliably diagnosed by 24 months, the average age of diagnosis is two years later. Waitlists for specialized developmental evaluations are one barrier to early diagnosis. The purpose of this study was to examine one potential approach to reducing wait time for an ASD diagnostic evaluation by examining the utility of using more than one threshold for an autism screening tool, the Screening Tool for Autism in Toddlers and Young Children (STAT). Participants included 171 children between 24 and 36 months of age who received a medical diagnostic evaluation through Illinois' Early Intervention Program. This study directly compared the performance of the STAT when scored: (a) using the original single threshold, (b) using seven equally weighted items using a single threshold, and (c) using all items differentially weighted based on how strongly that item predicts a later ASD diagnosis. In addition, this study explored the potential utility of using two thresholds rather than a single threshold for each scoring method. Results of this study suggest that using a two-threshold logistic regression method has potential psychometric advantages over a single threshold and categorical scoring. Using this approach may reduce the wait time for specialty ASD diagnostic evaluations by maximizing true negatives and true positives, such that specialty evaluations may be reserved for those cases that are more ambiguous or more complex.

Introduction

There is strong empirical evidence that access to early intensive behavior therapy as early in development as possible is critical to maximizing long-term outcomes for children with autism spectrum disorders (ASD; Warren et al., 2011). However, despite the fact that ASD can be reliably diagnosed by 24 months of age (Kleinman et al., 2008; Lord et al., 2006), the average age of diagnosis is 46 months (CDC, 2016). Furthermore, the time between an initial developmental evaluation and receipt of an ASD diagnosis is on average 13 months (Wiggins, Baio, & Rice, 2006). Given that many insurance companies require an ASD diagnosis for a child to receive ASD specific intervention services, such as applied behavior analysis therapy, reducing the average age of ASD diagnosis is critical. Waitlists for specialized developmental evaluations by child psychologists or developmental-behavioral pediatricians are one barrier to early diagnosis in the United States. The purpose of this study was to examine one potential approach to reducing wait time for an ASD diagnostic evaluation by examining the utility of using two thresholds to an ASD screening tool, the Screening Tool for Autism in Toddlers and Young Children (Stone, Coonrod, & Ousley, 2000).

The Autism Diagnostic Process

Recommendations regarding the ASD diagnostic process vary (Filipek et al., 2000; Johnson, et al., 2007; Molloy, Murray, Akers, Mitchell, & Manning-Courtney, 2011; Volkmar et al., 2014) but often include an interdisciplinary team using standardized and validated measures. Tools used to assess ASD symptoms are classified according to their purpose. Level 1 screeners are used to identify children at risk for ASD in the general population. These screeners are parent-report measures and are often implemented in primary care practices as part of well-child visits. Level 2 measures are designed to differentiate ASD risk from risk for other developmental

disorders. Many ASD level 2 screeners use direct observation by a trained clinician, given the limitations of parent-report measures (Norris & Lecavalier, 2010). Finally, measures such as the Autism Diagnostic Observation Schedule (ADOS-2; Lord, Rutter, Dilavore, Risi, Gotham, & Bishop, 2012) are used in conjunction with clinical judgment to make an ASD diagnosis.

In practice, Level 1 screeners for ASD are used 8% to 60% of the time (Arunyanart et al., 2012; Dosreis, Weiner, Johnson, & Newschaffer, 2006; Gillis, 2009; Self, Parham, & Rajagopalan, 2015). Furthermore, two-tiered screening (i.e., screening children who screened positive on a Level 1 screener with a Level 2 screener) for ASD is rarely used in the United States (Khowaja, Robins, & Adamson, 2017). Despite their lack of use, Level 2 screeners may help to reduce the long waitlists for specialty ASD diagnostic evaluations, by reducing the number of false positives. In addition, given that the majority of children (89%) are already enrolled in early intervention at the time of the ASD evaluation, Level 2 screeners may be implemented by early intervention providers (Monteiro et al., 2016).

The Screening Test for Autism in Toddlers and Young Children

Some Level 2 screeners include a structured observation of the child, reducing the reliance on parent-report of ASD-related behaviors that parents might not easily recognize. The Screening Test for Autism in Toddlers and Young Children (STAT; Stone et al., 2000) is the only structured observational Level 2 measure for children under 36 months of age (Johnson et al., 2007). However, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the STAT varies across studies from .47 (Khowaja et al., 2017) to .92 (Stone, Coonrod, Turner, & Pozdol, 2004) for sensitivity, from .70 (Newschaffer et al., 2017) to .86 (Stone et al., 2000) for specificity, from .77 (Khowaja et al., 2017; Stone et al., 2000) to .86

(Stone et al., 2004) for PPV and from .41 (Newschaffer et al., 2017) to .92 (Stone et al., 2004) for NPV.

This variability has prompted consideration of alternate scoring methods of the STAT. For example, an alternative 7-item scoring of the STAT equally weighting 7 of the 12 items that were most predictive of an ASD diagnosis increased sensitivity from .47 to .78 (Khowaja et al., 2017). These items were selected via a discriminant function analysis, which identified the relative strength of each item in its prediction of ASD diagnosis, which were then summed and assessed for an optimal screening threshold. However, researchers are unlikely to administer only seven of the STAT items, and the current availability of more sophisticated analytic methods removes the constraint of simple sum scores. If alternative scoring of the STAT can improve sensitivity, specificity, and performance in general, multiple modern statistical methods should be used and compared.

