Published in final edited form as:

J Autism Dev Disord. 2013 February ; 43(2): 371–381. doi:10.1007/s10803-012-1576-y.

Effects of child characteristics on the Autism Diagnostic Interview-Revised: Implications for use of scores as a measure of ASD severity

Vanessa Hus^{1,*} and Catherine Lord²

¹Department of Psychology, University of Michigan, 530 Church Street Ann Arbor, MI, 48109 USA

²Department of Psychiatry, Weill Cornell Medical College, White Plains, NY USA

Abstract

The Autism Diagnostic Interview-Revised (ADI-R) is commonly used to inform diagnoses of autism spectrum disorders (ASD). Considering the time dedicated to using the ADI-R, it is of interest to expand the ways in which information obtained from this interview is used. The current study examines how algorithm totals reflecting past (ADI-Diagnostic) and current (ADI-Current) behaviors are influenced by child characteristics, such as demographics, behavioral problems and developmental level. Children with less language at the time of the interview had higher ADI-Diagnostic and ADI-Current. ADI-Diagnostic totals were also associated with age; parents of older children reported more severe past behaviors. Recommendations are provided regarding the use of the ADI-R as a measure of ASD severity, taking language and age into account.

Keywords

Autism Spectrum Disorders; Autism Diagnostic Interview-Revised; severity; language level; chronological age

Many studies have used scores from the Autism Diagnostic Interview-Revised (ADI-R: Rutter, Le Couteur, & Lord 2003) to try to link biomarkers, such as genetic mutations or patterns of neural activity, to severity of ASD symptoms or an associated feature, such as language delay. Item-level ADI-R scores are now widely accessible in several large datasets that include genetic and behavioral information about children with ASD and their families (AGP 2009; AGRE 2011; NDAR 2010; Simons Foundation 2011). Strengths of the ADI-R include: 1) examiners are reliable on an item-level basis (which is unusual for a caregiver report), 2) summary scores are available for Reciprocal Social (SOC), Communication and Restricted, and Repetitive behaviors (RRBs), the three domains currently defining ASD in the most commonly used diagnostic frameworks (i.e., DSM-IV-TR: American Psychiatric Association 2000; ICD-10: World Health Organization 2011), and 3) scores are provided for both “most abnormal/ever” and “current” (in the last 3 months) diagnostic features. Previous research has indicated that certain information obtained from the ADI-R, such as ages of language milestones or presence of repetitive sensory motor behaviors, are affected by child characteristics, such as age and IQ of the child at the time of the interview (Hus, Pickles, Cook Jr., Risi, & Lord 2007). Nonetheless, ADI-R totals or domain scores are often used,

Tel: 734-273-9883; Fax: 734.615.0573; vhus@umich.edu.

Disclosure

C. Lord receives royalties for the ADI-R and ADOS; profits from this study were donated to charity.

without modification, as indicators of overall or domain-specific ASD severity (e.g., Coutanche, Thompson-Schill, & Schultz 2011; Meilleur & Fombonne 2009; Uddin et al. 2011). Given that nonspecific developmental and psychiatric characteristics have been shown to affect raw totals on other ASD-diagnostic instruments (e.g., the Social Responsiveness Scale (SRS): see Constantino, Hudziak, & Todd 2003 and Hus et al., *in press*; and the Autism Diagnostic Observation Schedule (ADOS): see Gotham, Pickles, & Lord 2009), a better understanding of how these factors influence ADI-R domain and total scores is important to inform their use as indicators of ASD severity. If ADI-R scores are confounded by non-ASD-specific child characteristics, and these factors are not carefully accounted for, conclusions from studies using ADI-R scores to reflect symptom severity may be inaccurate or misleading. The current study investigates how demographics, behavior problems, age, expressive language level, and IQ affect scores from the Diagnostic (ADI-Diagnostic, which focuses on past behavior) and Current Behavior (ADI-Current) algorithms.

To date, demographic factors, except for birth order in very young children, have not been shown to influence ADI-R scores in a predictable way (e.g., De Giacomo & Fombonne 1998; Holtmann, Bölte, & Poustka 2007). However, on the whole, few studies have had sufficient numbers of ethnic minorities, females or families with less than some college education to carry out meaningful investigation of these possible demographic influences.

Two of the major confounds in interpreting scores from diagnostic measures as indicators of ASD symptom severity are a child's current developmental level (e.g., chronological age, expressive language level) and the child's proportionate level of delay compared to other children of the same age (e.g., IQ). In general, when samples include children spanning the full range of IQ from profound intellectual disability to superior intellect, greater delays (i.e., lower IQs) tend to be associated with more severe impairment on most behavioral measures (e.g., Gotham et al. 2009; Mayes & Calhoun 2011). In contrast, associations with age are less consistent for studies including children spanning toddlerhood to adolescence. Some measures demonstrate higher scores for older children and adolescents (Gotham et al. 2009), while others report higher scores for younger children (Mayes & Calhoun 2011).

The ADI-Diagnostic algorithm attempts to reduce the effects of developmental level by focusing on the presence of ASD-related symptoms in the most abnormal form that they occurred in the past. "Positive" or atypical behaviors, such as unusual preoccupations or motor mannerisms, are described using *Ever* scores, which include behaviors that are *currently* present or have *ever* occurred for a period of 3 months or longer. For social and communication behaviors (e.g., sharing enjoyment or pointing to initiate joint attention) that are well-established in typically-developing children by 18 months of age, abnormality during the year between a child's fourth and fifth birthdays is scored. The "*most abnormal 4 to 5*" period was selected because even children with intellectual disability who are functioning at less than half their chronological age level should regularly exhibit these behaviors by the age of 4. Although the focus on past behavior in the ADI-Diagnostic algorithm was intended to control for developmental level to some extent, the degree to which algorithm scores are related to child characteristics at the time the parent completes the interview has not been systematically explored.

