BACKGROUND: Universal screening is recommended to reduce the age of diagnosis for autism spectrum disorder (ASD). However, there are insufficient data on children who screen negative and no study of outcomes from truly universal screening. With this study, we filled these gaps by examining the accuracy of universal screening with systematic follow-up through 4 to 8 years.

METHODS: Universal, primary care-based screening was conducted using the Modified Checklist for Autism in Toddlers with Follow-Up (M-CHAT/F) and supported by electronic administration and integration into electronic health records. All children with a well-child visit (1) between 16 and 26 months, (2) at a Children’s Hospital of Philadelphia site after universal electronic screening was initiated, and (3) between January 2011 and July 2015 were included (N = 25,999).

RESULTS: Nearly universal screening was achieved (91%), and ASD prevalence was 2.2%. Overall, the M-CHAT/F’s sensitivity was 38.8%, and its positive predictive value (PPV) was 14.6%. Sensitivity was higher in older toddlers and with repeated screenings, whereas PPV was lower in girls. Finally, the M-CHAT/F’s specificity and PPV were lower in children of color and those from lower-income households.

CONCLUSIONS: Universal screening in primary care is possible when supported by electronic administration. In this “real-world” cohort that was systematically followed, the M-CHAT/F was less accurate in detecting ASD than in previous studies. Disparities in screening rates and accuracy were evident in traditionally underrepresented groups. Future research should focus on the development of new methods that detect a greater proportion of children with ASD and reduce disparities in the screening process.
Although autism spectrum disorder (ASD) manifests in the first few years of life, the average age of diagnosis remains older than 4 years of age and is even later for children of color and those from rural and lower-income backgrounds. Improving early diagnosis is critical because it affords children access to earlier intervention, which has been shown to significantly improve outcomes. The American Academy of Pediatrics (AAP) recommends universal screening for ASD at 18 and 24 months to facilitate earlier identification. However, the US Preventive Services Task Force concluded that there is insufficient evidence to recommend universal screening, in part because of limited data on outcomes for children who screen negative and from diverse samples. Coupled with the lack of data on truly universal screening (ie, all children are screened rather than a selected subset), there are critical gaps in knowledge about the short- and long-term benefits of universal ASD screening.

The most widely used and studied screening tool is the Modified Checklist for Autism in Toddlers with Follow-Up (M-CHAT/F), a 2-stage tool that includes a 23-item parent questionnaire and a follow-up interview designed to reduce false-positives. The Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT), however, both have important limitations that restrict generalizability. A study conducted in Malaysian maternal-child health clinics yielded only a 0.2% screen-positive rate in a sample with a 0.3% ASD prevalence rate. A large study of population screening in Norway yielded a higher screen-positive rate (7.4%), but the ASD prevalence was still low (0.3%), suggesting that the study’s methods failed to detect many children with ASD. Despite large sample sizes, neither study achieved universal screening (ie, 65% and ~28%, respectively). Although there are limitations, these first estimates in samples that were systematically followed-up suggested low sensitivity and PPV (34%–36% and 2%–47%, respectively).

It is critical to assess screening accuracy within the intended population (eg, all children in a primary care population) to reduce bias and facilitate the generalization of findings to similar cohorts. Sensitivity and specificity are often presumed to be fixed and inherent to the measure, but in reality, these measures are strongly influenced by the sample and/or cohort in which they are estimated. Samples that are not ascertained through universal screening are likely to overrepresent children with parent and/or professional concern and/or underrepresent children of color and those from lower-income households. In addition, PPV and NPV are directly linked to the sample’s prevalence rate, such that PPV increases and NPV decreases with prevalence. Thus, without universal screening research and representative cohorts, we risk drawing incorrect conclusions about the accuracy of screening tools in real-world applications.

Our goal with this study was to examine the real-world accuracy of universal screening for ASD by using an epidemiological design and long-term follow-up through 4 to 8 years of age. Screening was conducted at the Children’s Hospital of Philadelphia (CHOP), a large pediatric network of primary and specialty care services with an integrated electronic health record (EHR). Our secondary goals were to examine the accuracy of repeated screenings and the effect of child and/or family characteristics on screening rates and accuracy.

