**BRIEF REPORT**

# Brief Report: Evaluation of the Short Quantitative Checklist for Autism in Toddlers (Q-CHAT-10) as a Brief Screen for Autism Spectrum Disorder in a High-Risk Sibling Cohort

Sarah Raza[1,2] · Lonnie Zwaigenbaum[1,2] · Lori-Ann R. Sacrey[1,2] · Susan Bryson[3] · Jessica Brian[4,5] · Isabel M. Smith[3] · Kyle Reid[2] · Wendy Roberts[6] · Peter Szatmari[5,7,8] · Tracy Vaillancourt[9] · Caroline Roncadin[10] · Nancy Garon[11]

## Abstract

This study examined the potential of the short form of the Quantitative Checklist for Autism in Toddlers (Q-CHAT-10) to identify autism spectrum disorder (ASD) in a high-risk sibling cohort. High-risk (HR; siblings of children diagnosed with ASD) and low-risk (LR; no family history of ASD) toddlers were assessed prospectively at 18 and 24 months of age using the Q-CHAT-10 and underwent blind diagnostic assessment for ASD at 36 months of age. The results indicated that at 18 and 24 months, total score differentiated between HR toddlers subsequently diagnosed with ASD from other HR and LR toddlers. The sensitivity at both time points was acceptable; however, the specificity was below the level recommended for clinical application.

**Keywords** Autism · Autism spectrum disorder · Early screening · Behavioral signs · Infant siblings · Classification accuracy

✉ Sarah Raza
sraza@ualberta.ca

1    Department of Pediatrics, University of Alberta, Edmonton, AB, Canada

2    Autism Research Centre (E209), Glenrose Rehabilitation Hospital, 10230-111 Avenue, Edmonton, AB T5G 0B7, Canada

3    IWK Health Centre, Dalhousie University, Halifax, NS, Canada

4    Bloorview Research Institute, Toronto, ON, Canada

5    University of Toronto, Toronto, ON, Canada

6    ISAND, Toronto, ON, Canada

7    The Hospital for Sick Children, Toronto, ON, Canada

8    Centre for Addiction and Mental Health, Toronto, ON, Canada

9    University of Ottawa, Ottawa, ON, Canada

10   McMaster Children's Hospital–Hamilton Health Sciences, Hamilton, ON, Canada

11   Mount Allison University, Sackville, NB, Canada

## Introduction

Autism spectrum disorder (ASD) is a behaviorally defined, neurodevelopmental disorder characterized by impairments in social-communication and the presence of restricted and/or repetitive patterns of behavior (American Psychiatric Association 2013). Although we are rapidly gaining knowledge about the early signs of ASD (Zwaigenbaum et al. 2013; Jones et al. 2014), there is often a significant delay between the time of first concern and eventual diagnosis of ASD. While many parents report concerns about their children's development between 6 and 24 months of age (Sacrey et al. 2015; Zwaigenbaum et al. 2015), ASD often goes undiagnosed until 4 years of age or later (Daniels and Mandell 2013). As a result, screening for early symptoms of ASD is important to facilitate earlier detection and implement targeted interventions to improve functional outcomes.

Routine universal screening has been recommended by the American Academy of Pediatrics to ensure optimal detection of the early signs of ASD in young children (Johnson and Myers 2007). Although an abundance of ASD-specific and broadband screening instruments are available to identify toddlers who may be on a developmental trajectory for ASD, as few as 8% of primary health-care professionals report screening for ASD (Dosreis

et al. 2006). Physicians often cite limited time available for patient encounters, cost of screening, lack of familiarity with ASD screening tools, and the process of making effective referrals for further evaluation as important barriers impeding early screening for ASD. In a community practice environment where optimizing time and productivity is crucial, the development and implementation of brief screening instruments for ASD is warranted. Brief screening tools would not only help parents articulate concerns about their children's development during pediatric wellness visits, but may also assist busy healthcare professionals with respect to decisions about referring a child for diagnostic assessment. Moreover, barriers against the time/cost to administer, score, and interpret results, as well as likelihood of parent completion, would be mitigated by implementation of brief screening instruments.

