Sex-differences in children referred for assessment: An exploratory analysis of the AMSE

Abstract
The autism mental status exam is an 8-item observational assessment that structures the way we observe and document signs and symptoms of ASD. Investigations of test performance indicate strong sensitivity and specificity using gold-standard assessment as reference standard. This study aims to explore potential sex differences in AMSE test performance and observations of 123 children referred for autism assessment. Results indicates more language deficits in females with ASD than in males with ASD and less sensory symptoms in females compared to males with ASD. The AMSE performance is similar in identifying ASD and non-ASD in females compared to males. Less disruptive behaviors in females, might cause a need for a bigger hit to other areas of development to raise concern.

Introduction
Prevalence studies consistently indicate that ASD is over-represented in boys compared to girls (Elsabbagh et al. 2012; Fombonne 2003; 2007; Fombonne et al. 2011; Loomes et al. 2017). However, the extent of the over-representation and its cause remain the topic of some debate (Werling and Geschwind 2013). Several groups have investigated potential biases inherent in diagnostic practices and measures as a potential cause of the skewed male-to-female ratio (Koenig and Tsatsanis 2005). Several hypotheses have been proposed to explain these sex differences in ASD prevalence. For example, this phenomenon may mirror more general differences between typically developing males and females (Halpern 1997; Zahn-Waxler et al. 2006), with ASD representing an extreme expression of male traits(Baron-Cohen 2002; 2009). The difference in prevalence could also reflect a genetic protective factor in females (Robinson et al. 2013), who may require a greater genetic load to manifest autistic behavioral impairments. (Gilman et al. 2011; Skuse 2007) Another theory suggests that subtle cases of ASD in females might go unrecognized due to average-
range IQ and fewer disruptive behavioral issues than their male peers (Dworzynski, Ronald, Bolton, and Happé 2012a).

For those individuals receiving a diagnosis, the existing literature remains somewhat inconsistent regarding the severity of ASD symptoms in males versus females (Carter, Black, Tewani, Connolly, Kadlec, and Tager-Flusberg 2007a; Kopp and Gillberg 2011; Lai et al. 2012; Mayes and Calhoun 2011; Zwaigenbaum et al. 2012). Studies have revealed that males tend to score higher on indices of externalizing behavior problems (Bolte et al. 2011; Hattier et al. 2011; Mandy et al. 2012; Szatmari et al. 2012), whereas females score higher on internalizing symptoms (Bolte et al. 2011; Mandy et al. 2012; Solomon et al. 2012; Szatmari et al. 2012). Recent research also suggest behavioral differences in the core domains of ASD such as better social skills (Dean et al. 2017; Head et al. 2014), less restricted and repetitive behaviors (Frazier et al. 2014), and joint attention (Øien et al. 2017). The aforementioned studies suggest that separately considering male and female profiles of performance on clinical tools may be helpful in order to more accurately characterized potentially sex-specific deviations from prototypical behavior. However, to date, no widely-used clinician-rated observational tools are designed to take into account sex differences.

The Autism Mental Status Exam (AMSE)(Grodberg et al. 2014) is a brief and free observational tool structuring the observation and documentation of signs and symptoms of ASD; it is designed to be used as a level two screener to be used after developmental concerns have already been raised. Initial test performance studies have indicated excellent sensitivity and specificity of the AMSE in children suspected of ASD, children suspected of ASD and ADHD, and in adults suspected of ASD. (Grodberg et al. 2014; 2012; Øien et al. 2016). The eight items comprise (1) eye contact, (2) interest in others, (3) pointing skills, (4) language, (5) pragmatics, (6) repetitive behaviors, (7) preoccupations, and (8) unusual sensitivities, which are aggregated to yield a total score. The AMSE has been translated into eight
languages, such as Italian and Norwegian, with other languages currently in the process of translation. Test performance has not been investigated cross translations yet, but several studies are currently examining test performance in other languages.

