Published in final edited form as:
J Autism Dev Disord. 2014 August ; 44(8): 1996–2012. doi:10.1007/s10803-014-2080-3.

The Autism Diagnostic Observation Schedule, Module 4: Revised Algorithm and Standardized Severity Scores

Vanessa Hus, MSc[1] and **Catherine Lord, PhD**[2]

[1]Department of Psychology, University of Michigan; Ann Arbor, MI; USA

[2]Center for Autism and the Developing Brain, New York Presbyterian Hospital & Weill Cornell Medical College; White Plains, NY; USA

Abstract

The Autism Diagnostic Observation Schedule, 2nd Edition includes revised diagnostic algorithms and standardized severity scores for modules used to assess children and adolescents of varying language abilities. Comparable revisions have not yet been applied to the Module 4, used with verbally fluent adults. The current study revises the Module 4 algorithm and calibrates raw overall and domain totals to provide metrics of ASD symptom severity. Sensitivity and specificity of the revised Module 4 algorithm exceeded 80% in the overall sample. Module 4 calibrated severity scores provide quantitative estimates of ASD symptom severity that are relatively independent of participant characteristics. These efforts increase comparability of ADOS scores across modules and should facilitate efforts to increase understanding of adults with ASD.

Keywords

Autism Spectrum Disorders; Autism Diagnostic Observation Schedule; Adults; Severity

The Autism Diagnostic Observation Schedule, Module 4: Revised Algorithm and Standardized Severity Scores

Defined by impairments in social-communication and the presence of restricted, repetitive, stereotyped behaviors and interests (American Psychiatric Association [APA] 2013), autism has historically been considered a childhood disorder. However, studies of young adults with autism spectrum disorders (ASD) report variable outcomes (see Levy & Perry 2011). Only 3 to 25% of individuals with ASD reportedly achieve “optimal outcomes” (generally referring to cognitive and adaptive abilities falling within the “average” range and the ability of these individuals to function independently in the community; Helt et al. 2008; Levy &

Correspondence concerning this article should be addressed to: Vanessa Hus, University of Michigan Department of Psychology, 530 Church Street, Ann Arbor, MI 48109; USA, Phone: 734.273.9883, vhus@umich.edu.

At the start of this study, Vanessa Hus and Catherine Lord were at the University of Michigan Department of Psychology and University of Michigan Autism & Communication Disorders Center in Ann Arbor, Michigan.
Vanessa Hus has remained at the University of Michigan Department of Psychology. Catherine Lord is now at the Center for Autism and the Developing Brain, New York Presbyterian Hospital & Weill Cornell Medical College, White Plains, New York.

Disclosure
CL receives royalties for the ADI-R and ADOS-2; profits from this study were donated to charity.

Perry 2011). Advances in early detection and intervention may contribute to higher rates of optimal outcomes in the future. Nonetheless, at present, the majority of individuals diagnosed with ASD require varying levels of life-long supports.

In the fall of 2010, 369,774 American children ages 6 through 21 received services under the special education classification of “Autism” (U.S. Department of Education, Office of Special Education Programs, Data Analysis System 2011). This provides a conservative estimate of the number of children who will be transitioning to adulthood over the next decade, as not all children diagnosed with ASD receive special education services. As a further reminder of this growing public health issue, the per capita lifetime incremental cost of autism is estimated at $3.2 million. Twenty-one percent is attributed to care for the adult with ASD and 30.7% to loss of the individual with ASD’s productivity during adult life (Ganz 2007).

These numbers underscore the pressing need for research to better understand the strengths and difficulties of adults with ASD, as well as factors that promote more positive outcomes (Interagency Autism Coordinating Committee 2011). Such information is critical to both develop and expand the available services and supports for adults, as well as to inform earlier intervention efforts and preparatory activities for the transition to adulthood that will promote positive long-term outcomes. With the help of longitudinal studies, we may begin to investigate trajectories in ASD that will inform prognosis when a child is diagnosed. Furthermore, a better understanding of the ASD phenotype is necessary for investigations seeking to link behavioral symptoms to differences in brain structure and function, which may contribute to the development of targeted interventions.

Recent studies have begun to explore predictors of outcome and the current needs of adolescents and adults with ASD (e.g., Farley et al. 2009; Howlin, Moss, Savage & Rutter 2013; see Henninger & Lounds-Taylor 2012 for review). Examination of development has been mostly limited to measures of cognitive and adaptive behavior, global ratings of outcome derived by authors and change in diagnostic classification on measures such as the *Autism Diagnostic Observation Schedule* (*ADOS*; Lord, Rutter, DiLavore, & Risi 1999) and *Autism Diagnostic Interview-Revised* (*ADI-R;* Rutter, LeCouteur, & Lord 2003). Although such analyses are informative, in order to fully understand the life course of the disorder, examination of trajectories in ASD symptom severity and how early ASD symptom severity predicts longer-term outcomes is needed. Several studies have used raw totals from measures such as the *ADI-R* or the *Childhood Autism Rating Scale* (*CARS;* Schopler, Reichler, DeVellis, & Daly, 1980) to predict adult outcomes or investigate change in ASD symptoms (e.g., Anderson, Liang & Lord, 2013; Eaves & Ho, 2008; Fein et al., 2013; Gillespie-Lynch et al. 2012; Howlin et al. 2013; Sigman & McGovern 2005; Piven, Harper, Palmer & Arndt 1996; Shattuck et al. 2007; see Levy & Perry 2011 for review)). However, scores from these measures may be confounded by strong associations with developmental level (e.g., Hus & Lord 2012; Perry, Condillac & Freeman, 2002) and reporting biases (Hus, Taylor & Lord 2007; Jones et al. 2013). Increasing the availability of instruments based on standardized protocols of observation that are less influenced by these factors, and that can be used to explore trajectories in ASD symptoms, will allow for a more thorough investigation of factors that predict adult outcomes.

Many longitudinal studies of ASD have included the *ADOS*, making it possible to use *ADOS* scores to examine developmental trajectories of ASD symptoms. Diagnostic algorithms for *ADOS* Modules 1–3, used to assess children and adolescents of varying language levels, were recently revised (*ADOS-2*; Lord, Rutter, DiLavore, Risi, Gotham, & Bishop 2012). Changes afforded improved diagnostic validity and increased item overlap across modules, thereby facilitating comparisons of scores across childhood and early adolescence. Furthermore, the revised algorithms were divided into two domains, consistent with the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA 2013) ASD diagnostic criteria: *Social Affect* (which combines social and communication behaviors) and *Restricted and Repetitive Behaviors*. There is a single diagnostic cut-off for the combined domain total. Algorithms for the recently published *ADOS-Toddler* module (Lord, Luyster, Gotham & Guthrie 2012) follow a similar structure to that of Modules 1–3, making it possible to examine symptom trajectories using the ADOS from a very young age.

Although the algorithms for Modules 1–3 were revised to be more comparable across modules, raw totals were significantly influenced by age and language level. Higher scores were associated with less language (i.e., Module 1>Module2>Module3). In addition, within Modules 1 and 2, older children tended to have higher scores than younger children (Gotham et al. 2007; Gotham, Pickles & Lord 2009). Intelligence, language and age are certainly important factors to consider when describing an individual's level of overall functional impairment. However, scores confounded by these issues make it difficult to examine the relative severity of autism-specific symptoms and cloud the interpretation of findings as specific to ASD. For example, an association between *ADOS-2* raw totals and a specific chromosomal abnormality may be misinterpreted as evidence of a causal mechanism for ASD when in fact it is a marker for more general cognitive impairment. To address this issue, Gotham and colleagues (2009) used a large sample of children with ASD to create the *ADOS-2* Calibrated Severity Metric (CSS), a standardized version of *ADOS-2* scores that is less strongly associated with age and language compared to raw *ADOS-2* totals. Importantly, the CSS is intended to be used a marker of ASD symptom severity relative to age and language level (not as an indicator of functional impairment). The CSS allow *ADOS-2* scores to be used to investigate relationships between behavioral symptoms and underlying biological mechanisms and study stability and change in symptom severity over time and across individuals of different developmental levels. Clinicians may also use the CSS to describe severity of a client's core symptoms relative to his or her age and language level (i.e., ADOS Module). However, for an adequate description of functional impairment, one needs to move beyond the CSS to include information regarding language, adaptive functioning and other behavioral features.

The diagnostic algorithm for *ADOS* Module 4, designed for assessment of verbally fluent adults, has not yet been revised. Recent studies have supported the validity of the original Module 4 algorithm (e.g., Brugha et al. 2011). However, the lack of comparability with algorithms for the *ADOS-Toddler* Module and Modules 1–3, along with, the absence of calibrated severity scores for Module 4, hinders comparisons of adult assessments to those conducted in childhood and early adolescence. Given that Module 4 is used with verbally fluent adults who are likely to demonstrate a wide range of abilities, revisions are particularly important to investigations seeking to understand how ASD severity interacts with other factors, such as verbal, cognitive, and adaptive ability, to predict functional outcomes. Thus, the purpose of this study is to revise the *ADOS* Module 4 algorithm to be more comparable to currently used algorithms for *ADOS-2* Modules 1–3 and to provide a calibrated score that can be used to quantify and compare the severity of core symptoms in adults with ASD.