One early screen of note is the Quantitative Checklist for Autism in Toddlers (Q-CHAT; Allison et al. 2008). The Q-CHAT, a 25-item parent-reported questionnaire scored on a 5-point Likert scale, was a revision of the original Checklist for Autism in Toddlers and marked the shift in early ASD screening from categorical to dimensional (i.e., quantitative) measurement. Demonstrating good test–retest reliability, internal consistency (0.83), and excellent power to detect more subtle manifestations of ASD and to discriminate young children (18 to 30 months of age) who may be on a developmental trajectory for ASD from those who are developing typically, the Q-CHAT showed promise as an early screen for ASD (Allison et al. 2008). Of important note, however, the full range of psychometric properties (i.e., sensitivity, specificity, positive and negative predictive validity) of the Q-CHAT has yet to be been examined. To meet the need for briefer instruments, an abbreviated version of the Q-CHAT was recently developed, the Q-CHAT-10 (Allison et al. 2012). Comprised of the 10 most discriminating items on the original Q-CHAT—including joint attention, language, pretend play, and social-communication behavior—the Q-CHAT-10 was effective at retrospectively discriminating preschool children with ASD from community controls in a case-control sample, providing preliminary support for the utility of this abbreviated questionnaire. Moreover, the test accuracy and psychometric properties of the Q-CHAT-10 were excellent (sensitivity of 0.91, specificity of 0.89, positive predictive validity of 0.58, internal consistency of 0.85). Although Allison et al. (2012) concluded that the Q-CHAT-10 might serve as a useful "red flag" or rapid screen for ASD risk and potentially assist frontline healthcare professionals in the referral pathway for ASD, the authors cautioned about the generalizability of the findings. The evaluation of the Q-CHAT-10 using prospective designs is warranted. As an initial step, assessing the Q-CHAT-10 in a high-risk (HR; has an older sibling with ASD diagnosis) cohort could be informative, and allows for a smaller

sample (thus, increased feasibility of outcome assessment of the total cohort) with an elevated rate of ASD.

The purpose of the present study was to assess the potential of the Q-CHAT-10 as a brief screen for ASD in a cohort of HR infant siblings. Parents of HR and low-risk (LR; no family history of ASD) toddlers completed the Q-CHAT-10 at 18 and 24 months of age, and toddlers underwent blinded diagnostic assessment for ASD at 36 months of age. To examine the screening ability of the Q-CHAT-10 in a HR risk context, our main objective was to first determine the diagnostic accuracy of the Q-CHAT-10, as measured by sensitivity and specificity of total score relative to 36-month diagnostic outcomes. It has been recommended that sensitivity and specificity of an early detection tool should exceed 0.70 to be considered for population screening (Volkmar et al. 1988; Cicchetti et al. 1995; Dumont-Mathieu and Fein 2005; Zwaigenbaum et al. 2015). A secondary objective was to determine whether the Q-CHAT-10 differentiated between typical and atypical development; thus, HR and LR groups were stratified by presence of ASD, atypical developmental features sub-threshold for ASD (Ozonoff et al. 2014), or typical development.

## Methods

### Participants

Infant siblings of children with ASD were recruited between 6 and 12 months of age as part of a larger ongoing prospective study of early development of ASD, conducted at four major ASD diagnostic centers in Canada (Glenrose Rehabilitation Hospital in Edmonton, Alberta; Holland Bloorview Kids Rehabilitation Hospital in Toronto, Ontario; McMaster Children's Hospital in Hamilton, Ontario; IWK Health Centre in Halifax, Nova Scotia). The research ethics boards at each site approved the research protocol, and written informed consent was obtained from the primary caregiver of each participant. All participants were born between 36 and 42 weeks gestation and had a birth weight greater than 2500 g. Diagnosis of ASD in the proband (i.e., older sibling) was confirmed by expert clinical judgment using DSM-IV criteria and review of clinical records. None of the probands or younger siblings had identifiable neurological disorders or genetic conditions, nor severe sensory or motor impairments that could potentially account for ASD. Low-risk (LR) controls were recruited from local communities between 6 and 12 months of age and did not have any first- or second-degree relatives an ASD diagnosis. All participants were born between 2001 and 2005, and 36-month follow-up was completed by early 2009.

Participants were included in this study if (1) they had a completed Q-CHAT-10 at 18 and/or 24 months of age and

(2) underwent their 36-month diagnostic assessment. Of the 191 toddlers with a 36-month follow-up, 116 high-risk (HR) siblings (62 boys and 54 girls) and 56 LR controls (27 boys and 29 girls) had complete Q-CHAT-10 data at 18 and/or 24 months of age and were included. Those that did not have complete Q-CHAT-10 data at either age were excluded from further analyses.

## Measures

Toddlers were assessed at 18, 24, and 36 months of age. Classification properties of the Q-CHAT-10 were examined at 18 and 24 months. ASD symptomology was assessed using the Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R) and cognitive development was assessed using the Mullen Scales of Early Learning (MSEL) at 36 months.

