
Karen Pierce, PhD; Vahid H. Gazestani, PhD; Elizabeth Bacon, PhD; Cynthia Carter Barnes, PhD; Debra Cha, PhD; Srinivasa Nalabolu, PhD; Linda Lopez, BS; Adrienne Moore, PhD; Sunny Pence-Stophaeros, MA; Eric Courchesne, PhD

IMPORTANCE Universal early screening for autism spectrum disorder (ASD) in primary care is becoming increasingly common and is believed to be a pivotal step toward early treatment. However, the diagnostic stability of ASD in large cohorts from the general population, particularly in those younger than 18 months, is unknown. Changes in the phenotypic expression of ASD across early development compared with toddlers with other delays are also unknown.

OBJECTIVES To examine the diagnostic stability of ASD in a large cohort of toddlers starting at 12 months of age and to compare this stability with that of toddlers with other disorders, such as developmental delay.

DESIGN, SETTING, AND PARTICIPANTS In this prospective cohort study performed from January 1, 2006, to December 31, 2018, a total of 2241 toddlers were referred from the general population through a universal screening program in primary care or community referral. Eligible toddlers received their first diagnostic evaluation between 12 and 36 months of age and had at least 1 subsequent evaluation.

EXPOSURES Diagnosis was denoted after each evaluation visit as ASD, ASD features, language delay, developmental delay, other developmental issue, typical sibling of an ASD proband, or typical development.

MAIN OUTCOMES AND MEASURES Diagnostic stability coefficients were calculated within 2-month age bands, and logistic regression models were used to explore the associations of sex, age, diagnosis at first visit, and interval between first and last diagnosis with stability. Toddlers with a non-ASD diagnosis at their first visit diagnosed with ASD at their last were designated as having late-identified ASD.

RESULTS Among the 1269 toddlers included in the study (918 [72.3%] male; median age at first evaluation, 17.6 months [interquartile range, 14.0-24.4 months]; median age at final evaluation, 36.2 months [interquartile range, 33.4-40.9 months]), the overall diagnostic stability for ASD was 0.84 (95% CI, 0.80-0.87), which was higher than any other diagnostic group. Only 7 toddlers (1.8%) initially considered to have ASD transitioned into a final diagnosis of typical development. Diagnostic stability of ASD within the youngest age band (12-13 months) was lowest at 0.50 (95% CI, 0.32-0.69) but increased to 0.79 by 14 months and 0.83 by 16 months (age bands of 12 vs 14 and 16 years; odds ratio, 4.25; 95% CI, 1.59-11.74). A total of 105 toddlers (23.8%) were not designated as having ASD at their first visit but were identified at a later visit.

CONCLUSIONS AND RELEVANCE The findings suggest that an ASD diagnosis becomes stable starting at 14 months of age and overall is more stable than other diagnostic categories, including language or developmental delay. After a toddler is identified as having ASD, there may be a low chance that he or she will test within typical levels at 3 years of age. This finding opens the opportunity to test the impact of very early-age treatment of ASD.
Autism spectrum disorder (ASD) is a common disorder of childhood, affecting 1 in 59 children. It is also becoming clear that ASD has its beginnings during prenatal life. Because many children with ASD have clinical signs within the first year, such as failure to respond to their name and reduced positive affect, there is a considerable demand for early detection, intervention, and services. Although several studies have shown that early signs of ASD can sometimes be detected using parent report screens as early as 12 to 18 months of age, the mean patient age at ASD detection is several years later, generally between 3 and 4 years of age. This late age of detection is a missed opportunity given the accelerated pace of brain development that occurs between birth and 3 to 4 years of age. Despite the appeal of the concept of early detection and treatment in ASD, there are many unknowns. Foundational questions regarding early-age diagnostic stability, age of clinical symptom onset, and overlap of early-age clinical symptoms between ASD and other disorders, such as language delay or global developmental delay, remain unanswered. A previous report by the US Preventive Services Task Force did not endorse early universal screening for ASD given the lack of clarity regarding the balance of benefits and harms of early screening and detection.

The months surrounding the first birthday are a remarkable time for a toddler’s development. At this age, toddlers learn to walk, speak their first word, and engage in a range of joint social attention behaviors, such as pointing and showing objects to others to share social attentional focus. The toddler stage is also the earliest age that ASD can be detected and treatment started yet the stability of an ASD diagnosis at this pivotal age is unknown.