Methods

Participantsh

The sample included data from 393 different *individual participants*. Some participants had repeated assessments, yielding a total of 437 *cases*. Each case was defined by an *ADOS* and best estimate clinical diagnosis; 29 participants provided data for multiple cases (*M*=2.52, *SD*=1.06, range=2–6) based on evaluations conducted at different points in time. The majority of participants (*n*=319) were research participants and clinic referrals for assessment of possible autism to the University of Chicago Developmental Disorders Clinic (UCDDC), the University of Michigan Autism and Communication Disorders Center (UMACC), or the New York Presbyterian Center for Autism and the Developing Brain (CADB) at Weill-Cornell. Seventy-four participants were evaluated as part of the Simons Simplex Collection (SSC; Fischbach & Lord 2011), a multi-site genetic study. Approximately 80% of the sample was male and 83% Caucasian. Inclusion/Exclusion criteria varied by research study. However, individuals with significant hearing, vision or motor problems that interfered with standardized testing or who were exhibiting active psychosis or uncontrolled seizures at the time of assessment were excluded from each study. Participants in the SSC were also required to meet Collaborative Programs for Excellence in Autism criteria for ASD (see Hus et al., 2013 for details) and were excluded if the individual had a diagnosis of Fragile X syndrome, tuberous sclerosis, Down syndrome or significant early medical history. Additionally, SSC participants could not have any first, second or third degree relatives with ASD or a sibling with substantial language or psychological problems related to ASD. Ages ranged from 9.92 to 62.25 years at the time of assessment (mean=21.56, standard deviation=8.62 years).

Of the 437 *cases*, 177 had clinical diagnoses of autism (40% of entire sample), 170 Other-ASD (i.e., PDD-NOS or Asperger's; 39%), and 90 Non-ASD diagnoses (21%). The Non-ASD sample was comprised of both clinical referrals and individuals recruited to research studies as controls. In addition to having first ruled-out an ASD diagnosis, 84% of non-ASD participants received a primary diagnosis of a non-ASD DSM-IV-TR disorder; 30% had a primary diagnosis of mood and/or anxiety disorders, 26% had non-specific intellectual disability, 14% had externalizing behavioral disorders (e.g., ADHD/ODD), 5% had Down syndrome or Fragile X, 4% had language disorders, 1% had Fetal Alcohol syndrome, 1% had Cerebral Palsy and 3% of cases had unspecified difficulties. The remaining 16% of Non-ASD sample did not meet criteria for a DSM-IV-TR diagnosis at the time of assessment; 64% of these individuals (n=9) had had a previous diagnosis of ASD and 36% (n=5) had had a previous Non-ASD diagnosis. There was no significant difference in *ADOS* totals between the 9 individuals with previous ASD diagnoses and the remaining non-ASD group (data available from authors upon request). Table 1 provides a more detailed sample description.

Measures

Autism Diagnostic Observation Schedule (ADOS)—The *ADOS* (Lord et al., 2000) is a standardized, semi-structured observational assessment used to assess communication, reciprocal social interaction, imagination/creativity and stereotyped behaviors and restricted interests to inform diagnosis of autism spectrum disorders. The *ADOS* is organized into four modules based on the individual's chronological age and expressive language level, ranging from preverbal to verbally fluent. Module 4 was designed for use with older adolescents and adults with fluent speech. The original Module 4 diagnostic algorithm provides separate cut-off values for the Communication and Social domains, as well as a cut-off for the sum of the two domains to provide instrument classifications of Autistic Disorder or Autism Spectrum Disorder. In the original validation study, internal consistency for all modules ranged from α=.74-.91 for the Communication and Social domains totals and α=.47 to .65 for the Stereotyped Behaviors and Restricted Interests domain totals. Interrater reliability of all Module 4 items exceeded 80% exact agreement, with $\kappa \geq .60$ for most items.

All *ADOSes* were administered and scored by a clinical psychologist or trainee (e.g., graduate student or research assistant) who met standard requirements for research reliability. Inter-rater reliability on the *ADOS* was monitored through joint administration and scoring by two different examiners or scoring of videotapes in 11% of cases. Exact item agreement was initially established at 80% and consistently exceeded 75%. Disagreements were resolved through discussion and consensus codes were used for analyses. Within this sample, 54 different examiners collected the data from the *ADOS* over 18 years (1994–2012).

Cognitive Assessment—Intelligence quotients (IQ) were derived from a developmental hierarchy of cognitive measures; most frequently the *Wechsler Abbreviated Scale of Intelligence* (Wechsler, 1999) and the D*ifferential Ability Scales* (Elliott 1990; Elliott 2007). The *Peabody Picture Vocabulary Test* (Dunn & Dunn 2007) and *Ravens' Progressive Matrices* (Raven 1960) were also sometimes used to estimate verbal IQ and nonverbal IQ, respectively.

Procedure

The ADOS was conducted as part of a clinical or research evaluation. Most commonly, evaluations began with the collection of a developmental history using the *Autism Diagnostic Interview- Revised*; Rutter, LeCouteur & Lord 2003 or the *Social Communication Questionnaire*; Rutter, Bailey, Lord & Berument 2003). This was then followed by direct assessment of the individual consisting of psychometric testing and the *ADOS* (see Gotham et al. 2009 and Fishbach & Lord 2011 for more detailed procedures). At the clinics (UCDDC, UMACC, CADB), best estimate clinical diagnoses based on DSMIV-TR criteria (APA 2000) were made by a supervising clinical psychologist, child psychiatrist and/or advanced graduate student after review of all assessment information. For the SSC, best estimate clinical diagnoses were assigned by an experienced clinician after reviewing all available information and viewing the child in person or on video (see Lord et al. 2011). The ADI-R or SCQ was available for 86% of cases. Verbal and/or nonverbal IQ estimates were available for 361 participants (91%). Clinic-referred participants received oral feedback and a written report of results without financial compensation. Participants recruited solely for research purposes received a written summary of results and financial compensation. Institutional review boards at all sites approved all procedures related to this project.

Design and Analysis

Analyses used to revise the Module 4 algorithm followed a similar procedure to that described for derivation of new algorithms for Modules 1 through 3 (Gotham et al. 2007). Calibration of Module 4 raw totals also followed the procedure described for standardization of the overall (Gotham et al. 2009) and domain (Hus et al. 2013) totals for Modules 1–3. These will be described below.

Analysis of Original Module 4 Algorithm—Item scores of 3 were collapsed with 2s. Domain total distributions were examined for floor or ceiling effects; variables contributing to the effects were identified. Correlations between Module 4 totals and participant characteristics (age and IQ) were examined to inform the need for different algorithms based on age or ability level. Item distributions were examined to select those items that best differentiated ASD vs. Non-ASD diagnoses. Two items, *Shared Enjoyment in Interaction* and *Self Injurious Behavior* were not included in an early, prepublication version of the *ADOS* and were missing data for some participants (n=56 and n=49, respectively). Because these items were not included in either the original or the revised algorithm, these participants were maintained in the dataset. The items were treated as missing for item-level analyses.

Development and Analysis of New Module 4 Algorithm—Social-communication items were labeled as "preferred" if no more than 20% of autism cases scored a zero and no more than 20% of Non-ASD cases scored a 2 or a 3. Inclusion criteria were not applied to RRB items. Although the presence of RRBs in a non-ASD participant may be clinically meaningful, the absence of RRBs during the *ADOS* is more difficult to interpret (i.e., while some *ADOS* "presses" may elicit RRBs from some individuals, the time-limited, standardized context of the *ADOS* may limit the number or intensity of RRBs exhibited by others). Exceptions were allowed for four items that were theoretically important and overlapped with items appearing on algorithms for other modules in order to promote conceptual uniformity that would enhance inter-module comparisons (Gotham et al. 2007). These items performed just outside one of the thresholds: *Unusual Eye Contact (15.8% Autism cases scored '0'; 32.2% Non-ASD cases scored a '2'), Emphatic Gestures (25.4%; 11.1%); Communication of Own Affect (33.9%; 8.8%); Amount of Reciprocal Social Communication (36.7%; 1.1%)*. One item that met preferred item criteria, *Responsibility*, was excluded due to high rates of nonzero scores (i.e., '1' and '2') in the Non-ASD group, which resulted in reduced specificity when the item was included.

Preferred items were entered into ordinal probit item response models, run with Mplus Version 6.1 software. This method of exploratory factor analysis was chosen to account for the ordinal nature of *ADOS* data. Factor loadings from promax oblique rotations were used to inform organization of items into domains. Root Mean Square Error Approximation (RMSEA) of 0.08 or less was used to indicate satisfactory fit (Browne & Cudeck 1993). Goodness-of-fit was assessed using confirmatory factor analysis. Comparative Fit Index above 0.9 was used to indicate a good fit (Skrondal & Rabe-Hesketh 2004). Logistic regression was used to examine the contributions of each domain to predict diagnosis.

Distributions of domain totals for the new algorithm were perused for floor and ceiling effects. Item correlations with the remainder of the domain (i.e., minus the item) and with participant characteristics (e.g., age and IQ) were examined. ROC curves were computed and the sensitivity and specificity of the revised algorithm was compared to that of the existing algorithm. Four participants were missing data from items used to compute the new algorithm total. These missing items were imputed using the average of the remaining domain items only for the purpose of computing the diagnostic algorithm.

Development and Analysis of Module 4 Overall Total and Domain Calibrated Severity Scores—Only scores from individuals with ASD diagnoses were used for raw score standardization. This included all assessments with a corresponding best estimate clinical diagnosis of Autism or Other-ASD. Participants were divided into 1-year chronological age groups to ensure that distributions of the overall and domain scores were comparable and then collapsed into 8 age groups based upon similar distributions (see Figure 2). Eventually, all participants aged 9–29 were collapsed into a single group because of similar distributions and minimal correlations between raw scores and age. Participants aged 40 and above were excluded from standardization due to differences in distribution compared to younger ages and too few older participants to calibrate separately.