*Short form of the Quantitative Checklist for Autism in Toddlers* (Q-CHAT-10; Allison et al. 2012). The Q-CHAT-10 is a 10-item parent-reported questionnaire abbreviated from the Quantitative Checklist for Autism in Toddlers (Q-CHAT; Allison et al. 2008). Designed as a brief screen for ASD for use in toddlers between 18 and 24 months of age, the Q-CHAT-10 assesses various ASD symptoms, including items related to joint attention, social-communication, and pretend play. The Q-CHAT-10 has been demonstrated to retrospectively reliably distinguish toddlers with and without ASD and possesses excellent test accuracy properties and internal consistency (Allison et al. 2012). Primary caregivers rate the degree to which their child exhibits certain types of behavior and characteristics using a 5-point rating scale ranging from 0 = *Always/very easy/very typical/many times a day*, 1 = *Usually/quite easy/quite typical/few times a day*, 2 = *Sometimes/quite difficult/slightly unusual/few times a week*, 3 = *Rarely/very difficult/very unusual/less than once a week*, 4 = *Never/impossible*. Ratings are converted to a binary scoring system, such that ratings of 0 or 1 correspond to a score of 0, and ratings of 2, 3, or 4 correspond to a score of 1 (reverse-scored for item 10). The suggested Q-CHAT-10 cut-point to flag risk of ASD is a total score of 3, as per Allison et al. (2012). Primary caregivers completed the Q-CHAT-10 at 18 and 24 months of age.

*Mullen Scales of Early Learning* (MSEL; Mullen 1995). The MSEL consists of five subscales assessing nonverbal, cognitive, language ability, and motor development in children between 0 and 69 months of age: visual reception, expressive language, receptive language, fine motor, and gross motor. It has been demonstrated to have excellent inter-rater and test–retest reliability (Mullen 1995). The MSEL was administered at 36 months of age.

*Autism Diagnostic Observation Schedule* (ADOS; Lord et al. 2000). The ADOS is a standardized, activity-based assessment designed to elicit communication, repetitive behavior, social interaction, and imaginative play behavior. It has been demonstrated to reliably distinguish children with ASD from typically developing children and shows excellent inter-rater reliability (Lord et al. 1989). The ADOS consists of four modules, each appropriate for individuals of differing language levels: module 1 = minimal or no language, module 2 = regular use of non-echoed three-word phrases, module 3 = child with fluent language, and module 4 = adolescent or adult with fluent language. Modules 1, 2, and 3 were used to assess the participants in the current study. The ADOS was administered at 36 months of age by examiners who had attained reliability according to the developers' criteria. To optimize comparability across modules (and thus, across language levels), severity indices for Social Affect, Restricted and Repetitive Behavior, and Total ADOS scores were calculated using the ADOS severity metrics at 36 months of age (Gotham et al. 2009). Specifically, total algorithm scores were converted into calibrated severity scores, which ranged from 1 to 10. Note that the ADOS Toddler Module was not available at the time of the study, as 36-month assessments were completed by early 2009 (i.e., prior to the publication of the ADOS Toddler Module).

*Autism Diagnostic Interview-Revised* (ADI-R; Lord et al. 1994). The ADI-R is an investigator-directed interview designed to elicit information regarding developmental history and ASD-related behavior in order to make a DSM-IV diagnosis of ASD. The ADI-R has been demonstrated to reliably distinguish children with ASD from other developmental disabilities and possesses excellent inter-rater reliability (Lord et al. 1994). The ADI-R was administered at 36 months of age and the ADI-R Total Algorithm Score was calculated (i.e., calculated from total score on the Social, Communication, and Restricted/Repetitive Behavior algorithm scores combined). ADI-R interviews were conducted by research-reliable examiners.

## Diagnostic Procedure

At 36 months, each participant underwent an independent diagnostic evaluation, conducted by an expert clinician (i.e., developmental pediatrician, child psychiatrist, or clinical psychologist) blind to prior study data and group membership. Diagnoses were assigned using the DSM-IV criteria and clinical judgment based on all available developmental measures, including the ADOS, ADI-R, and MSEL. Note that the DSM-5 was not available at the time of the study, as follow-up was completed prior to the publication of the DSM-5.

## Outcome Classification

For the purpose of evaluating the Q-CHAT-10 and its ability to differentiate between typical and atypical development,

toddlers were stratified into five groups based on risk status and 36-month outcome assessments, using the following definitions:

