

# Replication of Standardized ADOS Domain Scores in the Simons Simplex Collection

Vanessa Hus Bal and Catherine Lord

Raw totals from diagnostic and screening measures for autism spectrum disorder (ASD) are frequently used as dimensional measures of autism symptom severity without appropriate correction for confounding factors, such as developmental level or non-ASD-specific behavior problems. Although these associated features are important to consider when diagnosing ASD and developing intervention plans, both researchers and clinicians sometimes need metrics of ASD severity that are not influenced by these factors. The Autism Diagnostic Observation Schedule (ADOS) domain calibrated severity scores (CSS) were created to provide separate estimates of social affect (SA-CSS) and restricted, repetitive behaviors (RRB-CSS) that are relatively independent of child characteristics (Hus et al., 2014). Using a sample of 2,509 probands with ASD from the Simons Simplex Collection (SSC), this study provides the first replication of the ADOS domain CSS in an independent sample. Consistent with the original standardization study, when applied to existing SSC data, the ADOS domain CSS were less influenced by age and cognitive ability compared to raw domain totals. Domain CSS were also relatively independent of behavior problems. Use of the ADOS domain CSS to assess relationships between ASD symptoms and genetic risk factors will increase confidence that associations reflect domain-specific relationships. Scores also offer less developmentally-influenced estimates of ASD severity for future phenotypic explorations in the SSC. This independent replication provides support for the application of the ADOS domain CSS in other samples, though further replication in population-based samples will be an important next step. *Autism Res* 2015, 8: 583–592. © 2015 International Society for Autism Research, Wiley Periodicals, Inc.

**Keywords:** autism spectrum disorder; Autism Diagnostic Observation Schedule; severity; social affect; restricted and repetitive behaviors

## Introduction

Recent reports estimate that 1 in 68 children in the United States are diagnosed with an autism spectrum disorder (ASD) [DDM Network & CDC, 2014]. There are a growing number of studies seeking to elucidate genetic variants that enhance risk for autism and neurobiological mechanisms that underlie symptoms of this complex, developmental disorder. It is widely acknowledged that ASD is etiologically heterogeneous [Geschwind, 2011] and will require large samples to investigate potential contributing genetic factors.

The Simons Simplex Collection (SSC) [Fischbach & Lord, 2010] is a study of over 2,500 “simplex” families (i.e. families with one child with ASD who does not have first, second, or third degree relatives with the disorder). A strength of the SSC is the availability of phenotype data from a variety of behavioral measures that were carefully monitored for completion and reliability across sites [Lord, Petkova, et al., 2012]. Some of these measures have been used to explore genotype–pheno-

type correlations [e.g. Girirajan et al., 2013; Krumm et al., 2013; Sanders et al., 2011].

Many of these scales have names that suggest their scores should reflect a specific behavior or subset of behaviors, such as the Social Communication Questionnaire (SCQ) [Rutter, Bailey, Lord, & Berument, 2003] or the Social Responsiveness Scale (SRS) [Constantino & Gruber, 2005], or may capture autism severity more broadly (e.g. Autism Diagnostic Interview-Revised; ADI-R) [Rutter, Le Couteur, & Lord, 2003]. However, there is a growing body of research, including studies using behavioral data from the SSC, demonstrating that interpretation of raw totals from these measures is not straightforward [e.g. Warren et al., 2012]. For example, a recent SSC study [Hus, Bishop, Gotham, Huerta, & Lord, 2013] demonstrated that SRS scores of children with mild social impairments but high levels of behavior problems were indistinguishable from children with significant social impairments but few behavior problems. This comparison reflected findings that scores on the SRS, which is intended as an ASD screener and continuous measure of autism severity, were

From the Department of Psychology, University of Michigan, Ann Arbor, Michigan (V.H.B.); Department of Psychiatry, University of California San Francisco, San Francisco, California (V.H.B.); Center for Autism and the Developing Brain, Weill Cornell Medical College, White Plains, New York (C.L.)

Received August 21, 2014; accepted for publication February 04, 2015

Address for correspondence and reprints: Vanessa Hus Bal, Department of Psychiatry, University of California San Francisco, 1550 4th St, San Francisco, CA 94158. E-mail: Vanessa.HusBal@ucsf.edu

Published online 24 February 2015 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/aur.1474

© 2015 International Society for Autism Research, Wiley Periodicals, Inc.

strongly associated with several non-ASD-specific child characteristics. Effect sizes for behavior problems were similar or larger than effect sizes reflecting associations between SRS scores and measures of social competence or autism symptoms. A parallel study focusing on raw domain totals from the ADI-R found that its raw totals were strongly influenced by age and language, although not behavior problems [Hus & Lord, 2013]. Parents of older children or children with minimal language tended to report more ASD symptoms. Thus, ADI-R raw totals are not directly comparable across individuals of different developmental levels as a measure of ASD severity; scores will not distinguish older children or those with greater language impairment from children who truly have high levels of core ASD symptoms.

While factors such as language, cognitive impairment, or behavior problems are important to consider when diagnosing ASD and developing intervention plans, these features are not part of the core ASD diagnostic criteria. Thus, if the goal is to identify biomarkers for ASD, or that relate to dimensions of behavior, such as social-communication, it is important to control for these associated features to increase confidence in the specificity of the biological-behavioral association. Despite the significance of this issue, both behavioral and basic science researchers continue to use scores from these measures as if they are specific indices of autism severity or other specific behaviors (e.g. social impairment) without controlling for potentially confounding factors. For example, investigations of the pathophysiology of ASD often use scores from these measures to draw associations between ASD symptoms and genetic mutations or neurobiological differences [e.g. Connolly, Glessner, & Hakonarson, 2013; Coutanche, Thompson-Schill, & Schultz, 2011; Uddin et al., 2011]. Such studies rarely acknowledge the potential confounds to these measures [see Brune et al., 2006 for exception; also, Charman et al., 2007], which limits the interpretability of findings and may explain difficulties with replication (e.g. due to sample differences in age or cognitive level) [Hus, Pickles, Cook, Risi, & Lord, 2007; Jones & Lord, 2013]. This may continue in part because there has been a lack of measures available to investigate symptom severity that are not confounded by these other non-ASD-specific child characteristics.

