

Original Paper

# Online Concerns of Parents Suspecting Autism Spectrum Disorder in Their Child: Content Analysis of Signs and Automated Prediction of Risk

Ayelet Ben-Sasson<sup>1</sup>, ScD; Elad Yom-Tov<sup>2</sup>, PhD

<sup>1</sup>University of Haifa, Haifa, Israel

<sup>2</sup>Microsoft Research & Development, Herzelia, Israel

**Corresponding Author:**

Ayelet Ben-Sasson, ScD

University of Haifa

199 Aba Khoushy Ave

Haifa, 3498838

Israel

Phone: 972 48240364

Fax: 972 48249753

Email: [asasson@univ.haifa.ac.il](mailto:asasson@univ.haifa.ac.il)

## Abstract

**Background:** Online communities are used as platforms by parents to verify developmental and health concerns related to their child. The increasing public awareness of autism spectrum disorders (ASD) leads more parents to suspect ASD in their child. Early identification of ASD is important for early intervention.

**Objective:** To characterize the symptoms mentioned in online queries posed by parents who suspect that their child might have ASD and determine whether they are age-specific. To test the efficacy of machine learning tools in classifying the child's risk of ASD based on the parent's narrative.

**Methods:** To this end, we analyzed online queries posed by parents who were concerned that their child might have ASD and categorized the warning signs they mentioned according to ASD-specific and non-ASD-specific domains. We then used the data to test the efficacy with which a trained machine learning tool classified the degree of ASD risk. Yahoo Answers, a social site for posting queries and finding answers, was mined for queries of parents asking the community whether their child has ASD. A total of 195 queries were sampled for this study (mean child age=38.0 months; 84.7% [160/189] boys). Content text analysis of the queries aimed to categorize the types of symptoms described and obtain clinical judgment of the child's ASD-risk level.

**Results:** Concerns related to repetitive and restricted behaviors and interests (RRBI) were the most prevalent (75.4%, 147/195), followed by concerns related to language (61.5%, 120/195) and emotional markers (50.3%, 98/195). Of the 195 queries, 18.5% (36/195) were rated by clinical experts as low-risk, 30.8% (60/195) as medium-risk, and 50.8% (99/195) as high-risk. Risk groups differed significantly ( $P<.001$ ) in the rate of concerns in the language, social, communication, and RRBI domains. When testing whether an automatic classifier (decision tree) could predict if a query was medium- or high-risk based on the text of the query and the coded symptoms, performance reached an area under the receiver operating curve (ROC) curve of 0.67 (CI 95% 0.50-0.78), whereas predicting from the text and the coded signs resulted in an area under the curve of 0.82 (0.80-0.86).

**Conclusions:** Findings call for health care providers to closely listen to parental ASD-related concerns, as recommended by screening guidelines. They also demonstrate the need for Internet-based screening systems that utilize parents' narratives using a decision tree questioning method.

(*J Med Internet Res* 2016;18(11):e300) doi: [10.2196/jmir.5439](https://doi.org/10.2196/jmir.5439)

**KEYWORDS**

online queries; autistic disorders; parents; machine learning; early detection

## Introduction

The increasing rate of diagnosed autism spectrum disorders (ASD) [1] along with the public's growing awareness of the disorder leads more parents to suspect ASD in their child. These early concerns can arise months before parents decide to approach a professional [2]. As the majority of parents of children with ASD report having symptom-related concerns before the child reaches the age of 3 years [3-5], it stands to reason that reporting these concerns could potentially facilitate earlier evaluation and services. Parents of infants and toddlers seek information about their child's development on online community forums, where they expect to be able to verify or discuss their concerns [6-8]. However, as the resources of health care professionals are scarce, machine learning tools could be applied to perform prescreening of parental forums where ASD concerns are voiced.

On the one hand, parents' online descriptions of their concerns regarding their child's development offer the opportunity to facilitate early referrals; whereas on the other hand, the use of a free-text format makes it difficult to determine the degree of risk for a specific diagnosis. From a public health perspective, online queries are a window into the behaviors that parents recognize as alarming signs of ASD. In turn, this information can be used to design screening procedures to help parents detect signs that are not readily noticeable. Moreover, machine learning tools may offer a way to estimate the degree of ASD risk of the child whose symptoms were described in the online query. The goals of the study were two-fold: (1) To characterize the symptoms mentioned in online queries posed by parents who suspected that their child might have ASD and categorize queries according to the level of clinical risk and the age of the child; and (2) to test the efficacy of machine learning tools in classifying the child's risk of ASD based on the parent's narrative.

The entire process from the time the parents first suspect developmental problems until an ASD diagnosis is obtained can take years [2,9,10]. Studies have demonstrated that the time between parents' first concerns and the first professional ASD consultation was between 5 and 8 months on average [2,9,10], and the average time from concern to diagnosis was more than 32 months [2,11]. ASD is behaviorally diagnosed, based on Diagnostic and Statistical Manual of Mental Disorders (DSM-V) [12] criteria describing symptoms in the domains of social-communication and repetitive and restricted behaviors and interests (RRBI). Of all neurodevelopmental disorders, early identification of ASD is particularly difficult, for the following reasons: (1) the range of what is considered typical, healthy social-communication development is very wide; (2) certain ASD symptoms such as language patterns cannot be detected and evaluated before the child reaches the age of 2 years; (3) diagnosis depends on clinical expertise rather than on a biological marker; (4) early signs of ASD and other neurodevelopmental disorders partly overlap leading to complex diagnostic issues; and (5) ASD screening tools show moderate sensitivity and specificity. Consequently, parents are often left to cope with their worries, long before a diagnosis can be made, which in turn increases their urge to approach online

communities. The long-term goal of this study was to devise a computerized tool based on the parents' narrative of concerns that could estimate their child's ASD risk.

An algorithm for systematic screening of ASD, devised by the American Association of Pediatrics, is used by health care professionals to elicit parental concerns regarding their child's development to calculate a cumulative ASD-risk score and conduct closer monitoring of children who were found to be at risk [13]. Thus, parents' early concerns, which have been validated against ASD screening tools and subsequent diagnosis [14,15], constitute a source of valuable clinical information and often initiate the diagnostic process. Parental concerns are relatively easy to elicit, time-efficient, and bypass the challenges inherent in direct testing, which does not always reflect the child's true skill level. Notwithstanding these advantages, the widespread concerns of parents of young children pose a challenge for the differential prediction of ASD. In addition, parents may be biased in identifying certain signs, due to their emotional state, beliefs, and developmental knowledge. The validity of parental concerns is a subject of continuous research, and has implications for the clinical interpretation of parental concerns.