1.  HR-ASD (n = 25; 17 boys and 8 girls): High-risk toddlers who met diagnostic criteria for ASD using the ADOS, ADI-R, and expert clinical judgment.
2.  HR-AD (n = 30; 20 boys and 10 girls): High-risk toddlers who did not meet diagnostic criteria for ASD but had other evidence of atypical development (AD; delays or sub-threshold ASD symptoms), as defined by: (a) scores ≥ 1.5 standard deviations below the mean on two or more MSEL subtests, and/or (b) scores ≥ 2 standard deviations below the mean on one or more MSEL subtests, and/or (c) scores ≤ 3 points below the ASD cutoff on the ADOS (Ozonoff et al. 2014).
3.  HR-TD (n = 61; 25 boys and 36 girls): High-risk toddlers who did not meet criteria for ASD or atypical development as defined above (i.e., were typically developing).
4.  LR-AD (n = 13; 7 boys and 6 girls): Low-risk toddlers who did not have first- or second-degree relatives with an ASD diagnosis. These toddlers were also not diagnosed with ASD, but have AD using criteria as described above.
5.  LR-TD (n = 43; 20 boys and 23 girls): Low-risk toddlers who did not have first- or second-degree relatives with an ASD diagnosis and were not diagnosed with ASD (i.e., were typically developing).

## Analytic Approach

To assess utility of the Q-CHAT-10 as a rapid screen for ASD in a HR context, sensitivity, specificity, positive predictive validity, and negative predictive validity of the 18- and 24-month Q-CHAT-10 scores were examined relative to ASD diagnosis at 36 months. Analyses were limited to the HR cohort (i.e., HR-ASD versus HR-AD and HR-TD groups) to specifically examine the potential properties of the Q-CHAT-10 within that context, and since the only cases of ASD were from this group. Classification properties of the Q-CHAT-10 relative to predicting HR were determined based on an optimal cut-point of 3. Examination of the associated sensitivity and specificity of various cut-points in our sample (ranging from 1 to 10) provided evidence of the optimal cut-point to be 3 to flag ASD risk (sensitivity and specificity at 18 and 24 months were 0.75 and 0.71 for sensitivity, and 0.63 and 0.65 for specificity; there were no cut-points associated with sensitivity and specificity > 70%). Consistent with our sample, the cut-point of 3 was also recommended in previous research (Allison et al. 2012). Determinants of screening accuracy included: (1) sensitivity, defined as the proportion of HR toddlers with ASD correctly classified by the Q-CHAT-10; (2) specificity, defined as the proportion of HR toddlers not diagnosed with ASD correctly classified by the Q-CHAT-10; (3) positive predictive validity (PPV), the proportion of toddlers with ASD who are correctly identified as toddlers with ASD [(*true positive*/(*true positive + false positive*)]; (4) negative predictive validity (NPV), the proportion of toddlers without ASD who are incorrectly identified as toddlers with ASD [(*true negative*/(*true negative + false negative*)] and; (5) false positives, the proportion of toddlers who did not have ASD yet screened positive on the Q-CHAT-10; (Fischer et al. 2003).

Scores on the Q-CHAT-10 were compared between groups at 18 and 24 months using one-way ANOVAs. Group effects were explored using Benjamini and Hochberg (1995) corrections to reduce the chance of false positives. As our objective was to examine whether the Q-CHAT-10 was able to distinguish HR toddlers with ASD from other HR and LR toddler groups, planned comparisons were conducted on all group by age interactions.

## Results

### Participant Characteristics

Participant characteristics are summarized in Table 1. There was no significant sex difference by group ($\chi^2(4) = 8.64$, $p = 0.07$). There were significant group differences for the ADOS Social Affect, Restrictive Repetitive Behavior, and Total ADOS scores ($p$'s < 0.01), as well as the ADI-R Total Algorithm score ($p < 0.01$) at 36 months of age. The groups also differed on the 36-month MSEL Visual Reception, Expressive Language, Receptive Language, and Fine Motor subtests ($p$'s < 0.01).

### Individual Classification of the Q-CHAT-10 within the HR Group

The sensitivity and specificity of the Q-CHAT-10 were examined with respect to subsequent ASD diagnosis using the suggested cut-point of 3 (Allison et al. 2012) at 18 and 24 months. Estimates of sensitivity, specificity, PPV, and NPV for the Q-CHAT-10 were 0.75, 0.63, 0.36, and 0.90 at 18 months respectively, and 0.71, 0.65, 0.34, and 0.90 at 24 months, respectively. The proportions of false-positives for total Q-CHAT-10 scores at 18 and 24 months were 0.50 and 0.33 in the HR-AD group, and 0.33 and 0.33 in the HR-TD group, respectively.