Of course, sometimes the distance between genetic variation and observable behavior may also be too “far” to reasonably expect to draw strong phenotype–genotype associations [Kim & State, 2014]. Medical disorders with similar clinical presentations have demonstrated heterogeneous etiologies, whereas seemingly distinct syndromes may arise from the same pathophysiology [Insel et al., 2010]. As such, there has been movement toward greater consideration of dimensions of behavior that extend across the boundaries of the classic categorical diagnoses

in approaches such as NIMH’s Research Domain Criteria (RDoC) and APA’s DSM-5. With a primary goal of RDoC being to draw associations between neural circuitry and both clinical and genetic factors [Insel et al., 2010], it seems all the more important to put careful thought into the types of tools used to measure dimensions of behavior. Indeed, it is unlikely that neural circuits are likely to be correlated with measures encompassing many dimensions of behavior (e.g. the SRS or SCQ which include aspects of social and repetitive behaviors, as well as internalizing and externalizing behavior problems). Thus, constructing “purer” metrics of behavioral dimensions that are relatively independent of developmental level and other factors would appear to be an important contribution to both understanding the pathophysiology of ASD, as well as how variability of behavioral dimensions within ASD influence clinical outcomes.

One metric, the calibrated severity score (CSS) [Gotham, Pickles, & Lord, 2009] derived from the Autism Diagnostic Observation Schedule (ADOS) [Lord, Rutter, DiLavore, & Risi, 1999], is less influenced by age and language skills compared to raw ADOS totals. (Notably, in the recently revised ADOS-2 [Lord, Rutter, et al., 2012], the CSS was renamed the comparison score; however, we maintain use of the terms “ADOS-CSS” or “Overall CSS” to refer to the standardized overall total score to facilitate comparisons to the study by Hus, Gotham, & Lord [2014], which this manuscript seeks to replicate.) The ADOS-CSS offers an overall indicator of ASD severity, encompassing both core symptom domains: social-communication and restricted, repetitive behaviors (RRB), as observed during a standardized assessment. This metric has been used in some studies investigating possible ASD biomarkers. For example, Girirajan et al. [2013] used the ADOS-CSS to show that copy number variation (CNV) size was positively correlated with autism symptom severity in individuals with duplications but not deletions. Conversely, Nordahl et al. [2011] did not find a significant relationship between total cerebral volume and ADOS severity. Lack of associations are not unexpected given that the ADOS-CSS, like other measures mentioned earlier, combines a broad range of behaviors, encompassing both of the core domains of ASD-related symptoms.

Analyses of other ASD diagnostic instruments suggest that, consistent with DSM-5, ASD is best conceptualized by a model constituting two related, but distinct, dimensions: social-communication and repetitive behaviors [e.g. Mandy, Charman, Puura, & Skuse, 2014]. Indeed, researchers seeking to link biological mechanisms often focus on these domains separately (e.g. linking ADOS social domain scores to amygdala activation) [Dichter, Richey, Rittenberg, Sabatino, & Bodfish, 2011]. Thus, calibrated scores for each separate domain may be more useful in such investigations than the calibrated score

**Table 1. Sample Descriptives**

	Module 1, No Words ( <i>n</i> = 155)		Module 1, some words ( <i>n</i> = 299)		Module 2, under 5 ( <i>n</i> = 162)		Module 2, 5 or older ( <i>n</i> = 411)		Module 3 ( <i>n</i> = 1482)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	7.41	2.89	7.34	2.76	4.41	0.29	8.16	2.80	9.62	3.05
VIQ	22.03	11.90	43.41	19.08	87.02	15.05	63.07	20.79	94.06	20.60
NVIQ	40.71	13.89	58.71	20.83	92.93	16.97	75.13	20.63	96.38	18.57
VMA	1.47	0.65	2.66	0.85	3.60	0.87	4.35	1.12	8.90	3.61
NVMA	2.84	1.12	3.98	1.49	4.12	0.85	5.66	1.90	9.21	3.57
SA Raw	15.32	2.90	13.78	3.09	10.65	3.53	12.70	3.86	9.73	3.58
RRB Raw	5.97	1.87	4.98	1.75	4.33	1.75	5.19	1.84	3.29	1.81

*Note.* All ages in years; VIQ = Verbal IQ; NVIQ = Nonverbal IQ; VMA = Verbal Mental Age; NVMA = Nonverbal Mental Age; SA Raw = ADOS Social Affect Raw Total; RRB Raw = ADOS Restricted; Repetitive Behaviors Raw Total

combining both domains [Jones & Lord, 2013]. As such, in their 2014 paper, Hus and colleagues sought to separately standardize ADOS domain scores. Consistent with the overall CSS, the ADOS Social Affect (SA-CSS) and restricted, repetitive behavior (RRB-CSS) calibrated scores significantly reduced effects of child characteristics compared to domain raw totals.

In one SSC study, the overall and domain CSS were used to create subgroups in which to test the impact of subphenotyping on genetic homogeneity and ability to identify common genetic variants conferring ASD risk [Chaste et al., 2014]. Although the overall results of the study suggested that reducing phenotypic homogeneity was not a particularly fruitful approach for discovering genetic risk variants, the authors noted that probands with high repetitive behaviors (i.e. RRB-CSS  $\geq$  8) may be a more genetically homogenous group. Given that increased homogeneity was not observed in the overall-CSS group, one might interpret this as evidence that examining severity separately for each domain has some benefit over the overall-CSS encompassing both domains. Furthermore, score comparisons suggest that domain CSS are more informative than the overall-CSS for examining longitudinal trajectories of ASD symptoms in individual cases [Hus et al., 2014].