Evidence shows that the number and type of early parental concerns are predictive of a later developmental disorder [16] and specifically of ASD [15-18]. Types of concerns are classified as ASD-specific (ie, related to core ASD symptoms) and non-ASD-specific concerns. Research has shown that both types of concerns are associated with an eventual ASD diagnosis (reported in more than 18% of cases). Predictive ASD-specific concerns related to communication, language, social-emotional responses, and stereotyped behaviors; whereas non-ASD-specific concerns related to behavior or temperament, regression of skills, medical problems, or delay in milestones [9,10,15]. Studies of siblings of children with ASD indicate that the most frequent or first concern of parents involves communication development [9,15,19-21]. In a large retrospective study of parental concerns in ASD, social-emotional concerns (including nonverbal communication) were the most prevalent, followed by language and RRBI; however, non-ASD-specific concerns were mentioned as well by at least half of the sample [10]. Retrospective parent reports of motor problems, unusual sensory and repetitive behaviors, atypical play patterns, and behavioral problems differentiated children later diagnosed with ASD from those diagnosed with other developmental disabilities [22]. However, in another study [21], parental concerns unique for children with ASD were challenging behaviors and attention problems. Parental concerns that led to a differential diagnosis of atypical development (vs ASD) were related to motor and communication problems. Looking at the presence of a combination of concerns showed that parental concerns about behavioral problems and about cognitive delay in the absence of concerns about communication were not likely predictors of an ASD diagnosis [23]. This evidence underscores the need for an automated system, which can capture the combination of concerns in risk determination. This study analyzed the ASD- and non-ASD-specific signs reported online by parents who suspected ASD warning signs in their child's development.

Some studies associate the age of the child when the parents' concerns are first aroused with the type of signs and their predictive validity. Studies describing parental concerns in families at high risk for an ASD diagnosis concluded that the number of ASD-specific signs noted by concerned parents of 12-month-old children is a better predictor than the number of signs noted by concerned parents of 6-month-old children [15]. Research indicates that although communication was the most frequent first concern, it served to differentiate children with ASD from those with other disorders only among 24-month-old children [3]. Evidence shows that a second type of parental concern expressed frequently is in the behavior or temperament domains at 14 and 24 months and both behavior or temperament and social development domains at 36 months. This differentiation at 24 months and not earlier is in line with another study [17]. Non-ASD-specific concerns (in 53% of children) related to motor problems, anxiety, tantrums, and hyperactivity were associated with earlier parental awareness, whereas ASD-specific concerns, including social withdrawal, abnormal gaze, and poor social interaction, were associated with later parental concerns [10]. In this study, the types of concerns that alert parents of ASD were compared among different child age groups.

To summarize, the reviewed evidence supports the working hypothesis that early parental concerns from the child's second year of life predict later ASD diagnosis. However, most of the studies were based on retrospective reports of parents with children with ASD [10,24], or prospective reports in a high-risk sample [3,15,17,19,22,25]. In high-risk samples, parental concerns represent cases in which there is a truly elevated likelihood of ASD in a younger sibling, and they are already on alert for specific signs [26]. Previous evidence relied on parental responses to structured questions about early concerns in specific areas, sometimes followed by a textual description of the concern mentioned [21]. The Internet reflects the distribution of spontaneous parental ASD concerns in the general population, thus providing access to the data long before they are reported to a professional. The availability of such data makes it possible to explore the potential construction of an automated risk indicator, based on free-text descriptions.

There are several projects that use technology to enable automated early identification of developmental problems, including ASD. For example, the Modified-Checklist for Autism in Toddlers-Revised (M-CHAT-R/F), comprising an ASD screening questionnaire and follow-up interview, has been implemented electronically. The M-CHAT-R/F has proved efficient in lowering both false-positives and negatives compared with paper-format screening [27]. In another study, the online Ages and Stages Questionnaire screener was comparable with the paper version [28]. These studies support the reliance on an Internet-based screening platform, but they do not look at free-text analysis of parental concerns emerging prior to engaging in an ASD-specific screening process.

Regardless of ASD, parents, especially of young first-borns, frequently seek health- and developmental-related information online [6,29]. The Internet was identified as parents' third routine source for obtaining health information [6]. The Internet offers an anonymous round-the-clock platform for expressing

concerns. The increased trend in parents' online information seeking is related to sociodemographic changes, such as living at a geographical distance from parents, decrease in support from family and friends, information from the previous generation is perceived outdated, and there is a greater demand for experience-based information [8]. The clinical quality of answers provided online to parents suspecting ASD in their child varies greatly [30]. There is a need to develop online systems to support parents in interpreting their young child's behavior, so as to validate concerns and offer further guidance to parents when appropriate, while also minimizing the risks of acting solely on nonprofessional online advice.

Machine learning tools [31] have been previously applied for predicting health-related conditions from text, but never for predicting ASD. We note that online forums pose a challenge for analysis, given their unstructured nature and users' descriptions of their symptoms in nonmedical terms. De Choudhury et al [32] investigated the ability to detect clinical depression from social media postings. Search-engines queries were used to predict mood disorder episodes [33]. In another study, adolescents at risk of being bullied were identified using online texts from MTV's A Thin Line project. Thousands of teenagers' online posts were used to develop an algorithm for identifying offensive cyber bullying, based on topic modeling methods [34]. This study built upon these works to test the possibility of devising an automated ASD-risk estimator online, based on large amounts of unstructured textual data and a machine learning algorithm.

The remainder of the paper is organized as follows. The Methods section describes the nature of the sampled queries, coding procedures, and data analysis steps. The Results section describes the types of signs mentioned by parents and their comparison between ASD risk groups and age groups. This section ends with the description of automated prediction of ASD risk from the text. The Discussion section lays out the interpretations and implications of the study.

## Methods

### Sample

This study utilized the Yahoo Answers platform to examine queries of parents suspecting their child has ASD. On the Yahoo Answers platform, queries are posted using natural language, answers are submitted by users, and a community forms around this interaction. A query can elicit multiple answers; one of them is rated the best answer, either by the asker or by the community. Yahoo Answers queries—rather than search-engine queries—were selected for the purpose of this study, as they consist of anonymously posted queries on a public platform and hence are more likely to represent parents' true need for an answer.

We extracted all English-language queries from Yahoo Answers that were submitted between 6 June, 2006 and 12 December, 2013 and contained the words *autism*, *Asperger*, *ASD*, or *PDD*. A total of 8681 queries met these criteria. We used crowdsourcing [35] (using CrowdFlower) to differentiate between queries posted by parents of a child diagnosed with

autism (n=2412), queries posted by parents of a child diagnosed with autism whose sibling they suspected might have autism (n=41), queries posted by parents who suspected their child might have autism (n=1081), and queries posted by parents who did not match any of the above descriptions (n=5147).

Of the 1081 queries of suspecting parents, 195 were randomly selected and analyzed for this study. Among these, in 96.4% (188/195) of the queries that mentioned the age of the child, it was, on average, 3.2 years (SD 2.9; range 1.25 months-18 years; 60.3% [114/189] below 3 years of age). Boys were the subject of 84.7% (160/189) of the queries, girls were the subject of 15.3% (29/189), and the gender of the child was not specified in the rest.

In 4.6% (9/195) of the queries, parents reported a family history of ASD. In 30.8% (60/195) of the queries, parents did not mention reporting their concern to a health care provider.

