Molecular Autism

RESEARCH

Open Access

Harmonizing two measures of adaptive functioning using computational approaches: prediction of vineland adaptive behavior scales II (VABS-II) from the adaptive behavior assessment system II (ABAS-II) scores

Corinna Smith[1], Alexandra Lautarescu[2], Tony Charman[2], Jennifer Crosbie[3,4], Russell J. Schachar[3,4], Alana Iaboni[1], Stelios Georgiades[5], Robert Nicolson[6], Elizabeth Kelley[7,8,9], Muhammad Ayub[9], Jessica Jones[8,9], Paul D. Arnold[10], Jason P. Lerch[11,12,13], Evdokia Anagnostou[1,11,14] and Azadeh Kushki[1,15*]

Abstract

Background Very large sample sizes are often needed to capture heterogeneity in autism, necessitating data sharing across multiple studies with diverse assessment instruments. In these cases, data harmonization can be a critical tool for deriving a single dataset for analysis. This can be done through computational approaches that enable the conversion of scores across various instruments. To this end, our study examined the use of analytical approaches for mapping scores on two measures of adaptive functioning, namely predicting the scores on the vineland adaptive behavior scales II (VABS) from the scores on the adaptive behavior assessment system II (ABAS).

Methods Data from the province of Ontario neurodevelopmental disorders network were used. The dataset included scores VABS and the ABAS for 720 participants (autism n = 547, 433 male, age: 11.31 ± 3.63 years; neurotypical n = 173, 95 male, age: 12.53 ± 4.05 years). Six regression approaches (ordinary least squares (OLS) linear regression, ridge regression, ElasticNet, LASSO, AdaBoost, random forest) were used to predict VABS total scores from the ABAS scores, demographic variables (age, sex), and phenotypic measures (diagnosis; core and co-occurring features; IQ; internalizing and externalizing symptoms).

Results The VABS scores were significantly higher than the ABAS scores in the autism group, but not the neurotypical group (median difference: 8, 95% CI = (7,9)). The difference was negatively associated with age (beta = -1.2 ± 0.12, t = -10.6, $p < 0.0001$). All estimators demonstrated similar performance, with no statistically significant differences in mean absolute error (MAE) values across estimators (MAE range: 4.96–6.91). The highest contributing features to the prediction model were ABAS composite score, diagnosis, and age.

Limitations This study has several strengths, including the large sample. We did not examine the conversion of domain scores across the two measures of adaptive functioning and suggest this as a future area of investigation.

*Correspondence:
Azadeh Kushki
akushki@hollandbloorview.ca
Full list of author information is available at the end of the article

Conclusion Overall, our results supported the feasibility of harmonization. Our results suggest that a linear regression model trained on the ABAS composite score, the ABAS raw domain scores, and age, sex, and diagnosis would provide an acceptable trade-off between accuracy, parsimony, and data collection and processing complexity.

Background

Autism is a highly heterogeneous condition with large variability in etiology, neurobiology, and phenotype [6, 14], reflecting the multiplicity of mechanisms that may drive the type and intensity of differences in the core and co-occurring domains. For example, several high impact genetic variants, affecting multiple biological pathways, have been associated with autism, but these differences are present in less than 20% of autistic individuals [23]. Significant heterogeneity also exists in the neurobiology of autism, with mixed findings reported for differences in brain structure, function, and connectivity [5, 8, 12].

This heterogeneity significantly challenges statistical approaches traditionally used in autism research and can manifest as small effect sizes and difficulties with replicability [9, 16]. To understand and capture this heterogeneity, very large sample sizes, often involving thousands of participants, are needed [14]. At the same time, the obtainable sample size from a single study site or research study is limited by the cost and resources needed for data collection. To address this issue, there is significant interest in data sharing across multiple sites and studies to increase the sample size and diversity. Examples of data collection networks that engage in data sharing initiatives within the autism research community include the Province of Ontario Neurodevelopmental Disorders (OBI-POND) network, the Autism Brain Imaging Data Exchange Autism Innovative Medicine Studies-2-Trials (AIMS-2-TRIALS) Longitudinal European Autism Project (LEAP) ([6, 16], the Healthy Brain Network (HBN). More recently, the Autism Sharing Initiative (https://www.autismsharinginitiative.org/) has been established to create the first international platform for federation (a decentralized approach to analysis where data from multiple cohorts are processed locally and not exchanged or pooled [21]) of autism research data.