### Group Comparisons

#### 18 Months

There was a significant effect of group on total Q-CHAT-10 score at 18 months of age [$F(4,133) = 5.89$, $p < 0.01$]. Post hoc analysis was run using a Benjamini and Hochberg

**Table 1** Participant characteristics

| Characteristic | LR-TD, n (Total=43) | | LR-AD, n (Total=13) | | HR-TD, n (Total=61) | | HR-AD, n (Total=30) | | HR-ASD, n (Total=25) | | F test | p |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sex (M:F) | 20:23 | | 7:6 | | 25:36 | | 20:10 | | 17:8 | | χ²=8.64 | <0.07 |
| | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | Mean | (SD) | | |
| ADOS, 36 months | | | | | | | | | | | | |
| SA | 1.88 | 0.88 | 5.31 | 1.97 | 1.77 | 0.86 | 4.33 | 2.11 | 7.28 | 2.07 | 83.27 | <0.01[abcdefghj] |
| RRB | 3.28 | 2.4 | 5.46 | 2.76 | 4.08 | 2.44 | 5.80 | 2.80 | 8.20 | 1.68 | 19.22 | <0.01[abcdegi] |
| Overall severity | 1.42 | 0.59 | 4.77 | 1.69 | 1.59 | 0.76 | 4.37 | 1.83 | 7.48 | 2.14 | 118.39 | <0.01[abcdeghj] |
| ADI-R, 36 months | | | | | | | | | | | | |
| Total score | 0.63 | 1 | 1.23 | 1.96 | 0.80 | 1.76 | 1.59 | 1.32 | 12.12 | 7.17 | 70.88 | <0.01[abcd] |
| MSEL, 36 months | | | | | | | | | | | | |
| Visual reception | 64.81 | 8.59 | 63.23 | 9.85 | 61.75 | 9.94 | 50.32 | 16.09 | 40.88 | 17.45 | 20.50 | <0.01[abcdefg] |
| Receptive language | 60.38 | 9.66 | 53.15 | 10.56 | 55.36 | 8.02 | 47.63 | 8.49 | 36.83 | 13.49 | 26.18 | <0.01[abcdegij] |
| Expressive language | 59.91 | 7.40 | 55.23 | 8.66 | 55.71 | 7.40 | 47.61 | 8.87 | 38.84 | 14.05 | 25.53 | <0.01[abcdefgi] |
| Fine motor | 59.86 | 12.09 | 52.65 | 9.94 | 57.17 | 11.60 | 40.69 | 14.14 | 35.08 | 13.99 | 24.26 | <0.01[abcefg] |

*LR-TD* low-risk toddlers, *LR-AD*, low-risk toddlers who had atypical development, *HR-TD* high-risk toddlers who did not meet criteria for autism spectrum disorder or atypical development, *HR-AD* high-risk toddlers who did not meet criteria for autism spectrum disorder but have evidence of atypical development, *HR-ASD* high-risk toddlers diagnosed with autism spectrum disorder, *ADOS* Autism Diagnostic Observation Schedule, *ADI-R* Autism Diagnostic Interview-Revised, *MSEL* Mullen Scales of Early Learning, *SA* social affect, *RRB* restricted repetitive behaviors, *SD* standard deviation

[a]HR-ASD versus LR-TD
[b]HR-ASD versus LR-AD
[c]HR-ASD versus HR-TD
[d]HR-ASD versus HR-AD
[e]HR-AD versus LR-TD
[f]HR-AD versus LR-AD
[g]HR-AD versus HR-TD
[h]HR-TD versus LR-AD
[i]HR-TD versus LR-TD
[j]LR-AD versus LR-TD

correction ($q < 0.01$), showing that the total Q-CHAT-10 score was higher for HR toddlers with ASD compared to the HR-TD ($p < 0.01$) and LR-TD group ($p < 0.01$; Fig. 1a). Total scores did not differ between toddlers in the HR-AD, HR-TD, LR-AD, and LR-TD groups ($p$'s > 0.02).

### 24 Months

A significant effect of group on total Q-CHAT-10 score was revealed at 24 month of age [$F (4,125) = 8.64$, $p < 0.01$]. Post hoc analysis was run using a Benjamini and Hochberg correction ($q < 0.02$). As displayed in Fig. 1b, Q-CHAT-10 total score was higher for the HR-ASD toddlers compared to the other four groups ($p$'s < 0.01). Total score did not differ between toddlers in the HR-AD, HR-TD, LR-AD, and LR-TD groups ($p$'s > 0.05).

## Discussion

The objective of this study was to evaluate the potential of the short form of the Quantitative Checklist for Autism in Toddlers (Q-CHAT-10) as a brief screen for ASD in an HR cohort. Primary caregivers completed the Q-CHAT-10 when their toddlers were 18 and 24 months of age. There were two main findings. First, while sensitivity was above 70% at each age, specificity of the Q-CHAT-10 for detecting ASD within the HR group was below 70%, leading to over-identification of toddlers who were unlikely to be diagnosed with ASD (i.e., high rate of false-positive screens), limiting utility for



**Fig. 1** Total score on Q-CHAT-10 at (**a**) 18 and (**b**) 24 months of age. HR-ASD is significantly different from * = LR-TD, # = LR-AD, ^ = HR-TD, & = HR-AD

clinical application (Cicchetti et al. 1995; Dumont-Mathieu and Fein 2005). Second, parents of HR siblings with ASD endorsed more ASD symptoms compared to the HR-TD and LR-TD groups at 18 months, as well as to the HR-AD, HR-TD, LR-AD, and LR-TD groups at 24 months of age. Taken together, the results suggest that while elevated scores on the Q-CHAT-10 were associated with subsequent ASD diagnosis in the HR group, individual classification was not sufficient for the purpose of screening.