The present study aimed to investigate a) sex differences in observed/reported behaviors in children with ASD and b) the diagnostic ability of the AMSE in males and females suspected of ASD.

**Methods**

**Participants**

Participants (N=123 (28.5% females); mean age 5.74 years (S.D. = 2.88) were recruited from two sites: (1) The Seaver Autism Center for Research and Treatment at Mount Sinai and (2) Cincinnati Children’s Hospital Medical Center. At Mount Sinai, the sample included children receiving comprehensive autism-focused diagnostic evaluations, as part of their Assessment Core protocol from September 2013 through December 2014. The study was approved by the Mount Sinai Program for the Protection of Human Subjects. Informed written consent was obtained from all caregivers and assent was obtained when appropriate. At Cincinnati Children’s Hospital Medical Center, the sample included children receiving comprehensive autism-focused diagnostic evaluations at the Kelly O’Leary Center for ASD at Cincinnati Children’s Hospital Medical Center from April 2014-April 2015. This chart review project was part of a larger ongoing comprehensive assessment of clinical care of patients with ASD, and was approved by the CCHMC Institutional Review Board.
Sex-differences in children referred for assessment: An exploratory analysis of the AMSE

The AMSE is intended to guide clinicians in the context of diagnostic decision-making. Each item is scored on a 0–2 scale with possible total scores ranging from 0 to 14; higher scores reflecting greater symptom severity. Social items must be observed during the clinical examination, but communication and behavioral items may be reported or observed. Three items prompt the examiner to specify whether the item is reported or observed: pragmatics of language, encompassing preoccupations and unusual sensitivities. In these three items, the score is weighted if the item is observed. Scoring instructions for those three items also provide flexibility based on level of functioning. Further investigation of the AMSE’s sensitivity and specificity in detecting an independent DSM-5 (American Psychiatric Association 2014) diagnosis of ASD in adults also revealed excellent psychometric properties (SE: 0.91, SP: 0.93) (Grodberg et al. 2014). An online training curriculum provides the scoring manual and video simulation of clinical examinations based on individuals of varying ages and levels of functioning. In the present study, the total score of the AMSE was used as a measure of severity. The total score of AMSE has been shown to be significantly correlated with the ADOS calibrated severity scale (i.e. the comparison score, $r = .67, p < .001$) (Øien et al. 2016). As the present study had total AMSE score for all participants, it was chosen as a measure for severity.

<table>
<thead>
<tr>
<th>Table 1: Participants Characteristics</th>
<th>Total</th>
<th>Non-ASD</th>
<th>ASD</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>123</td>
<td>38</td>
<td>85</td>
</tr>
<tr>
<td>Male</td>
<td>88</td>
<td>26</td>
<td>62</td>
</tr>
<tr>
<td>Female</td>
<td>35</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>Mean age (S.D)</td>
<td>5.74(2.88)</td>
<td>5.68(2.55)</td>
<td>5.77(3.03)</td>
</tr>
<tr>
<td>Mean of total ASME score (S.D.)</td>
<td>6.43(2.63)</td>
<td>3.63(1.56)</td>
<td>7.68(1.95)</td>
</tr>
<tr>
<td>Mean of ADOS-2 Comparison score (S.D)*</td>
<td>6.74(4.03)</td>
<td>3.34(1.81)</td>
<td>8.21(3.84)</td>
</tr>
<tr>
<td>ADOS classification for ASD*</td>
<td>92</td>
<td>10</td>
<td>82</td>
</tr>
<tr>
<td>Comorbid Intellectual Disability (ID)</td>
<td>23</td>
<td>3</td>
<td>20</td>
</tr>
</tbody>
</table>
Clinical procedure