To establish the utility of this new ASD severity metric, it important to both compare results using overall- and domain-CSS, as well as to demonstrate that the domain CSS are replicable in other samples. The purpose of this study is to replicate the CSS for separate domains in an independent sample, the SSC, to demonstrate the validity of this metric for use in ongoing investigations. These scores may be useful for genotype-phenotype analyses and to elucidate the ASD behavioral phenotype in this rich dataset of over 2,500 families.

## Method

### Participants

Participants were drawn from a sample of 2,570 children with an ASD who participated in the SSC. All pro-

bands were required to meet Collaborative Programs of Excellence in Autism (CPEA) criteria for a diagnosis of Autism, Pervasive Developmental Disorder - Not Otherwise Specified (PDD-NOS), or Asperger disorder [Lainhart et al., 2006; see Hus & Lord, 2013]. Families were excluded if the proband had a nonverbal mental age below 18 months, significant sensory impairments that might affect standardized testing, or documentation of Fragile X, tuberous sclerosis, or Down syndrome [see Fishbach & Lord, 2010 for more information regarding inclusion and exclusion criteria]. Sixty-one children were excluded from the present analyses because they were outside the age range used in the original domain CSS calibration study (i.e.  $\geq$  15 years for Module 1 or  $\geq$  17 years for Modules 2 and 3) [Hus et al., 2014]. Participants were predominantly male (86.8%), White (78%), and from well-educated families (61% maternal education of Bachelor's degree or higher). Sample characteristics are provided in Table 1. Parents gave informed consent, approved by Institutional Review Boards at each of the 12 university-based sites.

### Procedure

The ADOS was conducted as part of the SSC's standard research battery. Briefly, this battery included, at minimum, a direct assessment with the child (ADOS, cognitive test), parent interview (ADI-R; Vineland Adaptive Behavior Scales, 2nd edition) [Sparrow, Cicchetti, & Balla, 2005] and several behavioral questionnaires. All ADOSes were administered and scored by a clinical psychologist or trainee who met standard requirements for research reliability and who maintained reliability with study consultants through semiannual workshops and video scoring. All children had verbal and nonverbal IQ scores derived from a developmental hierarchy of cognitive measures, most frequently the Differential Ability Scales, 2nd edition (85%) [Elliott, 2006] and Mullen Scales of Early Learning (11%) [Mullen, 1995]. Parents completed a battery of questionnaires, including the

**Table 2. Domain Raw Totals and CSS Means and Standard Deviations by Age/Language Cell**

Module	Age (years)	<i>n</i>	SA-Raw		SA-CSS		RRB-Raw		RRB-CSS	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Module 1, No Words	4–14	155	15.32	2.90	6.94	1.53	5.97	1.87	8.32	1.45
Module 1, some words	4	67	12.93	3.12	7.06	1.55	4.52	1.70	7.76	1.28
	5–14	232	14.03	3.04	7.10	1.39	5.11	1.75	8.06	1.49
Module 2	4	162	10.65	3.53	6.96	1.87	4.33	1.75	7.41	1.64
	5–6	188	11.37	3.75	7.10	1.79	5.04	1.81	8.02	1.59
	7–16	223	13.82	3.58	7.85	1.48	5.33	1.86	8.24	1.58
Module 3	4–5	155	10.06	3.52	7.14	1.86	3.53	1.87	8.06	1.93
	6–16	1327	9.69	3.58	7.22	1.83	3.26	1.81	7.66	2.01

Note. SA = Social Affect domain; RRB = Restricted, Repetitive Behavior domain; CSS = Calibrated Severity Score

Child Behavior Checklist (CBCL) [Achenbach & Rescorla, 2001]. CBCL internalizing and externalizing T-scores were used as an estimate of general behavior problems. See Lord, Petkova, et al. [2012] for more detailed procedures.

Procedures for deriving the Social Affect (SA) and RRB CSS are detailed in the original study [Hus et al., 2014]. Briefly, raw domain totals for participants with best estimate clinical diagnoses of ASD were compared across the 18 age and language groups used in the overall-CSS standardization [Gotham et al., 2009]. Groups with similar distributions were collapsed, yielding 12 age/language cells for derivation of domain CSS. Percentiles from mapping overall raw totals to the 10-point calibrated severity metric were used to inform mapping of SA and RRB raw totals. Mappings were adjusted so that 90% of participants with ADOS classifications of “Autism” had SA-CSS of  $\geq 6$ , 80% of participants with ADOS classifications of “Autism Spectrum” had SA-CSS  $\geq 4$ , and 80% of participants with a “Nonspectrum” ADOS classification had SA-CSS  $\leq 3$ . Given the lower sensitivity of repetitive behaviors in the limited context of the ADOS, a less stringent goal of 80% sensitivity was set for ADOS classification of “Autism” and RRB-CSS  $\geq 6$  and 80% specificity for “Nonspectrum” classification and RRB-CSS  $\leq 6$ . Notably, because the raw RRB total is comprised of only four items, the RRB-CSS includes a limited range of values (i.e. 1 and 5–10) [see Hus et al., 2014 for details].

For this study, ADOS raw totals were mapped on to the 10-point calibrated severity metric for the SA and RRB domains as outlined in the original study [Hus et al., 2014] (Table 2). This study includes replication of only 8 of the original 12 age/language cells because children under the age of 4 were not included in the SSC. Separate linear regression analyses were then conducted to examine the influences of child characteristics on raw domain totals and calibrated domain scores. As in Hus et al. [2014], verbal and nonverbal IQs and mental ages were entered into the first block; age, gender, maternal education, and race were entered into the second block. Signifi-

cant predictors were then entered into Forward Stepwise models to assess the relative contributions of child characteristics in predicting both raw domain totals and calibrated domain scores. Separate regression analyses exploring influences of internalizing and externalizing behaviors (controlling for demographics) were also conducted. Although the original study did not investigate whether ADOS-CSS were influenced by behavior problems, the availability of CBCL data in the SSC afforded an opportunity to examine these associations. These analyses were of interest because significant associations with behavior problems would limit interpretability of domain CSS as indicators of core ASD symptom severity.