**METHODS**

**Setting**

The CHOP network includes 31 pediatric primary care sites that serve a diverse patient population in Pennsylvania and New Jersey. Of these, 4 sites in urban Philadelphia serve a racially and economically diverse patient population (88% children of color; 74% public insurance/Medicaid), whereas suburban sites are less diverse (35% children of color; 24% public insurance/Medicaid). The M-CHAT/F is administered electronically at well-child visits between 16 and 26 months in accordance with Pennsylvania’s Early and Periodic Screening, Diagnostic, and Treatment program and is available in English and Spanish. Screening is automatically triggered at all well-child visits in this age range, regardless of previous screening results, to ensure that children are screened twice as recommended by the AAP. Given that not all children present for 18- and
24-month well-child visits, the M-CHAT/F can also be assessed manually at sick visits. Once the M-CHAT/F is completed, questionnaire results autopolulate into the child’s visit note along with a link for providers to complete the follow-up interview when children screen positive.

CHOP provides ASD diagnostic services through a multidisciplinary program that includes developmental pediatrics, psychology and psychiatry, and neurology clinics.

This study was approved by the CHOP Institutional Review Board with a waiver of consent.

**Study Design and Patient Population**

The cohort included all children who presented for a well-child visit (1) between 16 and 26 months of age, (2) at a CHOP site where universal screening had been initiated, and (3) between January 2011 and July 2015 to allow for longitudinal follow-up through ≥4 years. If the child’s first screening occurred before initiation of universal screening and the second occurred after universal screening initiation, the first screening was included to accurately represent first and second screens. Screenings at sick visits were also included (0.3% of M-CHAT/F administrations). When multiple M-CHAT/Fs were completed, the first administration was used unless otherwise noted.

Preliminary analyses of screening rates and results included the entire cohort. Children without a primary care visit at ≥4 years of age were excluded from diagnostic outcome analyses because of insufficient length of follow-up. Four years was chosen given recent estimates of the median age of diagnosis for ASD. Those without a documented language of English or Spanish (n = 82; 0.004%) were also excluded because the M-CHAT/F was not available in other languages.

The subgroup whose visits closely followed AAP guidelines was also examined; these children were screened during primary care visits at 18 months (±2) and 24 months (±2) with ≥3 months between screenings.

**Measures and EHR Methods**

**Demographic Data**

Demographics, gestational age, and insurance payer for the screening visit were extracted from the EHR. Federal information processing standard codes were linked to census tract–level data to generate estimates of median income, and a median split was performed. Language was coded as English only (ie, only English was documented) or other language (ie, documentation of any non-English language).

**Diagnostic Data**

Diagnoses were extracted from visit diagnoses and problem lists (a comprehensive list of active and relevant past diagnoses). All available data were used to determine diagnostic outcome. As a result, the length of follow-up period varied across children (although all children included in these analyses had follow-up through at least 4 years).

Children were considered to have ASD if an ASD diagnosis appeared in the EHR more than once or was provided by a specialist because these criteria have been associated with the greatest accuracy in other large health care systems. For example, when comparing EHR diagnostic codes and manual chart review, Coleman et al found that ≥2 ASD diagnoses yielded a PPV of 87%; specialist diagnoses were also associated with higher odds of a confirmed diagnosis.

Children were considered to have a non-ASD disorder and/or delay if they did not meet ASD classification criteria described above but had ≥1 code from 1 of the following categories: attention-deficit/hyperactivity disorder (ADHD) and related behaviors, anxiety disorder and related behaviors, disruptive behavior disorder and related behaviors, developmental delay, language disorder and/or delay, motor disorder and/or delay, sensory processing difficulty, and social delay without ASD. See Supplemental Table 5 for specific diagnostic codes used.

**Statistical Analyses**

Screening rates and results were summarized with percentages; χ² analyses and odds ratios (ORs) were calculated for subgroup comparisons. Sensitivity, specificity, PPV, and NPV were calculated from 2 × 2 contingency tables. Effect sizes for the proportion comparisons (Cohen’s h) and measures of statistical significance (2-sample tests of proportion) were used to estimate differences in M-CHAT/F accuracy across subgroups. Emphasis was placed on interpreting effect sizes rather than P values alone given the likelihood of statistical significance for trivial differences in this large cohort. Thus, only results with statistically significant comparisons and effect sizes ≥0.20 are reported as meaningful. Finally, Kaplan-Meier survival curves estimated the cumulative probability of ASD diagnosis across time since screening. The log-rank test was used to compare survival curves between children who screened negative and positive to detect differences in mean time to diagnosis by M-CHAT/F outcome.