A previous report stated that most studies examining the diagnostic stability of ASD before 3 years of age have involved slightly older, clinic-referred cohorts, usually at approximately 2 years of age. Stability coefficients within these studies have been high (mean, 88%, range, 63%-100%). Two studies examined stability at an even younger age (18 months) but examined this question from within multiplex families using the infant sibling design. One of these studies reported that 93% of siblings first diagnosed as having ASD at 18 months retained that diagnosis at a final age of 36 months, but only 69% of siblings first diagnosed as having ASD at 24 months did so (ie, 27 of 39 retained diagnosis). Although studies collectively suggest that an ASD diagnosis is moderately stable at young ages, there are several key questions remaining. First, it is unclear whether stability estimates from infant sibling designs would be found within a general population cohort. Second, none of the previous clinic-referred cohort studies included large groups of toddlers without ASD ascertained in the same manner as the toddlers with ASD. Such contrast groups are essential to understand how the ASD phenotype emerges from and overlaps with clinical expressions from other diagnostic groups, such as language and developmental delay, commonly found in clinical settings. Third, clinic-referred studies are small, usually containing 50 to 100 participants, and may generate less stable results. Moreover, children referred to a clinic because of already suspected ASD may generate artificially high stability rates relative to a community-ascertained sample. Fourth, despite the potential of the infant sibling design to study ASD from birth, stability estimates have only been reported starting at 18 months of age, leaving questions surrounding younger ages unanswered.

Interleaved with these gaps in knowledge is the recent finding from infant sibling studies that 50% to 80% of toddlers eventually diagnosed as having ASD at 3 years of age were not identified as having ASD by expert clinicians at 18 months of age. In short, despite extensive clinical testing that included the gold standard tool the Autism Diagnostic Observation Schedule (ADOS), these diagnoses were missed. A newer study, however, suggests that such so-called late-onset cases may be attributable to weaknesses inherent in standardized diagnostic tools at early ages, rather than a lack of observable ASD symptoms per se. Determining the degree to which such late-onset cases may be present in a general population cohort is essential, because if rates are as high as in infant sibling cohorts, it would strongly underscore the American Academy of Pediatrics recommendation for repeat screening at multiple ages. It would also add further urgency to the search for early behavioral or biological tests for ASD to more readily detect ASD during the earliest ages when detection is the most challenging. In this study, we sought to examine the diagnostic stability of ASD in a large cohort of toddlers starting at 12 months of age and to compare this stability with that of toddlers with other disorders, such as developmental delay.

### Methods

#### Participants

A total of 2241 toddlers 12 to 36 months of age were referred for a diagnostic evaluation to an autism expert evaluation center created at University of California, San Diego. Referrals were given through our early detection program, Get SET Early, which systematically screens for ASD and other disorders in the general population at 12-, 18-, and 24-month well-child checkups or through the general community. Typically developing (TD) toddlers were also recruited from the same pediatric offices participating in the Get SET Early program (eMethods in the Supplement). A total of 1269 of the 2241 toddlers were longitudinally evaluated 2 or more times and were...
the focus of this study. In this sample, approximately 75% came from the Get SET Early program and approximately 25% from community referral. Additional eligibility requirements included an interval of 6 months or longer between the first and last evaluations. Figure 1 and eFigure 1 and eFigure 2 in the Supplement show the cohort characteristics. This study was overseen by the institutional review board at the University of California, San Diego, and written informed consent was obtained from caregivers before study enrollment. At the data analysis phase of the study, the patient names were removed from our spreadsheets to protect their identity.

Diagnostic and Psychometric Testing
Highly experienced, licensed psychologists with PhD degrees performed diagnostic and psychometric tests, including the ADOS-2 (module T, 1, or 2 as appropriate), Mullen Scales of Early Learning, and Vineland Adaptive Behavior Scales. Toddlers who received their first diagnostic evaluation at younger than 36 months were diagnostically tested approximately every 12 months until 3 years of age. After each visit, psychologists filled out a diagnostic judgment form and entered it into a database. Psychologists were not masked to previous diagnoses during longitudinal test visits. A toddler was designated as having 1 of the following: ASD, ASD features, developmental delay, language delay (LD), other issue, TD, or typical sibling of an ASD proband. Parents were given feedback regarding their child’s performance after completion of testing and referred for treatment as appropriate. A description of psychologist training, diagnostic criteria used, data quality control process, and estimated Mullen T scores generated for 9% of toddlers who scored below a standard T score of 20 are given in the eMethods in the Supplement. The Table and eFigure 2 in the Supplement give information regarding the Diagnostic and Statistical Manual of Mental Disorders (DSM) version used.