## Results

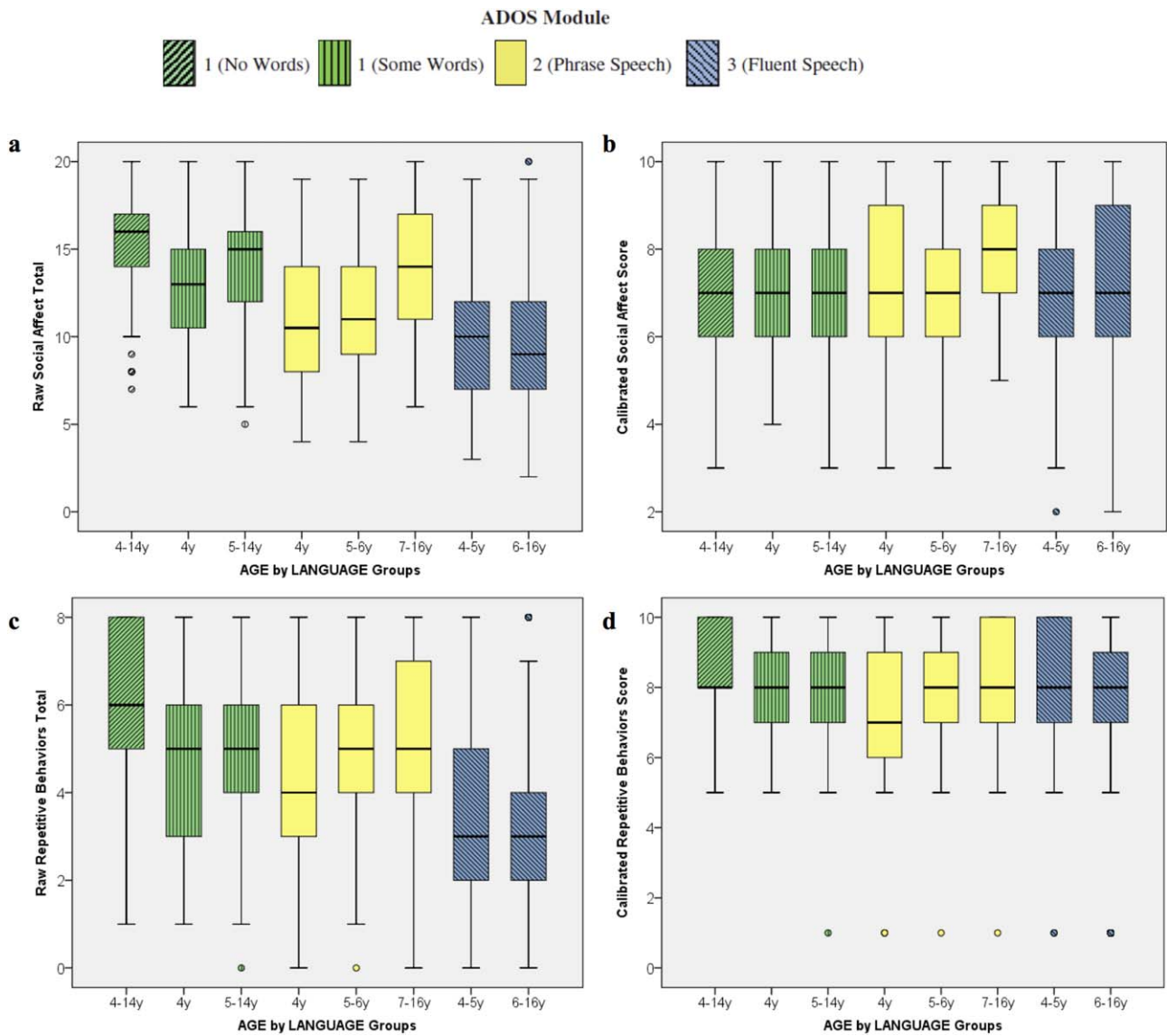
### *Comparison of Raw Domain Totals and Calibrated Domain Scores by Calibration Cell*

As observed in the original calibration sample, distributions of raw SA and RRB domain totals varied by age and language (Fig. 1a, c). Calibrated SA and RRB scores were more uniform, both across and within module groups, though some differences persisted (Table 2 and Fig. 1b, d). Most notably, as was observed with the raw SA totals, the older Module 2 group had significantly higher SA-CSS compared to other age and language groups ( $P < 0.05$  for all comparisons). Moreover, in the Module 1, No Words, the older Module 2 and the younger Module 3 groups, 28–32% of children received the highest RRB-CSS of 10, reflecting high levels of repetitive behaviors during the ADOS.

Mean SA-CSS and RRB-CSS distinguished between children grouped by clinician’s best estimate diagnosis (i.e. Autism vs. Other ASD; SA-CSS:  $t(1318.82) = 15.61$ , RRB-CSS:  $t(1107.18) = 13.83$ ,  $P < 0.001$ ), although as shown in Figure 2, there was marked overlap between the diagnostic groups.

### *Correlations Between Domain Calibrations and Overall CSS*

Correlations between the SA-CSS and RRB-CSS were significant but weak ( $r = 0.13$ ,  $P < 0.001$ ) [Cohen, 1988].



**Figure 1.** a: (top, left) Distributions of raw SA domain totals by age/language cells. b: (top, right) c: Distributions of calibrated SA domain scores by age/language cells. (bottom, left) Distributions of raw restricted and repetitive behavior domain totals by age/language cells. d: (bottom, right) Distributions of calibrated restricted and repetitive behavior domain scores by age/language cells.

Strong correlations between overall CSS and each domain CSS were observed, though relationships were stronger for SA-CSS ( $r = 0.86$ ) compared to RRB-CSS (0.51). This is likely due to the fact that the overall CSS is comprised of a greater proportion of SA items than RRB items.