Data sharing among multiple sites and studies comes with many challenges, including data harmonization (i.e., the process of combining data from multiple sources to derive a single and cohesive dataset for analysis purposes). An example of data harmonization is statistical corrections applied to neuroimaging data collected across multiple sites for effects of differences in scanning sequences/protocols [13]. Data harmonization can also be a significant effort for phenotypic measures, especially when different instruments or versions of the same instrument are used across studies to characterize symptoms within a single domain. For example, autism features can be quantified using various measures including the Social Communication Questionnaire (SCQ,e.g., POND) [17] and the Social Responsiveness Scale (SRS,e.g., LEAP) (Constantino et al., 2003). While these measures conceptually quantify features within similar domains, their scores, although correlated, may not be directly comparable [7]. This challenges the feasibility of combining data across multiple studies and cohorts, necessitating the development of computational approaches that enable the conversion of scores across various instruments. To this end, our study examined the use of different analytical approaches for mapping scores on two measures of adaptive functioning to each other.

Adaptive functioning quantifies the skills necessary for performing everyday tasks across domains of function [3]. Our study focused on two of the most commonly used instruments for adaptive functioning, namely the VABS [18] and the ABAS instruments used in LEAP and POND [11]. Both assessment systems provide a standardized composite score representing skills across the domains of adaptive function. Previous investigations have shown moderate to high correlation between VABS and ABAS scores, for both interview and questionnaire forms of VABS [7] (Sparrow et al., 2005). However, in two studies involving samples of autistic children, systematic differences between the VABS and ABAS were found. While the rank ordering of the participants was relatively preserved between the two measures, the ABAS composite scores were 9.2 points lower than the VABS composite, on average. Furthermore, the ABAS had relatively lower specificity for establishing adaptive functioning below the cut-off of 70. Despite this, scores on the two measures were highly correlated with a Pearson correlation of 0.78 for composite scores, and a range of 0.66 to 0.77 for conceptually similar subdomains [7, 15]. These systematic differences motivate the need for harmonization, while the high correlations support the feasibility of this approach. In this context, the objective of our study was to examine the utility of different computational approaches to map VABS to ABAS scores.

Methods

Participants

The data used in this study were collected by the Province of Ontario Neurodevelopmental Disorders (OBI-POND) network. POND participants are recruited across five sites in Ontario, Canada (Holland Bloorview Kids Rehabilitation Hospital, The Hospital for Sick Children, McMaster University, Lawson Health Research Institute, and Queen's University). Participants at all sites receive the same phenotypic measures. The POND data export was obtained on July 23, 2021 and included a subset of participants 6–21 years old who had completed both the ABAS and VABS assessments between March 2012 and November 2019. VABS assessments were not collected in POND after 2019 and as such our dataset include the full set of participants for whom both ABAS and VABS data are available. The age range was chosen to align with the age range for the ABAS-II school-age form, while minimizing the overlap with the preschool form. Participants had a primary diagnosis of autism confirmed using the Autism Diagnostic Observation Schedule–2 (ADOS) (Lord et al. 2000) and the Autism Diagnostic Interview–Revised (ADI-R) (Lord et al. 1994), or were neurotypical (NT; no history of a neurodevelopmental, psychiatric, or neurological diagnosis, born after 35 weeks gestation). The data were included for participants who had both VABS and ABAS measures available.

Instruments/measures

The vineland adaptive behavior scales II (VABS) scores used in this study were obtained by a researcher using a semi-structured interview with a parent/primary caregiver (Sparrow et al., 2005). The VABS items are scored on a 3-point scale: 0 = behavior never performed, 1 = behavior sometimes or partially performed, and 2 = behavior usually or habitually performed. This instrument provides a composite score as well as domain scores for communication, daily living skills, socialization, and maladaptive behaviors (optional) (Sparrow et al., 2005). In this analysis, we examined the prediction of the VABS composite score (age-normed standard scores; mean 100 ± 15).