As described by Gotham et al. 2009, for the overall total calibration, raw totals corresponding to a Module 4 *ADOS* diagnostic classification of “ASD” were mapped on to CSS ranging from 4 to 10. Totals corresponding to a “Nonspectrum” classification were distributed across CSS of 1–3. Ranges of raw totals assigned to each point on the 10-point severity scale were determined by the percentiles of available data within that classification range. Because there are not separate SA and RRB cut-offs for *ADOS* classifications, the percentiles used for mapping the overall totals (i.e., SA+RRB) were used to inform the mapping of the raw SA and RRB totals to each respective domain CSS. As with Modules 1–3, raw RRB domain totals were mapped on to CSS values of 0 and 5–10, due to the limited range of RRB raw totals (i.e., 0–10; Hus et al. 2013). Confidence intervals were computed for each scale as 95% [CSS +/− 1.96*(*SE*)], where $SE = SD * \sqrt{1-\alpha}$ (Brown 1999).

Linear regression analyses were conducted using the ASD participants who had contemporaneous demographic data to examine the influences of participant characteristics on raw totals and on the calibrated domain scores. Participants’ verbal and nonverbal IQs were entered into the first block, followed by participant chronological age, gender, maternal education and race in the second block. Only model R^2 are reported due to multicollinearity which limits interpretability of individual coefficients. Where there was more than one significant predictor, Forward Stepwise models were used to assess the relative contributions of those predictors in predicting raw totals and calibrated domain scores.

Four cases for whom longitudinal data were available were selected to demonstrate the utility of the calibrated total and domain scores for examining trajectories of ASD symptoms across the lifespan.

Results

Analysis of Original Module 4 Algorithm

Domain Total Distributions—First, domain total distributions were examined for each domain separately to assess the utility of the original Module 4 in discriminating between diagnostic groups and in describing severity of core symptoms within ASD participants. As expected, original Module 4 Communication totals differed significantly by diagnostic group; $F(2,434)=87.04$, $p<.001$. Participants with Autism had significantly higher scores than participants with Other-ASD diagnoses and both ASD groups had higher scores than the Non-ASD cases. (See Table 1.) As shown in Figure 1, Communication domain scores in the ASD sample were roughly normally distributed, with a slight right-skew; totals of 2–4 were the most frequent (22.5–23.1% each). Maximum Communication totals of 7 or 8 were rarely observed (a total of 2.9% of Module 4 ASD cases received either score). For 3 of the 4 items comprising the Communication domain (*Stereotyped/Idiosyncratic Use of Words or Phrases*, *Conversation* and *Emphatic or Emotional Gestures*), scores of ‘1’ were given in 45–56% of ASD cases. Over half of the participants scored a ‘0’ on the fourth item, *Descriptive, Conventional, Instrumental, or Informational Gestures*.

Social totals also differed across diagnostic groups, such that the Autism group scored significantly higher than the Other-ASD group which scored significantly higher than the Non-ASD group, $F(2,434)=128.65$, $p<.001$ (see Table 1). As shown in Figure 1, Social domain scores were approximately normally distributed in the ASD participants; Social totals of 6–8 were most frequent (12.1–16.1% each). Fewer than 9% of ASD cases received maximum Social domain scores of 12–14. This is partially explained by a high frequency of ‘1’ scores (i.e., 69–71%) for several items (*Facial Expressions Directed to Others, Quality of Social Overtures*, *Quality of Social Response*). The remaining 3 items (with the exception of *Unusual Eye Contact*, which does not have the option of a ‘1’) had 39–44% of cases with ‘1’.

Correlation with Participant Characteristics—Next, correlations between domain totals and participant characteristics were examined to inform the potential need for creation of algorithms based on ability level or age (i.e., as with the “Younger than 5 years” and “Greater or Equal to 5 years” algorithms for Module 2). Dividing Module 4 recipients by language level was not helpful because of limited variability in the *Overall Level of Language* item (88% scored a ‘0,’ indicating that the participant “Uses sentences in a largely correct fashion”). Divisions of groups according to other items were similarly unhelpful in Module 3 (Gotham et al. 2007). Among ASD participants, correlations between Social-Communication totals and verbal ($r=-.28$; n=324; $p<.001$) and nonverbal IQ ($r=-.21$; n=314; $p<.001$) were significant, but weak (Cohen 1988). Correlations were further reduced when only individuals without intellectual disability were included (i.e., IQ ≥ 70; n=303; VIQ: r=−.17, p=.003; NVIQ: r=−.09, p=.25). When the sample was limited to individuals with verbal IQ ≥ 85 (n=259), there was not a significant relation with IQ, even though this still represented a substantial range (85–148). Social-Communication totals were not significantly associated with age (r=−.07, p=.186).

Analysis to Develop New Module 4 Algorithm

Exploratory Factor Analysis—Exploratory Factor Analysis was performed with all preferred items included. As shown in Table 2, a 2-factor solution fit well, consistent with other modules (Gotham et al. 2007). Ten items loaded on to the Social Affect (SA) factor and 5 items loaded on the Restricted and Repetitive Behaviors (RRB) factor. Factors were significantly correlated (r=.46). As on Modules 1–3, some items from the Communication domain emerged on the SA domain (*Conversation and Emphatic Gestures*) and others on the RRB domain. *Stereotyped/Idiosyncratic Use of Words or Phrases* loaded on to the RRB factor, as had been the case for Modules 1–3. In addition, *Speech Abnormalities Associated with Autism*, previously a Communication domain item, also loaded on to the RRB factor. *Unusual Eye Contact* loaded solidly on both factors. When the sample was reduced to only participants with ASD, *Unusual Eye Contact* again loaded on both, but more strongly on the Social Affect factor (.43) compared to the RRB factor (.26). Because of the theoretical significance of this item and its inclusion as part of the SA domain for Modules 1–3 (Gotham et al. 2007) and the *ADOS*-T (Luyster et al. 2009), it was maintained in the same domain for Module 4.

Confirmatory factor analysis with each item assigned to one of two factors indicated good fit (CFI=.93); the 2-factor solution was a better fit than the 1-factor model (CFI=.91).

Logistic Regression Check on Weighting Domains—Logistic regression for ASD (i.e., Autism, PDD-NOS and Aspergers) versus Non-ASD cases indicated that both SA (B=.37, SE=.06, z=6.20, Exp(B)=1.45) and RRB (B=1.35, SE=.21, z=6.52, Exp(B)=3.85) totals were predictive of diagnosis. These results suggest a larger effect of RRB domain totals on predicting diagnosis compared to SA totals.

Correlations between Domain Totals, Items and Chronological Age and IQ—Correlations between each algorithm item and the domain scores minus that item were significant. Correlations ranged from r=.40 to r=.71 for the SA domain and from r=.25 to r=.57 for the RRB domain. Domain totals were also significantly correlated (r=.48). Internal consistency, measured by Cronbach's alpha (Cronbach 1951), was comparable to other modules for both the SA (α=.84) and RRB (α=.61) domains.

In contrast to the original Module 4 algorithm totals, the new Module 4 algorithm total (SA +RRB) demonstrated a significant, albeit weak, correlation with age (r=−.21, p<.001) but not with verbal or nonverbal IQ. New SA and RRB domain totals were also weakly correlated with age (r=−.20 and r=−.15, respectively, p<.001), but not IQ.

Correlations between each of the items comprising the new algorithm and chronological age and IQ were also examined. Ten items were significantly correlated with age, ranging from

r=−.10 (*Amount of Reciprocal Social Communication)* to r=−.19 (*Quality of Social Overtures)*. Only *Communication of Own Affect and Insight* were significantly, though weakly, correlated with verbal IQ (r=−.13, r=−.26, respectively) and nonverbal IQ (r=−.13, r=−.19).

Sensitivity and Specificity Comparison

Receiver Operating Characteristic (ROC) curves were computed to provide information regarding where to set cut-offs in order to maximize sensitivity and specificity of the old and new algorithms. For the new algorithm, ROC curves were run separately for the SA total and the combined SA + RRB total. Based on analyses indicating correlations between previous algorithm totals and IQ, sensitivity and specificity for the combined SA+RRB total was also examined for three verbal IQ groups (below average (<85), average (85–115), above average (>115). As in the past, scores of 3 were recoded to 2 for this procedure.

As shown in Table 3, the new algorithm performed better than the old algorithm. The combined SA+RRB total yielded somewhat higher sensitivity and considerably higher specificity than the SA total alone, both in the overall sample, as well as in each of the three IQ groups. Specificity was also generally higher in the new algorithm compared to the old algorithm, with the exception of the average IQ group. The difference here (i.e., between specificity of 77 vs. 82%) was accounted for by 2/44 fewer Non-ASD participants accurately classified by the revised algorithm compared to the old algorithm.

Development of Calibrated Severity Score

Examining overall total and domain score distributions across age groups— Although correlations between new algorithm totals and age were weak, score distributions were examined across age groups to confirm whether there was need for age-based calibration cells. As shown in Figure 2a, c, and e, distributions of total and domain scores were relatively similar across age groups, with the exception of the oldest two age groups (40–49 and 50–59 years), which included a total of only 9 ASD participants. Nonetheless, given the larger number of algorithm items in Module 4 (15) than Modules 1–3 (14), a calibrated severity metric was warranted in order to allow for comparison across modules. As such, it was decided to create a single calibrated severity metric for all Module 4 participants aged 9–39 years, including a total of 338 participants with ASD for calibration. Table 4 shows the mappings of raw overall, SA and RRB totals to the 10-point severity scale and confidence intervals for each scale.