Screening Thresholds

In order to calculate psychometric values such as sensitivity, specificity, PPV, and NPV, a single threshold (pass or fail) must be chosen. In order to achieve high sensitivity, the threshold is often set lower such that the number of false negatives are low. However, using a single threshold often requires the prioritization of sensitivity over specificity, PPV over NPV, or vice-versa. Furthermore, the use of a single threshold does not allow for the consideration that not every child who scores above a given threshold has the same probability of having ASD. Children who score at or just above the threshold are less likely to have ASD than children with the highest scores. As such, positive predictive value based on a single threshold may over- or underestimate a child's probability of having ASD. Sheldrick and Garfinkel (2017) suggest that test developers should consider two thresholds for different clinical decisions and note that such

thresholds are found in other pediatric disorders such as serum bilirubin monitoring to identify the likelihood of developing clinically significant hyperbilirubinemia (American Academy of Pediatrics, 2004).

Current Study

Given the wide variability in reported psychometric properties of the STAT and the potential utility of Level 2 screeners for streamlining the ASD diagnostic process, the goal of this study was to compare different methods of scoring the STAT in a clinically referred sample of toddlers. Specifically, this study directly compared the performance of the STAT when scored: (a) using the original single threshold (2.0 or higher), (b) using seven equally weighted items using a single threshold (3.0 or higher), and (c) using all items differentially weighted based on how well that item predicts an ASD diagnosis. In addition, this study explored the potential utility of using two thresholds rather than a single threshold for each scoring method. Using two thresholds takes into consideration that not all children above or below a single threshold are equally likely or not likely to have ASD. The use of two thresholds allows for the classification of three rather than two groups of children: (a) a lower group that is very unlikely to have ASD, (b) a middle group in which a diagnosis of ASD is more uncertain, and (c) an upper group that is very likely to have ASD. Identification of these groups may support the reduction of wait times such that the lower group is not referred to a specialty diagnostic program, and the upper group may be identified in primary care settings such as primary care pediatric practices or early intervention programs, leaving the middle group for specialty diagnostic programs.

Method

Study Design

Prediction accuracy of a regression-based scoring method for the STAT was calibrated on part of our dataset and was validated on another independent portion of our data. To do this, we randomly split our sample such that two-thirds of our data was used in the calibration sample, and one third in the validation sample. This split is a compromise between 50:50, 70:30, and 80:20 ratios (Nelles, 2001) that maximize calibration sample size without making the validation sample impracticably small. This study was approved by the Northwestern University Institutional Review Board (STU00204144).

Participants

The sample included 171 children between 24 and 36 months of age who were evaluated as part of a medical diagnostic evaluation through Illinois' Early Intervention Program. All children were referred by their service coordinators based on: (a) significant developmental delays, (b) lack of progress, (c) unexpected regression, or (d) atypical development that could not be explained based on known medical, developmental etiology. The majority of the entire sample (74%) was male, and the mean age of children was 31.80 months (SD = 3.16 months). Additional demographic information is provided in Table 1. In this sample, 119 (70%) children received a diagnosis of ASD.

Procedures and Measures

Each child was assessed by a multidisciplinary team that was composed of a developmental-behavioral pediatrician, a pediatric speech-language pathologist, a developmental therapist, and a pediatric audiologist. The evaluation process occurred over the course of three, appointments with each child and their caregivers. During the first visit, the STAT (Stone et al., 2000, 2004), the Mullen Scales of Early Learning (*MSEL*; Mullen, 1995), the Preschool Language Scale-5th Edition (*PLS-5*, Zimmerman, Steiner, & Pond, 2011) and a physical

examination were administered. The first visit lasted approximately 1 hour and 15 minutes (20 minutes to administer the STAT, 15 minutes to administer the Visual Reception Subscale of the MSEL, 20 minutes to administer the PLS-5, and 5 minutes for the physical examination). While therapists were administering these child assessments, the developmental-behavioral pediatrician was conducting a parent interview to gather additional information about the child's development and the parents' concerns. During the second visit, which lasted approximately one hour, the Autism Diagnostic Observation Schedule - Second Edition (*ADOS-2*; Lord et al., 2012) was administered. During the family's third, hour-long, visit to the clinic, the developmental-behavioral pediatrician explained the diagnostic decision and offered caregivers support for seeking additional resources.

Screening Tool for Autism in Toddlers and Young Children (*STAT*; (Stone et al., 2000, 2004)**.** The STAT is a screening tool that is scored based on clinical observation of a child's social and communication skills during a semi-structured, standardized assessment. The assessment, appropriate for use with children between 24 and 36 months, consists of 12 probes of skills that are characteristically impaired in toddlers with ASD. These target skills fall into the domains of Play, Imitation, Requesting, and Directing Attention. A child has multiple opportunities to receive credit for each probe. A score of 2 or greater on the STAT indicates a failed screening, the presence of ASD risk, and signals the need for an ASD-specific evaluation. The STAT was designed as a Level 2 screener, to be used to identify children who are specifically at high risk for ASD within a referred clinic sample, as opposed to identify at-risk children in the general population (Stone et al., 2004). In the current study, the developmental therapist administered the STAT to all children at their first assessment visit. The developmental therapist had completed the online STAT training.

Autism Diagnostic Observation Schedule - Second Edition (*ADOS-2*; Lord et al., 2012). The ADOS-2 is a standardized observational assessment of child behavior that is a common tool for diagnosing ASD. The ADOS-2 was designed to elicit the full range of a child's spontaneous social communication skills during semi-structured play interactions with an examiner in the presence of the child's caregiver. Activities within the 40- to 60-minute ADOS-2 administration are designed to probe for behaviors symptomatic of ASD per the two core diagnostic criteria specified in the DSM-5 (American Psychiatric Association, 2013).

ADOS-2 administration procedures and scoring criteria varied slightly depending on a child's chronological age. Children between 12-30 months were assessed using the Toddler Module of the ADOS-2, while Module 1 was administered with children who were greater than 31 months and who were not yet using spontaneous phrase speech (i.e., flexible three-word utterances). Administration differences exist between the two modules in order to tailor assessment activities to the needs and interests of children of different developmental ages. On both modules a child's behavior is coded for the presence of indicators of ASD. In the current study, the developmental therapist administered the ADOS-2 at the second assessment visit. The developmental therapist was research reliable on ADOS-2 administration and scoring.