Mount Sinai Site

Each subject first received a clinical evaluation by a psychiatrist with extensive experience in the diagnosis of ASD. The clinical evaluation included administration of the AMSE. Subjects were then administered an ADOS by an independent psychologist at the Center who was blind to the AMSE score and blind to the psychiatrists’ diagnostic impressions. An ADI-R was administered, when feasible (73.3% of the sample). All ADOS and ADI-R administrations were performed and scored by reliable raters. In order to ascertain clinical diagnosis in a way that was sufficiently independent from the AMSE score, a best estimate clinical diagnosis (BECD) protocol was implemented in which the supervising psychologist at the Center reviewed the full ADOS protocol, the ADI-R protocol, and the psychiatric history, which included the chief complaint, history of present illness, and past psychiatric, medical, and developmental histories. The AMSE was conducted blind to the rest of the assessment, and the trained clinician conducting the AMSE did not get any information about the gold standard assessment, and vice versa.

Cincinnati Site

Each participant first received a clinical evaluation by a psychologist and physician both with extensive experience in the diagnosis of ASD. The clinical evaluation included administration of the AMSE as well as an ADOS that was rated by a clinically-reliable rater. In order to ascertain the clinical diagnosis in a way that was sufficiently independent from the AMSE score, an independent expert clinician with expertise in diagnostic
Sex-differences in children referred for assessment: An exploratory analysis of the AMSE

assessment of ASD, blind to the AMSE score, reviewed all clinical material (excluding the AMSE) and assessed the patient.

Statistical analyses

In order to ensure that groups were comparable, we (a) examined males and females on age, rates of intellectual disability (ID), and rates of total score. One-way ANOVA were used for analysis of continuous variables and Fisher's exact test was used for categorical variables. Our subsequent analyses controlled for any variables for which there was a significant between-sex difference, with the assumption that if no significant differences were found that they would not be entered into the statistical model. In line with aim (a) an ordinal regression examined differences between ASD males and ASD females at item level. In line with aim (b) the diagnostic accuracy of the AMSE was examined by the nonparametric measure of area under a ROC curve for males and females separately. Cohen’s $d$ was used for measure of effect size (Cohen 1988). IBM SPSS 23 software was used for statistical analyses.

Results

No differences between ASD males and ASD females were found on Age ($(F (1, 83) = .433, p = .429, d = .197)$), Severity (Total AMSE score) $(F (1, 83) = .501, p = .481, d = .170)$ or ID $(t(82) = 1.61, p = .145, d = .371)$. Based on these findings, we did not consider statistical controls for these variables in subsequent analyses comparing males with ASD to females with ASD. However, differences were found between non-ASD and ASD on severity (Total AMSE score) $(F (1, 121) = 126.162, p < .001, d = 2.29)$ and ID $(t(120) = 2.1, p = .046, d = .439)$. No significant age differences between ASD and non-ASD subjects emerged $(F (1,121) = .024, p = .878, d = .030)$. 

<table>
<thead>
<tr>
<th>Table 2. Item analysis Autism Mental Status Exam – ASD males vs ASD females</th>
</tr>
</thead>
<tbody>
<tr>
<td>assessment of ASD, blind to the AMSE score, reviewed all clinical material (excluding the AMSE) and assessed the patient.</td>
</tr>
</tbody>
</table>

Statistical analyses

In order to ensure that groups were comparable, we (a) examined males and females on age, rates of intellectual disability (ID), and rates of total score. One-way ANOVA were used for analysis of continuous variables and Fisher's exact test was used for categorical variables. Our subsequent analyses controlled for any variables for which there was a significant between-sex difference, with the assumption that if no significant differences were found that they would not be entered into the statistical model. In line with aim (a) an ordinal regression examined differences between ASD males and ASD females at item level. In line with aim (b) the diagnostic accuracy of the AMSE was examined by the nonparametric measure of area under a ROC curve for males and females separately. Cohen’s $d$ was used for measure of effect size (Cohen 1988). IBM SPSS 23 software was used for statistical analyses.