Statistics and Data Visualization
Diagnostic Stability
Stability coefficients were first calculated within 2-month age bands by determining the proportion of toddlers with a particular diagnosis at their first diagnostic visit who retained that same diagnosis at their last visit. Diagnostic transition tables were created for overall and 2-month-interval age-binned data. Diagnostic stability was modeled using logistic regression, with sex, age at first diagnosis,
interval between first and last diagnosis, and diagnostic group at first visit as variables and results reported as odds ratios (ORs) (eTable 1 in the Supplement). To examine the association of age at first diagnosis with stability coefficients while optimizing statistical power, we binned age to 4 roughly equally populated groups: younger than 14 months, 14 to 17.99 months, 18 to 23.99 months, 24 months or older. No significant association between sex or interval with stability coefficients was found. Follow-up models with the 4 age bins as the only covariates were used to examine the association of age at first diagnosis with stability coefficients within each diagnostic group. A B-spline method

<table>
<thead>
<tr>
<th>Characteristic at Last Diagnostic Visit</th>
<th>ASD (n = 441)</th>
<th>ASD Features (n = 78)</th>
<th>DD (n = 89)</th>
<th>LD (n = 80)</th>
<th>Other (n = 91)</th>
<th>Typical Sibling (n = 51)</th>
<th>TD (n = 439)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male</td>
<td>361 (81.9)</td>
<td>68 (87.2)</td>
<td>66 (74.2)</td>
<td>58 (72.5)</td>
<td>61 (67.0)</td>
<td>26 (51.0)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>80 (18.1)</td>
<td>10 (12.8)</td>
<td>23 (25.8)</td>
<td>22 (27.5)</td>
<td>30 (33.0)</td>
<td>25 (49.0)</td>
</tr>
<tr>
<td>Age, mean (SD), mo</td>
<td>42.84 (20.28)</td>
<td>40.77 (17.61)</td>
<td>35.91 (10.15)</td>
<td>35.44 (11.42)</td>
<td>42.92 (13.08)</td>
<td>38.44 (13.76)</td>
<td>37.10 (9.84)</td>
</tr>
<tr>
<td>Final DSM diagnosis&lt;sup&gt;b&lt;/sup&gt;</td>
<td>DSM-IV</td>
<td>135</td>
<td>19</td>
<td>24</td>
<td>23</td>
<td>34</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>DSM-5</td>
<td>306</td>
<td>59</td>
<td>65</td>
<td>57</td>
<td>57</td>
<td>25</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Hispanic/Latino</td>
<td>128 (29.0)</td>
<td>17 (21.8)</td>
<td>36 (40.4)</td>
<td>38 (47.5)</td>
<td>20 (22.0)</td>
<td>15 (29.4)</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic/Latino</td>
<td>263 (59.6)</td>
<td>53 (67.9)</td>
<td>47 (52.8)</td>
<td>39 (48.8)</td>
<td>64 (70.3)</td>
<td>31 (60.8)</td>
</tr>
<tr>
<td></td>
<td>Not reported</td>
<td>50 (11.3)</td>
<td>8 (10.3)</td>
<td>6 (6.7)</td>
<td>3 (3.8)</td>
<td>7 (7.7)</td>
<td>5 (9.8)</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>237 (53.7)</td>
<td>51 (65.4)</td>
<td>48 (53.9)</td>
<td>47 (58.8)</td>
<td>60 (65.9)</td>
<td>31 (60.8)</td>
</tr>
<tr>
<td></td>
<td>Black/African American</td>
<td>9 (2.0)</td>
<td>1 (1.3)</td>
<td>2 (2.2)</td>
<td>2 (2.5)</td>
<td>5 (5.5)</td>
<td>2 (3.9)</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>48 (10.9)</td>
<td>7 (9.0)</td>
<td>7 (7.9)</td>