The adaptive behavior assessment system II (ABAS) was administered through the parent/primary caregiver forms for 5–21 years and completed by a parent or caregiver [11]. ABAS items are rated on a four-point scale ranging from 0 = is not able, 1 = never when needed, 2 = sometime when needed, 3 = always when needed. The ABAS measures adaptive function in 10 skill areas (communication, functional academics, self direction, community use, home living, health and safety, self care, leisure, social, work). The scaled skill area scores are aggregated into age-normed standard scores (mean 100 ± 15) for three domains (conceptual, practical, social) and a total composite score [11]. The work skill area was excluded from the present analyses given the participants' age. For the prediction task, we used both the raw skill area scores (sum of questionnaire items and the scaled skill area scores (age-normed, mean 10 ± 3. Unlike the VABS, ABAS composite scores have a floor of 40 and an age-dependent ceiling: 160 for 0–5 years, 130 for 5–7 years, and 120 for 8–89 years [11].

To further characterize the sample, IQ was measured using measures appropriate for the child's age and ability level (the Wechsler Abbreviated Scale of Intelligence, the Wechsler Intelligence Scale for Children, and the Stanford Binet Intelligence Scales). Autism-like traits were quantified using the Social Communication Questionnaire (SCQ)—Lifetime form [17]. ADHD-like traits were measured using the Strengths and Weaknesses of Attention-Deficit/Hyperactivity Disorder Symptoms and Normal Behavior Scale (SWAN) parent questionnaire which provides two subscale scores of inattentive and hyperactive/impulsive [20]. Emotional and behavioral symptoms were measured using the child behavior checklist (CBCL [1, 2]) parents/primary caregivers version, Internalizing and Externalizing scores, respectively. ADHD and emotional/behavioural symptoms were selected to characterize the sample given their relatively high prevalence in autism (Lai et al., 2019).

Analytic approach

Statistical tests were conducted in R version 4.3.0 and prediction analyses were performed using scikit-learn version 1.0.2. In this analysis, we predicted VABS scores from ABAS scores given that ABAS is a parent-reported measure and may offer advantages in terms of scalability and cost. We used six regression approaches, namely, ordinary least squares (OLS) linear regression, ridge regression [4], ElasticNet [24], LASSO [22], AdaBoost [10], and random forest (Breiman, 2001). The choice of linear regression models (OLS, ridge, ElasticNet, and LASSO) was motivated by the previously reported linear association between VABS and ABAS scores [7]. Noting that demographic and phenotypic predictor variables are highly correlated in our datasets (e.g., sex, age, autism/ADHD features), we employed ridge regression, ElasticNet, and LASSO, which provide relative strengths in handling multicollinearity compared to OLS regression (Variance inflation factors between 1.1 and 12.8). AdaBoost and random forest regression were used as two ensemble regressors (collection of several regressors) which can improve the bias and variance of estimation through the use of multiple estimators.

In all models, the VABS composite scores were predicted as outcomes. We have chosen to go from ABAS

to VABS given that the ABAS may be more scalable and cost effective in terms of administration as a parent-rated questionnaire (versus the VABS clinician interview), and as such may be more practical to use in a wider range of settings, especially in applications involving the creation of large data cohorts [7].

Four sets of predictor variables were examined. First, we established a minimum set of variables consisting of only the ABAS general composite score (set 1). The second set included the ABAS composite as well as age, sex, and diagnosis (set 2). Age was included in this set given previous findings of residual association with age in the VABS [7]. The third set included variables in set 2 in addition to ABAS raw scores (set 3). These were included to mitigate the impact of ABAS floor and ceiling. The fourth feature set included variables in set 3 as well as IQ, SCQ, SWAN, and CBCL internalizing and externalizing problems (set 4). The SCQ, SWAN, and CBCL variables were included in the prediction to examine the impact of core autism features and frequently co-occurring symptoms (inattention, hyperactivity, internalizing, externalizing) on the difference between VABS and ABAS scores. This was based on previous findings of residual association of VABS scores with these domains. For example, IQ was found to explain 46.6% of the variance in VABS composite scores, but only 36.4% in the ABAS composite scores [7].