**RESULTS**

**Screening Rates**

A total of 25,999 children had 42,973 eligible visits during the study period (see Table 1 and Fig 1). A total of 23,634 children (90.9%) were screened and 50.4% were screened more than once. White children were screened more often than other racial
groups (see Table 2). Children with English-only exposure, higher incomes, private insurance, and from suburban primary care sites were also screened more often, and premature children were screened less often. Only 47.8% were screened at 18 and 24 months largely because of failure to attend both well-child visits. Children who received 2 screenings according to the AAP schedule were more likely to be white and non-Hispanic, from a suburban site, and have English-only exposure, higher incomes, and private insurance (see Table 2).

**Screen-Positive Rates**

When considering the first M-CHAT/F, 9.5% \((n = 2256)\) screened positive on the 23-item questionnaire, a rate comparable to other large-scale US-based screening studies. Of those that screened positive, 88.7% \((n = 2002)\) required the follow-up interview (ie, scores of 3–7) and 41.2% \((n = 825)\) were administered it. Almost all (872; 94.8%) no longer screened positive after the follow-up interview. Of note, these numbers reflect all screened children (including those without follow-up data), so they differ somewhat from Fig 2, which only includes children included in accuracy analyses (ie, screened with follow-up data).

For accuracy analyses, children who screened negative after the questionnaire or follow-up interview were considered screen negatives. Those who continued to screen positive after the interview were considered screen positives. Children who screened positive on the questionnaire but did not receive the follow-up interview were also considered screen positives. Excluding this group would introduce substantial bias to the cohort because there were demographic and clinical differences between children who did and did not receive the follow-up interview. For note, these numbers reflect all screened children (including those without follow-up data), so they differ somewhat from Fig 2, which only includes children included in accuracy analyses (ie, screened with follow-up data).

For accuracy analyses, children who screened negative after the questionnaire or follow-up interview were considered screen negatives. Those who continued to screen positive after the interview were considered screen positives. Children who screened positive on the questionnaire but did not receive the follow-up interview were also considered screen positives. Excluding this group would introduce substantial bias to the cohort because there were demographic and clinical differences between children who did and did not receive the follow-up interview (see below). Furthermore, the positive questionnaire results were available to providers to base clinical action on (even in the absence of the follow-up interview), and many were referred after an incomplete M-CHAT/F screening (K.W., W.G., A.B., et al, unpublished data).

This approach resulted in a final screen-positive rate of 6.2%. Children of color, those from lower-income households, with public insurance/
Medicaid and non-English exposure, and those seen in urban practices screened positive more frequently, as did boys and premature children (see Table 2).

**Screen-Positive Rates by Age**

Older toddlers (21–26 months) screened positive more often (8.9%) than younger toddlers (16–20 months; 5.5%) on the first screening (see Table 3). However, screen-positive rates were somewhat higher at 18 months (4.2%) than at 24 months (3.5%) in the subgroup screened twice; 6.4% screened positive on 1 or both screenings.

**Outcome Diagnostic Data**

Most screened children (n = 20,437; 86.5%) continued to receive CHOP primary care at ≥4 years and were included in M-CHAT/F accuracy analyses because outcome diagnostic data were available. ASD prevalence was 2.2%, which is comparable to recent prevalence estimates in nearby New Jersey. A total 62.8% received an ASD diagnosis by a specialist (of these, 94.7% also had an ASD diagnosis documented by a primary care provider); the remaining 37.2% only had a diagnosis made or documented by a primary care provider.

**Accuracy Analyses**

The M-CHAT/F’s sensitivity to detect ASD was 38.8%, and its specificity was 94.9%. PPV was 14.6% and NPV was 98.6% (see Table 4, Fig 2, and Fig 3). The M-CHAT/F’s accuracy in detecting any documented delay and/or concern (including ASD) was as follows: sensitivity was 11.8%, specificity was 97.4%, PPV was 72.4%, and NPV was 65.9%.

**Effect of Age and Multiple Screenings**

Across all first-time screenings, screenings at older ages (21–26 months) were more sensitive than at younger ages (16–20 months; 48.8% vs 35.1%). In the AAP subgroup of children screened twice, the second screening at 24 months was marginally more sensitive (39.8% vs 31.8%) and had higher PPV (24.7%).
vs 16.4%) than the first screening at 18 months. Combining results from 18- and 24-month screenings yielded greater sensitivity than either screening alone (51.1%). See Table 4.