The sensitivity, specificity, PPV, and NPV of Q-CHAT-10 total score at 18 and 24 months were 0.75 and 0.71 for sensitivity, 0.63 and 0.65 for specificity, 0.36 and 0.34 for PPV, and 0.90 and 0.90 for NPV, respectively, using the suggested cut-point of 3 (Allison et al. 2012). Our sensitivity, specificity, and PPV estimates were lower relative to the case-control sample examined by Allison et al. (2012), who reported sensitivity and specificity as high as 0.91 and 0.89, and PPV of 0.58. However, it is important to note that in that study the Q-CHAT-10 was used as a retrospective screen in a convenience sample of preschool children with confirmed diagnoses of ASD compared to typically developing peers. This case-comparison may have excluded children with more ambiguous presentations (including evidence of atypical development sub-threshold for ASD), potentially inflating sensitivity, specificity, and PPV estimates. Therefore, in the present study, toddlers were sampled consecutively in a systematic way and HR toddlers who did not meet diagnostic criteria for ASD were stratified into HR-AD and HR-TD subgroups to account for this variation. Inclusion of the HR-AD subgroup, in particular, lowered specificity estimates as 33% and 50% of HR-AD and 33% and 33% of HR-TD toddlers falsely screened positive for ASD at 18- and 24-month time points, respectively. These data are consistent with previous findings that other screeners for ASD are associated with higher false-positive rates prior to 24 months (Chawarska et al. 2007; Pandey et al. 2008; Zwaigenbaum et al. 2015).

We considered the potential clinical utility of the Q-CHAT-10 for the purpose of ASD screening in HR infants with reference to guidelines for indices of diagnostic accuracy (sensitivity, specificity, PPV, and NPV) proposed by Cicchetti and colleagues (1995). The guidelines are as follows: *poor* = < 0.70; *fair* = 0.70–0.79; *good* = 0.80–0.89; *excellent* = 0.90–1.00 (Cicchetti et al. 1995). Applying these criteria, only sensitivity and NPV exceeded 0.70 at 18 and 24 months in our HR sample. Specificity was lower (0.63 and 0.65 at 18 and 24 months, respectively), as was PPV (0.36 and 0.34), indicating that about 2/3 of screen positive children were not ultimately diagnosed with ASD. The implication is that the data from the current study do not support using the Q-CHAT-10 as a stand-alone screen for ASD in HR infants, but it is possible that the Q-CHAT-10 combined with a follow-up measure with higher specificity could yield acceptable PPV. Notably, other ASD screens
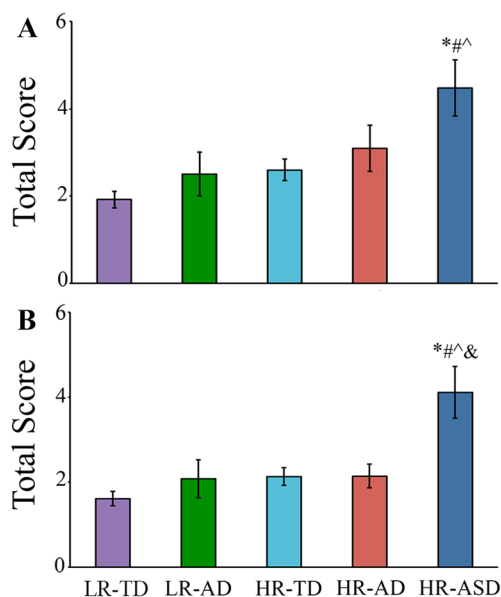
currently recommended by the American Academy of Pediatrics (Johnson and Myers 2007) based on adequate accuracy involve multiple steps (e.g., the M-CHAT- R/F, which includes a follow-up interview for parents endorsing a set number of items on an initial questionnaire). Such a process might be further explored for the Q-CHAT-10 if future studies report adequate sensitivity but insufficient specificity and/or PPV in other HR or LR samples.