*Predictors of SA-Raw and SA-CSS*

The final model including all child characteristics as predictors explained 27.0% of the variance in the SA-Raw total. Verbal and nonverbal IQ, nonverbal mental age, chronological age, and maternal education (mothers with graduate/professional degrees vs. all others) were significant predictors of raw SA totals. In contrast,

the same model explained only 3.1% of the variance in the SA-CSS. Verbal mental age and chronological age made small but significant contributions to the SA-CSS.

Next, verbal and nonverbal IQ, chronological age, and maternal education were entered into a Forward Stepwise model to assess the relative contributions of each of these variables in predicting SA-Raw and SA-CSS (see Table 3). Verbal IQ accounted for 26% of the variance in SA-Raw, whereas chronological age (0.7%) and maternal education (0.2%) made minimal contributions; nonverbal IQ was excluded from the model indicating it did not significantly predict SA-Raw. In the forward model predicting SA-CSS, verbal IQ accounted for 2% of the variance and chronological age an additional 0.3%. Nonverbal IQ and maternal education

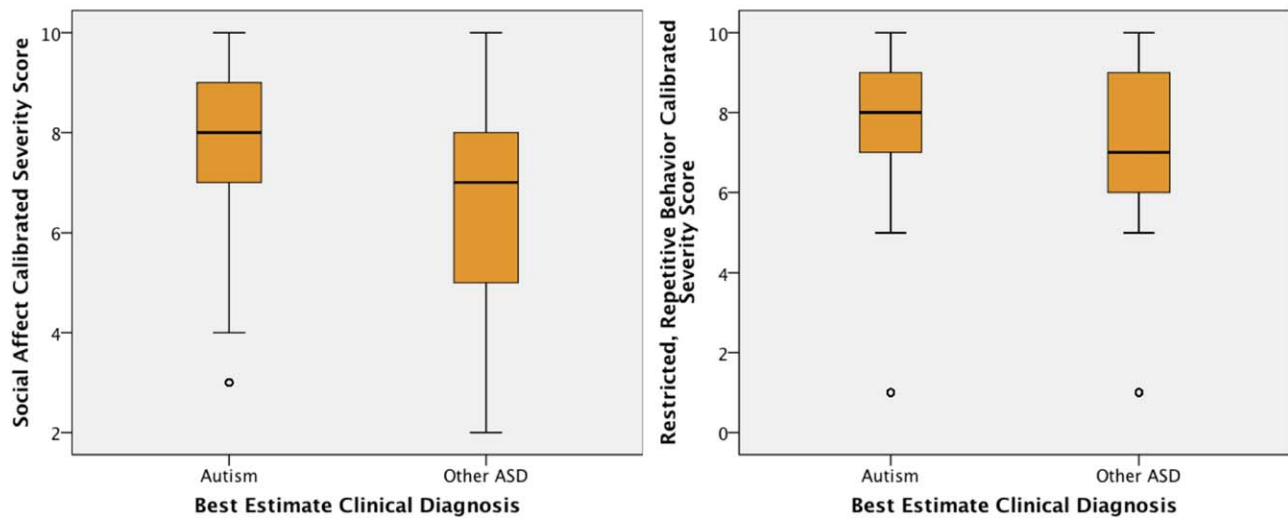


Figure 2. Calibrated scores by best estimate clinical diagnosis collapsed (Autism vs. Other ASD)

were not significant predictors of SA-CSS. Because verbal and nonverbal IQ were highly correlated ( $r = .83$ ), when verbal IQ was removed from the model, nonverbal IQ predicted 18.8% of variance in SA-Raw and 1.2% in SA-CSS. Maternal education was excluded as a predictor from both models.

Finally, CBCL internalizing and externalizing T-scores were entered into separate models predicting SA-Raw and SA-CSS. CBCL internalizing behaviors emerged as a significant predictor but accounted for < 1% of the variance of SA-Raw (overall  $R^2 = 0.011$ ;  $r_{\text{part}} = -0.076$ ;  $P < 0.001$ ). Behavior problems were not significant predictors of SA-CSS.

*Predictors of RRB-Raw and RRB-CSS*

Child characteristics explained 22.8% of variance in the RRB-Raw total. Verbal and nonverbal IQ, verbal and nonverbal mental age and chronological age emerged as significant predictors of raw RRB totals. In contrast, only 4.5% of variance in RRB-CSS was explained by the same model. Nonverbal IQ, verbal and nonverbal mental age, and gender were significant predictors of RRB-CSS.

As shown in Table 3, verbal and nonverbal IQ, chronological age and gender were entered into Forward Stepwise models to assess relative contributions of these child characteristics in predicting RRB-Raw and RRB-CSS. RRB-Raw totals were significantly predicted by verbal IQ (18.5% of variance), chronological age (3.9%), and nonverbal IQ (0.3%). Calibrated RRB scores reduced the influence of child characteristics, with verbal IQ explaining only 2.9% of the variance and chronological age and gender contributing 0.9% and 0.2%, respec-

tively. Again, if verbal IQ was removed from the model, nonverbal IQ predicted 14.9% of variance in RRB-Raw and 2.4% of variance in RRB-CSS.

Behavior problems explained just under 2% of variance in RRB-Raw, with CBCL-internalizing emerging as a small, but significant predictor (overall  $R^2 = 0.024$ ;  $r_{\text{part}} = -0.135$ ,  $P < 0.001$ ). This association was reduced in the model predicting RRB-CSS (overall  $R^2 = 0.012$ ;  $r_{\text{part}} = -0.096$ ,  $P < 0.001$ ).

**Discussion**

Standardized ADOS domain scores have recently been proposed to reduce effects of child characteristics on raw social affect and restricted repetitive behavior totals [Hus et al., 2014]. This is particularly important for geneticists and neuroscientists interested in using scores from the ADOS as dimensions of severity of social-communication and repetitive behaviors. Associations made between ADOS calibrated domain scores and genetic or neurobiological mechanisms are more likely to indicate that the mechanism is influencing social-communication skills or RRBs than associations with raw ADOS totals, which may reflect sample differences in cognitive level or age.