**Effect of Child and/or Family Characteristics**

Despite comparable sensitivities across racial groups, specificity and PPV were higher in white children (97.9%; 24.0%) compared with black children (91.7%; 11.7%), Asian children (90.4%; 10.8%), and those from other or multiple racial groups (93.8%; 13.4%). Differences were not observed between black, Asian, and other or multiple racial groups, or by ethnicity.

Higher specificity and PPV were observed in children with English-only exposure compared with children with non-English exposure (95.2% vs 86.9%; 15.3% vs 8.5%) as well as children from higher- versus lower-income families (97.0% vs 92.3%; 20.4% vs 11.8%). The same pattern was observed for insurance payer because children with private insurance had higher specificity (97.6% vs 91.0%) and PPV (22.1% vs 11.3%). This pattern was also observed for practice type because specificity was higher in children screened in suburban sites (96.8% vs 91.5%) as was PPV, although this effect size fell below the cutoff (h = 0.18; 18.5% vs 12.0%).

PPV was higher in boys than in girls (19.9% vs 7.7%). Sensitivity was higher in children born premature (54.3% vs 35.8%), but specificity was lower (89.3% vs 95.4%).

**Effect of Follow-up Interview Administration**

Children requiring the follow-up interview (ie, scores of 3–7) who received it were more likely to be full term (P = .007; OR = 1.41), have lower incomes (P < .001; OR = 1.65), and be from urban practices (P < .001; OR = 1.60) than those who did not receive the follow-up interview. Black children were also more likely than white (P = .001; OR = 1.48), Asian (P = .06; OR = 1.40), or children of other or multiple races (P < .001; OR = 1.64) to receive the follow-up interview. Children without ASD were also more likely to receive the follow-up interview than children with ASD (P = .02; OR = 1.54).

PPV was examined separately on the basis of the presence or absence of follow-up interview results. PPV was

---

**FIGURE 2**

M-CHAT/F results for screened cohort with outcome data.
34.8% in children with a questionnaire score of 8+ (ie, follow-up interview bypassed, \(n = 201\)). PPV was 38.2% in children who continued to screen positive after receiving the follow-up interview \(n = 34\). PPV was 9.6% in those that did not receive the interview \(n = 967\). Other metrics could not be calculated separately by follow-up results because they are calculated by using screen negatives, and it is not possible to know how many of the 967 children who did not receive the interview would have screened negative.

Excluding all follow-up interview data (ie, M-CHAT/F questionnaire results), sensitivity was 45.2%, specificity was 91.7%, PPV was 11.0%, and NPV was 98.7%.

**M-CHAT/F Results and Time to Diagnosis**

Kaplan-Meier survival curves revealed that the mean time to diagnosis was significantly shorter for children with ASD who screened positive than for those who screened negative (mean difference = 7.45 months; \(P < .001\)).

**DISCUSSION**

In this study we examined the M-CHAT/F’s accuracy within a universally screened cohort. The results revealed high screening rates, achieved through robust EHR support for screening. Systematic follow-up of children who screened positive and negative allowed an estimation of sensitivity, specificity, PPV, and NPV. The M-CHAT/F’s sensitivity to detect ASD was just 39%, indicating that the majority of children later diagnosed with ASD screened negative. PPV was just 15%, an estimate consistent with recent large (but not universally screened) studies conducted in Norway and Malaysia.\(^{14,15}\) However, this estimate was much lower than that found in some US-based studies conducted in research settings, likely partially because of different prevalence rates across studies.\(^{9,21}\) PPV improved substantially when any diagnosis or concern was considered (72%). Specificity and NPV for ASD were high (95% and 99%, respectively), but it is important to remember that with low prevalence and screen-positive rates, specificity and NPV will tend toward high rates.
Although the M-CHAT/F identified fewer children with ASD than expected, those who did screen positive received an ASD diagnosis 7 months earlier than those who screened negative. This suggests that a positive M-CHAT/F screening may have contributed to an earlier diagnosis for children with ASD in this cohort. However, continued research is needed to understand the specific effect of a positive screen on age of diagnosis.