Comparison of New Module 4 Algorithm Totals and Calibrated Domain Scores Across Age Groups—As shown in Figure 2b, d, and f, distributions of overall calibrated scores remained relatively comparable across age groups. The 19–20 year old group had a somewhat narrower distribution of overall CSS compared to the other ages. Examination of SA-CSS and RRB-CSS distributions suggest this may be due to the higher proportion of 19–20 year olds that exhibited few repetitive behaviors during the *ADOS* (4% with raw RRB-CSS of 1, reflecting no RRB during the *ADOS*; 17% with RRB-CSS of 5, reflecting that RRBs during the *ADOS* were rare and unclear).

As noted above, the *ADOS* classification, based on the raw overall total (SA+RRB) was used to anchor the raw total-to-overall severity score mapping (i.e., "Autism Spectrum" classification mapped to CSS of 4–10). Using this approach, mean overall CSS distinguished between individuals grouped by clinicians' best estimate clinical diagnoses (i.e., Autism vs. Other-ASD vs. Non-ASD diagnoses); $F(2,409)=191.45$, $p<.001$.

Next, percentiles from the raw total-to-overall CSS mapping were used to map raw domain totals-to-domain severity score mappings. Ninety-nine percent of participants with an *ADOS* classification of "Autism Spectrum" had an SA-CSS of 4 or higher and 76% of individuals with an *ADOS* classification of "Nonspectrum" had an SA-CSS less than or equal to 3. With regard to RRB-CSS, 84% of participants with an *ADOS* classification of "Autism Spectrum" had an RRB-CSS of 6 or higher and 79% with a "Nonspectrum" classification had an RRB-CSS of 5 or lower. Mean domain CSS also distinguished between best estimate clinical diagnostic groups (SA-CSS: $F(2,409)= 132.68$, $p<.001$; RRB-CSS: $F(2,409)= 150.63$, $p<.001$). Nonetheless, there was marked overlap in the distribution of scores (see Figure 3).

Correlations Between Domain Calibrations and Overall Calibrated Score—In the ASD sample, associations between SA-CSS and RRB-CSS were significant, but weak ($r=.26$). Although correlations between each of the domain calibrated scores and the overall CSS were strong, the association between SA-CSS and CSS ($r=.90$) was greater than that observed for RRB-CSS and CSS (r=.60).

Comparison of Raw Overall and Domain Totals to Calibrated Scores—The final model including all predictors explained a total of 9% of variance in the overall Raw Total. Verbal IQ emerged as the only significant predictor of the Raw Total. The same model accounted for 7% of the variance in the overall CSS, with verbal IQ again making the only significant contribution to CSS. As there was only one significant predictor of the overall raw total and CSS, a Forward Stepwise model was not run.

Participant characteristics explained 10% of the variance in the raw Social Affect total, with Verbal IQ and chronological age the only significant predictors. This model accounted for approximately 8% of variance in the SA-CSS, with Verbal IQ and chronological age again emerging as significant predictors. Verbal IQ and age were entered into a Forward Stepwise model to assess the relative contributions of each of these variables in predicting SA-CSS. In the forward model, Verbal IQ explained 4.9% of the variance; $F(1,270)=13.89$, $p<.001$. Age was excluded by the model, indicating that it was not significant.

In the model predicting RRB-Raw, participant characteristics accounted for 5% of the variance. Race (Caucasian vs. all others) was the only significant predictor of RRB-Raw. The overall model predicting RRB-CSS also accounted for approximately 5% of the variance, with race remaining the only significant predictor reflecting somewhat higher RRB-CSS for Caucasian participants compared to participants of other races.

Case Summaries

Four children were chosen to demonstrate the utility of the newly calibrated Module 4 scores for examining severity of ASD symptoms over time (see Table 5 for child characteristics at first and last assessments). Each child's overall CSS, SA-CSS, and RRB-CSS are plotted by age in Figure 4.

Case 1. "John," a Caucasian male, was seen at 2 years of age as part of a clinical research study. He received a diagnosis of Autism at that time. During his first *ADOS*, John exhibited severe social-communication symptoms (i.e., an SA-CSS=9) and somewhat milder repetitive behaviors (RRB-CSS=7). A period of stability in toddlerhood was marked by limited use of nonverbal communication behaviors (i.e., gestures, facial expressions and eye contact) and poor quality social overtures. John showed an apparent decrease in severity of social-communication symptoms at age 10, reflecting an increased range of gestures and facial expressions and improved quality of overtures. At age 18, however, John's SA-CSS increased by one point. At this age, he maintained gains in nonverbal communication, but exhibited more unusual overtures and responses. In contrast, John's severity of RRB symptoms increased from ages 2 to 3, and then remained relatively stable across childhood and adolescence. Scores reflect persistent sensory interests, preoccupations with objects and references to highly specific topics during the *ADOS*. In John's case, the overall CSS showed an apparent increase in symptoms during toddler years, followed by apparent stability across childhood and adolescence. It did not capture the possible divergence in social-communication and repetitive behavior symptom severity in middle childhood.

Case 2. "Parker," a Caucasian male, was diagnosed with Autism at 2 years of age when seen as part of a clinical research study. At first assessment, Parker showed severe repetitive behaviors (RRB-CSS= 10) and moderately severe social-communication symptoms (SA-CSS=7). However, his overall CSS demonstrated an apparently steady decrease in symptoms across childhood and adolescence. Separate examination by domain indicates that Parker demonstrated a decrease in repetitive behaviors (i.e., sensory interests and unusual preoccupations) during his *ADOS* at age 3, but persistent challenges in social-communication behaviors (e.g., limited eye contact, facial expressions and mildly inappropriate social overtures). Through middle-childhood, the severity of his repetitive symptoms appeared stable. However, he showed significant improvements in social-communication, particularly in nonverbal communication and quality of social overtures. By his final assessment at age 18, Parker demonstrated only subtle difficulties with social reciprocity and no repetitive behaviors.

Case 3. "Emily," an African American female, was seen as part of a clinical research study and diagnosed with Autism just before her 3rd birthday. Emily's overall and domain CSS scores followed very similar trajectories, demonstrating relative stability of symptom severity in toddlerhood, followed by an apparent decrease in symptoms at age 5. At age 11, she showed a possible worsening of both social-communication and repetitive behaviors; however, at her last assessment just before her 20th birthday, severity scores again decreased, returning to the milder range that was observed at age 5. The apparent increase in symptoms at age 11 may reflect a shift in her skill set. During her first three assessments, the Module 1, No Words algorithm was employed and Emily demonstrated improvements in initiation of social-communication behaviors, such as pointing, showing and initiation of joint attention, and a decrease in repetitive interests. At age 11, Emily gained sufficient language to be assessed with a Module 3, during which she exhibited limited social reciprocity, inappropriate overtures, repetitive speech and frequent references to unusual or highly specific topics. These symptoms were notably improved at her last assessment.

Case 4. “Robert,” an African American male was assessed at age 3 for a clinical research study and diagnosed with PDD-NOS. Robert’s overall CSS suggests a steady worsening of symptoms across childhood and adolescence. Separate examination of symptoms by domain indicates that Robert exhibited an increase in social-communication severity (SA-CSS=4 at age 3 and 10 at age 19). Across time, Robert exhibited limited eye contact and mildly unusual social overtures. As he grew older, his social overtures and social reciprocity decreased and he displayed a flatter affect than previously observed. In contrast, Robert’s repetitive behaviors remained relatively stable (RRB-CSS=5 at ages 3 and 19 reflecting mild speech abnormalities and brief repetitive interests), with the exception of his age 11 *ADOS*, during which he did not exhibit any repetitive behaviors.

Discussion

In the current study, the original Module 4 algorithm domain totals discriminated between diagnostic groups (i.e., Autism vs. Other-ASD vs. Non-ASD) and provided good sensitivity (89.6%) and adequate specificity (72.2%). However, the somewhat restricted range of the Communication and Social domain totals suggested that the original Module 4 algorithm was not the best combination of items to describe severity of core symptoms within the ASD group. Although items were originally designed to capture a higher proportion of ‘1’ scores (i.e., approximately 50%), the finding that several items received scores of ‘1’ more than two-thirds of the time supported the need to consider items with more variability in the creation of new algorithms.

This larger, more diverse sample provided the opportunity to revise the Module 4 algorithm using items corresponding or equivalent to the revised algorithms for Modules 1–3 (Gotham et al., 2007). The two-domain Module 4 algorithm is consistent with DSM-5 criteria for ASD. Moreover, the addition of the RRB domains improves the diagnostic utility of the *ADOS* to discriminate between individuals with ASD and Non-ASD diagnoses. The new algorithm yields improved sensitivity and specificity, both above 80% in the overall sample.

In light of changes in DSM-5 ASD diagnostic criteria, a single cut-off score that yields a good combination of sensitivity and specificity is provided to differentiate between ASD and Non-ASD classifications. This contrasts to other modules, which have separate cut-offs available for Autism and ASD. For researchers who may be interested in achieving a higher level of specificity, at the cost of somewhat lower sensitivity (i.e., equivalent to an *ADOS-2* classification of “Autism” only on other modules), a cut-off of 10 may be useful. This cut-off yields an overall specificity of 91.1%, but there is also a sizeable reduction in sensitivity (79.3% overall; 71.3% for individuals with above average IQs; data not shown, additional information available from authors upon request). It is also noteworthy that, although DSM-5 criteria now require that an individual exhibit deficits in both social-communication and restricted and repetitive behaviors, separate domain cut-offs are not provided on the *ADOS* algorithm. As noted above, the time-limited, standardized nature of the *ADOS* may influence the extent to which some individuals exhibit RRBs during the assessment. Thus, it is likely that implementing a separate RRB-domain cut-off would reduce sensitivity of the instrument (i.e., some individuals with ASD will exhibit few or no RRBs during the 40 minute observation period). For example, in this sample, the SA cut-off of 6 shown in Table 3 and an RRB cut-off of 1, specificity improves to 96.7%; however, sensitivity is reduced to 75.5% [data not shown, additional information available from authors upon request]. Nonetheless, inclusion of both domains (SA+RRB) in the overall total results in considerably better specificity (and somewhat higher sensitivity; see Table 3) than relying on the SA domain alone. While the *ADOS* is not designed for use in isolation as a DSM “checklist” to determine a clinical diagnosis of ASD, it provides highly valid instrument classifications and a useful context in which to observe behaviors relevant to clinical diagnosis.