### Procedures

The content analysis of Yahoo Answers queries was conducted using the NVIVO software. Two types of content analysis processes were applied:

First, one type of content analysis was used to rate a child's risk of ASD as either low, medium, or high. To this end, a set of rules was devised for defining levels of ASD-risk from text. Medium-risk was defined by concerns related to one type of ASD-specific sign, general description of developmental delay, non-ASD-specific concern requiring evaluation and mentioning a risk factor for ASD such as a family member with ASD but no ASD-specific sign. For example:

*My son is 19 months old & sometimes flaps his arms when excited or dancing, could he be autistic? he was born 3 months early & i do not know if this is normal behavior in a toddler or not...please help!*

High-risk was defined as concerns related to at least two types of ASD-specific sign, 1 from the RRBI domain and another from the Social and Communication domains. High-risk rating also considered the severity of the described signs and urgency expressed by the parent:

*My son is 2 1/2 years old and he can count to 15 and sing the whole abc song but he is not speaking with meaning asking me for things like juice and so. He also repeats long sentences from cartoons all day long but has nothing to do with what he is doing at that moment...*

Low-risk was defined, by default, as queries that did not meet the above criteria, for example:

*My child has a bent index finger on both hands. Sometimes it straightens out does this mean she has autism?*

Then, 2 clinical experts in ASD separately rated the risk level of children described in 38 (19.49%) queries and reached kappa of .72 in their differentiation of at risk queries. Finally, ASD risk in the remaining queries was rated by 1 clinical expert.

Second, another process of content analysis was conducted to identify the types of warning signs noted by parents. This

process, which involved deductive and inductive analysis methods, was conducted by a different clinical expert. The deductive method implied coding the warning signs according to domains and subdomains that match DSM-V criteria for ASD. An inductive method was used to identify concerning signs that did not correspond to the DSM-V diagnosing criteria such as describing cognitive impairment or language delay. The resulting taxonomy of the concerning signs mentioned in parents' queries consisted of 12 domains and 72 subdomains (Multimedia Appendix 1). The first author and clinical expert coded 38 queries and obtained an inter-rater agreement as measured by significant kappa values between .54 and 1 across subdomains, with .71-1 values for the 12 domains. Due to challenges in agreeing on the number of manifestations pertaining to a single type of warning sign, each subdomain of sign (ie, the lowest hierarchy level within a subdomain) was coded for its presence in the query, regardless of the number of times that type was mentioned in the query (eg, "rocks, swings and sways body" is coded once for rocking). While coding, the clinical expert and the first author discussed coding dilemmas and refined the coding rules accordingly. To summarize, each query received an ASD global risk score and was coded for either presence or absence of each sign domain and its subdomains.

### Data Analysis

Data regarding the age and gender of the child were extracted using a combination of automated text analyses, followed by crowdsourcing rating to correct for errors. The sample with available age data was divided into 4 age groups: 0-2 (n=62), 2-3 (n=52), 3-6 (n=58), and  $\geq 6$  years (n=18).

Only signs with an occurrence of more than 5% were included in the analyses. Alpha level (Type 1 error) was corrected using Bonferroni for multiple comparisons, such that the threshold  $P$  value (.05) was divided by the number of comparisons. The sign domains differentiating the 3 ASD risk groups were analyzed using chi-square tests, given their dichotomous nature and Fisher's exact tests for pairwise comparisons. The null hypothesis was that the distribution of sign domains across groups is equal. The association between the child's gender and age group and the ASD-risk level was determined using chi-square tests. The length of a query differed significantly between ASD-risk groups ( $F_{2,192}=10.93, P<.001$ ; mean number of words associated with each risk level was as follows: low=132.57, medium=225.32, high=281.62). Bonferroni post-hoc tests indicated that the number of words was significantly ( $P<.05$ ) higher in the high-risk group relative to the low-risk and medium-risk groups.

Data analysis was then conducted to assess whether children at risk of ASD could be detected from the queries using the text-coding method devised. The goal of this analysis was to differentiate between low-risk, medium-risk, and high-risk queries. To this end, each query was analyzed to determine (1) the number of times each word or word pair (bigram) appeared in the text, and, separately, (2) the warning signs, coded according to domains and subdomains, as previously explained.

Additional attributes analyzed included the child's age and gender, the number of words in the query, as well as the length

of the query, as measured by the number of characters in it. We trained a linear classifier [31] (with a least-squares criterion) and estimated its performance in terms of the receiver operating curve (ROC) using Leave-One-Out, that is, for each query  $i$ , we trained a classifier using all other queries, and tested on the  $i$ -th query [36].

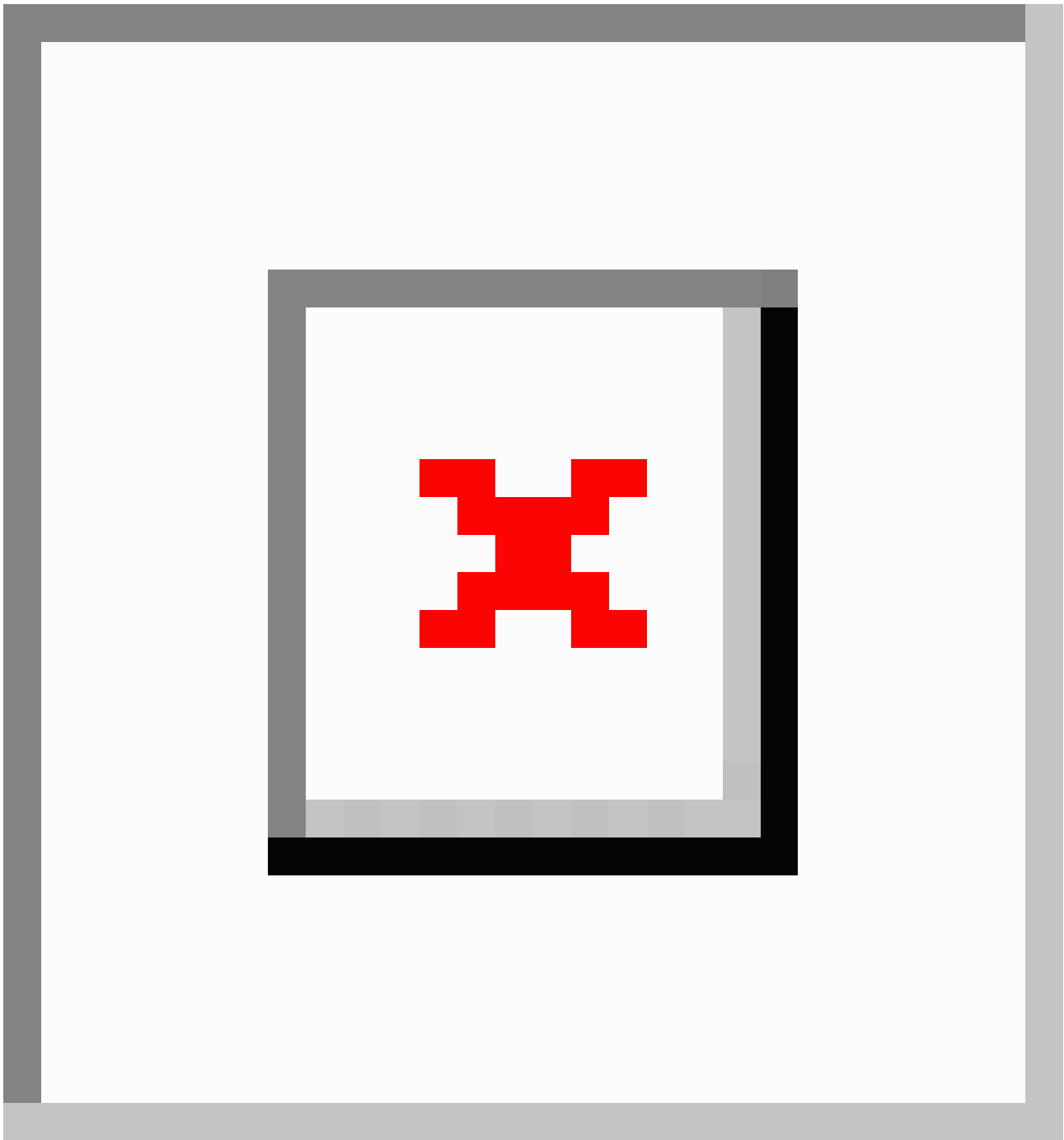
## Results

More than one third of the concerns were in ASD-specific domains: RRBI 75.4% (147/195), social 48.21% (94/195), and communication 42.05% (82/195); as well as non-ASD-specific domains: language 61.5% (120/195), emotional 50.3% (98/195), and cognitive 26.7% (52/195). Other sign domains that were mentioned in at least 5% of the queries were attention deficit hyperactivity disorder (ADHD) 18.97% (35/195), medical conditions 15.9% (31/195), motor 12.3% (24/195), activities of daily living (ADL) 11.8% (23/195), eating 8.7% (17/195), and sleeping problems 6.7% (13/195). The distribution of the types of warning signs mentioned in queries is presented in [Figure 1](#).