Relatively little research has made use of ADI-Current items, except for studies describing changes in autistic features over time (e.g., Piven, Harper, Palmer, & Arndt 1996; Seltzer et al. 2003) and examining developmental effects on RRBs (Bishop, Richler, & Lord 2006; Richler, Huerta, Bishop, & Lord 2010). This is somewhat surprising, given that scores reflecting a child's current behavior would be useful in treatment planning and measuring treatment outcomes. Theoretically, one might also expect that ADI-Current would

demonstrate stronger associations with measures of current brain function than ADI-Diagnostic. Considering the potential uses of ADI-Current, investigation of how child characteristics affect ADI-Current scores is warranted.

Another recurring concern is that caregiver-report measures of ASD symptoms are sometimes strongly related to caregiver reports of general behavior problems. The influences of behavior problems is most notable for the SRS (Constantino et al. 2003; Hus et al. *in press*), but Charman and colleagues (2007) also found reduced specificity of the Child Communication Checklist-2 (CCC-2; Bishop 2003) and the Social Communication Questionnaire (SCQ; Rutter, Bailey, Lord, & Berument 2003) in children with elevated behavior problems. It is striking that the SCQ, which is based on the ADI-Diagnostic and asks about the presence of behaviors either *ever* in the child's life or *between 4 and 5 years of age*, was significantly correlated with the Strengths and Difficulties Questionnaire (Goodman 2001), a measure of the child's current emotional, peer, conduct, and hyperactivity problems. These results suggest that it is also important to examine associations between measures of ASD symptoms and behavior problems. The association between general levels of behavior and emotional problems and ADI-R scores, either Diagnostic or Current, is not yet well-understood.

Separating out the role of child characteristics, such as demographics, general behavior problems, age, and language level, in measures that are portrayed as continuous indices of ASD-specific symptoms is particularly important for studies attempting to link brain function or genetic abnormalities with specific behavioral domains in autism (e.g., social behavior; see Duvall et al. 2007; Kaiser et al. 2010). Failure to account for developmental level may lead researchers to misinterpret significant associations as evidence of a causal mechanism for a subset of ASD-related symptoms (e.g., social deficits) when they really are markers for more global delays. Recently, using cross-sectional and longitudinal data from large samples of children with scores on the ADOS (Lord, Rutter, DiLavore, & Risi 1999), severity scores standardized across developmental levels and ages were generated (Gotham et al. 2009). These scores place a child's current functioning in the context of the behavior of other children with ASD of similar ages and language levels. Studies have begun to use the ADOS severity metric as a treatment outcome measure (e.g., Dawson et al. 2010), though further research is needed to assess the sensitivity of this measure to detect such changes (Cunningham 2011). Before similar strategies can be applied to the ADI-R, a more thorough understanding of the factors that affect ADI-Diagnostic and ADI-Current scores is needed.

In the present paper, a large sample of children and adolescents with ASD recruited to the Simons Simplex Collection was used to investigate relationships between ADI-Diagnostic and ADI-Current scores and child characteristics. Based on previous research, it was hypothesized that demographics (gender, race, maternal education) would not be related to ADI-R totals, but higher ADI-R scores (indicating greater ASD-related impairment) would be associated with more behavior problems, older age, and lower expressive language level and IQ. It was also expected that these factors would exhibit stronger effects on ADI-Current scores compared to ADI-Diagnostic.

Methods

Participants

Participants were drawn from a sample of 2,442 families with a proband between 4 and 18 years of age. Families participated in the Simons Simplex Collection (SSC), a multi-site genetic study of families with one child with ASD who did not have any first-, second- or third-degree relatives with ASD. All probands met Collaborative Programs of Excellence in

Autism (CPEA) criteria for a diagnosis of Autism, ASD, or Aspergers Disorder (Lainhart et al. 2006; see Supplementary Materials) and had a nonverbal mental age of at least 18 months. Families were excluded if probands had any significant sensory impairments (e.g., blindness) that might affect standardized test administration, a significant early medical history (e.g., very low birth rate), or a diagnosis of Fragile X syndrome, Tuberous Sclerosis or Down syndrome. Families were also excluded if the sibling had substantial language or psychological problems related to ASD. Additional details regarding the SSC are provided elsewhere (Fischbach & Lord 2010; Lord et al. 2011). Parents provided informed consent and children provided assent, approved by Institutional Review Boards at each university.

Forty-four probands were excluded from the present analyses because they were missing at least one ADI-R item used to compute the ADI-Current Total (described below). Sixty-four probands administered an ADOS Module 4 were also excluded because Calibrated Severity Scores (CSS) are not available for Module 4. This yielded a total of 2,334 probands, of which 2326 overlap with a parallel study on the Social Responsiveness Scale (see Hus et al. *in press*). Sample demographics are provided in Table 1.

Measures

Autism Symptoms—The Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord 2003) is a semi-structured diagnostic interview administered to a caregiver. All items on the ADI-R are coded for current and past behavior. The Diagnostic algorithm is comprised of *Most Abnormal 4–5/Ever* codes divided into three domains based on the ICD-10 and DSM-IV criteria for autism: SOC, Communication, and RRB. There is also a fourth domain to indicate whether the child meets criteria for age of onset prior to 3 years of age. The Communication Domain is split into Verbal (VC) and Nonverbal Communication (NVC) Totals. The VC total is used for children using at least 3 word phrases which sometimes include a verb on a daily basis. It includes 6 items assessing conversation and stereotyped speech and 7 items assessing nonverbal communication (e.g., gestures) and play skills. For children who do not have sufficient language to be considered “verbal,” the NVC total includes only the 7 nonverbal communication/play items. A Current Behavior algorithm based on *Current* scores is also available for clinical uses, such as treatment planning; items comprising the three domains vary by age.