We ran the six models with each feature set for participants who had complete scores for all assessments included in the analysis as well as the autism group only (6 models × 3 features × 2 groups). Additionally, sensitivity analysis was conducted by running models for feature sets 2 and 3 using the subset of participants included in feature set 4. We also examined the effect of family nesting, by comparing the performance of the linear regression model with and without accounting for family nesting by including family as a random effect in the model (set 2). Furthermore, we ran the models on a subset of the participants that included one, randomly chosen child from each family.

To evaluate model performance, tenfold stratified cross validation was used. Folds were stratified based on VABS composite scores (10 point bins were used for scores between 60 and 110), sex, and diagnosis. Median absolute error (MAE) was used as a measure of the residuals magnitude to quantify the difference between predicted and true VABS scores. To understand the sources of prediction error, we examined the association between the error from the above linear regression model and the covariates. Residuals were computed as the difference between the predictions of VABS scores and their true values. Predicted scores were obtained from using a linear regression model with coefficient obtained through 1000 bootstrap iterations on feature set 2.

The contribution of each variable to the model was quantified using the permutation feature importance approach which computes feature importance as the decrease in the model R-squared when a given feature is randomly shuffled (*Permutation Feature Importance*, n.d.).

Results

Sample characteristics

The POND sample consisted of 1025 participants with Vineland scores available who were neurotypical or had a diagnosis of autism. Of these participants, 920 had completed ABAS scores. Exclusion of participants younger than 6 years old and having more than two year difference between ABAS and VABS assessment resulted in the sample size of 720. The final sample consisted of a total of 720 participants (autism n = 547, neurotypical n = 173), as described in Table 1. The dataset contained data from the following sites: Holland Bloorview Kids Rehabilitation Hospital (n = 377), Hospital for Sick Children (n = 1), Lawson Health Sciences (n = 25), McMaster University (n = 183), and Queen's University (n = 134). VABS assessments occurred after ABAS assessments for the majority of the sample (median difference = 8 days; IQR = 61 days). Feature sets 1, 2, 3, and 4 contained n = 720, n = 720, n = 719, and n = 615, respectively. Characterization of the excluded participants is provided in Supplementary Table 1.

Characterization of the ABAS-VABS difference

Figure 1 depicts the difference between VABS and ABAS scores for male and female participants in the autism and neurotypical groups, with linear regression lines fitted individually for each group. The VABS-ABAS difference was significantly higher for the autism group (difference mean = −8.0, SE = 1.1, t = −7.45, $p < 0.0001$). The conclusions remained unchanged after accounting for family nesting (difference mean = −8.3, SE = 1.1, t = −7.48, $p < 0.0001$). For the autistic group, the VABS scores were significantly greater than ABAS scores (Wilcoxon signed rank test; Z = 13.20, $p < 0.0001$, estimated median difference: 8.0, 95% CI = (7,9)), but this difference was not statistically significant for the neurotypical group (Wilcoxon signed rank test; Z = 1.08, $p = 0.30$, estimated median difference: −1.0, 95% CI = (−3.5,1.0)). The difference was significantly associated with age (beta = −1.1, SE = 0.1, t = -9.70, $p < 0.0001$), but not with sex (beta = 0.7, SE = 1.0, t = 0.72, $p = 0.47$). The conclusions remained unchanged after accounting for family nesting (age: beta = −1.1, SE = 0.1, t = −9.84, $p < 0.0001$; sex: beta = 0.7, SE = 1.0, t = 0.74, $p = 0.46$). Floor effects were observed for the