The 4 most prevalent types of signs were: repetitive movements (40.5% [79/195]), speech delay (36.4% [71/195]), sensory issues (34.9% [68/195]), and difficulties making friends (30.8% [60/195]).

The distribution of queries among the 3 ASD-risk levels, as determined by clinical experts, was: low-risk (n=35), medium-risk (n=60), and high-risk (n=100). The distribution of sign domains mentioned in queries is presented in [Table 1](#) according to the level of ASD risk; results from chi-square tests comparing groups are also shown. Fisher's exact pairwise comparisons reflected the significant difference between the high-risk and the other 2 groups in terms of ASD-specific sign domains. A surprising finding was the lack of significant difference in the types of ASD-specific concerns mentioned in low versus medium ASD-risk groups. There were significantly fewer ( $P<.001$ ) language concerns in the low-risk group than in the other groups, which did not differ with regard to this domain. Exploratory analysis indicated that within each domain there were individual signs that contributed to the domain differences between the groups ([Table 1](#)).

**Figure 1.** Percentage of domains and sub-domains of warning signs mentioned in  $\geq 5\%$  of queries. RRBI: repetitive and restricted behaviors and interests; ADHD: attention deficit hyperactivity disorder; ADL: activities of daily living



**Table 1.** Distribution of domains of warning signs mentioned in queries presented according to ASD-risk levels.

Domain of signs	Risk group, n (%)			$\chi^2$	Differentiating subdomains <sup>a</sup>
	Low-risk (n=35)	Medium-risk (n=60)	High-risk (n=100)		
RRBI <sup>b</sup>	15 (42.9) <sup>c</sup>	37 (61.7) <sup>c</sup>	95 (95.0) <sup>d</sup>	46.78 <sup>d</sup>	1.3 Sensory issues 1.3.1 Over-responsivity 1.3.1.3 Tactile sensitivity 1.2 Stereotyped or repetitive use of objects 1.4 Unusual and narrow interests 1.1.13 Stimming 1.5.1. Idiosyncratic language
Social	3 (8.6) <sup>c</sup>	18 (30.0) <sup>c</sup>	73 (73.0) <sup>d</sup>	54.61 <sup>d</sup>	2.1 Difficulties in making friends, 2.1.3 Social play
Communication	9 (25.7) <sup>c</sup>	15 (25.0) <sup>c</sup>	58 (58.0) <sup>d</sup>	21.43 <sup>d</sup>	
Language	8 (22.9) <sup>c</sup>	40 (66.7) <sup>d</sup>	72 (72.0) <sup>d</sup>	27.42 <sup>d</sup>	
Emotional	14 (40.0)	29 (48.3)	55 (55.0)	2.46	
Cognitive <sup>a</sup>	3 (8.6)	13 (21.7)	36 (36.0)	11.08	
Medical conditions	5 (14.3)	11 (18.3)	15 (15.0)	0.40	
Motor <sup>e</sup>	1 (2.9)	6 (10.0)	17 (17.0)	5.23	
ADL <sup>f</sup>	2 (5.7)	7 (11.7)	14 (14.0)	1.71	
ADHD <sup>g</sup>	2 (5.7)	11 (18.3)	24 (24.0)	5.66	
Sleeping <sup>e</sup>	2 (5.7)	2 (3.3)	9 (9.0)	2.00	
Eating	3 (8.6)	3 (5.0)	11 (11.0)	1.70	

<sup>a</sup> $P \leq .001$ .

<sup>b</sup>RRBI: repetitive and restricted behaviors and interests.

<sup>c,d</sup>Risk groups with different subscripts differed significantly in Fisher's exact pairwise comparisons. For the significantly different domains, the signs differentiating risk groups were determined using chi-square tests with  $P < .001$ . Note that all hierarchies of signs with 5% occurrence and above were analyzed.

<sup>e</sup>Warning signs pertaining to these domains were observed in less than 5% of the queries.

<sup>f</sup>ADL: activities of daily living.

<sup>g</sup>ADHD: attention deficit hyperactivity disorder.

Looking at the association between the child's age and level of risk indicated that the percentage of queries from each risk group did not differ between the 4 age groups ( $\chi^2_6 = 11.39$ ,  $P = .08$ ). There was a significant difference in the frequency of parents reporting language-related signs (see [Figure 2](#) for pairwise comparison results between age groups using Fisher's exact tests). Very few parents reported language signs in the oldest age group versus the other groups (16.7% [3/18] relative to 59.7% [37/62] to 75.0% [39/52] in the other age groups,  $\chi^2_3 = 20.87$ ,  $P < .001$ ). Note that no significant difference was found in terms of the percentage of queries pertaining to boys versus girls in each risk group (eg, high-risk was 50% [80/160] and 58.6% [17/29], respectively,  $\chi^2_2 = 6.17$ ,  $P = .04$ ).

Next, we tested the efficacy of an automated, text-based ASD-risk estimator. When distinguishing high-risk queries from low- and medium-risk queries, the Area under the ROC curve (AUC) was found to be 0.67 (0.50-0.78). The AUC for this task

using coded signs was 0.82 (0.80-0.86; see [Figure 3](#)). The latter result is significantly better than that obtained using the text alone ( $P = .002$ ). Using both textual descriptions and coded signs reduced the AUC compared with using coded signs alone, probably because of the high dimensionality of the data, relative to the number of queries. Distinguishing low-risk from medium- and high-risk queries, the AUC using text was 0.54 and using signs was 0.84. Thus, text was a poor predictor for this task, compared with both signs and to the low- and medium-risk versus high-risk text classification.