To investigate the ADI-R’s use as a metric of ASD-severity, scores from the three domains were summed to provide a single total for the Diagnostic algorithm (i.e., ADI-Diagnostic), and the Current-Behavior algorithm (ADI-Current). Following ADI-R conventions, 2s and 3s were collapsed. To make totals comparable across participants of different ages and language levels, ADI-Diagnostic included totals from the SOC, NVC, and the RRB domains; the Verbal Rituals item, an item on the RRB domain that is only scored for verbal participants, was subtracted from the total, leaving 27 items (totals ranged from 9–54). The ADI-Current included 10 items from the SOC, 4 from the NVC, and 5 from the RRB domain which are scored on the Current Behavior Algorithm for participants of all ages and language levels (totals ranged from 1–35). Predictors of the combined SOC+NVC and RRB domains were also examined separately for both Diagnostic and Current algorithms. When the sample was limited to only children who were considered “verbal” on the ADI-R (88%) and VC items were included in the overall totals, predictors were nearly identical; the variance explained by language level was somewhat reduced due to fewer children with Module 1s being included in the limited sample. Given the similar results, only analyses using ADI-R totals comparable across all children are reported below.

The Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi 1999) is a standardized observational assessment that is organized into four modules, based on the child’s expressive language level. The ADOS CSS (Gotham et al. 2009) was used as

an indicator of ASD severity. The ADOS CSS was established on a sample of children ranging from ages 2–14 for Module 1 and ages 2–16 for Modules 2 and 3. For the purposes of this study, ADOS-CSS for children outside of these age ranges ($n=53$) were assigned based on the raw total-to-CSS mapping for the oldest age group provided by Gotham et al., 2009. When those 53 children were excluded from analyses, results were nearly identical; therefore the entire sample was included in the models reported below.

Each site was overseen by an off-site consultant who was an experienced trainer on the ADI-R and ADOS and two on-site clinical supervisors at each site who met research reliability standards for both instruments. All ADI-R and ADOS were administered by clinicians who achieved initial reliability with a site supervisor; each examiner also submitted a reliability video to their site consultant on a quarterly basis. In addition, the consultants reviewed data for all cases and requested to see ADI-R and ADOS videos for any questionable cases (e.g., appropriateness of module selection; see Lord et al., 2011 for additional details).

Developmental Level—All probands were between the ages of 4 and 18 years of age at the time of assessment. Age in years was used as a continuous predictor of ADI-Diagnostic and ADI-Current.

ADOS Module was used as a categorical indicator of expressive language in the regression analyses. Nineteen percent of probands received Module 1, designed for children who are nonverbal or using single words; 23% received a Module 2, for children with phrase speech; 58% received a Module 3, for children and adolescents who are verbally fluent.

Nonverbal IQ (NVIQ) scores were used as a continuous measure of cognitive ability. Scores were derived using the following tests: the Wechsler Abbreviated Scale of Intelligence (Wechsler 1999; 2%), Wechsler Intelligence Scale for Children (Wechsler 2003; 2%), Differential Scales of Ability (Elliott 2007; 85%), or Mullen Scales of Early Learning (Mullen 1995; 11%).

Behavior Problems—The Child Behavior Checklist (CBCL; Achenbach & Rescorla 2001) is a parent-completed questionnaire of behavior and emotional problems in children. Children ages 18 months to 5 years and children ages 6 to 18 years receive different forms, each of which yields standard T-scores for Internalizing (CBCL-I) and Externalizing (CBCL-E) domains. These were used as estimates of behavior problems in regression models.

Data Analysis

Linear regression models were analyzed separately for ADI-Diagnostic and ADI-Current totals and domain scores to investigate relationships with demographics (gender, race, maternal education), behavior problems (CBCL-I, CBCL-E), and developmental level (chronological age, language level indicated by ADOS Module and NVIQ), while controlling for ASD symptoms (ADOS-CSS). Results were identical when analyses were performed as mixed models including site as a random effect (data not shown). All variables were centered at the mean. Cohen's R^2 was computed to assess the effect of each set of predictors while controlling for all other variables in the model; R^2 of .02, .15, and .35 reflect small, medium, and large effect sizes, respectively (Cohen 1988).

All analyses were conducted using SPSS 17.0. Given the large sample size and multiple comparisons, significance level was set at $p .001$.

Results

Preliminary analyses

ADOS-CSS was selected as a measure of ASD symptoms because these scores have been standardized to account for age and developmental level differences. To verify that the ADOS-CSS was indeed relatively independent of these factors, separate regression models were run predicting ADOS-Raw Totals and ADOS-CSS (data available upon request). The model predicting ADOS-Raw explained 34% of the variance. Language level (i.e., ADOS Module) and NVIQ accounted for the majority of variance in ADOS-Raw (30%) and a significant age-by-ADOS Module interaction explained an additional 2%. In contrast, only 6% of variance in ADOS-CSS was explained by the same predictors, with language and age accounting for approximately 4%. These weaker associations suggested that ADOS-CSS was an indicator of ASD-severity that was relatively independent of other non-ASD-specific factors.

Notably, when regression models included ADOS-Raw as a predictor, there was a higher proportion of variance explained by ADOS-Raw compared to ADOS-CSS (e.g., $R^2_{\text{Raw}}=.14$ vs. $R^2_{\text{CSS}}=.04$, for ADI-Diagnostic), but the proportion of variance explained by developmental level was reduced ($\Delta R^2=.13$ for the model including ADOS-Raw vs. $\Delta R^2=.22$ for the model including ADOS-CSS). When ADOS-Raw was entered last, after controlling for developmental level and other factors, associations were much smaller (e.g., $\Delta R^2_{\text{Raw}}=.04$). Changes to the significance of other predictors (Demographics, Behavior Problems) and overall variance explained by the model were minimal (e.g., $R^2=.29$ vs. $R^2=.28$). These findings further demonstrate the influence of developmental level on ADOS-Raw. Consequently, in order to investigate the effects of developmental level on ADI-R scores, analyses reported below use ADOS-CSS.