The new Module 4 algorithm totals were weakly, but significantly correlated with age. In the cross-sectional design of this study, it is not clear if these differences are due to recruitment effects or if they reflect true developmental variation. The new algorithm totals were not significantly correlated with IQ. However, correlations between previous algorithm totals and IQ suggested a need to examine performance across different levels of cognitive ability. Although an estimate of IQ or developmental level is recommended as a key part of a diagnostic evaluation (Hus & Lord 2011), we acknowledge that such information is not always available. As such, it did not seem practical to make separate algorithms for individuals of different cognitive levels, all of whom had fluent, complex language. Overall, IQ appears to be considerably less influential for Module 4 scores than was observed for Modules 1–3 (Gotham et al. 2007). Nonetheless, researchers are encouraged to be attentive to the fact that specificity may be poorer in individuals with average IQs. Because the Non-ASD samples within IQ ranges are relatively small, it is difficult to know exactly what this means. It is notable that, though this group demonstrated lower specificity than other IQ groups for Module 4, specificity was actually comparable to that observed for most other algorithms (i.e., Modules 1–3). This contrast reflects that, in fact, Module 4 specificity is generally quite good. This finding also highlights the fact that, as with all of the ASD diagnostic instruments, the *ADOS* is best used as one measure of behavior in combination with other sources of information.

Associations between the new Module 4 raw totals and participant characteristics were small (i.e., 5–10% variance explained by participant characteristics not specific to ASD, such as verbal IQ, age and race). Calibrated severity scores were derived in order to facilitate comparison to Modules 1–3 used with younger children and individuals with limited language skills, for which greater influences of participant characteristics on algorithm totals are observed (Gotham et al. 2009; Hus et al. 2013). It was also important to take into account differences in number of items across modules on the RRB domain, given the decision to include both *Stereotyped/Idiosyncratic Use of Words or Phrases* and *Speech Abnormalities* on the new Module 4 algorithm, while also keeping theoretically meaningful items capturing sensory interests, hand mannerisms and excessive references to highly specific topics. In the end, this yielded a total of 5 items on the Module 4 RRB domain, compared to only 4 items for Modules 1–3.

As observed for Modules 1–3, there was overlap in the distributions of domain calibrated scores across diagnostic groups. Overlap of the Non-ASD group with the Autism and Other-ASD groups may reflect recruitment bias (i.e., some of our Non-ASD sample had been referred for assessment of ASD, but received a clinical Non-ASD diagnosis). However, this also reflects the intention of the calibrated score to provide a continuous, quantitative dimensional measures of social-communication and repetitive behaviors extending beyond diagnostic categorization, consistent with the collapsing of diagnostic categories in DSM-5.

It is hoped that the newly revised Module 4 algorithm and CSS will help to expand research efforts to better understand the specific strengths and difficulties in social-communication and repetitive behaviors experienced by adults with ASD. The Module 4 CSS offers the opportunity for comparisons to scores obtained from Modules 1–3, used with younger and less verbally fluent children and adolescents. Thus, the CSS allows examination of longitudinal trajectories of ASD symptoms across childhood and into young adulthood.

Module 4 scores can also be used to further our understanding of how ASD symptom severity interacts with other factors, such as verbal, cognitive and adaptive ability to predict functional outcomes for adults with ASD. Moreover, the Module 4 CSS may be a useful phenotyping measure for neurobiological studies seeking to draw associations between dimensions of ASD and differences in brain structure or function.

Clinically, the Module 4 revisions yield scores that provide a more accurate summary of ASD symptoms, with an algorithm that is more closely aligned with DSM-5 criteria than the original algorithm. It also affords good sensitivity and improved specificity compared to the original Module 4 algorithm. Although it is always recommended that the *ADOS* be used as one source of information in a diagnostic battery, good specificity is particularly important in the assessment of adults, for whom parents are not always available to provide the comprehensive developmental history that is often helpful in making differential diagnoses. Finally, clinicians may use the Module 4 CSS to monitor symptom severity (relative to age and language level) during the course of treatment. However, it is important to remember that the *ADOS'* primary use is as a *diagnostic instrument* and the CSS is intended to capture severity of *core symptoms* that may not be expected to remit in the same way symptoms of depression or anxiety are reduced in response to treatment. Moreover, because the CSS is *not* intended as a measure of functional impairment, it may not be as sensitive to more subtle changes as measures of adaptive social functioning. Thus, while a significant reduction in scores over time may be viewed as evidence of improvement, stability of scores should not be viewed as discouraging. Notably, confidence intervals (shown in Table 4) should be taken into account when assessing the clinical significance of a change in score.

Limitations

Sensitivity and specificity of the algorithm may vary in different clinical and research settings as a consequence of differences in examiner skill, sequence of administration and other factors (Gotham et al. 2007). While the Non-ASD group is the largest to-date used in the validation of the *ADOS*, it is a diagnostically diverse group. Future studies examining the diagnostic utility of the *ADOS* in more specific comparison samples (e.g., individuals with mood disorders) would be useful to inform understanding of the behavioral patterns

NIH-PA Author Manuscript

NIH-PA Author Manuscript

observed in other groups and the Module 4's ability to differentiate between ASD and Non-ASD diagnoses.

Results of a recent study examining the validity of the *ADOS* in a sample of adults suggest that our revisions to the Module 4 algorithm will increase discriminative validity in difficult to differentiate psychiatric groups (Bastiaansen et al., 2011). This study demonstrated good overall specificity of the original Module 4 algorithm (.82) in adults with Psychopathy, Schizophrenia or typical development. Domain totals discriminated between the ASD vs. the Psychopathy and typically developing groups, but did not discriminate between the ASD and Schizophrenia group. This was thought to be due to the overlap in negative symptoms observed in both ASD and Schizophrenia (e.g., limited range of directed facial expressions and lack of asking the examiner for information). Application of the revised Module 3 algorithm (Gotham et al., 2007) differentiated between the ASD and each of the three groups, including the Schizophrenia group. Examination of individual items suggested that only three of the 22 Module 4 items distinguished the ASD from the Schizophrenia group: *Stereotyped Language*, *Quality of Social Response* and *Overall Quality of Rapport*. All three of these items are included in the revised Module 4 algorithm, in addition to seven items found to differentiate ASD from the psychopathy and typically developing groups. Given that our changes to the Module 4 algorithm have increased comparability to the revised Module 3 algorithm and that the new Module 4 algorithm comprises many items shown to differentiate groups in Bastiaansen's study, we would expect that the revised Module 4 algorithm will better differentiate between ASD and these Non-ASD groups than the original algorithm. Examination of Module 4 performance with these and other diagnostic groups will be an important future direction for validating the revised Module 4 algorithm.

New Module 4 totals were weakly associated with age and there was some variability in score distributions across ages, such that 19–20 year olds in this sample had a somewhat narrower distribution of scores compared to other age groups, which showed more similar distributions. It is likely that this difference reflects recruitment bias in this clinical sample. As such, it will be critical that the validity of the Module 4 revised algorithm and calibrated severity metric be replicated in other samples. In addition, older adults (>40 years) in this sample tended to have considerably lower *ADOS* raw totals (Mean=5.4) compared to other age groups, which tended to vary around means of 10–14. This is likely to also reflect sample recruitment biases (i.e., older individuals in our sample were generally self-referred to one of the clinics for a first-time diagnosis and may have had more subtle symptom patterns than those referred at younger ages). Because there were only 9 participants in this older age group, it was decided to exclude them from the calibrated severity score derivation. A larger sample of participants over 40 years of age is needed to explore these differences in symptom presentation for older individuals.

Effects of race on both raw and calibrated RRB totals are also likely to be an artifact of recruitment bias in this predominantly Caucasian sample. As noted for Module 1–3, studies of the *ADOS* in population-based or clinical samples recruited outside of North America will be crucial (Hus et al. 2013).

Finally, it is noteworthy that Module 4 is intended for adults who are verbally fluent. Adaptations to increase the validity of Modules 1 and 2 for use with adults who have significant language impairments are underway (Hus et al., 2011). These will include changes to both the tasks administered, as well as items used to score behaviors. Diagnostic algorithms will also be derived to provide instrument classifications of ASD and Nonspectrum.

Conclusion

The revised Module 4 algorithm provides improved sensitivity, while maintaining or increasing specificity across individuals of different cognitive levels. The revised algorithm is consistent with the revised DSM-5 two-domain criteria for ASD and offers increased comparability to recently published algorithms for *ADOS* Modules T and 1–3 (Lord, Rutter, et al. 2012; Lord, Luyster, et al. 2012). Module 4 calibrated severity scores provide quantitative estimates of severity of social-communication and repetitive behaviors that are relatively independent of participant characteristics. The new severity scores also extend the ability to compare domains and overall totals across modules. These changes will facilitate future research efforts to increase understanding of the strengths and challenges experienced by adults with ASD.