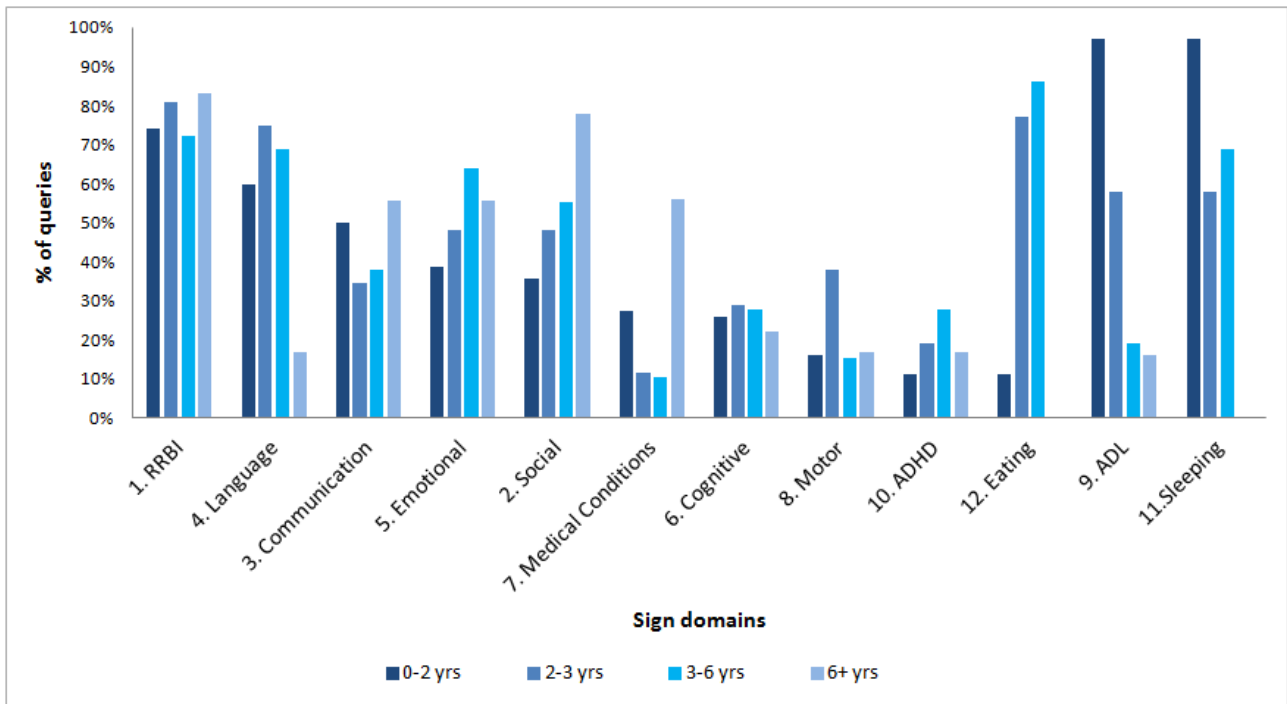
Finally, we created a regression model to predict the actual risk score from the text and (separately) from signs. The Spearman correlation using text was .29 ( $P < .001$ ), whereas the same using signs was .61 ( $P < .001$ ). Thus, the actual risk score can also be deduced with higher accuracy from signs than from the text.

[Figure 4](#) represents the decision-tree classifier for distinguishing low- and medium-risk from high-risk queries. Each node shows

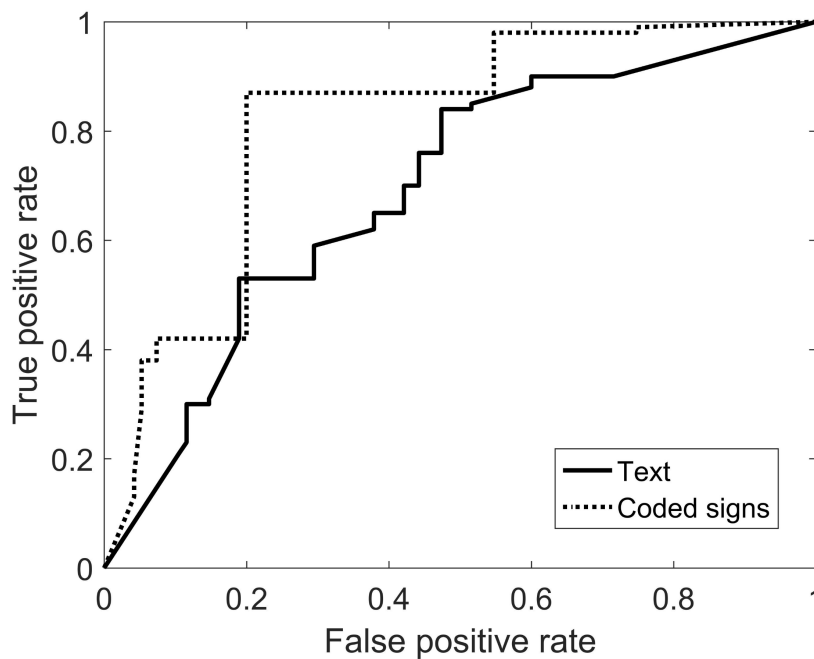
the coding variable used for decision and the fraction of high-risk queries at the node. The numbers at the end of the branch indicate the likelihood of high-risk for that branch. Social, RRBI, communication, cognitive, and motor concern domains entered the final model. Children who were mentioned

to have a social problem had a 78% (73/94) chance of being at high-risk, compared with a 27% (27/101) chance in the rest of the sample. If the parent also reported an RRBI and motor delay, they had a 100% (13/13) chance of being in the high-risk group.

**Figure 2.** Percentage of sign domains mentioned in queries by age groups. RRBI: repetitive and restricted behaviors and interests; ADHD: attention deficit hyperactivity disorder; ADL: activities of daily living.

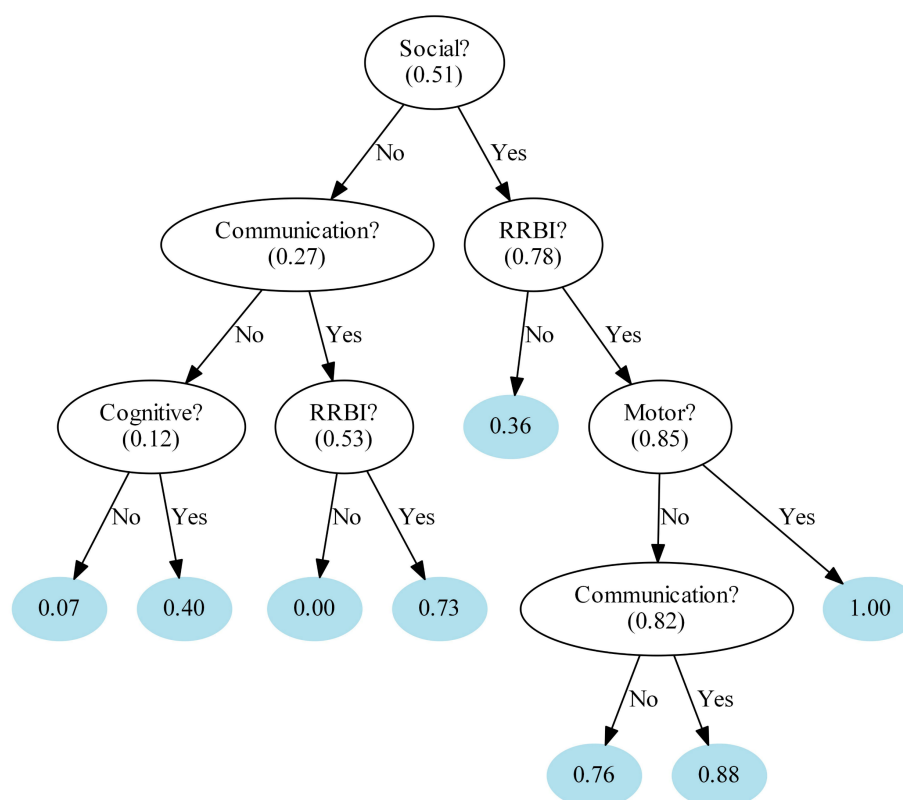


**Figure 3.** Receiver operating curve (ROC) plots predicting risk from text versus coded signs.





**Figure 4.** Decision tree classifier for distinguishing low-risk queries from medium- and high-risk queries. RRBI: repetitive and restricted behaviors and interests.