Overall total and domain scores from the ADI-Diagnostic and ADI-Current were generally highly correlated (see Table 2). Associations with ADOS-CSS were small, though significant. [Table 2 about here]

Predictors of ADI-Diagnostic

ADI-Diagnostic Total—As shown in Table 3, the final model explained just under 28% of the variance in ADI-Diagnostic. ADOS-CSS accounted for approximately 4%; adding demographics only explained another 1% (see Model 1). Including behavior problems did not significantly improve the model (R^2 change=.002, $p=.12$). However, the addition of developmental level explained an added 22% of variance (see Model 3). Older children and children with more limited language skills (i.e., Module 1 or 2 compared to Module 3) or lower NVIQ had higher ADI-Diagnostic. Significant age-by-ADOS Module interactions indicated that the difference in ADI-Diagnostic between children with phrase speech and those who were verbally fluent was more pronounced for older children; the interaction between age and language was marginal for children who were nonverbal or using single words (i.e., Module 1).

Although the ADOS-CSS accounted for a relatively small amount of variance in ADI-Diagnostic, one concern might be that entering these scores into the model first would result in over-matching on autism symptom severity. Thus, models were also fit including ADOS-CSS last. Significant predictors of ADI-Diagnostic remained the same, with effect sizes for demographics ($f^2=.010$), behavior problems ($f^2=.001$), developmental level ($f^2=.238$) and ADOS-CSS ($f^2=.039$) similar to those in the model reported above (in which ADOS-CSS was included first).

To demonstrate the effects of age and language, ADI-Diagnostic totals were plotted by age and ADOS Module. As shown in Figure 1, children who were nonverbal or using single words (i.e., Module 1) had consistently higher ADI-Diagnostic totals than children who were verbally fluent (Module 3), irrespective of age. Young children with simple phrases (Module 2) had similar scores to verbally fluent children; however, older children administered a Module 2 looked more similar to their same-aged peers administered a Module 1.

ADI-Diagnostic Domains—As shown in Table S1, the final regression model predicting the combined Diagnostic-SOC+NVC total explained 29% of variance. Predictors of the Diagnostic-SOC+NVC were nearly identical to those observed for the overall ADI-Diagnostic total. Age, language level, and an age-by-ADOS Module interaction accounted for nearly 24% of variance in Diagnostic-SOC+NVC.

The model predicting the Diagnostic-RRB domain total explained only 4.6% of variance. Significant predictors included ADOS-CSS, gender, CBCL-I, and an age-by-Module interaction; each explained less than 1% of variance in the Diagnostic-RRB.

Predictors of ADI-Current

ADI-Current Total—The overall model accounted for approximately 33% of variance in ADI-Current (see Table 4). ADOS-CSS explained only 2.8% of variance in ADI-Current. Adding demographics hardly improved the model, explaining less than 1% of additional variance (see Model 1), whereas behavior problems accounted for a further 3.7% (see Model 2). In contrast, indicators of developmental level had a large effect, accounting for an added 26% of variance in ADI-Current (see Model 3). Similar to ADI-Diagnostic, children with less language (nonverbal to simple phrases) had significantly higher ADI-Current than children who were verbally fluent. Unlike the ADI-Diagnostic, interactions between ADOS Module and age were nonsignificant and therefore dropped from the model. Although age was not a significant predictor of ADI-Current, for comparison to ADI-Diagnostic, ADI-Current scores were also plotted by ADOS Module and age group. As shown in Figure 1, while there was virtually no effect of age, children with less language had higher ADI-Current than children who were verbally fluent.

When ADOS-CSS was entered last into the model predicting ADI-Current, NVIQ emerged as a significant predictor ($B = -.03$, $SE B = .01$, $r_{part} = -.07$), compared to the above model in which NVIQ was only marginally significant. Effect sizes for each set of predictors remained highly similar (demographics $f^2 = .007$), behavior problems ($f^2 = .037$), developmental level ($f^2 = .386$) and ADOS-CSS ($f^2 = .036$).

ADI-Current Domains—The final model explained approximately 34% of variance in the Current-SOC+NVC total (see Table S2). As observed for the overall ADI-Current total, ADOS-CSS, CBCL-E, CBCL-I and ADOS Module were significant predictors. Language level had a strong effect on Current- SOC +NVC, explaining an additional 28% of variance, while ASD symptoms and behavior problems accounted for only 2–3% of variance in the Current-SOC+NVC total.

Predictors of Current-RRB domain totals were similar to those for the ADI-Current overall total and SOC+NVC, however, the proportion of variance explained was much less (just over 8%). Language level explained approximately 3% of variance in Current-RRB totals, while ASD symptoms and behavior problems explained 1 and 4%, respectively.

Discussion

Domain scores and overall totals from the ADI-Diagnostic and ADI-Current algorithms were strongly correlated with each other, though correlations between different domains (i.e., SOC+NVC and RRB) tended to be smaller. Small, but significant correlations between ADI-R totals and the ADOS-CSS are a reminder that method variance (in this case, semi-structured caregiver interviews vs. semi-structured observations by experienced clinicians) can have a significant influence on the types of behaviors being described, even when the two methods are aimed at measuring similar behaviors. The weak correlations between ADI-R totals and ADOS-CSS also demonstrate that ADOS-CSS have been calibrated to reduce the effects of age and language. This is supported by the somewhat stronger correlations observed between ADI-R totals and the ADOS-Raw Total. It is also reflected in the reduction in ADI-R score variance explained by developmental level when ADOS-Raw were included as a predictor in the model. In addition, because all children were required to meet diagnostic criteria on both the ADI-R and the ADOS to be included in the SSC, score ranges were restricted in range. In studies without such requirements, correlations between raw totals from the two measures have been much stronger (e.g., deBiltdt et al. 2004; Risi et al. 2006).