Identifying HR toddlers with symptoms of ASD versus milder developmental difficulties at 18 and 24 months of age proved complex, due to variable patterns of symptom expression and emergence. As expected, total scores were higher for toddlers with ASD compared to the group of LR-TD toddlers both at 18 and 24 months of age (as per Allison et al. 2012). Surprisingly, scores on the Q-CHAT-10 did not discriminate between toddlers with ASD and those with atypical development until the 24-month time-point. This is contrary to previous reports, in which HR toddlers with such features were distinguishable from toddlers with ASD as early as 12–18 months of age (as defined by scores on diagnostic and developmental assessments; Chawarska et al. 2014; Ozonoff et al. 2014). These findings illustrate the difficulties involved in screening for early signs of ASD in an HR context, where variable symptom onset patterns and clinical profiles may overlap amongst toddlers who are later diagnosed with ASD versus other developmental concerns.

This is the first known prospective examination of the performance of the Q-CHAT-10 in a HR context. We acknowledge that screens may be more effective in certain contexts and not in others; thus, caution should be exercised when generalizing findings from HR cohorts to community samples and vice versa. Whereas Allison et al. (2012) supported the utility of Q-CHAT-10 as a rapid "red flag" ASD screener in a case-control sample, predictive utility in our high-risk cohort was not at the recommended level for clinical application. It is possible that our high-risk findings may be conservatively biased, as HR toddlers diagnosed with ASD may display fewer or less severe symptoms and better adaptive skills than toddlers from community referral (Sacrey et al. 2017). Ascertainment method (i.e., infant sibling cohorts versus community samples) may contribute to diversity within ASD and should be considered in the assessment of early detection and screening tools.

Limitations to this study must be acknowledged. First, parents of HR toddlers have at least one child already diagnosed with ASD. Increased awareness and vigilance of the early behavioral signs associated with ASD may have influenced how parents scored Q-CHAT-10 questionnaires (i.e., endorsing the presence of behavioral anomalies in their toddlers). Second, during each visit (prior to and following 18 months), parents received ongoing feedback concerning their children's development. This insight may

have affected how parents subsequently reported their toddlers' behavior on the Q-CHAT-10. Third, the analyses were based on the original ADOS and DSM-IV criteria (these editions were available at the time of study onset and completion, and prior to the release of newer editions). The ADOS has since between modified (i.e., ADOS-2) and now includes a Toddler Module—a new, standardized module that extends the application of the ADOS to children under 30 months of age who have minimal speech and nonverbal mental age of at least 12 months (Luyster et al. 2009). Moreover, the DSM is currently in its fifth edition, now encompassing the four previously separate DSM-IV disorders (i.e., autistic disorder, Asperger's disorder, childhood disintegrative disorder, and pervasive developmental disorder not otherwise specified) into two categories of symptoms (American Psychiatric Association 2013). Thus, the results of this study may not be generalizable nor hold true for the later editions of the ADOS and DSM. That is, these changes may impact the number of Q-CHAT-10 items endorsed in HR siblings, potentially increasing or decreasing total score. Future research should evaluate the potential of the Q-CHAT-10 using the ADOS Toddler Module and DSM-5 criteria. Finally, variable expression patterns, enhanced surveillance, and increased prevalence of ASD may complicate early ASD screening in a HR context, as well as in the general population.

In summary, our current data do not support using the Q-CHAT-10 as a stand-alone screening instrument in a high-risk context, with lower sensitivity and specificity estimates than previously reported in a case-control community sample (Allison et al. 2012). Based on its sensitivity, the Q-CHAT-10 may be useful as one component of an overall early detection strategy for ASD in HR infants (combined with a follow-up measure with higher specificity), but this approach remains to be evaluated in future research.

**Author Contributions** SR made substantial contributions to the conception and design of the paper, analyzed data, and prepared all drafts of the paper. LZ, LRS, SB, JB, IMS, KR, WR, PS, TV, CR, and NG contributed to the conception of the project, provided critical review of the manuscript, and approved the final draft.

## Compliance with Ethical Standards

**Conflict of interest** The authors declare that they have no conflicts of interest.

**Ethical Approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## References

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "Red Flags" for autism screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist for autism in toddlers in 1000 cases and 3000 controls. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*, 202–212.e7.

Allison, C., Baron-Cohen, S., Wheelwright, S., Charman, T., Richler, J., Pasco, G., & Brayne, C. (2008). The Q-CHAT (Quantitative Checklist for Autism in Toddlers): A normally distributed quantitative measure of autistic traits at 18–24 months of age: Preliminary report. *(2008). Journal of Autism and Developmental Disorders, 38*, 1414–1425.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th edn.). Arlington: American Psychiatric Publishing.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological), 57*, 289–300.

Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry, 48*(2), 128–138.