## Discussion

### Principal Findings

Our study examined the nature of online queries of parents who were concerned that their child might have ASD. The Internet offers parents a venue for expressing and verifying their concerns anonymously at any time and place. The analysis of these narratives highlights signs that alert parents, in the general public, of the possibility of their children having ASD. These concerns mirror parental developmental knowledge, awareness, and expectations, as well as levels of parenting anxiety. Most of the Yahoo Answers queries were judged by clinical experts as reflecting high-risk for ASD or as medium-risk, validating parents' concerns. Differentiating ASD early in life is important, so that the child with ASD can gain the most from targeted interventions [37]. Parental concerns expressed online offer a new method for facilitating earlier screening and referral for evaluation. In the long run, an online tool which provides a gross estimate of ASD risk based on textual descriptions of warning signs and guided questions could prompt parents to approach a professional. Such a tool can harness social media to support worried parents and minimize the risk of acting on nonprofessional advice or disregarding worries.

### Comparison With Prior Work

Our findings that parents with mostly no family history of ASD (95.4% [186/195]) associated a broad range of ASD-specific signs with the disorder were encouraging. The prevailing signs that parents found worrisome were within the DSM-V [12] ASD core domain of RRBI, followed by the social-communication domain. Nevertheless, more than a third of the parents sampled,

expressed concerns related to domains of language, cognitive, and emotional, which according to the DSM-V are non-ASD-specific; however, they are highly prevalent in ASD. Other non-ASD-specific concerns mentioned in some of the queries related to motor development, ADL, and medical conditions. Interestingly, concerns related to motor problems entered as a meaningful domain in the decision-tree classifier. The mix of concerns from ASD and non-ASD-specific domains is consistent with previous studies documenting the types of first parental concerns of children later diagnosed with ASD [9,10,15].

The average child age at which an online concern was raised was 38.03 months, close to the average age of ASD diagnosis [38,39]. Nonetheless, this age is greater than the average age of the first concerns that was reported in ASD research, which is 10-18 months [5,9,19,21,24]. The majority of queries described sons, although analysis of queries did not reveal different types of signs for sons versus daughters, and the rate of ASD risk was not significantly higher for boys than for girls. Comparisons between age groups showed that language concerns differed in prevalence across age groups. Language problems were mentioned most frequently in queries regarding children 0-3-year-old and less in the oldest group. It is likely that parents of a nonverbal 7-year-old have reached the stage beyond suspicion and hence would be less likely to query this. Note that the peak of emotional concerns was in the 3-6 years old age group, particularly descriptions of outbursts or extreme shyness. Social concerns increased with highest prevalence in the older group. These differences between age groups can be explained in the light of the changes in developmental expectations of parents from children at different ages.

At the extremes, there were parents who raised concerns too early to determine risk, for example:

*My 8-week-old son has yet to flash his first true smile. I know that all babies develop differently, but I am worried about autism...*

At the other end of the spectrum, there were parents of older children who were either questioning a non-ASD diagnosis their child received or were never evaluated but always felt something was different, for example:

*(regarding an 18-year-old child) we always knew he was different, he displays some—but not all—of the common symptoms, but we just put it down to him being different and an introvert...apart for the virtual world of computer games he is turning into a hermit...and has no social skills.*

When designing an online screening tool, the age of the child must be considered and, based on the findings, the need to continue monitoring such a child is warranted.

Although there were differences in rates of RRBI concerns across risk groups, they were mentioned in at least 75.4% [147/195] of the queries. The RRBI domain included the largest number of hierarchy levels of subdomain coding as well as individual signs (the subdomains are: repetitive movements, stereotypes and repetitive use of objects, sensory issues, unusual and narrow interests, repetitive speech, eating, difficulty with change, rigid thinking, and rituals). This reflects the diversity in types of symptoms described in the DSM-V criteria. The most prevalent RRBI mentioned were repetitive movements, repetitive use of objects, and sensory abnormalities. The RRBI characterizing queries of medium- or high-risk were repetitive speech, sensory issues (particularly tactile over-reactivity), unusual use of objects, repetitive speech (particularly idiosyncratic language), and repetitive interest. Interestingly, research shows that RRBI are not the most prevalent first concerns of parents of children later diagnosed with ASD [18,22]. It may be the most reported domain in online queries, as parent's attention is more easily drawn to atypical behaviors or socially inappropriate behaviors, compared with their ability to recognize a delay in attaining a milestone. Signs within this domain can be intense and can interfere with play and participation, and thus are noticeable. The presence of some RRBI in typical development (eg, head banging, noise sensitivity) presents a further challenge for relying on this domain to verify risk status. There is a need to increase parental awareness of the typical manifestations of RRBI during the first 3 years of life, to help parents understand when such concerns may be warranted. Online concerns reflect a parent's call for help regardless of whether a child has ASD. Therefore, there is clearly a need for interactive parenting education materials aimed at interpreting and coping with RRBI.

Our results indicated that it is possible to predict risk from the text using machine learning methods once the text is classified into sign domains, whereas using text alone provided insufficient information (at least in our corpus) for accurate identification of children at risk. Developing an ASD-specific flowchart into which parents could insert their narratives related to certain types of concerns may provide a basis for a more accurate

prediction of ASD risk from text. An automated screening tool in online forums will benefit from starting with a social concern question, and if not present then ask about communication while if present ask about the presence of RRBI. While ASD specific questions will need to dominate such a tool, probing about cognitive and motor markers, which are not ASD-specific, is also important. Results from the decision tree indicate that the combination of signs from the social, RRBI, and motor domains predicted the highest likelihood of ASD risk from the coded text. This is in line with the evidence showing that parental concerns pertaining to a combination of several domains predicted an ASD diagnosis [3]. The prediction of high risk from text or text combined with coded signs was better for the high ASD risk group alone rather than predicting for both medium- and high-risk groups. Future research relying on a larger corpus could robustly test different combinations of concerns in predicting ASD risk. As parents seek first-level support online, developing Internet- and mobile-health tools to automate ASD screening relying on the decision-tree classification described in this study may reduce screening time, increase response to screening, and increase accuracy.

The distribution of the types of signs mentioned in online queries did not fully correspond to the signs most commonly reported in previous studies. For instance, in an ASD study, parental language and communication concerns were found to be the most prevalent early concerns, followed by social, RRBI, medical, and emotional domains [15]. Communication signs were highly prevalent among parental concerns noted in other studies as well; however, they did not consistently differentiate those later diagnosed with ASD [3,9,10,17,19,20,22]. These differences in the most common ASD concerns of parents may be explained by the fact that previous studies relied on retrospective reports and samples characterized by high genetic risk for ASD [3,15,19,22,25], as opposed to a sample with little familiarity with ASD symptomatology and a lower likelihood for the child to have ASD.