<td>1 (1.3)</td>
<td>2 (2.2)</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td></td>
<td>Pacific Islander</td>
<td>4 (0.90)</td>
<td>3 (3.8)</td>
<td>4 (4.5)</td>
<td>2 (2.5)</td>
<td>2 (2.2)</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td></td>
<td>Native American/Alaska</td>
<td>2 (0.50)</td>
<td>0</td>
<td>1 (1.1)</td>
<td>3 (3.8)</td>
<td>1 (1.1)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mixed race</td>
<td>57 (12.9)</td>
<td>7 (9.0)</td>
<td>9 (10.1)</td>
<td>1 (1.3)</td>
<td>11 (12.1)</td>
<td>7 (13.7)</td>
</tr>
<tr>
<td></td>
<td>Not reported</td>
<td>84 (19.0)</td>
<td>9 (11.5)</td>
<td>18 (20.2)</td>
<td>24 (30.0)</td>
<td>10 (11.0)</td>
<td>9 (17.6)</td>
</tr>
<tr>
<td>Mullen T score, mean (SD)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Visual reception</td>
<td>38.0 (14.9)</td>
<td>51.4 (13.5)</td>
<td>35.3 (13.4)</td>
<td>49.6 (11.7)</td>
<td>54.1 (13.4)</td>
<td>61.0 (9.7)</td>
</tr>
<tr>
<td></td>
<td>Fine motor</td>
<td>34.0 (12.6)</td>
<td>43.8 (11.6)</td>
<td>31.2 (11.0)</td>
<td>46.4 (10.4)</td>
<td>45.3 (13.2)</td>
<td>53.1 (10.3)</td>
</tr>
<tr>
<td></td>
<td>Receptive language</td>
<td>32.1 (15.0)</td>
<td>46.1 (12.0)</td>
<td>33.4 (12.0)</td>
<td>40.8 (10.6)</td>
<td>48.7 (11.0)</td>
<td>52.7 (10.3)</td>
</tr>
<tr>
<td></td>
<td>Expressive language</td>
<td>30.6 (16.9)</td>
<td>48.6 (12.0)</td>
<td>30.8 (13.9)</td>
<td>33.9 (9.4)</td>
<td>49.2 (12.4)</td>
<td>54.2 (8.8)</td>
</tr>
<tr>
<td></td>
<td>ELC</td>
<td>71.5 (22.1)</td>
<td>94.7 (22.6)</td>
<td>68.6 (17.7)</td>
<td>86.0 (15.1)</td>
<td>99.4 (18.7)</td>
<td>110.5 (14.7)</td>
</tr>
<tr>
<td>Vineland standard score, mean (SD)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Communication</td>
<td>72.1 (25.0)</td>
<td>96.3 (21.2)</td>
<td>78.3 (21.2)</td>
<td>84.6 (19.5)</td>
<td>98.2 (17.2)</td>
<td>101.2 (18.0)</td>
</tr>
<tr>
<td></td>
<td>Daily living</td>
<td>75.2 (22.5)</td>
<td>95.1 (18.2)</td>
<td>83.8 (19.0)</td>
<td>94.9 (18.6)</td>
<td>96.3 (15.7)</td>
<td>98.7 (16.8)</td>
</tr>
<tr>
<td></td>
<td>Socialization</td>
<td>72.6 (21.5)</td>
<td>95.2 (18.6)</td>
<td>85.9 (18.5)</td>
<td>92.3 (18.4)</td>
<td>97.0 (15.4)</td>
<td>103 (16.9)</td>
</tr>
<tr>
<td></td>
<td>Motor skills</td>
<td>76.2 (27.0)</td>
<td>92.5 (20.4)</td>
<td>80.4 (20.3)</td>
<td>91.9 (23.9)</td>
<td>91.3 (17.9)</td>
<td>95.8 (15.9)</td>
</tr>
<tr>
<td></td>
<td>Adaptive behavior composite</td>
<td>73.3 (21.8)</td>
<td>95.5 (16.2)</td>
<td>80.5 (13.1)</td>
<td>91.23 (11.4)</td>
<td>96.0 (13.4)</td>
<td>100.7 (10.5)</td>
</tr>
<tr>
<td>ADOS (module T, 1, or 2) score, mean (SD)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>ADOS SA/CoSo score</td>
<td>12.9 (4.1)</td>
<td>4.4 (2.7)</td>
<td>3.8 (3.3)</td>
<td>2.4 (2.1)</td>
<td>3.1 (2.4)</td>
<td>2.0 (1.8)</td>
</tr>
<tr>
<td></td>
<td>ADOS RRB score</td>
<td>4.6 (1.9)</td>
<td>2.6 (1.5)</td>
<td>1.4 (1.5)</td>
<td>0.6 (0.9)</td>
<td>0.7 (0.8)</td>
<td>0.3 (0.7)</td>
</tr>
<tr>
<td></td>
<td>ADOS total score</td>
<td>17.6 (4.8)</td>
<td>7.0 (3.1)</td>
<td>5.2 (3.1)</td>
<td>3.0 (2.3)</td>
<td>3.8 (2.6)</td>
<td>2.4 (1.8)</td>
</tr>
</tbody>
</table>