Chawarska, K., Shic, F., Macari, S., Campbell, D. J., Brian, J., Landa, R., Hutman, T., Nelson, C. A., Ozonoff, S., Tager-Flusberg, H., Young, G. S., Zwaigenbaum, L., Cohen, I. L., Charman, T., Messinger, D. S., Klin, A., Johnson, S., & Bryson, S. (2014). 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: A baby siblings research consortium study. *Journal of the American Academy of Child and Adolescent Psychiatry, 53*, 1317–1327.

Cicchetti, D. V., Volkmar, F., Klin, A., & Showalter, D. (1995). Diagnosing autism using ICD-10 criteria: A comparison of neural networks and standard multivariate procedures. *Child Neuropsychology, 1*, 26–37.

Daniels, A. M., & Mandell, D. S. (2013). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism, 18*, 583–597.

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 & Behavioral Pediatrics, 27*, 88–94.

Dumont-Mathieu, T., & Fein, D. (2005). Screening for autism in young children: the Modified Checklist for Autism in Toddlers (M-CHAT) and other measures. *Mental Retardation and Developmental Disabilities Research Reviews, 11*(3), 253–262.

Fischer, J. E., Bachmann, L. M., & Jaeschke, R. (2003). A readers' guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care Medicine, 29*, 1043–1051.

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

Johnson, C. P., & Myers, S. M. (2007). American Academy of Pediatrics Council on Children with disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*(5), 1183–1215.

Jones, E. J. H., Gliga, T., Bedford, R., Charman, T., & Johnson, M. H. (2014). Developmental pathways to autism: A review of prospective studies of infants at risk. *Neuroscience and Biobehavioral Reviews, 39*, 1–33.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H. Jr., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19*, 185–212.

Lord, C., Rutter, M., & LeCouteur, A. (1994). Autism diagnostic interview-revised: A revised version of the diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659–685.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Pertrak, R., Pierce, K., Bishop, S., Esler, A., Hus, V., Oti, R., Richler, J., Risi, S., & Lord, C. (2009). The Autism Diagnostic Observation Schedule – Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(9), 1305–1320.

Mullen, E. (1995). *Mullen scales of early learning: American guidance services*. Circle Pines: American Guidance Service.

Ozonoff, S., Young, G. S., Belding, A., Hill, M., Hill, A., Hutman, T., Johnson, S., Miller, M., Rogers, S. J., Schwichtenberg, A. J., Steinfeld, M., & Iosif, A. M. (2014). The broader autism phenotype in infancy: When does it emerge? *Journal of the American Academy of Child and Adolescent Psychiatry, 53*(4), 398–407.e2.

Pandey, J., Verbalis, A., Robins, D., Boorstein, H., Klin, A., Babitz, T., Chawarska, K., Volkmar, F., Green, J., Barton, M., & Fein, D. (2008). Screening for autism in older and younger toddlers with the modified checklist for autism in toddlers. *Autism, 12*(5), 513–535.

Sacrey, L. R., Zwaigenbaum, L., Bryson, S., Brian, J., Smith, I. M., Roberts, W., Szatmari, P., Roncadin, C., Garon, N., Novak, C., Vaillancourt, T., McCormick, T., MacKinnon, B., Jildera, S., & Armstrong, V. (2015). Can parents' concerns predict autism spectrum disorder? A prospective study of high-risk siblings from 6 to 36 months of age. *Journal of the American Academy of Child and Adolescent Psychiatry, 54*(6), 470–478.

Sacrey, L. R., Zwaigenbaum, L., Szatmari, P., Bryson, S., Georgiades, S., Brian, J., Smith, I. M., Vaillancourt, T., Garon, N., Roncadin, C., & Elsabbagh, M. (2017). Brief report: Characteristics of preschool children with ASD vary by ascertainment. *Journal of Autism and Developmental Disorders, 47*(5), 1542–1550.

Volkmar, F. R., Cicchetti, D. V., Dykens, E., Sparrow, S. S., Lechman, J. F., & Cohen, D. J. (1988). An evaluation of the autism behavior checklist. *Journal of Autism and Developmental Disorders, 18*, 81–97.

Zwaigenbaum, L., Bauman, M. L., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., McPartland, J. C., Natowicz, M. R., Choueiri, R., Fein, D., Kasari, C., Pierce, K., Buie, T., Carter, A., Davis, P. A., Granpeesheh, D., Mailloux, Z., Newschaffer, C., Robins, D., Roley, S. S., Wagner, S., & Wetherby, A. (2015). Early

identification of autism spectrum disorder: Recommendations for practice and research. *Pediatrics, 136*. S10-S40.

Zwaigenbaum, L., Bryson, S., & Garon, N. (2013). Early identification of autism spectrum disorders. *Behavioral Brain Research, 251*, 133–146.