### Limitations

The limitation of the current exploratory study is the lack of clinical testing for the actual ASD-risk status of the child as opposed to other neurodevelopmental disorders. The next step would be to study the external validity of risk status using standardized developmental measures, develop a structured format for parents to enter their concerns and, test our algorithm for predicting risk from text with a new corpus. Extracting signs from text has its own limitations, as in some cases it requires the clinical inference of meaning, as opposed to the ability to probe a parent and thus extract the meaning behind the concern. We attempted to minimize such bias by developing coding rules and testing coding reliability. The anonymous nature of online queries makes online parental concerns a unique resource that offers an authentic snapshot of parental ASD-related concerns, unaffected by issues of social desirability or other emotional biases.

### Conclusions

Early parental concerns constitute a valuable component of early childhood screening. There is accumulating evidence that early parental concerns regarding specific ASD markers are

associated with a higher likelihood of an eventual ASD diagnosis [3,15,40]. Our study is the first to investigate online queries describing the types of signs that lead parents to suspect ASD in their child. The fact that the clinical experts found that the majority of queries corresponded to either medium- or high-risk for ASD validated the need to facilitate the parents' earlier consultation with a professional. We showed the potential of utilizing machine learning methods for ASD screening based on parental concerns. Findings also highlight the need for designing parent-education tools regarding behaviors that are age appropriate, particularly those pertaining to the RRBI domain. Finally, it is important to empower parents' confidence in their concerns and increase their awareness of the

disadvantages of relying solely on an online community for determining ASD-risk status. This study's findings support the call for health care providers to closely listen to parental ASD-related concerns, as recommended by screening guidelines [13]. Results also demonstrate the need for Internet-based screening systems that utilize parents' narratives combined with a hierarchical screening questioning. Worried parents approach online communities, comprised mostly lay person answers, to obtain opinions regarding their child's likelihood of having an ASD diagnosis. A more efficient mechanism for supporting worried parents online is important for prompting a clinical evaluation when needed and reducing parental anxiety.

## Acknowledgments

This study was funded by a Caesarea Rothschild Institute grant. ABS was funded by a Y Alon Fellowship. We thank Yael Schwartz-Klein and Neta Katz for their assistance with text analysis in this study and Danielle Wolinski for her assistance with manuscript preparation.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Taxonomy domains and sub-domains of warning signs coded from Yahoo queries.

[\[PDF File \(Adobe PDF File\), 43KB-Multimedia Appendix 1\]](#)

## References

1. French LR, Bertone A, Hyde KL, Fombonne E. Epidemiology of autism spectrum disorders. In: Buxbaum JD, Hof PR, editors. *The Neuroscience of Autism Spectrum Disorders*. Amsterdam: Elsevier/AP; 2013:3-24. .
2. Siklos S, Kerns KA. Assessing the diagnostic experiences of a small sample of parents of children with autism spectrum disorders. *Res Dev Disabil* 2007;28(1):9-22. [doi: [10.1016/j.ridd.2005.09.003](https://doi.org/10.1016/j.ridd.2005.09.003)] [Medline: [16442261](https://pubmed.ncbi.nlm.nih.gov/16442261/)]
3. Hess CR, Landa RJ. Predictive and concurrent validity of parent concern about young children at risk for autism. *J Autism Dev Disord* 2012 Apr;42(4):575-584 [FREE Full text] [doi: [10.1007/s10803-011-1282-1](https://doi.org/10.1007/s10803-011-1282-1)] [Medline: [21584850](https://pubmed.ncbi.nlm.nih.gov/21584850/)]
4. Jónsdóttir SL, Saemundsen E, Antonsdóttir IS, Sigurdardóttir S, Ólason D. Children diagnosed with autism spectrum disorder before or after the age of 6 years. *Res Autism Spectrum Disord* 2011;5(1):175-184. [doi: [10.1016/j.rasd.2010.03.007](https://doi.org/10.1016/j.rasd.2010.03.007)]
5. Kishore MT, Basu A. Early concerns of mothers of children later diagnosed with autism: implications for early identification. *Res Autism Spectrum Disord* 2011;5(1):157-163. [doi: [10.1016/j.rasd.2010.03.005](https://doi.org/10.1016/j.rasd.2010.03.005)]
6. Khoo K, Bolt P, Babl FE, Jury S, Goldman RD. Health information seeking by parents in the Internet age. *J Paediatr Child Health* 2008;44(7-8):419-423. [doi: [10.1111/j.1440-1754.2008.01322.x](https://doi.org/10.1111/j.1440-1754.2008.01322.x)] [Medline: [18564080](https://pubmed.ncbi.nlm.nih.gov/18564080/)]
7. Morahan-Martin JM. How Internet users find, evaluate, and use online health information: a cross-cultural review. *Cyberpsychol Behav* 2004 Oct;7(5):497-510. [doi: [10.1089/cpb.2004.7.497](https://doi.org/10.1089/cpb.2004.7.497)] [Medline: [15667044](https://pubmed.ncbi.nlm.nih.gov/15667044/)]
8. Plantin L, Daneback K. Parenthood, information and support on the internet: a literature review of research on parents and professionals online. *BMC Family Practice* 2009 May 18;10(1):1. [doi: [10.1186/1471-2296-10-34](https://doi.org/10.1186/1471-2296-10-34)]
9. De Giacomo A, Fombonne E. Parental recognition of developmental abnormalities in autism. *Eur Child Adolesc Psychiatry* 1998 Sep;7(3):131-136. [Medline: [9826299](https://pubmed.ncbi.nlm.nih.gov/9826299/)]
10. Guinchat V, Chamak B, Bonniau B, Bodeau N, Perisse D, Cohen D, et al. Very early signs of autism reported by parents include many concerns not specific to autism criteria. *Res Autism Spectrum Disord* 2012;6(2):589-601.
11. Noterdaeme M, Hutzelmeyer-Nickels A. Early symptoms and recognition of pervasive developmental disorders in Germany. *Autism* 2010 Nov;14(6):575-588. [doi: [10.1177/1362361310371951](https://doi.org/10.1177/1362361310371951)] [Medline: [20923894](https://pubmed.ncbi.nlm.nih.gov/20923894/)]
12. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. 5th edition.. Washington, DC: American Psychiatric Association Publishing; 2013.
13. Johnson CP, Myers SM, American Academy of Pediatrics Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 2007 Nov;120(5):1183-1215. [doi: [10.1542/peds.2007-2361](https://doi.org/10.1542/peds.2007-2361)] [Medline: [17967920](https://pubmed.ncbi.nlm.nih.gov/17967920/)]
14. Glascoe FP, MacLean WE, Stone WL. The importance of parents' concerns about their child's behavior. *Clin Pediatr (Phila)* 1991 Jan;30(1):8-11. [Medline: [1995206](https://pubmed.ncbi.nlm.nih.gov/1995206/)]