Mullen Scales of Early Learning (*MSEL*; Mullen, 1995). The MSEL is a standardized assessment normed for use with children between birth and 68 months that measures five developmental domains: expressive and receptive language, gross and fine motor, and visual reception. The developmental therapist administered only the Visual Reception Scale to all children in the current study during their first visit to the diagnostic program. This scale has a mean T-score of 50 and a standard deviation of 10. This scale was used to evaluate a child's

nonverbal cognition and involved the child's completion of problem-solving activities such as nesting cups, a simple puzzle, and matching tasks.

Preschool Language Scales - Fifth Edition (*PLS-5*; Zimmerman, Steiner, & Pond, 2011). The PLS-5 measures receptive and expressive communication skills. During this assessment, which is appropriate for use with children through age seven, children were guided through structured tasks such as identifying objects, completing simple instructions, and labeling pictures of nouns and actions. The speech-language pathologist administered the PLS-5 to all children in the current study during their first visit to the diagnostic program. Auditory Comprehension and Expressive Communication Subscales each have a mean standard score of 100 and a standard deviation of 15. Items designed to assess communication skills that are developmentally expected of children up to 30-month olds can be scored through observation, caregiver report, or elicitation.

Consensus Clinical Diagnosis. Following each child's first and second assessment appointments, the multidisciplinary team of expert clinicians met to collaboratively determine an appropriate diagnosis for the child. Diagnoses were made based on DSM-5 criteria for ASD (American Psychiatric Association, 2013), requiring that a child present with both (1) impairments in social communication, such as maintaining reciprocal social interactions, and (2) restricted or repetitive behaviors, such as atypical sensory interests or sensitivities. Consensus diagnostic discussions allowed for a nuanced, global interpretation of the child's behavior across multiple testing days and with various clinicians. Clinicians considered the child's performance during naturalistic, play-based assessments as well as during structured assessments, such as the ADOS-2. In addition, caregiver reports during semi-structured interviewing, developmental and medical histories, and observational impressions of the child were also considered.

Multidisciplinary diagnostic decisions are an optimal approach to diagnosing young children with ASD given the frequent complexity of the behavioral presentation of ASD before three years of age (Charman & Baird, 2002). The multidisciplinary diagnostic process minimizes the diagnostic uncertainty that would be inherent in using a single assessment or a single discipline to differentially diagnose ASD, especially in the presence of expressive-receptive language delays and significant problem behaviors. Notably, this approach has been demonstrated to produce diagnoses that are stable between two and three years of age (Lord, 1995; Stone et al., 1999).

Scoring Methods. Three methods of scoring the STAT were used: two previously existing methods and one that was developed using this dataset. First, we used the original scoring method (Stone et al., 2000, 2004) which groups individual items into four domains, assigns each item a weighted score within its domain, such that the maximum score in a given domain totals 1.0, and then sums each domain score for an overall STAT score, such that the possible range of scores is 0 (lowest risk of ASD) to 4 (highest risk of ASD). The pass-fail threshold using the original scoring method is 2 or higher. In the second, we used an alternative scoring 7-item scoring method (Khowaja et al., 2017), which uses a sum of seven of the twelve STAT items to create a total score ranging from zero to seven with a threshold of three or higher for risk for ASD. In addition to these existing methods, a new regression-based method was also evaluated.

Data Analysis

For the regression-based method, we used a logistic regression of all twelve items, which yielded a predicted probability of having ASD for each individual in the sample and weighted each item on the STAT based on how well that item predicted an ASD diagnosis. We used

LASSO regression (Tibshirani, 1996), an alternative to stepwise model selection that includes a parameter penalty to reduce overfitting and dependence on a single sample. A single threshold was calculated using best fitting thresholds as defined by Youden's J statistic (Youden, 1950).

All three methods were assessed for their predictive accuracy using two types of statistics using receiver operator characteristic (ROC) analyses (Fawcett, 2006). First, we calculated prediction accuracy for a given cut-off value for single and two thresholds using statistics including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR-), and accuracy. These values were computed for each of the three scoring methods. We also summarized overall fit of each prediction model using the area under the curve (AUC) statistic, providing a measure of overall prediction accuracy for all possible thresholds. Finally, we explored the potential utility of using two thresholds rather than one threshold, in order to maximize the number of true negative cases while minimizing false negatives (i.e., NPV close to 1.00) and to maximize true positive cases while minimizing false positives (i.e., PPV close to 1.00). All analyses were conducted using the R software package (R Core Team, 2017).

Results

Original STAT Scoring Method

Results for the original published scoring method for the STAT for the entire sample are presented in Table 2. Possible scores on the STAT range from 0 to 4 in increments of 0.25. The original published scoring cutoff of 2.0 (a STAT sum score at or above 2 indicates risk for ASD) yielded sensitivity of 0.93, specificity of 0.69, a PPV of 0.87, an NPV of 0.82, an LR+ of 3.03, an LR- of 0.10, and accuracy of 86%. Ranging across all possible cutoffs, this scoring method yielded an AUC of 0.89 in the full sample.

7-Item Scoring Method

Results for the 7-item scoring method method (Khowaja et al., 2017) are shown in Table 3. The suggested scoring cutoff of three out of seven failed items yielded sensitivity of 0.98, specificity of 0.58, a PPV of 0.84, an NPV of 0.91, an LR+ of 2.30, an LR- of 0.04, and accuracy of 85.4%. This model had an AUC of 0.89 in the full sample.