Table 1 Demographic characteristics of the sample included in analyses

	All (n = 720)	autism (n = 547)	neurotypical (n = 173)
Site (HB:OC:QU:LHS:SK)	377:183:134:25:1	313:183:25:25:1	64:0:109:0:0
Sex (male:female)	528:192	433:114	95:78
Age at ABAS (years)	11.8 (9.0,14.7)	11.5 (8.8, 14.5)	13.2 (10.1, 15.4)
Age at VABS (years)	11.9 (9.0, 14.8)	11.6 (8.8, 14.7)	13.1 (9.8, 15.4)
Families with multiple children (1: 2: 3)	596:53:6	454:39:5	142:14:1
Race/ethnicity n (%)			
Minoritized	152 (30)	110 (32)	42 (25)
White	360 (70)	234 (68)	126 (75)
Household income n (%)			
< $74,999	158 (31)	114 (33)	44 (26)
$75,000–$199,000	219 (43)	133 (39)	86 (51)
> $200,000	52 (10)	35 (10)	17 (9)
Don't know	25 (5)	18 (5)	7 (4)
Prefer not to answer	60 (12)	45(13)	15 (9)
ABAS composite score	68.0 (54.0, 90.0) n = 720	61.0 (48.0, 73.0) n = 547	105.0 (95.0,117.0) n = 173
VABS composite score	76.0 (64.0, 92.0) n = 720	71.0 (62.0, 79.0) n = 547	104.0 (97.0, 111.0) n = 173
IQ	99.0 (77.0, 111.0) n = 666	90.0(69.0, 107.0) n = 493	108.0 (101.0, 118.0) n = 173
SCQ	17.0 (7.0, 24.5) n = 715	21.0 (15.0, 26.0) n = 545	2.0 (1.0, 3.0) n = 170
CBCL—Internalizing	60.5 (52.0, 68.0) n = 693	65.0 (57.0, 70.0) n = 547	48.0 (39.0, 53.8) n = 146
CBCL—Externalizing	54.0 (46.0, 63.0) n = 693	58.0 (50.0, 65.0) n = 547	41.0 (34.0, 48.0) n = 146
SWAN—Inattentive	3.0 (0.0, 6.0) n = 708	5.0 (2.0, 7.0) n = 539	0.0 (0.0, 0.0) n = 169
SWAN—Hyperactive/Impulsive	2.0 (0.0, 5.0) n = 708	3.0 (1.0, 7.0) n = 539	0.0 (0.0, 0.0) n = 169

The reported values are median (IQR) for continuous measures and absolute numbers for categorical variables

Minorized race/ethnicities include Indigenous, Arab, Black, Chinese, East Asian, Filipino, Japanese, Jewish, Korean, Latin American/Hispanic, South Asian, Southeast Asian, and West Asian. Race/ethnicity and household income data were available for 532 and 534 participants, respectively

HB Holland Bloorview Kids Rehabilitation Hospital, *OC* Offord Centre, *QU* Queen's University, *LHS* London Health Sciences, *SK* Hospital for Sick Children, *ABAS* adaptive behavior assessment system, *VABS* vineland adaptive behavior scales, *SCQ* social communication questionnaire, *CBCL* child behaviour checklist, *SWAN* strengths and weaknesses of attention-deficit/hyperactivity disorder symptoms

ABAS at composite scores of 40 for 63 of participants in the autism group (9%). Ceiling effects were found at an ABAS composite score of 120 for one participant in the autism group (0.1%) and 34 neurotypical participants (28%).

Prediction of VABS scores

Prediction performance is reported for the six computational models and four feature sets in Table 2. All estimators demonstrated similar performance on a given feature set, with no statistically significant differences in MAE values across estimators (Wilcoxon test across cross-validation folds with Bonferroni correction; MAE range: 4.96–6.91). Although feature sets 2–4 resulted in lower error values, the differences were smaller than the clinically significant difference of 2–3.75 points for the VABS (Chatham et al., 2018). When using the autism-only subsample, the MAE decreased across all estimators (MAE range: 4.80–6.83). Again, the differences were smaller than the clinically significant difference. The results of sensitivity analysis, including MAE for models 2 and 3 for the subset of participants included in feature set 4 are provided in Supplementary Table 2.

For the linear regression model, the coefficient estimates and their standard error for feature set 2 were: intercept: mean = 58.10, SD = 2.66; ABAS GAC: mean = 0.57, SD = 0.02, age: mean = − 1.05, SD = 0.10; sex: mean = − 0.96, SD = 0.91; diagnosis: mean = − 11.35,