Acknowledgments

This research was supported by a Dennis Weatherstone Predoctoral Fellowship to VH and Simons Foundation and National Institute of Mental Health grant R01MH081873 to CL. We thank Andrew Pickles for his statistical consultation. We are also grateful to the families, as well as SSC principal investigators (A.Beaudet, R.Bernier, 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 by applying at https://base.sfari.org.

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Washington, DC: American Psychiatric Association; 2013.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed., Text Revision. Washington, DC: American Psychiatric Association; 2000.

Anderson DK, Liang JW, Lord C. Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. Journal of Child Psychology and Psychiatry. 2013 Advanced Online Publication.

Brown JD. Statistics Corner questions and answers about language testing statistics: Standard error vs. standard error of measurement. Shiken: JALT Testing & Evaluation SIG Newsletter. 1999; 3(1):20–25.

Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing Structural Equation Models. Newbury Park, CA: Sage; 1993. p. 136-162.

Brugha TS, McManus S, Smith J, Scott FJ, Meltzer H, Purdon S, Berney T, Tantam D, Robinson J, Radley J, Bankart J. Validating two survey methods for identifying cases of autism spectrum disorder among adults in the community. Psychological medicine. 2012; 42(3):647–656. [PubMed: 21798110]

Dunn, LM.; Dunn, LM. Peabody Picture Vocabulary Test. 4 th ed. Bloomington: Pearson; 2007.

Eaves LC, Ho HH. Young adult outcome of autism spectrum disorders. Journal of Autism and Developmental Disorders. 2008; 38(4):739–747. [PubMed: 17764027]

Elliott, CD. Differential Ability Scales—Second Edition (DAS-II). San Antonio, TX: The Psychological Corporation; 2006.

Elliott, Colin D. Differential Ability Scales. San Antonio, TX: Psychological Corporation; 1990.

Farley MA, McMahon WM, Fombonne E, Jenson WR, Miller J, Gardner M, Block H, Pingree CB, Ritvo ER, Ritvo RA, Coon H. Twenty-year outcome for individuals with autism and average or near-average cognitive abilities. Autism Research. 2009; 2(2):109–118. [PubMed: 19455645]

Fein D, Barton M, Eigsti I-M, Kelley E, Naigles L, Schultz RT, Stevens M, Helt M, Orinstein A, Rosenthal M, Troyb E, Tyson K. Optimal outcome in individuals with a history of autism. Journal of Child Psychology and Psychiatry. 2013; 54(2):195–205. [PubMed: 23320807]

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]

Ganz ML. The lifetime distribution of the incremental societal costs of autism. Archives of Pediatrics & Adolescent Medicine. 2007; 161(4):343–349. [PubMed: 17404130]

Gillespie-Lynch K, Sepeta L, Wang Y, Marshall S, Gomez L, Sigman M, Hutman T. Early childhood predictors of the social competence of adults with autism. Journal of Autism and Developmental Disorders. 2012; 42(2):161–174. [PubMed: 22187106]

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]

Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. Journal of Autism and Developmental Disorders. 2007; 37(4):613–627. [PubMed: 17180459]

Helt M, Kelley E, Kinsbourne M, Pandey J, Boorstein H, Herbert M, Fein D. Can children with autism recover? If so, how? Neuropsychology Review. 2008; 18(4):339–366. [PubMed: 19009353]

Henninger NA, Taylor JL. Outcomes in adults with autism spectrum disorders: a historical perspective. Autism. 2013; 17(1):103–116. [PubMed: 22914775]

Howlin P, Moss P, Savage S, Rutter M. Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. Journal of the American Academy of Child & Adolescent Psychiatry. 2013; 52(6):572–581. e1. [PubMed: 23702446]

Hus V, Bishop S, Gotham K, Huerta M, Lord C. Factors influencing scores on the social responsiveness scale. Journal of Child Psychology and Psychiatry. 2013; 54(2):216–224. [PubMed: 22823182]

Hus, V.; Lord, C. Evaluation & Testing. In: Hollander, Eric; Kolevzon, Alexander; Coyle, Joseph T., editors. Textbook of autism spectrum disorders. Arlington, VA: American Psychiatric Publishing, Inc.; 2011. p. 49-65.

Hus V, Lord C. Effects of child characteristics on the Autism Diagnostic Interview-Revised: Implications for use of scores as a measure of ASD severity. Journal of Autism and Developmental Disorders. 2013; 43(2):371–381. [PubMed: 22729382]

Hus, V.; Maye, M.; Harvey, L.; Guthrie, W.; Liang, J.; Lord, C. The Adapted ADOS - Preliminary findings using a modified version of the ADOS for adults who are nonverbal or have limited language. Poster presented at the International Meeting for Autism Research; San Diego, CA. 2011 May.

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]

Interagency Autism Coordinating Committee. The 2011 Interagency Autism Coordinating Committee Strategic Plan for Autism Spectrum Disorder Research. 2011 Jan 18. Retrieved from http://iacc.hhs.gov/strategic-plan/2011/index.shtml

Jones RM, Risi S, Wexler D, Anderson D, Corsello C, Pickles A, Lord C. How interview questions are placed in time influences caregiver description of social communication symptoms on the ADI-R. 2013 *under review*.

Levy A, Perry A. Outcomes in adolescents and adults with autism: A review of the literature. Research in Autism Spectrum Disorders. 2011; 5(4):1271–1282.

Lord, C.; Luyster, R.; Gotham, K.; Guthrie, W. Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module. Torrance, CA: Western Psychological Services; 2012.

Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM, Ousley O, Risi S. A multisite study of the clinical diagnosis of different autism spectrum disorders. Archives of General Psychiatry. 2012; 69(3):306–313. [PubMed: 22065253]

Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S.; Gotham, K.; Bishop, S. Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1–4. Torrance, CA: Western Psychological Services; 2012.

Luyster R, Gotham K, Guthrie W, Coffing M, Petrak R, Pierce K, Bishop S, Esler A, Hus V, Oti R, Richler J, Risi S, Lord C. The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. Journal of Autism and Developmental Disorders. 2009; 39(9):1305–1320. [PubMed: 19415479]

Perry A, Condillac RA, Freeman NL. Best practices and practical strategies for assessment and diagnosis of autism. Journal on Developmental Disabilities. 2002; 9(2):61–75.

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]

Raven, JC. Guide to the standard progressive matrices: sets A, B, C, D and E. HK Lewis; 1960.

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.

Schopler E, Reichler RJ, DeVellis RF, Daly K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). Journal of Autism and Developmental Disorders. 1980; 10(1):91–103. [PubMed: 6927682]

Shattuck PT, Seltzer MM, Greenberg JS, Orsmond GI, Bolt D, Kring S, Lounds J, Lord C. Change in autism symptoms and maladaptive behaviors in adolescents and adults with an autism spectrum disorder. Journal of Autism and Developmental Disorders. 2006; 37(9):1735–1747. [PubMed: 17146700]

Sigman M, McGovern CW. Improvement in cognitive and language skills from preschool to adolescence in autism. Journal of Autism and Developmental Disorders. 2005; 35(1):15–23. [PubMed: 15796118]

Skrondal, A.; Rabe-Hesketh, S. Generalized latent variable modeling. Boca Raton, FL: Chapman & Hall/CRC Press; 2004.

U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS. Table 1–3. Number of students ages 6 through 21 served under IDEA, Part B, by disability category and state: Fall 2010. 2011 Jul 15. Retrieved from https://www.ideadata.org/TABLES34TH/AR_1-3.pdf

Wechsler, D. Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: Psychological Corporation; 1999.

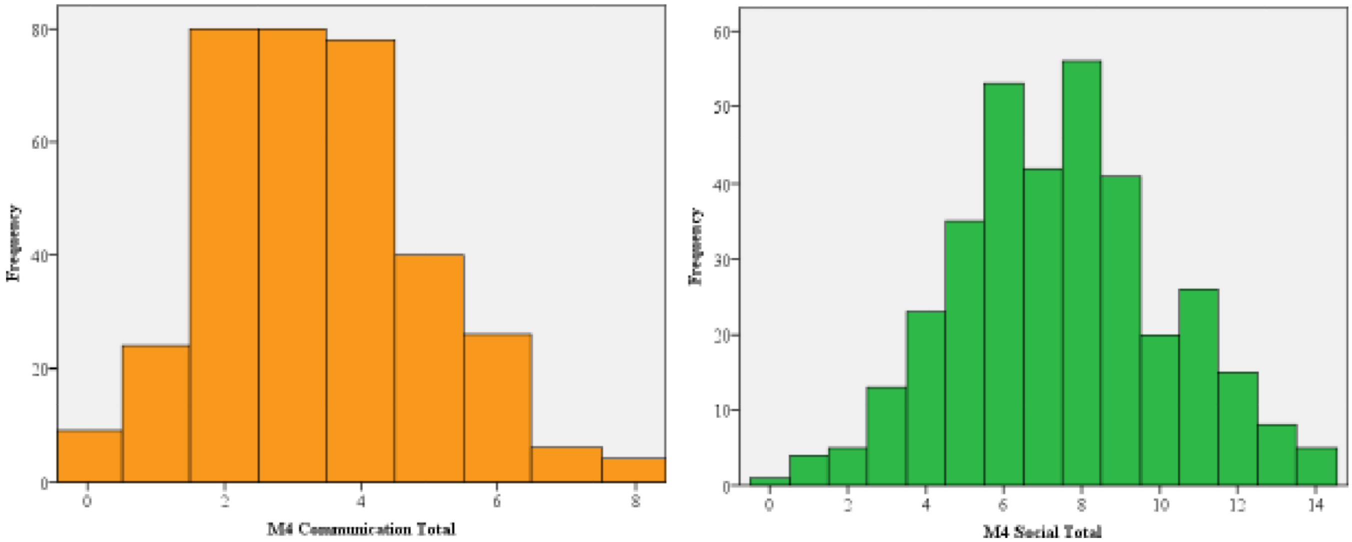

Fig. 1.
Distribution of original Module 4 algorithm domain totals.