The M-CHAT/F was significantly more sensitive at older ages (49% at 21–26 months) than at younger ages (35% at 16–20 months). However, PPV did not improve with age (16% vs 14%), which is in contrast to findings from some previous studies. PPV did improve for repeated screenings; 25% of children who screened positive at the second screening had ASD (regardless of results of the first screening) compared with 17% for the first screen. These results highlight the potential importance of screening twice as well as the difficulty of accurate screening at ~18 months.

Sensitivity was higher, but specificity was lower for premature children compared with those born full term, consistent with several other studies in very premature children. PPV was lower in girls (8%) than in boys (20%), but it is unclear whether this was because of delayed diagnosis for some girls or poorer M-CHAT/F performance in girls, but future research should be used to examine sex differences in each stage of detection, from screening to diagnosis.

Although electronic screening yielded high screening rates, when children were missed, they were more likely to be children of color, from lower-income households, seen in an urban practice, receive public insurance/Medicaid, and be exposed to a language other than English. These same children were also less likely to present for 2 well-child visits, resulting in disparities in receiving care according to AAP guidelines.

### TABLE 3 M-CHAT/F Screen-Positive Rates by Child and Family Characteristics

<table>
<thead>
<tr>
<th>Screened Positive on M-CHAT/F, No. (%)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screened cohort (n = 23,628)</td>
<td>1474 (6.2)</td>
</tr>
</tbody>
</table>

#### By age at first screening
- 16–20 mo (n = 18,585) 1027 (5.5) —
- 21–26 mo (n = 5043) 447 (8.9) 1.66 (1.48–1.87) a

#### By multiple screenings
- First screening at ~18 mo (n = 11,624) 518 (4.2) —
- Second screening at ~24 mo (n = 11,624) 440 (3.5) 0.84 (0.74–0.96) a

#### By sex
- Female (n = 11,375) 635 (5.6) —
- Male (n = 12,253) 839 (6.8) 1.24 (1.12–1.38) a

#### By gestational age
- Full term, 37+ wk (n = 18,587) 1022 (5.5) —
- Premature, <37 wk (n = 2429) 302 (12.4) 2.44 (2.13–2.80) a

#### By language documented in the EHR
- English only (n = 22,221) 1301 (5.9) —
- Language other than English (n = 1063) 159 (15.0) 2.83 (2.37–3.38) a

#### By race
- White (n = 10,781) 319 (3.0) —
- Black (n = 7924) 752 (9.5) 3.45 (3.01–3.94) a
- Asian (n = 1016) 108 (10.6) 3.90 (3.11–4.90) a
- Other or multiple races (n = 3884) 295 (7.6) c 2.70 (2.29–3.16) a

#### By ethnicity
- Not Hispanic or Latino (n = 21,811) 1305 (6.0) —
- Hispanic or Latino (n = 1743) 167 (9.6) 1.67 (1.41–1.97) a

#### By insurance payer at screening visit
- Private insurance (n = 13,437) 427 (3.2) —
- Public insurance/Medicaid (n = 10,086) 1038 (10.3) 3.50 (3.11–3.93) a

#### By primary care site type
- Suburban (n = 14,667) 610 (4.2) —
- Urban (n = 8961) 864 (9.6) 2.46 (2.21–2.74) a

#### By median income
- Higher income (n = 12,571) 496 (3.9) —
- Lower income (n = 11,049) 977 (8.8) 2.36 (2.11–2.64) a

The first group listed is the reference group for ORs. Some subgroups do not add to 23,628 participants because of missing data. A comparison of the first and second screening was conducted in subsample with 2 screenings according to AAP guidelines (n = 11,624 children). CI, confidence interval; —, not applicable.

a Indicates screen-positive rates that are different by subgroup at P < .05 after Bonferroni correction.
Despite these disparities, screening rates among children from traditionally underrepresented groups were still relatively high (>80%). Children of color and those from lower-income households screened positive at 2 to 3 times the rates of white, higher-income, privately insured, and suburban children. Elevated screen-positive rates resulted in somewhat higher sensitivity to detect ASD but also higher false-positive rates (ie, lower PPV and specificity) in these groups.

It is suggested by these data that disparities in age of diagnosis are likely preceded by disparities in screening rates and differential accuracy of screening tools for children from underrepresented and underresourced groups. Future
research should attempt to disentangle the effects of race and/or ethnicity, income, language, and primary care setting to better understand the role of child-, family-, and practice-level variables on screening.