Logistic Regression Scoring Method

Results for the logistic calibration regressions are presented in Table 4. Univariate logistic and logistic regression coefficients for the calibration sample are provided in Table 5. We used Youden's J statistic (Youden, 1950) to identify an optimal probability threshold. The optimal threshold cutoff on the model predicted probability was 81.6% which yielded sensitivity of 0.88, specificity of 0.77, a PPV of 0.90, an NPV of 0.72, an LR+ of 3.72, an LR- of 0.16, and accuracy of 84.2% for the validation sample. Similar AUCs were observed between all three samples (full, calibration, and validation), with an AUC of .91 for the full sample, 0.92 for the calibration sample, and 0.88 for the validation sample. See Table 6 and Figure 1 for a comparison of the predictive accuracy across all scoring methods.

Two Thresholds

We then considered the benefits of using two thresholds on the STAT; one to maximize the number of true negative cases while minimizing false negatives (i.e. NPV close to 1.00) and another to maximize true positive cases while minimizing false positives (i.e. PPV close to 1.00) for each of the scoring methods. For our sample, using the original scoring and a lower threshold of 1.0 or lower on the STAT, 15 out of 15 children were correctly classified as not having ASD, with a false negative rate of 0. An upper threshold of 4.0 correctly classified 9 out of 9 cases of

ASD, with a false positive rate of 0. Taken together, a lower threshold of 1 and an upper threshold of 4 yielded 100% accuracy in ASD diagnosis for 14% of the entire sample.

Two thresholds for the 7-item scoring method yielded similar results. Using a lower threshold of 1 or lower, 7 out of 7 children were correctly identified as not having ASD, with a false negative rate of 0. An upper threshold of 7 correctly classified 23 of 24 cases of ASD, with only one false positive. Using 1 as the lower and 7 as the upper threshold yielded 97% accuracy for 18% of the entire sample.

Our logistic regression method revealed the strongest potential for utilizing two thresholds when determining children at the highest and lowest risk for ASD. A lower threshold at or below 25% from the logistic regression method correctly identified nine truly negative cases for ASD with no false negatives. An upper threshold at 95% or higher from the logistic regression method correctly classified 21 out of 22 children with AS, with only one false positive. Using 25% as the lower threshold and 95% as the upper threshold yielded 97% accuracy for 54% of the validation sample. See Table 7 for a comparison of thresholds across STAT scoring methods.

Discussion

This study compared three different methods for scoring the STAT. All three methods performed similarly, providing full sample AUCs of approximately 0.89 to 0.90, a maximum of 86% accuracy, and better sensitivity than specificity at common thresholds. It is important to note that the logistic regression (Table 5) yielded a regression coefficient in the opposite direction than would be expected for item 11 (Imitation: Drum Hands), indicating that failing this item (i.e., a child did not imitate drumming hands) was associated with reduced ASD risk controlling for all other STAT items. A similar pattern was also present in the discriminant

analyses of Khowaja and colleagues (2017). While this partial coefficient goes in the opposite direction, two additional factors are worth discussion. First, all of the simple associations are in the expected direction, indicating that increases in any single symptom are associated in increased ASD risk. Second, the item that has a negative relationship is from the Imitation subdomain, which includes four of the twelve STAT items. As these regression coefficients are partial coefficients, they are the effect of a particular item holding all other items constant. Because all of the items in a particular domain are highly correlated, evaluating a single item while holding three other highly-correlated items constant can lead to unstable results.

The primary difference between these methods is in how they approach final categorization of ASD risk. The two existing methods provide a discrete binary categorization, with no further differentiation between individuals or any measure of certainty. The logistic regression approach provides a predicted probability of ASD, which can be interpreted on its own or categorized based on a clinician's needs. Simple categorical scoring of the STAT does not adequately address the fact that not all children above the threshold have the same probability of having ASD. A logistic regression approach provides a continuous score that can be transformed into a simple probability: children with a score of 0.50 have a 50% chance of having ASD, which may be more clinically meaningful than a conventional STAT score of 1.75.

In addition, by using two thresholds within the context of a logistic regression framework, clinicians may be able to triage which children require further testing (those in between the upper and lower thresholds). Clinicians may choose to use relatively high thresholds to remove false positives, low thresholds to omit false negatives, or both. A predicted probability lower threshold at or below 25% from the logistic model correctly identified nine truly negative cases for ASD with no false negatives and a predicted probability upper threshold at 95% or

higher correctly identified 21 out of 22 children as having autism. Thus, 54% of participants from our validation sample could be classified with a high degree of accuracy using a logistic regression method and two thresholds. This two-threshold logistic regression approach may be applied to other screening methods such as developmental surveillance and two-stage screening approaches. Future research should examine the utility of using these statistical approaches on other screening measures such as level 1 surveillance screeners (e.g., the Modified Checklist of Autism in Toddlers-Revised) or two-stage screening approaches (Khowaja et al., 2017; Robins, Casagrande, Barton, Chen, Dumont-Mathieu, & Feinl 2014).

Given the waitlists for specialized developmental evaluations and the potential utility of using two thresholds to distinguish between children for whom there is high versus low certainty of an ASD diagnosis, future research should focus on the development of open-source, two-threshold ASD diagnostic measures that may be used by general pediatricians. If half of children referred for an ASD diagnosis (i.e. those below the lower threshold, those above the upper threshold) could receive an initial diagnosis or rule out of ASD by their pediatrician, developmental-behavioral pediatricians would have increased time to devote to the more complex or ambiguous cases. In fact, pediatric specialists report that between 20 to 40% of specialty cases may be managed by a pediatrician (Corso & Greenspan, 2015). Using general pediatricians as part of a team in behavioral/developmental access clinics reduced wait time by over five months compared to a visit with a developmental-behavioral pediatrician (Harrison, Jones, Sharif, & Di, 2017). In addition, this two-threshold approach could facilitate children in the high-certainty group receiving earlier access to specialized autism-specific services.