This replication study confirms earlier independent findings [Hus et al., 2014] that the ADOS calibrated domain scores effectively reduced associations with child characteristics compared to raw domain totals in the Simon’s Simplex Collection (SSC). Twenty-seven percent of the variance in SA-Raw totals was explained by non-ASD-specific child characteristics; standardization of scores reduced relationships to 3.1% for SA-CSS. Similarly, associations were reduced from 22.8% of

**Table 3. Forward Stepwise Linear Regression Models for Domain Raw Totals and Calibrated Domain Scores**

	$R^2$	$\Delta F$	df	$B$	SE $B$	$\beta$		$R^2$	$\Delta F$	df	$B$	SE $B$	$\beta$
SA-Raw							SA-CSS						
Step 1	0.26	881.50	1, 2507				Step 1	0.02	51.69	1, 2507			
Constant				16.35	0.19		Constant				7.85	0.10	
Verbal IQ				-0.07	0.00	-0.51	Verbal IQ				-0.01	0.00	-0.14
Step 2		24.12	1, 2506				Step 2	0.02	6.39	1, 2506			
Constant				17.26	0.26		Constant				7.61	0.13	
Verbal IQ				-0.07	0.00	-0.51	Verbal IQ				-0.01	0.00	-0.14
Age				-0.10	0.02	-0.08	Age				0.03	0.01	0.05
Step 3	0.27	6.39	1, 2505										
Constant				17.19	0.27								
Verbal IQ				-0.07	0.00	-0.51							
Age				-0.10	0.02	-0.08							
Mat Ed				0.40	0.16	0.04							
RRB-Raw							RRB-CSS						
Step 1	0.19	569.34	1, 2507				Step 1	.029	74.13	1, 2507			
Constant				6.30	0.10		Constant				8.63	0.10	
Verbal IQ				-0.03	0.00	-0.43	Verbal IQ				-0.01	0.00	-0.17
Step 2	0.22	125.89	1, 2506				Step 2	.038	23.53	1, 2506			
Constant				7.39	0.14		Constant				9.11	0.14	
Verbal IQ				-0.03	0.00	-0.43	Verbal IQ				-0.01	0.00	-0.17
Age				-0.01	0.00	-0.20	Age				-0.06	0.01	-0.10
Step 3	0.23	10.83	1, 2505				Step 3	.040	5.17	1, 2505			
Constant				7.65	0.16		Constant				9.15	0.14	
Verbal IQ				-0.02	0.00	-0.34	Verbal IQ				-0.01	0.00	-0.17
Age				-0.01	0.00	-0.20	Age				-0.05	0.01	-0.09
Nonverbal IQ				-0.01	0.00	-0.10	Gender				-0.24	0.11	-0.04

Note. SA = Social Affect domain; RRB = Restricted, Repetitive Behavior domain; CSS = Calibrated Severity Score. Mat Ed = Maternal Education (graduate/professional degrees vs. all others)

variance in RRB-Raw to 4.5% of RRB-CSS; Verbal IQ emerged as the strongest predictor, explaining only 2–3% of variance of either domain calibrated score. In addition to effects of developmental level, this study explored the potential influence of both internalizing and externalizing behaviors (as measured by the CBCL) on ADOS Raw totals and domain CSS. Only internalizing behaviors emerged as a significant predictor, explaining less than 1% of variance in RRB-CSS.

It is noteworthy that, compared to this study, the original paper by Hus et al. [2014] reported that regression models using the same predictors explained a greater proportion of the variance in SA-Raw (45%) and SA-CSS (15%). In contrast, models from that study predicted lower levels of variance in RRB-Raw (15%) and were comparable for RRB-CSS (5.5%). Differences in the proportions of variance explained by child characteristics in the present replication sample compared to the validation sample may reflect differences in age and cognitive level, as well as differences in the distribution of raw totals for each algorithm group. This variation is most likely a reflection of differences in the purposes for which these samples were ascertained (simplex genetic study vs. primarily clinical referral and research participants) or the time period during which they were collected (SSC participants seen between approximately

2007–2010 vs. original sample collected in the 1990s and early 2000s).

SA-CSS was actually less influenced by child characteristics in the replication sample than the original validation study. Results for RRB-CSS were highly similar. Most important, distributions of domain CSS across age/language cells were more uniform than Raw totals in this sample. In spite of sample differences, the distributions of SA-Raw domain totals followed a similar pattern across studies—higher scores for children with more impaired language (i.e. Module 1 vs. 3) and for older children with similar language (i.e. Module 2 7–16 year olds vs. Module 2 4 year olds). RRB-Raw distributions also followed the same general pattern, though there was greater overlap between children with the greatest language impairment (Modules 1 and 2) in both studies. Replication of domain CSS in population-based samples will be important to assess whether similarities and differences in patterns of score distributions and associations with child characteristics actually reflect meaningful variations in symptom severity across study samples or are reflections of ascertainment differences.

These findings stand in contrast to previous reports examining the influences of child characteristics on scores from parent report measures that are frequently

used to approximate ASD severity. For example, while 22–26% of the variance in ADI-R totals was explained by developmental level (i.e. language and IQ) [Hus & Lord, 2013], SRS scores were strongly associated with behavior problems (as indicated by the CBCL,  $\Delta R^2 = 0.20 - 0.26$  in probands and  $\Delta R^2 = 0.22$  in typical siblings) and more modestly influenced by developmental level ( $\Delta R^2 = 0.12$ ) [Hus et al., 2013]. Influences of non-ASD-specific child characteristics on the SRS and other screening measures, such as the SCQ and Child Communication Checklist have also been demonstrated in other samples [e.g. Charman et al., 2007; Constantino, Hudziak, & Todd, 2003; Kanne, Abbacchi, & Constantino, 2009].