Results

No differences between ASD males and ASD females were found on Age ($(F (1, 83) = .433, p = .429, d = .197)$), Severity (Total AMSE score) $(F (1, 83) = .501, p = .481, d = .170)$ or ID $(t(82) = 1.61, p = .145, d = .371)$. Based on these findings, we did not consider statistical controls for these variables in subsequent analyses comparing males with ASD to females with ASD. However, differences were found between non-ASD and ASD on severity (Total AMSE score) $(F (1, 121) = 126.162, p < .001, d = 2.29)$ and ID $(t(120) = 2.1, p = .046, d = .439)$. No significant age differences between ASD and non-ASD subjects emerged $(F (1,121) = .024, p = .878, d = .030)$. 

<table>
<thead>
<tr>
<th>Table 2. Item analysis Autism Mental Status Exam – ASD males vs ASD females</th>
</tr>
</thead>
<tbody>
<tr>
<td>assessment of ASD, blind to the AMSE score, reviewed all clinical material (excluding the AMSE) and assessed the patient.</td>
</tr>
</tbody>
</table>
As concerns aim (a) ordinal regression analyses showed that individuals with ASD were more likely to score higher on all items except pragmatics, indicating more symptoms than non-ASD individuals. ASD females were more likely to score higher than ASD males on the language item of the AMSE. However, ASD females were less likely to score higher (e.g. more symptoms) than ASD males on the oversensitivity item of the AMSE. To reiterate, ASD males and ASD females generally expressed similar levels of impairment compared to non-ASD peers. Comparing ASD females to ASD males revealed greater language impairment in ASD females. However, ASD females were less likely to have increased or unusual sensitivities or heightened threshold for pain.

As concerns aim (b), ROC curve analysis was used to determine the optimal cutoff for the AMSE compared with the BECD based on *DSM-5*. For males, sixty-two participants (70.4%) met BECD for ASD using *DSM-5* criteria. The area under the ROC curve (AUC) was .95 (95% CI = [.913, .992]; Figure 1). For females, twenty-three participants (65.7%) met BECD for ASD using *DSM-5* criteria. The area under the ROC curve (AUC) was .95 (95% CI = [.893, 1.000]; Figure 2).
Figure 1. ROC Curve analysis – Males, AMSE total score x diagnosis
Sex-differences in children referred for assessment: An exploratory analysis of the AMSE

Discussion

The present study revealed marked sex differences, in that ASD males and ASD females differed in their responses to the AMSE. Specifically, ASD females were more likely to have a selective impairment in language compared to ASD males, but tended to express fewer issues related to unusual sensitivities such as heightened sensitivity to noise, touch, smell or taste, and having a high pain threshold. Examples of oversensitivity include a child covering his or her ears in response to noise, or a child’s discomfort with clothing labels, fabric texture or other surfaces. The item measuring language deficits were restricted to nonverbal, undeveloped sentences, single word use, and the use of less than three words.

Previous studies have shown a relationship between the early identification of ASD in females and the increased presence of language issues (Carter, Black, Tewani, Connolly,
Kadlec, and Tager-Flusberg 2007b; Hartley and Sikora 2009; Volkmar et al. 1993). These findings have been linked to the assumption that females need a stronger genetic loading to be diagnosed with ASD (Robinson et al. 2013; Skuse 2007). In line with previous work, the present study found that within the ASD sample females’ language performance was worse than their male counterparts. Selective language impairment might be reflected when ID is present. If these findings are considered alongside the relatively reduced presence of sensory issues, and a tendency towards less disruptive internalizing symptomatology in females, it could be suggested that either the presence or lack of a language deficiency, might influence the age at which females are detected and diagnosed. While it is important to stress that this view is still theoretical, and behavioral patterns within ASD are often nuanced, this perspective is in line with other studies, which have reported that females often remain unrecognized for significantly longer periods of time compared to males due to having more complex language skills (Lai et al. 2012; Salomone et al. 2015). Those findings are not yet well-integrated with differentiating features of girls with ASD relative to boys, such as a tendency towards less disruptive behavioral symptoms (Dworzynski, Ronald, Bolton, and Happé 2012b). Further, there might be a potential for some degree of male bias in diagnostic criteria and instruments, due to constructing and validating them in mainly predominantly male populations (Koenig and Tsatsanis 2005). This could potentially play an important role in sex-linked diagnostic asymmetries (Dworzynski, Ronald, Bolton, and Happé 2012b).