As found in previous studies, demographics had minimal, if any, effect on ADI-Diagnostic and ADI-Current totals. Gender, race and maternal education never explained more than 1% of the variance in ADI totals. CBCL-Internalizing and Externalizing T-scores explained 2–4% of variance in ADI-Current, with CBCL-I accounting for most of the association. Although the CBCL-I scale emerged as a significant predictor of ADI-Diagnostic in the final model, it explained less than 0.5% of variance. Thus, while the effects of behavior problems on ADI totals were significant, they were much smaller than correlations reported between measures of behavior problems and scores from screening questionnaires such as the SRS, CCC-2 and SCQ (e.g., Charman et al. 2007; Constantino et al. 2003; Hus et al. *in press*).

More problematic is the strong influence of developmental level, particularly expressive language at the time of reporting (here, represented by ADOS module), on both ADI-Diagnostic and Current totals. In line with our hypothesis, effects of language level were somewhat stronger for ADI-Current than ADI-Diagnostic. Contrary to our predictions, however, age at the time of the interview was significantly associated with ADI-Diagnostic scores but not related to ADI-Current. For ADI-Diagnostic, there was also a significant interaction between age and language level, a pattern seen in studies of other measures (e.g., Gotham et al. 2009). Across both domains, younger children who speak in simple phrases (Module 2) were more similar to same-aged peers with fluent speech (Module 3), whereas older children who have phrase speech but are not yet verbally fluent (Module 2) are more like children who are nonverbal or using single words (Module 1). The lack of association between age and ADI-Current may be due to our exclusion of items that were not applicable to children of all ages in the calculation of the ADI-Current total. Although every item comprising the ADI-Diagnostic was asked for all children, the tendency for parents of older, more impaired children to retrospectively recall their child's early development as being more delayed may influence scores (Hus, Taylor, & Lord 2011). This is consistent with the stronger relationship between age and ADI-Diagnostic totals for children with more impaired language compared to those who were verbally fluent (see Figure 1). Nonverbal IQ (NVIQ) was also a significant predictor of ADI-Diagnostic and Current totals, though the influence of NVIQ was much weaker due to the relationship between language level and IQ. If ADOS Module was omitted from the model, NVIQ accounted for approximately 11% of variance in ADI-Diagnostic and 14% in ADI-Current (data not shown).

In line with the overall totals, demographics and behavior problems had small effects on both current and diagnostic (i.e., based on past behaviors) Social and Nonverbal Communication (SOC+NVC) and Repetitive Behavior (RRB) domain totals. While developmental level had a strong effect on both SOC+NVC totals, influences of developmental level on the RRB domain totals were considerably smaller. In part, this may reflect the restricted range in RRB totals, which are based upon 5 items, compared to the SOC+NVC domain, comprised of 14 (Current) to 22 (Diagnostic) items. The small association between developmental level and the current RRB domain may also reflect that different RRB items comprising the current RRB total have different relationships to IQ (Bishop et al., 2006). Thus, when these items are combined into one total, items not strongly correlated with IQ may reduce the correlation and positive and negative correlations may be offset by each other. However, the lack of correlation between the diagnostic RRB total and developmental level in this sample was striking, compared to small, but significant correlations with age and, moderate to strong correlations with both verbal and nonverbal IQ, reported in other studies (e.g., Hus et al., 2007). Although the RRB domain was not included in the CPEA criteria for study inclusion, the more restricted age range and the relatively higher IQ in this sample may have contributed to this difference. Future studies using the RRB domain totals should explore the effects of developmental level on RRB totals carefully, as these may vary from one sample to another.

Although associations with child characteristics limit our ability to interpret raw scores from the ADI-R as indices of ASD severity, there are relatively straightforward ways to control for the effects of language level, with or without age, through statistical analyses. From data available through the large genetic consortiums, it should be possible to calibrate ADI-R scores by language level and age, as has been done for raw ADOS totals (Gotham et al. 2009). These would yield much “purer” measures of ASD symptom severity, as reported for children with ASD of similar age and language ability. However, because new diagnostic systems for DSM-5 (American Psychiatric Association 2012) and ICD-11 (World Health Organization 2011) are in preparation, it is likely that the ADI-R algorithms will need to be reconfigured to match revisions to the diagnostic criteria for ASDs. Thus, before attempting to standardize ADI-R scores, it may make the most sense to wait for ADI-R algorithms to be revised in response to final versions of these frameworks.

Meanwhile, researchers who want to use ADI-R scores as a metric of ASD severity can reduce the confounding influences of age, IQ and language level by statistically adjusting for these effects to yield more accurate measures of the core features of autism. This requires information concerning expressive language level at the time of data collection. Here we have demonstrated the utility of using ADOS Module as a proxy for expressive language level; other possibilities include using the *overall level of language* item from the ADI-R, an age equivalent from the Vineland-II Expressive Communication subdomain or an estimated verbal mental age (from any measure that addresses expressive language). However, researchers should be aware that each of these measures may group children somewhat differently. For example, in the present sample, 40% of children administered a Module 1 were scored as verbal (i.e., received a ‘0’) on the ADI-R overall level of language item.

Although ADI-Current scores did not demonstrate effects of age in this sample, age effects should always be explored, as this may vary by sample and the measures used to estimate language. For example, age effects may be reduced in many neuroimaging samples because participants are recruited within restricted age ranges. This should be directly tested before ADI-Current scores are used without controlling for age. Additionally, while ADI-Current scores may be less affected by age than ADI-Diagnostic scores, ADI-Current may be somewhat more strongly influenced by non-ASD-specific factors, such as behavior and

emotional problems. Such effects may also be statistically controlled using the CBCL or another measure of general behavior problems.