Abbreviations: ADOS, Autism Diagnostic Observation Schedule; ASD, autism spectrum disorder; CoSo, Communication Social Score; DD, developmental delay; DSM, Diagnostic and Statistical Manual of Mental Disorders; ELC, early learning composite; LD, language delay; RRB, restricted and repetitive behavior; SA, social affect; TD, typical development.

<sup>a</sup> Data are presented as number (percentage) of toddlers unless otherwise indicated.

<sup>b</sup> Version of the DSM used at the final diagnostic evaluation (eMethods and eFigure 2 in the Supplement).

<sup>c</sup> A total of 9% percent of the overall sample had a chronologic or mental age that exceeded the validated age range for use with the Mullen scales at their last diagnostic evaluation visit and received a Wechsler Preschool and Primary Scale of Intelligence instead.

<sup>d</sup> Administered ADOS module depended on the age and language ability of the toddler at the time of testing. For these individuals, their most recent available Mullen scores were used.

<sup>e</sup> A total of 9% percent of the overall sample had a chronological or mental age that exceeded the validated age range for use with the Mullen scales at their last diagnostic evaluation visit and received a Wechsler Preschool and Primary Scale of Intelligence instead.
Diagnostic Stability of the Early Autism Spectrum Disorder Phenotype in the General Population Starting at 12 Months

Diagnostic Stability

Overall stability was 0.84 (95% CI, 0.80-0.87) for an ASD diagnosis and 0.79 (95% CI, 0.74-0.83) for a TD diagnosis (Figure 2A). Results from the overall logistic regression model showed a significant association of age and diagnosis at first visit with stability (eTable 1 in the Supplement). No significant differences were found in stability based on sex (OR, 0.76; 95% CI, 0.56-1.04) or interval between first and last diagnostic evaluations (OR, 0.99; 95% CI, 0.98-1.00). Logistic regression analyses showed a nonsignificant difference in stability coefficients between ASD and TD (OR, 0.86; 95% CI, 0.57-1.29). In contrast, significant differences were found between ASD and the remaining diagnostic groups (OR, 0.11 [95% CI, 0.03-0.32] vs ASD features; OR, 0.15 [95% CI, 0.09-0.25] vs DD; OR, 0.04 [95% CI, 0.03-0.06] vs LD; and OR, 0.16 [95% CI, 0.09-0.28] vs other) (eTable 1 in the Supplement). For ASD, stability was weakest at 12 to 13 months of age (stability coefficient, 0.50; 95% CI, 0.32-0.69). Stability of an ASD diagnosis increased to 0.79 by 14 months of age and 0.83 by 16 months of age (age bands of 12 vs 14 and 16 months; OR, 4.25; 95% CI, 1.59-11.74) (Figure 3 and eFigure 4, eFigure 5, eTable 2, and eTable 3 in the Supplement). When toddlers with ASD features were considered to have ASD, the stability coefficients increased to 0.70 (95% CI, 0.52-0.85) at 12 months of age, 0.85 (95% CI, 0.71-0.94) at 14 months of age, and 0.94 (95% CI, 0.81-0.99) at 16 months of age. Given the transient nature of many early delays,25 overall stability was low for the remaining delay groups (Figure 2 and Figure 3 and eFigure 4 and eTable 4 in the Supplement). Exclusion of 73 toddlers (34 with ASD, 1 with ASD features, 24 with DD, 7 with other disorders, 1 with a typical sibling, and 6 with TD) whose nonverbal mental age based on the visual reception component of the Mullen scale was younger than 12 months (mean nonverbal mental age, 9.6 months) did not improve the stability coefficient of ASD at 12 to 13 months (eTable 5 and eFigure 6 in the Supplement).