While effects on measures such as the ADI-R and SRS can be statistically controlled if information about developmental level and behavior problems is available, researchers looking for a somewhat more straightforward estimation of ASD severity may wish to turn to the calibrated domain severity scores, which can be computed using existing ADOS data. These findings lend support to the validity of the ADOS domain calibrated scores and suggest that this metric provides a relatively independent measure of social-communication and repetitive behavior dimensions that may be useful for genotype–phenotype analyses and other behavioral explorations. Because the SSC is limited to children above the age of 4, replication in a younger sample is also warranted.

Consistent with the original ADOS domain calibration study, SA and RRB calibrated scores distinguished between children with Autism vs. Other ASD diagnoses; however, there was marked overlap between the two groups. This is not surprising given earlier findings that the designation of categorical diagnoses (i.e. Autism vs. PDD-NOS vs. Aspergers) were unreliable across SSC sites and did not consistently reflect differences in symptom severity [Lord, Petkova, et al., 2012]. It is hoped that domain CSS will capture the heterogeneity in symptom severity that characterizes ASD [Hus et al., 2014]. How the domain CSS relate to DSM-5 severity specifiers has not yet been explored; however, because the ADOS provides behavioral information in a single context, this metric would need to be used in conjunction with other assessment modalities (e.g. parent report, school observation) to appropriately describe the level of support a given individual requires [Hus & Lord, 2014]. Regardless, these findings support the decision to collapse diagnostic categories in DSM-5 to provide a single diagnosis of ASD.

## Limitations

As noted above, given the stringent inclusion and exclusion criteria used by the SSC, this sample may not

be representative of children with ASD in the general population, particularly outside of North America. Moreover, because the SSC only included probands 4 years or older, we were not able to investigate the replicability of the domain CSS in 2 and 3 year olds. Nonetheless, replication in this sample is useful to demonstrate that the ADOS domain CSS effectively reduce effects of child characteristics in another large sample, as well as being of particular interest to researchers using the SSC data.

It is also noteworthy that there were RRB-CSS ceiling effects for three groups: children who were nonverbal or had fewer than five words (Module 1 No Words), older children and adolescents with phrase speech (Module 2 7–16 year olds) and verbally fluent preschool children (Module 3 4–5 year olds). Compared to the original sample used to calibrate ADOS domain scores [Hus et al., 2014], children in the SSC tended to have somewhat higher raw ADOS restricted repetitive behavior totals. This may be related to the SSC's focus on clear cases of ASD, even though SSC study criteria did not require that children demonstrate evidence of restricted and repetitive behaviors on any diagnostic instrument (i.e. ADOS-cutoffs are based upon the overall total, which could be exceeded by high scores on the SA domain and CPEA ADI-R criteria include only cut-offs for the social and communication domains). As noted above, it would be useful to replicate the ADOS domain CSS in a population-based sample to determine how sampling bias may have influenced these distributions or if there are true differences in repetitive behavior severity in the SSC, or in simplex families more broadly, compared to other clinically ascertained samples.

It is also recognized that, while separately calibrated ADOS domain scores provide somewhat more specific indications of social-communication and repetitive behaviors, the ADOS was designed to be a diagnostic instrument (as opposed to providing a dimensional metric of symptoms). Thus, ADOS scores continue to encompass a range of ASD-related behaviors, including specific constructs (social-communication) and subconstructs (e.g. production of facial and nonfacial communication) [NIMH, 2014] proposed in the RDoC framework, as well as other dimensions of behavior that may be separable (e.g. repetitive sensory motor behaviors and insistence on sameness behaviors) [Bishop et al., 2006; 2013]. As such, the ADOS domain calibrated scores are not proposed to be the only, or even the “best,” way to measure dimensions of social-communication and repetitive behaviors for scientists aiming to elucidate the pathophysiology of ASD and other neurodevelopmental disorders. Their development represents an effort to increase the utility of already widely available data in large-scale databases such as the SSC.



## Conclusion

This study provides the first replication of the ADOS domain CSS [Hus et al., 2014]. The ADOS SA and RRB CSS provide separate estimates of severity consistent with studies [e.g. Mandy et al., 2014] showing ASD is best conceptualized as two core dimensions of symptoms: social-communication deficits and RRBs. Behavioral studies often highlight the need for basic science researchers to exercise caution in their selection of measures used to investigate their phenotype of interest. For example, failure to take into account non-ASD-specific influences on various metrics may lead to misleading interpretations of associations between scores and biological mechanisms. In contrast to other phenotype measures that have appropriate-sounding names (e.g. SCQ, SRS) but have clear associations with factors that may confound interpretation as dimensional measures of ASD severity, ADOS domain calibrated scores are relatively independent of child characteristics, such as age, language, cognitive ability, and other behavior problems. It is hoped that the newly calibrated domain scores will be useful in studies investigating the complex links between biology and behavior.

## Acknowledgments

This work was supported by a University of Michigan Lohr Fellowship and Rackham Graduate School Fellowship to V.H.B. and 1R01 MH81873-01A1 (NIMH) to C.L. We gratefully acknowledge the Simons Foundation, as well as the SSC families and 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 (v.14) by applying at <https://base.sfari.org>. Catherine Lord acknowledges receipt of royalties for the ADOS; profits from this study were donated to charity.

## References

Achenbach, T.M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: ASEBA.

Bishop, S.L., Richler, J., & Lord, C. (2006). 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*, 12(4–5), 247–267.

Brune, C.W., Kim, S.J., Salt, J., Leventhal, B.L., Lord, C., & Cook, E.H., Jr. (2006). 5-HTTLPR Genotype-Specific Pheno-

type in Children and Adolescents With Autism. *American Journal of Psychiatry*, 163(12), 2148–2156.

Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., et al. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *The British Journal of Psychiatry*, 191(6), 554–559.

Chaste, P., Klei, L., Sanders, S.J., Murtha, M.T., Hus, V., Lowe, J.K., et al. (2013). Adjusting head circumference for covariates in autism: Clinical correlates of a highly heritable continuous trait. *Biological Psychiatry*, 74(8), 576–584.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Psychology Press

Connolly, J.J., Glessner, J.T., & Hakonarson, H. (2013). A Genome-Wide Association Study of Autism Incorporating Autism Diagnostic Interview–Revised, Autism Diagnostic Observation Schedule, and Social Responsiveness Scale. *Child Development*, 84(1), 17–33.