Many studies have used the ADI-Diagnostic total and domain scores as indicators of severity (e.g., Meilleur & Fombonne 2009). However, because these scores include past behaviors (i.e., *Most Abnormal 4–5* and *Ever*), they will provide somewhat different information than measures of current behavior, such as ADI-Current or ADOS. Comparisons of associations between ADI-Diagnostic and ADI-Current or ADOS, while controlling for the appropriate factors, may be a useful approach for considering how the “severity” of past behavior might be related to current brain function vs. how current brain function predicts “severity” of current behavior. Despite the fact that the ADI-R is not intended to assess change in symptoms over time, such comparisons may be useful for generating hypotheses and informing research questions to be more appropriately tested in longitudinal studies.

Limitations

The present sample was large and collected from multiple sites. However, this study is limited by the fact that it essentially consists of 12 different convenience samples. While inclusion of site as a random effect did not alter results, we cannot determine if some associations (e.g., the finding that older children are reported as having more severe problems in the past than younger children) are due to referral bias within the study or the effects of memory on parent reports for older children. It may also be that increased awareness of ASD and improved diagnostic practices over the past 5–10 years have enhanced the identification of young children with milder symptoms, suggesting that the younger children in our sample may truly be less severely impaired. The sample was also quite homogeneous in ethnicity and maternal education, and was comprised entirely of children with ASD from simplex families. Moreover, the sample had very few older children with minimal verbal skills and because there is not an ADOS-CSS for Module 4, older, verbally fluent adolescents were also excluded. Thus these findings may be most relevant to U.S. and Canadian samples of middle-class children ranging from preschool through secondary school age. Whether similar effects are observed in samples collected in other countries, or samples which include older individuals or a greater proportion of low-income families, remains to be explored. It will also be important for future studies to investigate whether similar effects of child characteristics on ADI-R scores are observed for children with non-ASD developmental disabilities or psychiatric diagnoses. Similarly, influences of child characteristics on ADI-R scores should be examined for children who come from multiplex families, in which heritability of child characteristics may alter the association between developmental level and ASD severity and parent perception and report of behaviors may be influenced by having more than one child with an ASD.

An additional limitation to using the ADI-R as a measure of ASD severity is that approximately 25% of items are not scored for children who do not meet the ADI-R’s “verbal” criteria. In order to make scores more comparable across children of all language levels, only items administered for all participants were included in the ADI-R totals used in this study. Although predictors of totals including the verbal items were nearly identical (data not shown), it is possible that exclusion of verbal items may underestimate ASD-severity for verbal children. Future efforts to standardize ADI-R scores to account for age and language effects should consider including items from the Verbal Communication domain in the calibrated scores for verbal children. When the ADI-R algorithm is revised to reflect DSM-5 and ICD-11 changes, it will also be important to explore whether and how the factor structure of ADI-R items differs for different subgroups of individuals, such as children who are verbal or adults who have intellectual disability. Although some studies have suggested that the factor structure of the measure is relatively stable across age (e.g.,

Frazier et al., 2008), less is known about the associations between items for different subgroups of individuals. Thus, although the present study included items from the existing ADI-R algorithms to reflect how ADI-R scores are currently used, it is possible that factor analysis would suggest that different subsets of items may more accurately reflect ASD severity in different subgroups. As a result, the associations between ADI-R score and child characteristics may vary for each subgroup.

Conclusions

Findings suggest that, with statistical adjustment for expressive language level (and in some cases, age) at the time of interview, ADI-Diagnostic and ADI-Current domain and total scores can be used as estimates of the severity of core ASD symptoms. More sophisticated ways of calibrating scores to provide a more easily interpretable metric seem quite possible once new diagnostic systems are in place.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by a graduate fellowship from the Simons Foundation and a Dennis Weatherstone Predoctoral Fellowship to VH and Simons Foundation and National Institute of Mental Health grants R01MH081873 and RC1MH089721 to CL. We are grateful to the families, as well as SSC principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, D. Grice, A. Klin, D. Ledbetter, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC dataset described in this study ([https://ordering.base.sfari.org/~browse_collection/archive\[sfari_collection_v12\]/ui:view\(\)](https://ordering.base.sfari.org/~browse_collection/archive[sfari_collection_v12]/ui:view())) by applying at <https://base.sfari.org>.

References

- Achenbach, TM.; Rescorla, L. Manual for the ASEBA school-age forms & profiles. ASEBA; Burlington, VT: 2001.
- AGP. The Autism Genome Project. 2009. Retrieved March 31, 2010, from <http://autismgenome.org>
- AGRE. The Autism Genetic Resource Exchange. 2011. Retrieved March 31, 2010, from <http://agre.org>
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: American Psychiatric Association; 1994. Text Revision: DSM-IV-TR
- American Psychiatric Association. DSM-5 Development. 2012. Retrieved April 4, 2012, from <http://www.dsm5.org>
- Bishop SL, Richler J, Lord C. Association between restricted and repetitive behaviors and nonverbal IQ in children with autism spectrum disorders. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*. 2006; 12(4–5):247–267. [PubMed: 16911971]
- Charman T, Baird G, Simonoff E, Loucas T, Chandler S, Meldrum D, Pickles A. Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *The British Journal of Psychiatry*. 2007; 191(6):554–559. [PubMed: 18055961]
- Constantino JN, Hudziak JJ, Todd RD. Deficits in reciprocal social behavior in male twins: evidence for a genetically independent domain of psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2003; 42(4):458–467. [PubMed: 12649633]
- Coutanche MN, Thompson-Schill SL, Schultz RT. Multi-voxel pattern analysis of fMRI data predicts clinical symptom severity. *NeuroImage*. 2011; 57(1):113–123. [PubMed: 21513803]
- Cunningham AB. Measuring change in social interaction skills of young children with autism. *Journal of Autism and Developmental Disorders*. 2011 Advance online publication. 10.1007/s10803-011-1280-3

- Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, Donaldson A, et al. Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*. 2010; 125(1):e17–e23. [PubMed: 19948568]
- De Giacomo A, Fombonne E. Parental recognition of developmental abnormalities in autism. *European Child & Adolescent Psychiatry*. 1998; 7:131–136. [PubMed: 9826299]
- deBildt A, Sytema S, Ketelaars C, Kraijer D, Mulder E, Volkmar F, Minderaa R. Interrelationship between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders*. 2004; 34(2):129–137. [PubMed: 15162932]
- Duvall JA, Lu A, Cantor RM, Todd RD, Constantino JN, Geschwind DH. A quantitative trait locus analysis of social responsiveness in multiplex autism families. *American Journal of Psychiatry*. 2007; 164(4):656–662. [PubMed: 17403980]
- Fischbach GD, Lord C. The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron*. 2010; 68(2):192–195. [PubMed: 20955926]
- Frazier TW, Youngstrom EA, Kubu CS, Sinclair L, Rezai A. Exploratory and confirmatory factor analysis of the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders*. 2008; 38(3):474–480. [PubMed: 17619129]
- Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2001; 40(11):1337–1345. [PubMed: 11699809]
- Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2009; 39(5):693–705. [PubMed: 19082876]
- Holtmann M, Bölte S, Poustka F. Autism spectrum disorders: sex differences in autistic behaviour domains and coexisting psychopathology. *Developmental Medicine & Child Neurology*. 2007; 49(5):361–366. [PubMed: 17489810]
- Hus V, Bishop S, Gotham K, Huerta M, Pickles A, Lord C. Factors influencing scores on the Social Responsiveness Scale. *Journal of Child Psychology and Psychiatry*. (in press).
- Hus V, Pickles A, Cook EH Jr, Risi S, Lord C. Using the Autism Diagnostic Interview--Revised to increase phenotypic homogeneity in genetic studies of autism. *Biological Psychiatry*. 2007; 61(4):438–448. [PubMed: 17276746]
- Hus V, Taylor A, Lord C. Telescoping of caregiver report on the Autism Diagnostic Interview – Revised. *Journal of Child Psychology and Psychiatry*. 2011; 52(7):753–760. [PubMed: 21410473]
- Kaiser MD, Hudac CM, Shultz S, Lee SM, Cheung C, Berken AM, Deen B, et al. Neural signatures of autism. *Proceedings of the National Academy of Sciences*. 2010; 107(49):21223–21228.
- Lainhart JE, Bigler ED, Bocian M, Coon H, Dinh E, Dawson G, Deutsch CK, et al. Head circumference and height in autism: A study by the collaborative program of excellence in autism. *American Journal of Medical Genetics Part A*. 2006; 140A(21):2257–2274. [PubMed: 17022081]
- Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM, Ousley O, et al. A Multisite Study of the Clinical Diagnosis of Different Autism Spectrum Disorders. *Archives of General Psychiatry*. 2011 Advance online publication. 10.1001/archgenpsychiatry.2011.148
- Lord, C.; Rutter, M.; DiLavore, PS.; Risi, S. *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles, CA: Western Psychological Services; 1999.
- Mayes SD, Calhoun SL. Impact of IQ, age, SES, gender, and race on autistic symptoms. *Research in Autism Spectrum Disorders*. 2011; 5(2):749–757.
- Meilleur AAS, Fombonne E. Regression of language and non-language skills in pervasive developmental disorders. *Journal of Intellectual Disability Research*. 2009; 53(2):115–124. [PubMed: 19054269]
- NDAR. National Database for Autism Research. 2010. Retrieved March 31, 2010, from <http://ndar.nih.gov/ndarpublicweb/standards.go>
- Piven J, Harper J, Palmer P, Arndt S. Course of behavioral change in autism: a retrospective study of high-IQ adolescents and adults. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1996; 35(4):523–529. [PubMed: 8919715]

- Richler J, Huerta M, Bishop SL, Lord C. Developmental trajectories of restricted and repetitive behaviors and interests in children with autism spectrum disorders. *Development and Psychopathology*. 2010; 22(1):55–69. [PubMed: 20102647]
- Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, Cook EH, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2006; 45(9):1094–1103. [PubMed: 16926617]
- Rutter, M.; Bailey, A.; Lord, C.; Berument, S. Social Communication Questionnaire. Los Angeles, CA: Western Psychological Services; 2003.
- Rutter, M.; Le Couteur, A.; Lord, C. Autism Diagnostic Interview-Revised. Los Angeles, CA: Western Psychological Services; 2003.
- Seltzer MM, Krauss MW, Shattuck PT, Orsmond G, Swe A, Lord C. The symptoms of autism spectrum disorders in adolescence and adulthood. *Journal of Autism and Developmental Disorders*. 2003; 33(6):565–581. [PubMed: 14714927]
- Simons Foundation. The Simons Simplex Collection. 2011. Retrieved March 31, 2010 from <http://sfari.org/sfari-initiatives/simons-simplex-collection>
- Uddin LQ, Menon V, Young CB, Ryali S, Chen T, Khouzam A, Minshew NJ, et al. Multivariate Searchlight Classification of Structural Magnetic Resonance Imaging in Children and Adolescents with Autism. *Biological Psychiatry*. 2011 Online First(0).
- World Health Organization. International statistical classification of diseases and related health problems, 11th revision. 2011. Retrieved November 27, 2011 from <http://www.who.int/classifications/icd/revision/en/>

[Figure 1 Top]

Fig. 1.
ADI-Diagnostic and ADI-Current Totals plotted by age and ADOS module

Table 1

Proband characteristics

	Mean	(SD)
Male N, [%]	2024	[86.72]
White N, [%]	1822	[78.10]
Grad or BA N, [%]	1420	[60.84]
Age (years)	8.73	(3.35)
ADI-Current	16.11	(6.82)
ADI-Diagnostic	35.25	(9.18)
ADOS-CSS	7.43	(1.68)
CBCL-I	60.21	(9.53)
CBCL-E	56.53	(10.62)
NVIQ	84.71	(25.82)

Note: N=2334; Ns vary due to missing data. Grad or BA=Maternal Education Graduate or Bachelor degree; CBCL=Child Behavior Checklist; I=Internalizing; E=Externalizing; NVIQ=Nonverbal IQ.