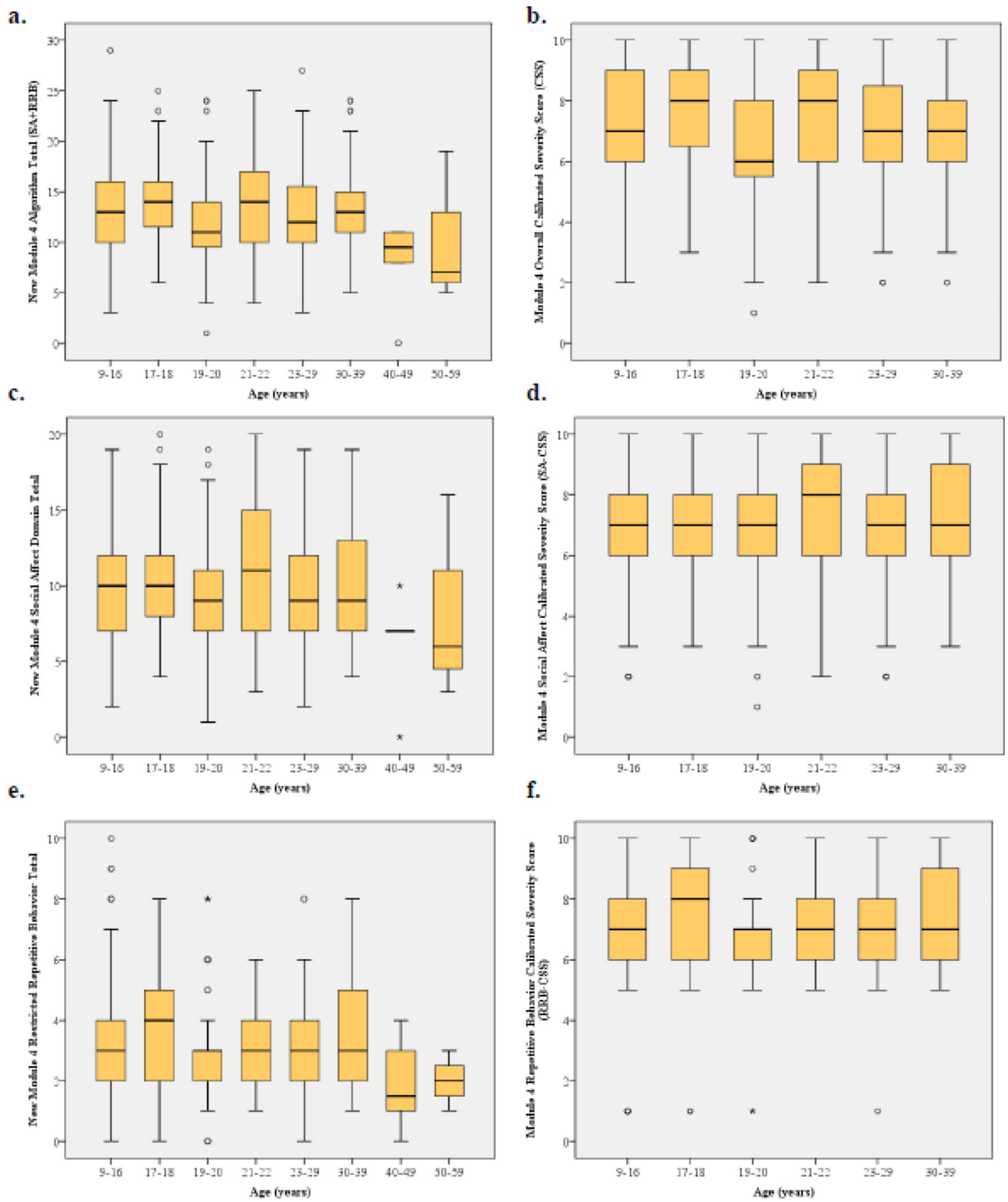

Fig. 2.

a (top, left) Distributions of new Module 4 Algorithm raw overall totals by age cells; **b** (top, right) Distributions of Module 4 calibrated overall scores by age cells; **c** (middle, left) Distributions of raw Social Affect domain totals by age cells; **d** (middle, right) Distributions of calibrated Social Affect domain scores by age cells; **e** (bottom, left) Distributions of raw Repetitive Behavior domain totals by age cells; **f** (bottom, right) Distributions of calibrated Repetitive Behavior domain scores by age cells.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

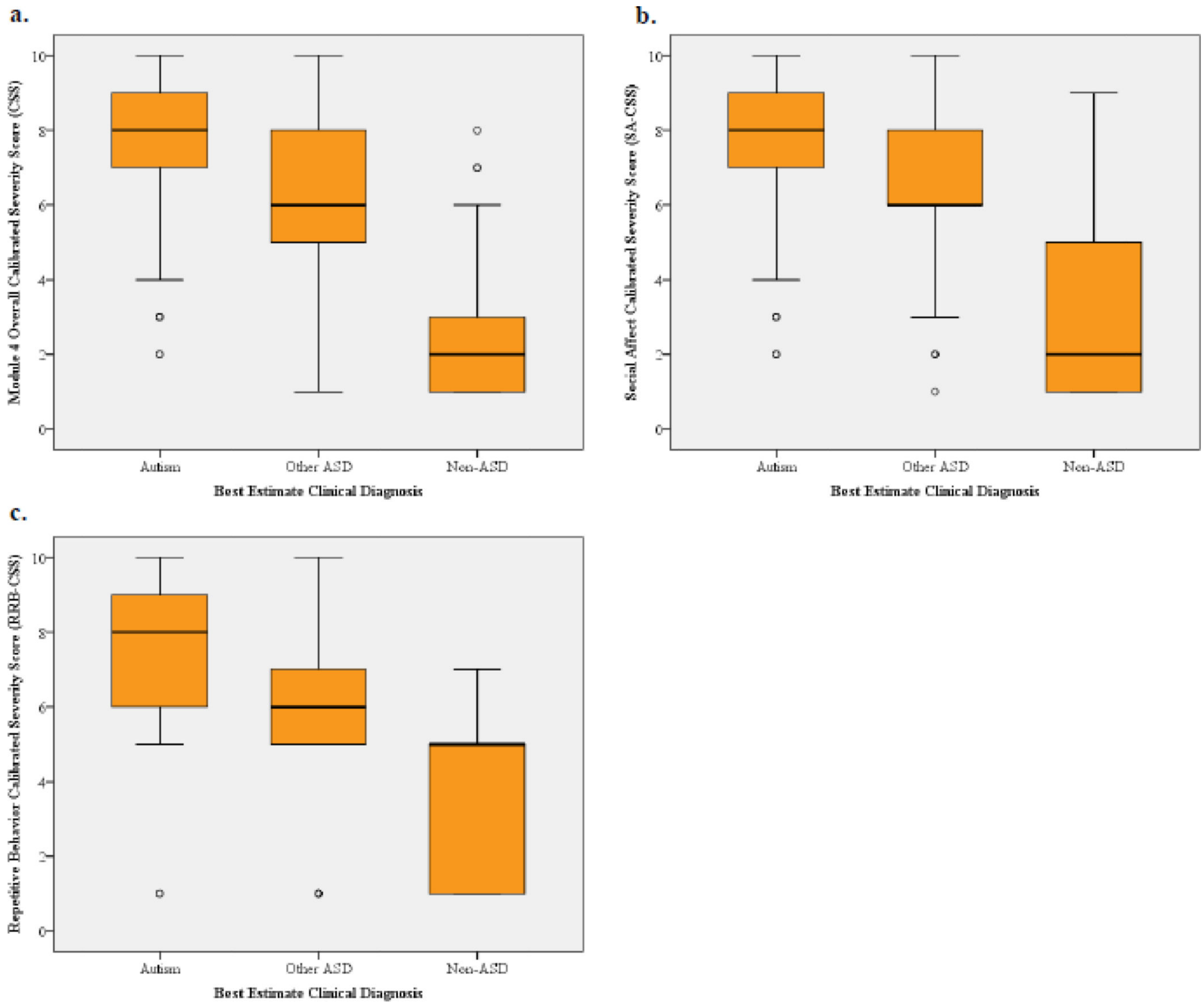

Fig. 3.

a (top, left) Distributions of Module 4 Overall Calibrated Severity Score by best estimate clinical diagnosis; **b** (top, right) Distributions of calibrated Social Affect domain scores by best estimate clinical diagnosis; **c** (bottom, right) Distributions of calibrated Repetitive Behavior domain scores by best estimate clinical diagnosis.