Constantino, J.N., Hudziak, J.J., & Todd, R. D. (2003). 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*, 42(4), 458–467.

Constantino, J.N., & Gruber, C. (2005). *The Social Responsiveness Scale*. Los Angeles, CA: Western Psychological Services.

Coutanche, M.N., Thompson-Schill, S.L., & Schultz, R.T. (2011). Multi-voxel pattern analysis of fMRI data predicts clinical symptom severity. *NeuroImage*, 57(1), 113–123.

Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, & Centers for Disease Control and Prevention (CDC). (2014). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly Report. Surveillance Summaries* (Washington, D.C.: 2002), 63(2), 1–21.

Dichter, G.S., Richey, J.A., Rittenberg, A.M., Sabatino, A., & Bodfish, J.W. (2011). Reward circuitry function in autism during face anticipation and outcomes. *Journal of Autism and Developmental Disorders*, 42(2), 147–160.

Elliott, C.D. (2006). *Differential ability scales—second edition (DAS-II)*. San Antonio, TX: The Psychological Corporation.

Fischbach, G.D., & Lord, C. (2010). The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron*, 68(2), 192–195.

Geschwind, D.H. (2011). Genetics of autism spectrum disorders. *Trends in Cognitive Sciences*, 15(9), 409–416.

Girirajan, S., Dennis, M.Y., Baker, C., Malig, M., Coe, B.P., Campbell, C.D., et al. (2013). Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *The American Journal of Human Genetics*, 92(2), 221–237.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(5), 693–705.

Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the Social Responsiveness Scale. *Journal of Child Psychology and Psychiatry*, 54(2), 216–224.

Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the Social Responsiveness Scale. *Journal of Child Psychology and Psychiatry*, 54(2), 216–224.

- Hus, V., & Lord, C. (2014). The Autism Diagnostic Observation Schedule, Module 4: Revised Algorithm and Standardized Severity Scores. *Journal of Autism and Developmental Disorders*, 44(8), 1996–2012.
- Hus, V., & Lord, C. (2013). 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*, 43(2), 371–381.
- Hus, V., Pickles, A., Cook, E.H., Risi, S., & Lord, C. (2007). Using the Autism Diagnostic Interview-Revised to increase phenotypic homogeneity in genetic studies of autism. *Biological Psychiatry*, 61(4), 438–448.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., et al. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167(7), 748–751.
- Jones, R.M., & Lord, C. (2013). Diagnosing autism in neurobiological research studies. *Behavioural Brain Research*, 251, 113–124.
- Kanne, S.M., Abbacchi, A.M., & Constantino, J.N. (2009). Multi-informant ratings of psychiatric symptom severity in children with autism spectrum disorders: The importance of environmental context. *Journal of Autism and Developmental Disorders*, 39(6), 856–864.
- Kim, Y.S., & State, M.W. (2014). Recent challenges to the psychiatric diagnostic nosology: A focus on the genetics and genomics of neurodevelopmental disorders. *International Journal of Epidemiology*, 43(2), 465–475.
- Krumm, N., O’Roak, B.J., Karakoc, E., Mohajeri, K., Nelson, B., Vives, L., et al. (2013). Transmission disequilibrium of small CNVs in simplex autism. *The American Journal of Human Genetics*, 93(4), 595–606.
- Lainhart, J.E., Bigler, E.D., Bocian, M., Coon, H., Dinh, E., Dawson, G., et al. (2006). Head circumference and height in autism: A study by the collaborative program of excellence in autism. *American Journal of Medical Genetics Part A*, 140A(21), 2257–2274.
- Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D. M., et al. (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*, 69(3), 306–313.
- Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule: ADOS-2*. Western Psychological Services Torrance.
- Lord, C., Rutter, M., DiLavore, P.S., & Risi, S. (1999). *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles, CA: Western Psychological Services.
- Mandy, W., Charman, T., Puura, K., & Skuse, D. (2014). Investigating the cross-cultural validity of DSM-5 autism spectrum disorder: Evidence from Finnish and UK samples. *Autism*, 18(1), 45–54.
- Mullen, E. (1995). *The Mullen scales of early learning*. Circle Pines, MN: American Guidance Service, Inc.
- National Institute of Mental Health. (2014). Development and definitions of the RDoC domains and constructs. Retrieved November 4, 2014, from <http://www.nimh.nih.gov/research-priorities/rdoc/development-and-definitions-of-the-rdoc-domains-and-constructs.shtml>
- Nordahl, C.W., Lange, N., Li, D.D., Barnett, L.A., Lee, A., Buonocore, M.H., et al. (2011). Brain enlargement is associated with regression in preschool-age boys with autism spectrum disorders. *Proceedings of the National Academy of Sciences*, 108(50), 20195–20200.
- Rutter, M., Bailey, A., Lord, C., & Berument, S. (2003). *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.
- Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised*. Los Angeles, CA: Western Psychological Services.
- Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., et al. (2011). Multiple recurrent De Novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*, 70(5), 863–885.
- Sparrow, S.S., Cicchetti, D.V., & Balla, D.A. (2005). *Vineland adaptive behavior scales* (2nd ed.). Circle Pines, MN: America.
- Uddin, L.Q., Menon, V., Young, C.B., Ryali, S., Chen, T., Khouzam, A., et al. (2011). Multivariate searchlight classification of structural magnetic resonance imaging in children and adolescents with autism. *Biological Psychiatry*, 70(9), 833–841.
- Warren, Z., Vehorn, A., Dohrmann, E., Nicholson, A., Sutcliffe, J.S., & Veenstra-VanderWeele, J. (2012). Accuracy of phenotyping children with autism based on parent report: What specifically do we gain phenotyping “rapidly”? *Autism Research*, 5(1), 31–38.