Table 2

Correlations ADOS-CSS, ADI-Diagnostic and ADI-Current totals.

	1	2	3	4	5	6	7
1. ADI-Diagnostic Total	-						
2. ADI-Diagnostic SOC+NVC	0.97						
3. ADI-Diagnostic RRB	0.42	0.21					
4. ADI-Current Total	0.67	0.63	0.37				
5. ADI-Current SOC+NVC	0.63	0.64	0.13	0.93			
6. ADI-Current RRB	0.39	0.23	0.80	0.54	0.26		
7. ADOS-CSS	0.21	0.20	0.09	0.17	0.15	0.13	
8. ADOS-Raw Total	0.38	0.38	0.10	0.40	0.41	0.23	0.82

Note: All correlations significant $p < .001$; SOC=Reciprocal Social Interaction; NVC=Nonverbal Communication; RRB=Restricted, Repetitive and Stereotyped Behavior; CSS=Calibrated Severity Score.

Table 3

Predictors of ADI-Diagnostic Totals

	Model 1				Model 2				Model 3			
	<i>B</i>	<i>SE B</i>	95% CI		<i>B</i>	<i>SE B</i>	95% CI		<i>B</i>	<i>SE B</i>	95% CI	
			Lower	Upper			Lower	Upper			Lower	Upper
Constant	35.21	.19	34.85	35.58	35.21	.19	34.85	35.58	35.62	.17	35.28	35.96
ASD-Symptoms												
ADOS-CSS	1.12	.11	.91	1.34	1.13	.11	.92	1.35	.94	.10	.74	1.14
Demographics												
Gender	-.32	.55	-1.40	.76	-.32	.55	-1.40	.76	-.01	.49	-2.22	-.32
Race	1.54	.46	.64	2.44	1.56	.46	.67	2.46	.30	.41	-.50	1.10
Mat Educ	1.43	.38	.68	2.18	1.36	.38	.61	2.12	.60	.34	-.06	1.27
Behavior Problems												
CBCL-E					.01	.02	-.03	.05	.02	.02	-.02	.06
CBCL-I					.03	.02	-.01	.08	.07	.02	.03	.12
Developmental Level												
Age									.78	.06	.67	.89
M1vsM3									8.22	.63	6.98	9.45
M2vsM3									3.87	.49	2.91	4.84
Nonverbal IQ									-.03	.01	-.05	-.01
M1vsM3*age									.33	.13	.07	.59
M2vsM3*age									.97	.13	.71	1.23
R ²			.053				.055				.277	
R ² change			.011 ^a				.002				.223	
Effect size (<i>f</i> ²)			.012				.002				.308	

Note: **bold** = $p < .001$; *italics* = $p < .05$;

^aR² = .042 with only ADOS-CSS as a predictor; R² change reflects additional variance explained when Demographics are added to model; Mat Educ = Maternal Education; ADOS CSS = Autism Diagnostic Observation Schedule Calibrated Severity Score CBCL = Child Behavior Checklist; I = Internalizing; E = Externalizing; M1 = ADOS Module 1; M2 = ADOS Module 2; M3 = ADOS Module 3

Table 4

Predictors of ADI-Current Totals

	Model 1				Model 2				Model 3			
	<i>B</i>	<i>SE B</i>	95% CI		<i>B</i>	<i>SE B</i>	95% CI		<i>B</i>	<i>SE B</i>	95% CI	
			Lower	Upper			Lower	Upper			Lower	Upper
Constant	16.12	.14	15.84	16.39	16.12	.14	15.85	16.38	16.13	.12	15.90	16.35
ASD-Symptoms												
ADOS-CSS	.68	.08	.52	.84	.71	.08	.55	.87	.64	.07	.50	.78
Demographics												
Gender	.72	.41	-.09	1.53	.62	.41	-.17	1.42	.03	.35	-.64	.72
Race	.73	.34	.06	1.40	.81	.34	.15	1.46	.05	.29	-.12	-.04
Mat Educ	.90	.29	.34	1.46	.61	.28	.05	1.16	.04	.13	-.34	.61
Behavior Problems												
CBCL-E					.08	.02	.05	.11	.06	.01	.04	.09
CBCL-I					.07	.02	.03	.10	.12	.01	.09	.15
Developmental Level												
Age									.10	.04	.03	.18
M1vsM3									9.02	.44	8.15	9.89
M2vsM3									2.82	.33	2.17	3.47
Nonverbal IQ									-.01	.01	-.03	-.03
R ²			.035 ^a				.072				.332	
R ² change			.007 ^a				.037				.260	
Effect size (<i>f</i> ²)			.007 ^a				.040				.389	

Note: **bold** = $p < .001$; *italics* = $p < .05$;

^aR² = .028 with only ADOS-CSS as a predictor; R² change reflects additional variance explained when Demographics are added to model; Mat Educ = Maternal Education; ADOS CSS = Autism Diagnostic Observation Schedule Calibrated Severity Score CBCL = Child Behavior Checklist; I = Internalizing; E = Externalizing; M1 = ADOS Module 1; M2 = ADOS Module 2; M3 = ADOS Module 3