According to aim (c), the ROC curve analysis revealed that the AMSE discriminated well between ASD females and non-ASD females, as well as between ASD males and non-ASD males. This indicates that the AMSE performs well in the identification of both males and females referred for ASD specific assessment. It can be questioned whether females, due to different behavioral symptoms, go unidentified due to e.g. less parental concern or gender biases in the development of diagnostic instruments.
It is important to keep in mind the potential selection biases in the present study, as there are implications when utilizing a sample of children that are already referred for assessment based on concern. For example, the base-rate of ASD in the present sample is much higher than in clinical samples where inclusion is based on positive screening, as children in the present sample were referred to ASD-specific assessment by clinicians. This selection bias is important to note, as it might impact the observed psychometric properties of the AMSE. This could further impact the number of males and females that receive an ASD specific assessment. For example, parents of males with ASD and more disruptive behavior might be more likely to seek clinical assessment because of the disruptive behavior, while parents of females with less language issues might have a hard time recognizing or understanding the specific phenomenology of ASD specific traits. This might cause a selection bias where more males with disruptive behavior and females with more significant language issues are present. Another potential limitation to the present study, is the lack of full IQ information on each participant. Even though information on comorbid ID were available for most of the participants, there is a risk that one of the groups could potentially have higher IQ than the other. Further studies should aim to investigate language and sensory in samples of children with matched IQ. The present study reveals the differences between males and females that were referred to ASD specific assessment, and received an ASD diagnosis. The results regarding psychometric properties of the AMSE indicate that the AMSE performs just as well in discriminating females with ASD from females without ASD that are at high-risk of ASD, an effect also seen in males. As noted, this might be affected by the fact that females with more complex language abilities(Salomone et al. 2015) and less disruptive behaviors(Dworzynski, Ronald, Bolton, and Happé 2012a) are later identified and therefore not present due to not raising enough concern to be considered for ASD specific assessment. On the other hand, such differences could be put forward as hypotheses for the
Sex-differences in children referred for assessment: An exploratory analysis of the AMSE predominant male prevalence, as it could ultimately cause females to go under the radar for an ASD diagnosis for longer than males. The common consensus male-to-female ratio of ASD is often reported as 4:1 (Werling and Geschwind 2013), while a recent meta-analysis has revealed that there appears to be a gender bias, ultimately indicating that females meeting criteria for ASD are at disproportionate risk of not receiving a clinical diagnosis (Loomes et al. 2017).

It has yet to be explored if different developmental trajectories and/or behaviors are causing females to fall below the threshold for diagnosis more often than males. Even if a child falls slightly below the threshold for ASD, failure to fully meet the diagnostic criteria for ASD does not preclude the potential benefits they might receive from intervention and appropriate levels of support.

Conclusion

This exploratory analysis indicates that the AMSE’s test performance is comparable in girls at high-risk of ASD compared to boys at high-risk of ASD. Additionally, among children in our study sample who met diagnostic criteria for ASD, females have significantly more language impairment but fewer sensory symptoms than males as reflected on the AMSE. Further studies are needed to understand more about the disproportional prevalence of males and females receiving an ASD diagnosis, and more prospective studies are needed to understand the phenomenology of those identified later in life.

References

American Psychiatric Association. (2014). Diagnostic and Statistical Manual of Mental Disorders. APA DSM.
Sex-differences in children referred for assessment: An exploratory analysis of the AMSE


Sex-differences in children referred for assessment: An exploratory analysis of the AMSE