Additional strengths of this study include the use of a LASSO prediction model and a community-based sample. The LASSO prediction model is theoretically superior, as it includes a

smoothing parameter that helps compensate for model complexity and increases generalizability. However, model-implied probabilities correlate $r > 0.99$, indicating that the relationship between STAT items and ASD is relatively robust to modeling assumptions. Future applications should seek to replicate regression-based prediction and utilize alternatives to our simple calibration/validation sample split. The study sample included children referred for a medical diagnostic evaluation through the state of Illinois' Early Intervention Program, rather than self-referred research participants. Given that participants are likely to be more representative of all children referred for a medical diagnostic evaluation in Illinois, the external validity of these results are high.

The findings of this study should be considered within the context of the study's limitations. First, our cross-validation approach is relatively simple, consisting of a single split of the sample into calibration and validation subsamples. Despite our stratification to ensure that each sample had the same frequency of ASD diagnosis, the randomness inherent to this design meant that the two samples were slightly different. The full sample regression results should yield an estimate of these aggregated results, albeit without a measure of cross-validation error. Pooling STAT data from multiple sources would allow the regression results to be improved, yielding a more accurate and better replicated prediction model. Second, the sample only included children who spoke English and lived in Illinois, and, as such, it is unclear if these results would generalize to children from other cultures or who speak languages other than English. Furthermore, all children in this sample were already enrolled in Illinois' Early Intervention Program and the specific reason for the referral for a medical diagnostic evaluation is unknown. Thus, these results may not generalize to different samples of children (e.g., children not already enrolled in early intervention). Third, while the autism diagnosis was

determined by the multidisciplinary diagnostic team using information from all assessments and parent report, the same person administered both the STAT and the ADOS-2, which may have influenced the scoring of the ADOS-2. Finally, the STAT was designed as a level 2 screener rather than a diagnostic measure. While the use of two thresholds on the STAT shows high levels of diagnostic accuracy, the items on the STAT do not represent all DSM-5 criteria for an ASD diagnosis; specifically the STAT excludes restricted and repetitive behaviors. Future research should explore using the STAT with an observation system that measures all of the DSM-5 criteria such as the Systematic Observation of Red Flags (Dow, Guthrie, Stronach, & Wetherby, 2016), or the Childhood Autism Rating Scale (Schopler, Van Bourgondien, Wellman, & Love, 2010).

Results of this study suggest that using a two-threshold, logistic regression model has potential psychometric advantages over a single threshold and categorical scoring. Using such an approach may reduce the wait time for specialty ASD diagnostic evaluations by maximizing true negatives and true positives, such that specialty evaluations may be reserved for those cases that are more ambiguous or more complex. These findings highlight the importance of developing ASD screening and diagnostic measures that use probability-based estimates with two thresholds and that may be implemented by generalists rather than specialists.

References

Agresti, A., & Coull, B. (1998). Approximate is better than "exact" for interval estimation of binomial proportions, *American Statistician, 52*, 119-126.

American Academy of Pediatrics (2004). Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. *Pediatrics*, *114*(1), 297–316. https://doi.org/10.1542/peds.114.1.297

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

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), 175–187. https://doi.org/10.1097/IYC.0b013e31825a5a42

CDC. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years: Autism and Developmental Disabilities Monitoring Networks. *Surveillance Summaries*, *63*, 1–23.

Charman, T., & Baird, G. (2002). Practitioner review: Diagnosis of autism spectrum disorder in 2- and 3-year-old children. *Journal of Child Psychology and Psychiatry*, *43*(3), 289–305. https://doi.org/10.1111/1469-7610.00022

Corso, P., & Greenspan, J. S. (2015). New patient access for pediatric specialties: Some tools and challenges. *The Journal of Pediatrics*, *166*(6), 1333–1334. https://doi.org/10.1016/j.jpeds.2015.02.057

DeLong, E., DeLong, D., & Clarke-Pearson, D. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*, 837-845.

Dosreis, S., Weiner, C. L., Johnson, L., & Newschaffer, C. J. (2006). Autism spectrum disorder screening and management practices among general pediatric providers. *Journal of Developmental and Behavioral Pediatrics, 27*(2), S88-94.

Dow, D., Guthrie, W., Stronach, S. T., & Wetherby, A. M. (2017). Psychometric analysis of the Systematic Observation of Red Flags for autism spectrum disorder in toddlers. *Autism*, *12*(3). 301-309. https://doi.org/10.1177/1362361316636760

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Filipek, P. A., Accardo, P. J., Ashwal, S., Baranek, G. T., Cook, E. H., Dawson, G., … Volkmar, F. R. (2000). Screening and diagnosis of autism: Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology*, *55*(4), 468–479.

Gillis, J. M. (2009). Screening practices of family physicians and pediatricians in 2 southern states. *Infants and Young Children*, *22*(4), 321–331.

Harrison, M., Jones, P., Sharif, I., & Di, G. (2017). General pediatrician-staffed Behavioral/Developmental Access Clinic decreases time to evaluation of early childhood developmental disorders. *Journal of Developmental and Behavioral Pediatrics*, *38*(6), 353–357. https://doi.org/10.1097/DBP.0000000000000448

Johnson, C. P., Myers, S. M., & The American Academy of Pediatrics Council on Children with Disabilities. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*(5), 1183–1215. https://doi.org/10.1542/peds.2007-2361

Khowaja, M., Robins, D. L., & Adamson, L. B. (2017). Utilizing two-tiered screening for early detection of autism spectrum disorder. *Autism*. Advance online publication. https://doi.org/10.1177/1362361317712649

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S., … Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(4), 606–615. https://doi.org/10.1007/s10803-007-0427-8

Lord, C. (1995). Follow-up of two-year-olds referred for possible autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *36*(8), 1365–1382.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition*. Torrance, CA: Western Psychological Services.