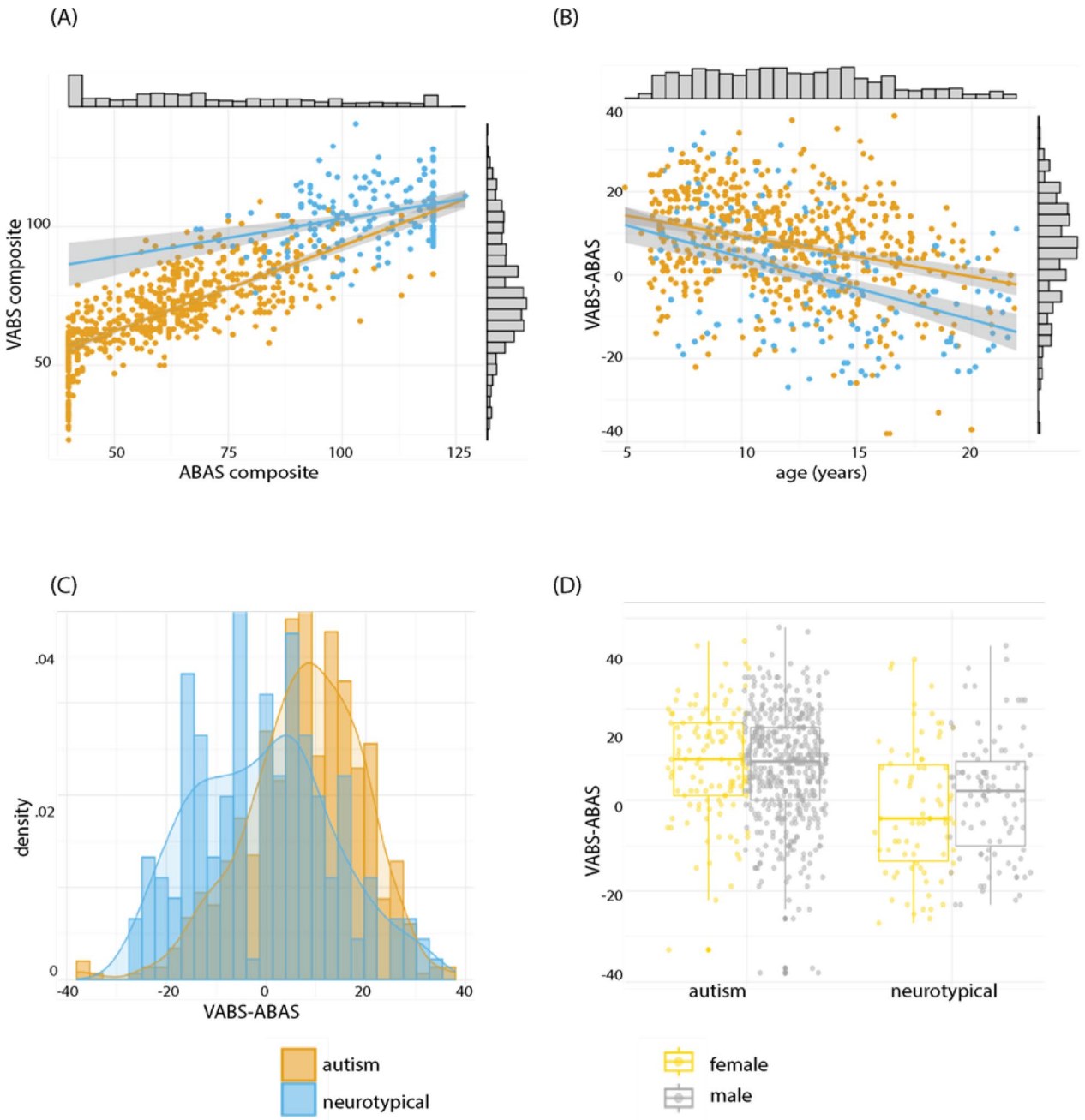

Fig. 1 **A** Association between VABS and ABAS composite scores, with linear regression lines fitted individually for each group; **B** association of the VABS-ABAS difference with age by diagnosis, with linear regression lines fitted individually for each group; **C** histograms of the VABS-ABAS difference demonstrating a diagnosis effect; **D** VABS-ABAS difference by diagnosis and sex. No significant effect of sex revealed

SD = 1.24. The model accounted for 77% of the variability in VABS scores (adjusted R-squared). To examine the effect of family nesting, we compared the performance of the linear regression model with and without accounting for family nesting for set 2. No significant differences were found (difference = 0.00, CI = (−0.04, 0.04), t = −0.03, df = 19, $p = 1.0$). Additionally, we ran the models on a subset of participants that included one, randomly chosen child from each family. The results of these exploratory analyses are included in Supplementary Table 3.