Transition Patterns

Diagnostic heat maps (Figure 2B) illustrate diagnostic transition patterns for toddlers who were evaluated 2, 3, or 4 or more times. The transition from an initial diagnosis of LD or developmental delay to ASD was the most common transition type. Transitioning from an initial designation of ASD to a final diagnosis of TD was rare and occurred in only 1.8% of overall cases (ie, 7 toddlers of 400 toddlers initially designated as ASD). However, 5 of these 7 toddlers with false-positive results were initially evaluated at the youngest ages (12-13 months of age) (eFigure 4 in the Supplement).

Results

Participant Characteristics

Among the 1269 toddlers, 918 (72.3%) were male, median age at first evaluation was 17.6 months (interquartile range, 14.0-24.4 months), mean number of diagnostic visits was 2.7 (interquartile range, 2-3), and median age at final evaluation was 36.2 months (interquartile range, 33.4-40.9 months). The Table gives the demographic information and clinical test scores for each diagnostic group.

Diagnostic Stability

Overall stability was 0.84 (95% CI, 0.80-0.87) for an ASD diagnosis and 0.79 (95% CI, 0.74-0.83) for a TD diagnosis (Figure 2A). Results from the overall logistic regression model showed a significant association of age and diagnosis at first visit with stability (eTable 1 in the Supplement). No significant differences were found in stability based on sex (OR, 0.76; 95% CI, 0.56-1.04) or interval between first and last diagnostic evaluations (OR, 0.99; 95% CI, 0.98-1.00). Logistic regression analyses showed a nonsignificant difference in stability coefficients between ASD and TD (OR, 0.86; 95% CI, 0.57-1.29). In contrast, significant differences were found between ASD and the remaining diagnostic groups (OR, 0.11 [95% CI, 0.03-0.32] vs ASD features; OR, 0.15 [95% CI, 0.09-0.25] vs DD; OR, 0.04 [95% CI, 0.03-0.06] vs LD; and OR, 0.16 [95% CI, 0.09-0.28] vs other) (eTable 1 in the Supplement). For ASD, stability was weakest at 12 to 13 months of age (stability coefficient, 0.50; 95% CI, 0.32-0.69). Stability of an ASD diagnosis increased to 0.79 by 14 months of age and 0.83 by 16 months of age (age bands of 12 vs 14 and 16 months; OR, 4.25; 95% CI, 1.59-11.74) (Figure 3 and eFigure 4, eFigure 5, eTable 2, and eTable 3 in the Supplement). When toddlers with ASD features were considered to have ASD, the stability coefficients increased to 0.70 (95% CI, 0.52-0.85) at 12 months of age, 0.85 (95% CI, 0.71-0.94) at 14 months of age, and 0.94 (95% CI, 0.81-0.99) at 16 months of age. Given the transient nature of many early delays,25 overall stability was low for the remaining delay groups (Figure 2 and Figure 3 and eFigure 4 and eTable 4 in the Supplement). Exclusion of 73 toddlers (34 with ASD, 1 with ASD features, 24 with DD, 7 with other disorders, 1 with a typical sibling, and 6 with TD) whose nonverbal mental age based on the visual reception component of the Mullen scale was younger than 12 months (mean nonverbal mental age, 9.6 months) did not improve the stability coefficient of ASD at 12 to 13 months (eTable 5 and eFigure 6 in the Supplement).
Discussion

Children with ASD are detected and treated nationally at approximately 4 years of age.\(^1\) However, we found that within the context of an early detection program,\(^6\) children can be reliably diagnosed with ASD several years earlier, as young as 14 months. The implications of this finding extend beyond information that relates to diagnostic stability and may open new opportunities to consider if and how treatments started at this early age are associated with long-term outcomes of affected children.