15. Ozonoff S, Young GS, Steinfeld MB, Hill MM, Cook I, Hutman T, et al. How early do parent concerns predict later autism diagnosis? *J Dev Behav Pediatr* 2009 Oct;30(5):367-375 [[FREE Full text](#)] [Medline: [19827218](#)]
16. Glascoe FP. Early detection of developmental and behavioral problems. *Pediatr Rev* 2000 Aug;21(8):272-280. [Medline: [10922024](#)]
17. Veness C, Prior M, Bavin E, Eadie P, Cini E, Reilly S. Early indicators of autism spectrum disorders at 12 and 24 months of age: a prospective, longitudinal comparative study. *Autism* Jul 6 2012;16(2):163-177. [doi: [10.1177/1362361311399936](#)] [Medline: [21733958](#)]
18. Sacrey LA, Zwaigenbaum L, Bryson S, Brian J, Smith IM, Roberts W, et al. Can parents' concerns predict autism spectrum disorder? A prospective study of high-risk siblings from 6 to 36 months of age. *J Am Acad Child Adolesc Psychiatry* 2015 Jun;54(6):470-478. [doi: [10.1016/j.jaac.2015.03.014](#)] [Medline: [26004662](#)]
19. Chawarska K, Paul R, Klin A, Hannigen S, Dichtel LE, Volkmar F. Parental recognition of developmental problems in toddlers with autism spectrum disorders. *J Autism Dev Disord* 2007 Jan;37(1):62-72. [doi: [10.1007/s10803-006-0330-8](#)] [Medline: [17195921](#)]
20. Young RL, Brewer N, Pattison C. Parental identification of early behavioural abnormalities in children with autistic disorder. *Autism* 2003 Jun;7(2):125-143. [Medline: [12846383](#)]
21. Kozlowski AM, Matson JL, Horovitz M, Worley JA, Neal D. Parents' first concerns of their child's development in toddlers with autism spectrum disorders. *Dev Neurorehabil* 2011;14(2):72-78. [doi: [10.3109/17518423.2010.539193](#)] [Medline: [21410398](#)]
22. Coonrod EE, Stone WL. Early concerns of parents of children with autistic and nonautistic disorders. *Infants Young Child* 2004 Jul 1;17(3):258-268 [[FREE Full text](#)]
23. Turygin N, Matson JL, Williams LW, Belva BC. The relationship of parental first concerns and autism spectrum disorder in an early intervention sample. *Res Autism Spectrum Disord* 2014 Feb 28;8(2):53-60. [doi: [10.1016/j.rasd.2013.10.008](#)]
24. Howlin P, Asgharian A. The diagnosis of autism and Asperger syndrome: findings from a survey of 770 families. *Dev Med Child Neurol* 1999 Dec;41(12):834-839 [[FREE Full text](#)] [Medline: [10619282](#)]
25. Wetherby AM, Brosnan-Maddox S, Peace V, Newton L. Validation of the Infant-Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism* 2008 Sep;12(5):487-511 [[FREE Full text](#)] [doi: [10.1177/1362361308094501](#)] [Medline: [18805944](#)]
26. McMahon CR, Malesa EE, Yoder PJ, Stone WL. Parents of children with autism spectrum disorders have merited concerns about their later-born infants. *Res Pract Persons Severe Disabilities* 2007 Jun 1;32(2):154-160.
27. Harrington JW, Bai R, Perkins AM. Screening children for autism in an urban clinic using an electronic M-CHAT. *Clin Pediatr (Phila)* 2013 Jan;52(1):35-41. [doi: [10.1177/0009922812463957](#)] [Medline: [23117237](#)]
28. Yovanoff P, Squires J, McManus S. Adaptation from paper-pencil to Web-based administration of a parent-completed developmental questionnaire for young children. *Infants Young Child* 2013 Oct 1;26(4):318-332. [doi: [10.1097/IYC.0b013e31829f95f3](#)]
29. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. *J Med Internet Res* 2004 Mar 1;6(1):e7 [[FREE Full text](#)] [doi: [10.2196/jmir.6.1.e7](#)] [Medline: [15111273](#)]
30. Ben-Sasson A, Pelleg D, Yom-Tov E. The quality of online answers to parents who suspect that their child has an autism spectrum disorder. 2016 Mar 31 Presented at: Tenth International AAAI Conference on Web and Social Media; May 17-20, 2016; Cologne, Germany.
31. Duda RO, Hart PE, Stork DG. *Pattern Classification*. New Jersey: John Wiley & Sons; Nov 9, 2012.
32. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. *J Gen Intern Med* 2011 Mar 1;26(3):287-292. [Medline: [20945113](#)]
33. Yom-Tov E, White RW, Horvitz E. Seeking insights about cycling mood disorders via anonymized search logs. *J Med Internet Res* 2014 Feb 25;16(2):e65 [[FREE Full text](#)] [doi: [10.2196/jmir.2664](#)] [Medline: [24568936](#)]
34. Lieberman H, Dinakar K, Jones B. Crowdsourced ethics with personalized story matching. New York, NY Presented at: CHI '13 Conference on Human Factors in Computing Systems; April 27 - May 02, 2013; Paris, France: ACM; 2013 p. 709-714.
35. Lease M, Yilmaz E. Crowdsourcing for information retrieval. 2012 Jan 9 Presented at: ACM SIGIR Forum; August 12-16, 2012; Portland, Oregon, USA p. 66-75.
36. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982 Apr;143(1):29-36. [doi: [10.1148/radiology.143.1.7063747](#)] [Medline: [7063747](#)]
37. Dawson G. Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Dev Psychopathol* 2008 Jun 1;20(3):775-803. [doi: [10.1017/S0954579408000370](#)] [Medline: [18606031](#)]
38. Goin-Kochel RP, Mackintosh VH, Myers BJ. How many doctors does it take to make an autism spectrum diagnosis? *Autism* 2006 Sep 1;10(5):439-451. [doi: [10.1177/1362361306066601](#)] [Medline: [16940311](#)]
39. Mandell DS, Novak MM, Zubritsky CD. Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics* 2005 Dec 1;116(6):1480-1486 [[FREE Full text](#)] [doi: [10.1542/peds.2005-0185](#)] [Medline: [16322174](#)]

40. Talbott MR, Nelson CA, Tager-Flusberg H. Diary reports of concerns in mothers of infant siblings of children with autism across the first year of life. *J Autism Dev Disord* 2015 Jul 1;45(7):2187-2199 [[FREE Full text](#)] [doi: [10.1007/s10803-015-2383-z](https://doi.org/10.1007/s10803-015-2383-z)] [Medline: [25703030](https://pubmed.ncbi.nlm.nih.gov/25703030/)]

## Abbreviations

**ADHD:** attention deficit hyperactivity disorder

**ADL:** activities of daily living

**ASD:** autism spectrum disorders

**AUC:** area under the curve

**DSM-V:** Diagnostic and Statistical Manual of Mental Disorders

**M-CHAT-R/F:** Modified Checklist for Autism in Toddlers-Revised with Follow-Up

**MTV:** music television

**ROC:** receiver operating curve

**RRBI:** repetitive and restricted behaviors and interests

*Edited by G Eysenbach; submitted 16.12.15; peer-reviewed by D Cohen, D Lee, A Venkatakrisnan; comments to author 03.04.16; revised version received 29.06.16; accepted 22.09.16; published 22.11.16*

*Please cite as:*

*Ben-Sasson A, Yom-Tov E*

*Online Concerns of Parents Suspecting Autism Spectrum Disorder in Their Child: Content Analysis of Signs and Automated Prediction of Risk*

*J Med Internet Res* 2016;18(11):e300

URL: <http://www.jmir.org/2016/11/e300/>

doi: [10.2196/jmir.5439](https://doi.org/10.2196/jmir.5439)

PMID: [27876688](https://pubmed.ncbi.nlm.nih.gov/27876688/)

©Ayelet Ben-Sasson, Elad Yom-Tov. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.11.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.