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry, 63*(6), 694–701. https://doi.org/10.1001/archpsyc.63.6.694

Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T., & Manning-Courtney, P. (2011). Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism: The International Journal of Research and Practice*, *15*(2), 143–162. https://doi.org/10.1177/1362361310379241

Monteiro, S. A., Dempsey, J., Broton, S., Berry, L., Goin-Kochel, R. P., & Voigt, R. G. (2016). Early intervention before autism diagnosis in children referred to a regional autism clinic.

Journal of Developmental and Behavioral Pediatrics, *37*(1), 15–19. https://doi.org/10.1097/DBP.0000000000000241

Mullen, E. (1995). *Mullen Scales of Early Learning*. San Antonio, TX: Pearson.

Nelles, O. (2001). *Nonlinear system identification: From classical approaches to neural networks and fuzzy models*. Berlin: Springer-Verlag.

Newschaffer, C. J., Schriver, E., Berrigan, L., Landa, R., Stone, W. L., Bishop, S., … Warren, Z. E. (2017). Development and validation of a streamlined autism case confirmation approach for use in epidemiologic risk factor research in prospective cohorts. *Autism Research*, *10*(3), 485–501. https://doi.org/10.1002/aur.1659

Norris, M., & Lecavalier, L. (2010). Screening accuracy of Level 2 autism spectrum disorder rating scales. A review of selected instruments. *Autism: The International Journal of Research and Practice*, *14*(4), 263–284. https://doi.org/10.1177/1362361309348071

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Robins, D., Casagrande, K., Barton, M., Chen, C., Dumont-Mathieu, T., & Fein, D., (2014). Validation of the Modified Checklist for Autism in Toddlers, Revised with Follow-Up. *Pediatrics, 133*(1), 37-45.

Schopler, M., Van Bourgondien, M., Wellman, G., & Love, S. (2010). *Childhood Autism Rating Scale, Second Edition*. Los Angeles, CA: Western Psychological Services.

Self, T. L., Parham, D. F., & Rajagopalan, J. (2015). Autism spectrum disorder early screening practices: A survey of physicians. *Communication Disorders Quarterly*, *36*(4), 195–207. https://doi.org/10.1177/1525740114560060

Sheldrick, R. C., & Garfinkel, D. (2017). Is a positive developmental-behavioral screening score sufficient to justify referral? A review of evidence and theory. *Academic Pediatrics*, *17*(5), 464–470. https://doi.org/10.1016/j.acap.2017.01.016

Stone, W. L., Lee, E. B., Ashford, L., Brissie, J., Hepburn, S. L., Coonrod, E. E., & Weiss, B. H. (1999). Can autism be diagnosed accurately in children under 3 years? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *40*(2), 219–226.

Stone, W. L., Coonrod, E. E., & Ousley, O. Y. (2000). Brief report: Screening Tool for Autism in Two-Year-Olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders*, *30*(6), 607–612.

Stone, W. L., Coonrod, E. E., Turner, L. M., & Pozdol, S. L. (2004). Psychometric properties of the STAT for early autism screening. *Journal of Autism and Developmental Disorders*, *34*(6), 691–701. https://doi.org/10.1007/s10803-004-5289-8

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

Volkmar, F., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., State, M., & American Academy of Child and Adolescent Psychiatry Committee on Quality Issues. (2014). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(2), 237–257. https://doi.org/10.1016/j.jaac.2013.10.013

Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-Vanderweele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics*, *127*(5), e1303-1311. https://doi.org/10.1542/peds.2011-0426

Wiggins, L. D., Baio, J., & Rice, C. (2006). Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *Journal of Developmental and Behavioral Pediatrics*, *27*(2), S79-87.

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

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool Language Scale (5th ed.)*. San Antonio: Psychological Corp.

Table 1

Participant demographics

	Whole Sample		Calibration Sample		Validation Sample	
	ASD (n=119)	Not ASD (n=52)	ASD (n=79)	Not ASD (n=35)	ASD (n=40)	Not ASD (n=17)
Age (months)	31.83 (3.16)	31.7 (3.25)	31.98 (3.17)	31.84 (3.21)	31.83 (3.35)	30.82 (3.49)
Male (%)	78%	63%	78%	63%	78%	65%
Expressive Communication, PLS-5[a]	64.44 (11.05)	80.27 (13.59)	65.32 (11.74)	80.94 (12.29)	62.77 (9.54)	79.00 (16.08)
Auditory Comprehension, PLS-5[a]	55.10 (7.99)	80.60 (18.85)	56.03 (9.14)	79.87 (17.28)	53.35 (4.82)	81.88 (21.85)
Visual Reception, MSEL[b]	25.33 (8.32)	35.94 (16.13)	25.62 (9.03)	35.85 (16.75)	24.77 (6.87)	36.12 (15.31)
STAT[c] - Sum of Domain Scores	3.04 (0.69)	1.46 (0.96)	2.95 (0.72)	1.43 (0.90)	3.21 (0.61)	1.51 (1.10)
Diagnoses						
ASD	100%	0%	100%	0%	100%	0%
Speech Sound Disorder		19%		17%		24%
Developmental Delay		52%		54%		47%
Mixed Receptive-Expressive Language Impairment		27%		26%		29%
No Diagnosis		2%		3%		0%

[a] Preschool Language Scales-Fifth Edition, Standard Score (mean = 100; SD = 15)

[b] Mullen Scales of Early Learning, Visual Reception Scale T score (mean = 50; SD = 10)