An initial ASD diagnosis was more stable than any other diagnosis, including TD. In our cohort, 84% of toddlers initially
A. Plots show diagnostic stability per group across 2-month age intervals based on the age at first diagnostic evaluation. Age intervals with missing data points reflect an absence of toddlers who received their first diagnostic evaluation at that age. B-spline logistic regression line is shown; bands represent 95% CIs for the fit line. Overall stability was highest in toddlers initially designated as having autism spectrum disorder (ASD) or typical development as illustrated by the relatively tight CI bands, and the largely consistent stability coefficients within each age band. B. Diagnostic stability coefficients in the 4 age bins used in the logistic regression model across diagnostic groups. The lines represent 95% CIs. Coefficients were estimated based on within group logistic regression models. eFigure 5 and eFigure 6 in the Supplement give complementary analyses.

diagnosed with ASD at their first visit retained this diagnosis at 3 to 4 years of age. Most toddlers within the remaining 16% did not lose their delays entirely but instead presented with milder delays at their final diagnostic visit. The most common transition was ASD to ASD features, a diagnostic category used for toddlers with signs of ASD but not enough to meet DSM criteria. The least common transition was ASD to TD (ie, only 1.8% of toddlers initially designated as having ASD transitioned to TD). Because all toddlers were immediately referred for treatment once any delay was detected, improvements in symptom severity...
Figure 4. Comparison of Clinical Features in Toddlers With Autism Spectrum Disorder (ASD) Stratified by Identification Age

Violin plots show differences in Autism Diagnostic Observation Schedule (ADOS) total scores (A), Mullen Expressive Language T scores (B), Vineland Adaptive Behavior Composite scores (C), and Mullen Receptive Language T scores (D) at the first diagnostic evaluation between toddlers with ASD identified at 12 to 18 months of age (early-age persistent ASD diagnosis), toddlers with ASD identified after 18 months (middle-age persistent ASD diagnosis), or toddlers not identified as having ASD at their first diagnostic visit (late-identified ASD). Black dots represent the mean. The width of the shape represents patient density, and the length illustrates the range of the scores. Data from 270 toddlers with typical development (TD) identified at their first diagnostic visit and retaining that diagnosis at their last visit are shown for comparison. Note that scores from the late-identified group were significantly different from toddlers with TD across all clinical domains, suggesting that symptoms were already present at the first diagnostic visit in a large fraction of late-identified ASD cases. Also note that 39 toddlers in the late-identified group (37%) did fall within the range of concern on the Autism Diagnostic Observation Schedule toddler module (cutoff score for concern using the few to no words algorithm = 10), however, were designated as non-ASD based on practitioner judgment, underscoring the challenges in differential diagnoses particularly at the youngest ages. Effect sizes are reported as Cohen d (95% CI). eFigure 3 in the Supplement gives an expanded figure that includes all diagnostic groups.
Diagnostic Stability of the Early Autism Spectrum Disorder Phenotype in the General Population Starting at 12 Months

S. Pierce, J. Gazestani, N. Bacon, E. Courchesne.

ARTICLE INFORMATION
Accepted for Publication: January 15, 2019.
Published Online: April 29, 2019.
Open Access: This is an open access article distributed under the terms of the CC-BY License. © 2019 Pierce K et al. JAMA Pediatrics.

Author Contributions: Dr Pierce had full access to all the data and takes responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: Pierce, Bacon, Courchesne. Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Pierce, Gazestani, Bacon, Courchesne. Critical revision of the manuscript for important intellectual content: Pierce, Gazestani, Bacon, Carter Barnes, Cha, Lopez, Pence-Stophaeross, Courchesne. Statistical analysis: Gazestani. Obtained funding: Pierce, Courchesne. Administrative, technical, or material support: Bacon, Carter Barnes, Cha, Nalabolu, Lopez, Moore, Pence-Stophaerus, Courchesne. Supervision: Pierce, Courchesne. Conflict of Interest Disclosures: No disclosures were reported.

Funding/Support: This work was supported by grant NINR01016517 from the Novo Nordisk Foundation through Center for Biosustainability at the Technical University of Denmark (Dr Gazestani), grants R01-MH0104446 and R01-MH08134 from the National Institute of Mental Health (Dr Pierce), a grant from the National Foundation for Autism Research (Dr Pierce), grant PS0-MH081755 from the National Institute of Mental Health (Dr Courchesne), and a Progenity grant (Dr Courchesne).

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank the parents and children in San Diego who participated in our research. Pediatricians and family practice physicians spanning a range of medical groups, including University of California, San Diego, Sharp Rees-Stealy, Scripps, Rady Children’s Primary Care Medical Group, Chula Vista Pediatrics, Graybill.

E9
Medical Group, Grossmont Pediatrics, Linda Vista Health Care Center, Mills Pediatrics, North County Health Services, San Diego Family Care, and Sea Breeze Pediatrics provided support for Get SET Early.

REFERENCES