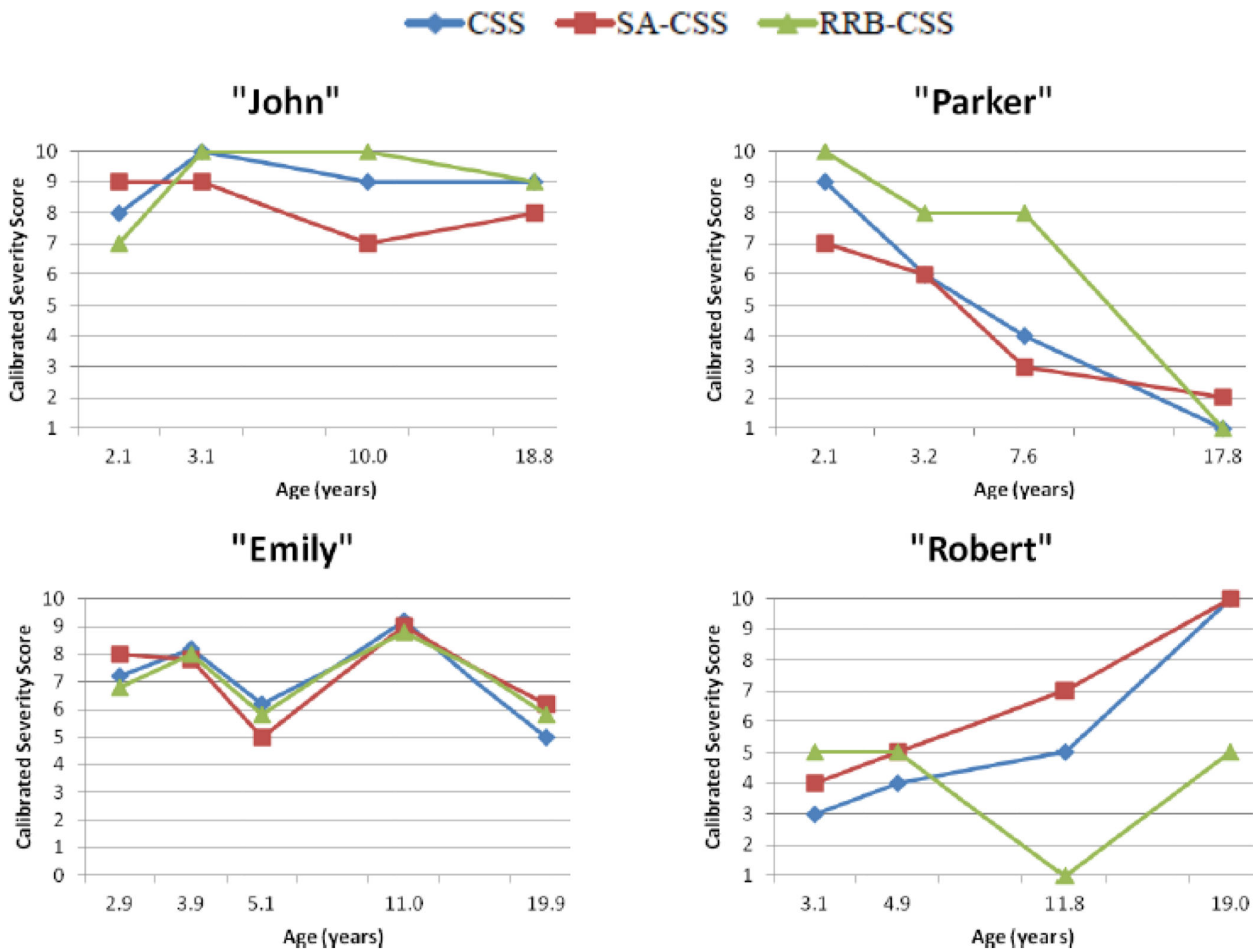

Fig. 4.
Case summaries of longitudinal severity scores. 95% CI: CSS +/− 2.16, SA-CSS +/− 1.94; RRBCSS +/− 2.99

NIH-PA Author Manuscript
NIH-PA Author Manuscript

J Autism Dev Disord. Author manuscript; available in PMC 2015 August 01.

Table 1

Sample Description

	Autism					Other-ASD					Non-ASD				
	N	**Mean**	**SD**	**Min**	**Max**	**N**	**Mean**	**SD**	**Min**	**Max**	**N**	**Mean**	**SD**	**Min**	**Max**
age (years)	177	20.12	6.30	10.33	55.58	170	21.14	7.79	9.92	54.00	90	25.17	12.35	13.33	62.25
VIQ	145	96.79	23.52	34	148	136	107.43	20.27	38	150	82	91.94	27.37	23	144
NVIQ	144	94.42	21.99	32	140	135	103.44	19.60	49	153	81	94.48	26.61	33	147
ADI Social	139	19.30	6.38	0	30	122	16.41	6.31	2	28	66	9.67	6.55	0	24
ADI Comm-V	139	15.73	4.99	4	24	122	13.12	5.13	0	24	66	8.18	5.82	0	24
ADI Comm-NV	139	8.58	3.80	0	14	122	7.43	3.85	0	14	66	4.18	3.80	0	14
ADI-RRB	139	6.32	2.62	0	12	122	4.87	2.49	0	12	66	3.52	2.71	0	11
ADOS Comm	177	3.85	1.58	0	8	170	2.85	1.45	0	6	90	1.32	1.34	0	6
ADOS Social	177	8.38	2.63	3	14	170	6.50	2.49	0	14	90	3.09	2.50	0	11
ADOS RRB	177	1.75	1.55	0	8	170	.91	1.00	0	4	90	.33	0.65	0	3

Note. VIQ = verbal IQ; NVIQ = nonverbal IQ; ADI=Autism Diagnostic Interview-Revised; Comm-V=Verbal Communication; Comm-NV=Nonverbal Communication; RRB=Restricted, Repetitive Behavior; ADOS=Autism Diagnostic Observation Schedule

Table 2

Revised Module 4 algorithm mapping

Domains	Module 4 N=437	Factor Loadings	Module 3 N=398*	Factor Loadings
Social Affect	Unusual Eye Contact**	0.37	Unusual Eye Contact	0.51
	Amount of Social Communication	0.86	Amount of Social Communication	0.81
	Facial Expressions	0.54	Facial Expressions	0.67
	Quality of Rapport	0.67	Quality of Rapport	0.72
	Communication Own Affect	0.65	Shared Enjoyment	0.84
	Quality of Social Overtures	0.57	Quality of Social Overtures	0.65
	Conversation	0.74	Conversation	0.73
	Emphatic Gestures	0.57	Descriptive Gestures	0.71
	Quality of Social Response	0.49	Quality of Social Response	0.6
	Insight	0.59	Reporting of Events	0.65
Eigen Value	6.4		6.2	
Restricted Repetitive Behaviors	Speech Abnormalities	0.59		
	Stereotyped Language	0.81	Stereotyped Language	0.60
	Unusual Sensory Interest	0.60	Unusual Sensory Interest	0.44
	Highly Specific Topics	0.46	Highly Specific Topics	0.63
	Hand Mannerisms	0.54	Hand Mannerisms	0.48
Eigen Value	1.7		1.7	
RMSEA	0.08		0.06	
Rho	0.46		0.38	

Note.

*reproduced from Gotham et al., 2007 for comparison

**loads .45 on RRB when all sample (ASD & nonASD) included; loading is higher for SA when only ASD included in EFA

RMSEA = Root Mean Square Error (values 0.08 or less indicate a good fit).

Rho = correlation between Social Affect & Restricted Repetitive Behaviors factors.

Items from the 2000 algorithm not included in new Module 4 algorithm: Descriptive Gestures, Responsibility Loadings from FA including all participants (ASD & Non-ASD; N=437)

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Sensitivity and specificity of previously used and revised algorithms

		Overall		VIQ <85		VIQ 85–115		VIQ >115	
		Sens	**Spec**	**Sens**	**Spec**	**Sens**	**Spec**	**Sens**	**Spec**
		ASD=347	**NS=90**	**ASD=67**	**NS=26**	**ASD=156**	**NS=44**	**ASD=101**	**NS=14**
2000 algorithm	Met 3 domains*	89.6	72.2	89.5	73.1	84.0	81.8	79.2	85.7
New algorithm	SA only (cut=6)	89.0	72.2	91.0	65.4	90.4	68.2	86.1	85.7
	SA+RRB (cut=8)	90.5	82.2	94.0	80.8	91.7	77.3	87.1	92.9

Note.

*Met or exceeded cut-offs on Social, Communication and Social+Communication domains; VIQ=Verbal IQ; Sens=Sensitivity; Spec=Specificity; SA=Social Affect; RRB=Restricted, Repetitive Behavior

Table 4

Mapping of ADOS Module 4 raw overall and domain totals to calibrated severity scores for ages 9–39 years

	Raw totals		
CSS	**Overall Total**	**SA Domain**	**RRB Domain**
1	0–2	0–1	0
2	3–5	2–3	-
3	6–7	4	-
4	8	5	-
5	9	6	1
6	10–11	7–8	2
7	12–13	9–10	3
8	14–15	11–12	4
9	16–19	13–15	5
10	20–29	16–20	6–10

Note. CSS=Calibrated Severity Score; SA=Social Affect; RRB=Restricted and Repetitive Behaviors; 95% CI: CSS +/− 2.16, SA-CSS +/− 1.94; RRB-CSS +/− 2.99

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 5

Case summary characteristics

	Demographics		First Assessment					Last Assessment				
	Gender	**Race**	**Age**	**VIQ**	**NVIQ**	**ADOS Module**	**Diagnosis**	**Age**	**VIQ**	**NVIQ**	**ADOS Module**	**Diagnosis**
John	male	White	2.1	26	68	1	Autism	18.8	188	85	4	Autism
Parker	male	White	2.1	95	92	1	Autism	17.75	108	99	4	No Dx
Emily	female	Af. Amer.	2.9	48	85	1	Autism	19.9	72	81	4	Autism
Robert	male	Af. Amer.	3.1	53	80	1	PDD-NOS	19.8	72	78	4	PDD-NOS

Note. All ages in years; VIQ = verbal IQ; NVIQ = nonverbal IQ; Diagnosis at Last Assessment was based on presentation of symptoms at the time of assessment (examiners were blind to previous diagnosis)