[c]Screening Tool for Autism in Toddlers and Young Children

Table 2

Original STAT[a] scoring method

STAT[a] Score	**PPV[b]**	**NPV[c]**	**Sensitivity**	**Specificity**	**Accuracy**	**LR+[d]**	**LR-[e]**	**TP[f]**	**FP[g]**	**FN[h]**	**TN[i]**
0.00	0.70	NA	1.00	0.00	0.70	1.00	NA	119	52	0	0
0.25	0.71	1.00	1.00	0.06	0.71	1.06	0.00	119	49	0	3
0.50	0.73	1.00	1.00	0.14	0.74	1.16	0.00	119	45	0	7
0.75	0.73	1.00	1.00	0.17	0.75	1.21	0.00	119	43	0	9
1.00[j]	0.76	1.00	1.00	0.29	0.78	1.41	0.00	119	37	0	15
1.25	0.79	0.95	0.99	0.39	0.81	1.61	0.02	118	32	1	20
1.50	0.83	0.90	0.98	0.54	0.84	2.11	0.05	116	24	3	28
1.75	0.86	0.89	0.97	0.64	0.87	2.64	0.05	115	19	4	33
2.00[k]	**0.87**	**0.82**	**0.93**	**0.69**	**0.86**	**3.03**	**0.10**	**111**	**16**	**8**	**36**
2.25	0.92	0.75	0.88	0.83	0.87	5.10	0.14	105	9	14	43
2.50	0.93	0.66	0.81	0.87	0.83	5.99	0.22	96	7	23	45
2.75	0.94	0.60	0.74	0.89	0.78	6.41	0.29	88	6	31	46
3.00	0.94	0.52	0.63	0.90	0.71	6.55	0.41	75	5	44	47
3.25	0.93	0.47	0.55	0.90	0.66	5.68	0.50	65	5	54	47
3.50	0.94	0.40	0.39	0.94	0.56	6.70	0.65	46	3	73	49
3.75	0.89	0.34	0.21	0.94	0.43	3.64	0.84	25	3	94	49
4.00	1.00	0.32	0.08	1.00	0.36	NA	0.92	9	0	110	52

[a]Screening Tool for Autism in Toddlers and Young Children; [b]Positive Predictive Value; [c]Negative Predictive Value; [d]Positive Likelihood Ratio; [e]Negative Likelihood Ratio; [f]True Positives; [g]False Positives; [h]False Negatives; [i]True Negatives; [j]Highlighted rows indicate potential cutoffs for the proposed two threshold method; [k]Bolded row indicates originally-published cutoff.

Table 3

7-item scoring method of the STAT[a]

7-item cutoff	PPV[b]	NPV[c]	Sensitivity	Specificity	Accuracy	LR+[d]	LR-[e]	TP[f]	FP[g]	FN[h]	TN[i]
0	0.70	NA	1.00	0.00	0.70	1.00	NA	119	52	0	0
1[j]	0.73	1.00	1.00	0.14	0.74	1.16	0.00	119	45	0	7
2	0.77	0.94	0.99	0.31	0.78	1.43	0.03	118	36	1	16
3[k]	**0.84**	**0.91**	**0.98**	**0.58**	**0.85**	**2.30**	**0.04**	**116**	**22**	**3**	**30**
4	0.88	0.80	0.92	0.71	0.86	3.20	0.11	110	15	9	37
5	0.93	0.65	0.80	0.87	0.82	5.93	0.23	95	7	24	45
6	0.93	0.47	0.56	0.90	0.66	5.77	0.49	66	5	53	47
7	0.96	0.35	0.19	0.98	0.43	10.05	0.82	23	1	96	51

[a]Screening Tool for Autism in Toddlers and Young Children; [b]Positive Predictive Value; [c]Negative Predictive Value; [d]Positive Likelihood Ratio; [e]Negative Likelihood Ratio; [f]True Positives; [g]False Positives; [h]False Negatives; [i]True Negatives; [j]Highlighted rows indicate potential cutoffs for the proposed two threshold method; [k]Bolded row indicates suggested cutoff by Khowaja et al. (2017).

Table 4

Logistic-regression method scoring of the STAT[a] for the validation sample

Logistic Probability	**PPV[b]**	**NPV[c]**	**Sensitivity**	**Specificity**	**Accuracy**	**LR+[d]**	**LR-[e]**	**TP[f]**	**FP[g]**	**FN[h]**	**TN[i]**
0%	0.70	NA	1.00	0.00	0.70	1.00	NA	40	17	0	0
5%	0.76	1.00	1.00	0.24	0.77	1.31	0.00	40	13	0	4
10%	0.78	1.00	1.00	0.35	0.81	1.55	0.00	40	11	0	6
15%	0.83	1.00	1.00	0.53	0.86	2.13	0.00	40	8	0	9
20%	0.83	1.00	1.00	0.53	0.86	2.13	0.00	40	8	0	9
25%	0.83	1.00	1.00	0.53	0.86	2.13	0.00	40	8	0	9
30%	0.83	0.90	0.98	0.53	0.84	2.07	0.05	39	8	1	9
35%	0.83	0.82	0.95	0.53	0.83	2.02	0.09	38	8	2	9
40%	0.84	0.83	0.95	0.59	0.84	2.31	0.09	38	7	2	10
45%	0.84	0.83	0.95	0.59	0.84	2.31	0.09	38	7	2	10
50%	0.84	0.83	0.95	0.59	0.84	2.31	0.09	38	7	2	10
55%	0.86	0.85	0.95	0.65	0.86	2.69	0.08	38	6	2	11
60%	0.86	0.79	0.93	0.65	0.84	2.62	0.12	37	6	3	11
65%	0.86	0.73	0.90	0.65	0.83	2.55	0.15	36	6	4	11
70%	0.88	0.75	0.90	0.71	0.84	3.06	0.14	36	5	4	12
75%	0.88	0.71	0.88	0.71	0.83	2.98	0.18	35	5	5	12
80%	0.90	0.72	0.88	0.77	0.84	3.72	0.16	35	4	5	13
81.6%	**0.90**	**0.72**	**0.88**	**0.77**	**0.84**	**3.72**	**0.16**	**35**	**4**	**5**	**13**
85%	0.91	0.61	0.78	0.82	0.79	4.39	0.27	31	3	9	14
90%	0.93	0.50	0.63	0.88	0.70	5.31	0.43	25	2	15	15
95%	0.96	0.46	0.53	0.94	0.65	8.92	0.50	21	1	19	16
100%	NA	0.30	0.00	1.00	0.30	NA	1.00	0	0	40	17

[a]Screening Tool for Autism in Toddlers and Young Children; [b]Positive Predictive Value; [c]Negative Predictive Value; [d]Positive Likelihood Ratio; [e]Negative Likelihood Ratio; [f]True Positives; [g]False Positives; [h]False Negatives; [i]True Negatives; [j]Highlighted rows indicate potential cutoffs for the proposed two threshold method; [k]Bolded row indicates single cutoff suggested by Youden's J.