With regard to the follow-up interview, pediatricians did not complete the interview systematically but instead may have used clinical judgement when deciding when to administer. Children who received the follow-up interview were significantly less likely to have ASD, and when children received the follow-up, almost all (95%) screened negative. Thus, pediatricians may have used the interview to confirm a clinical opinion of “not ASD” and chosen to skip the interview and go straight to referral when they suspected ASD. Additional research is needed to disentangle differences in follow-up interview rates by race and income; this may have been a clinical adaptation to artificially high screen-positive rates or it may incorrectly delay diagnosis for these children. However, the follow-up interview did appear to reduce false-negatives (ie, improve PPV), underscoring the importance of this step of the screening process.

There are limitations to the current study, largely surrounding the real-world nature of this cohort. Although this epidemiological study represented all children within the CHOP primary care network, findings may not generalize to other populations, particularly those with less access to ASD specialty care. Diagnoses were given in real-world clinical settings rather than through rigorous research studies. Diagnostic information was also only available through 4 to 8 years of age, so the M-CHAT/F’s accuracy may differ as children in this cohort age.

The M-CHAT/F, rather than the M-CHAT–R/F, was used, although accuracy of these 2 versions is comparable.9 As indicated above, not all eligible children received the follow-up interview, which likely downwardly biased specificity and PPV and upwardly biased sensitivity and NPV in that some would have screened negative if they had received the follow-up interview. Thus, this study cannot estimate the accuracy of perfect M-CHAT/F administration but instead provides critical information on how well this tool detects ASD in a real-world, universally screened cohort.

CONCLUSIONS

These results suggest that universal screening in primary care is possible when supported by electronic administration and EHR integration. However, universal screening and systematic follow-up revealed low accuracy of the M-CHAT/F, particularly for children of color and those from lower-income households. Although some may interpret these findings as evidence against universal screening, we caution against this interpretation given the earlier age of diagnosis for screen positives. Instead, results suggest that augmentative screening methods should be developed to detect more children through universal screening efforts and reduce disparities.

Promising new methods include parent-report tools that are supported by picture or video models26 and direct data gathering methods that leverage technological advances in computing and machine learning.27,28 However, any new method should be tested in cohorts that are universally screened and systematically followed-up to reduce the bias associated with screening and following selected populations.

ACKNOWLEDGMENTS

We thank R. Christopher Sheldrick for his helpful feedback on this article, as well as the providers and families who contributed data to this project through clinical care at CHOP.

ABBREVIATIONS

AAP: American Academy of Pediatrics
ADHD: attention-deficit/hyperactivity disorder
ASD: autism spectrum disorder
CHOP: Children’s Hospital of Philadelphia
EHR: electronic health record
M-CHAT: Modified Checklist for Autism in Toddlers
M-CHAT/F: Modified Checklist for Autism in Toddlers with Follow-Up
M-CHAT–R/F: Modified Checklist for Autism in Toddlers, Revised, with Follow-Up
NPV: negative predictive value
OR: odds ratio
PPV: positive predictive value
REFERENCES


Accuracy of Autism Screening in a Large Pediatric Network
Whitney Guthrie, Kate Wallis, Amanda Bennett, Elizabeth Brooks, Jesse Dudley, Marsha Gerdes, Juhi Pandey, Susan E. Levy, Robert T. Schultz and Judith S. Miller
Pediatrics 2019;144;
DOI: 10.1542/peds.2018-3963 originally published online September 27, 2019;
Accuracy of Autism Screening in a Large Pediatric Network
Whitney Guthrie, Kate Wallis, Amanda Bennett, Elizabeth Brooks, Jesse Dudley, Marsha Gerdes, Juhi Pandey, Susan E. Levy, Robert T. Schultz and Judith S. Miller
Pediatrics 2019;144;
DOI: 10.1542/peds.2018-3963 originally published online September 27, 2019;

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://pediatrics.aappublications.org/content/144/4/e20183963

Data Supplement at:
http://pediatrics.aappublications.org/content/suppl/2019/09/18/peds.2018-3963.DCSupplemental

Pediatrics is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. Pediatrics is owned, published, and trademarked by the American Academy of Pediatrics, 345 Park Avenue, Itasca, Illinois, 60143. Copyright © 2019 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 1073-0397.

American Academy of Pediatrics
DEDICATED TO THE HEALTH OF ALL CHILDREN®