Table 5

Univariate logistic and multivariate LASSO logistic regression coefficients for each STAT[a] item for the calibration sample

Item	Item Description	Univariate Logistic Coefficient	12-item Logistic Coefficient
1	Directing Attention: Balloon	2.84	1.96
2	Requesting: Food	2.37	1.35
3	Directing Attention: Bag of Toys	2.32	0.27
4	Directing Attention: Puppet	2.27	0.49
5	Requesting: Bubbles	1.89	0.57
6	Imitation: Hop Dog	1.63	0.31
7	Play: Turn-Taking	1.54	0.78
8	Imitation: Roll Car	1.50	0.41
9	Play: Doll	1.45	1.43
10	Directing Attention: Noisemaker	1.15	0.75
11	Imitation: Shake Rattle	1.09	0.52
12	Imitation: Drum Hands	0.88	-0.11

[a]Screening Tool for Autism in Toddlers and Young Children

Table 6

Comparison of different STAT[a] scoring methods

		Original[b]	7-Item[c]	Logistic Regression[d]
Calibration Data	**Threshold:**	**2**	**3**	**0.82**
	PPV[e]:	0.87 (0.78, 0.92)[h]	0.83 (0.74, 0.89)	0.97 (0.89, 0.99)
	NPV[f]:	0.75 (0.58, 0.87)	0.86 (0.67, 0.95)	0.69 (0.53, 0.79)
	Sensitivity:	0.90 (0.81, 0.95)	0.96 (0.89, 0.99)	0.81 (0.70, 0.87)
	Specificity:	0.69 (0.52, 0.81)	0.54 (0.38, 0.70)	0.94 (0.81, 0.98)
	Accuracy:	0.83 (0.75, 0.89)	0.83 (0.75, 0.89)	0.85 (0.76, 0.90)
	AUC[g]:	0.89 (0.82, 0.96)[f]	0.88 (0.81, 0.95)	0.92 (0.86, 0.98)
Validation Data	**Threshold:**	**2**	**3**	**0.79**
	PPV:	0.89 (0.77, 0.95)	0.87 (0.74, 0.94)	0.90 (0.76, 0.96)
	NPV:	1.00 (0.76, 1.00)	1.00 (0.74, 1.00)	0.72 (0.49, 0.88)
	Sensitivity:	1.00 (0.91, 1.00)	1.00 (0.91, 1.00)	0.88 (0.74, 0.95)
	Specificity:	0.71 (0.47, 0.87)	0.65 (0.41, 0.83)	0.76 (0.53, 0.90)
	Accuracy:	0.91 (0.81, 0.96)	0.89 (0.79, 0.95)	0.84 (0.73, 0.91)
	AUC:	0.88 (0.75, 1.00)	0.90 (0.79, 1.00)	0.88 (0.77, 0.98)
Entire Sample	**Threshold:**	**2**	**3**	**0.82**
	PPV:	0.87 (0.81, 0.92)	0.84 (0.77, 0.89)	0.94 (0.88, 0.97)
	NPV:	0.82 (0.68, 0.90)	0.91 (0.76, 0.97)	0.69 (0.55, 0.77)
	Sensitivity:	0.93 (0.87, 0.97)	0.97 (0.93, 0.99)	0.82 (0.73, 0.87)
	Specificity:	0.69 (0.56, 0.80)	0.58 (0.44, 0.70)	0.89 (0.77, 0.95)
	Accuracy:	0.86 (0.80, 0.90)	0.85 (0.79, 0.90)	0.84 (0.77, 0.88)
	AUC:	0.89 (0.83, 0.96)	0.89 (0.83, 0.95)	0.91 (0.85, 0.96)

[a]Screening Tool for Autism in Toddlers and Young Children; [b]Original STAT scoring method has a threshold of 2 to indicate ASD risk. [c]The 7-item scoring method (Khowaja et al., 2017) has a threshold of 3 to indicate ASD risk. [d]Threshold for the logistic regression was found by maximizing Youden's J statistic. [e]Positive Predictive Value; [f]Negative Predictive Value; [g]Area under the curve; [h]Wilson Score 95% confidence intervals (Agresti & Coull, 2002) were computed independently in R using the binomconf function from the Hmisc package. For this reason, rounding may be different than by-hand transformations of these otherwise equivalent statistics. [f]Confidence intervals for AUC values were calculated using the method described by DeLong, Delong, and Clark (1988) and using the ci.auc function from the pROC package.

Table 7

Comparison of two thresholds by different STAT[a] scoring methods

	Original		**7-Item**		**Logistic Regression**	
	Threshold	*% of sample*	*Threshold*	*% of sample*	*Threshold*	*% of sample*
Lower threshold (no ASD)	0 to 1.00	9%	0 to 1	4%	0 to 25	16%
Middle (unknown)	1.25 to 3.50	86%	2 to 6	82%	25 to 95	46%
Upper threshold (has ASD)	4.00	5%	7	14%	95 to 100	38%

[a]Screening Tool for Autism in Toddlers and Young Children

Figure 1

Receiver Operator Characteristic curves for each scoring method