As an exploratory analysis, we ran three additional linear models including interactions terms with age, diagnosis, and sex. The MAE results over 10 folds of cross

Table 2 Median absolute error for the computational models, quantifying the magnitude of the residuals for each model

	Features	Linear regression	Ridge	LASSO	ElasticNet	Random forest	AdaBoost
All	Set 1: ABAS GAC	6.49 ± 0.95[a]	6.49 ± 0.95[a]	6.50 ± 0.93[a]	6.50 ± 0.94[a]	6.91 ± 1.39[b]	6.48 ± 0.94[a]
	Set 2: ABAS GAC + demographics	5.57 ± 0.63[b]	5.54 ± 0.63	6.20 ± 0.74[a]	5.95 ± 0.81[b]	6.41 ± 1.23	5.65 ± 0.67
	Set 3: ABAS GAC + demographics + ABAS raw	5.24 ± 0.71	5.24 ± 0.71	5.23 ± 0.69	5.68 ± 0.49	5.94 ± 0.63	5.30 ± 0.70
	Set 4: ABAS GAC + demographics + ABAS raw + phenotype	4.97 ± 0.36	4.96 ± 0.34	5.11 ± 0.46	5.16 ± 0.68	5.36 ± 1.26	5.05 ± 0.55
Autism only	Set 1: ABAS GAC	5.80 ± 2.03	5.80 ± 2.03	5.80 ± 2.03	5.78 ± 2.03	6.83 ± 1.19[a]	5.77 ± 1.79
	Set 2: ABAS GAC + demographics	5.36 ± 0.50	5.36 ± 0.51	5.42 ± 0.71	5.30 ± 0.59	5.65 ± 1.10	5.44 ± 0.73
	Set 3: ABAS GAC + demographics + ABAS raw	5.11 ± 0.33	5.11 ± 0.34	4.98 ± 0.42	4.94 ± 0.41	5.51 ± 0.73	5.11 ± 0.28
	Set 4: ABAS GAC + demographics + ABAS raw + phenotype	4.76 ± 0.51	4.74 ± 0.50	4.73 ± 1.31	4.75 ± 1.06	5.48 ± 1.10	4.64 ± 0.97

Median error across all participants was reported due to the skewed nature of the error distribution. The values provided are median (Q1, Q3) over the 10 folds of cross-validation

ABAS adaptive behavior assessment system, *GAC* general adaptive composite

[a] Error significantly higher than sets 3 and 4

[b] Error significantly higher than set 4

[c] Error significantly higher than sets 2–4

validation results were not significantly different from the base model (age interaction: Wilcoxon signed rank test; Z = 0, $p = 1.0$, estimated median difference: 0.00, 95% CI = (-0.59,0.73); diagnosis interaction: Wilcoxon signed rank test; Z = -0.19, $p = 1.0$, estimated median difference: −0.12, 95% CI = (−0.82,0.55); sex interaction: Wilcoxon signed rank test; Z = −0.19, $p = 1.0$, estimated median difference: -0.07, 95% CI = (−0.70, 0.70)). The adjusted R2 for the models were 0.776 (base model), 0.776 (age interaction), 0.781 (diagnosis interaction), 0.776 (sex interaction).

Further, we ran the models for prediction of the VABS subdomains of socialization, practical, and conceptual, using the respective ABAS subdomain score and age, sex, and diagnosis as covariates. The results are presented in Supplementary Table 4.

Feature importance values for sets 2–4, averaged over 10 cross-validation folds, are provided in Fig. 2. Across all feature sets, the largest importance values corresponded to the ABAS composite score, diagnosis, and age. Among the ABAS domain scores and phenotypic features, ABAS communication and IQ contributed the most to the model.

Correlates of prediction error

There was no statistically significant effect of diagnosis, sex, age, or site on the residuals.

Discussion

Data harmonization is a key challenge for data sharing across multiple sites and studies. In this paper, we reported the results of a demonstration project, aiming to harmonize scores on two measures of adaptive functioning. To this end, we used computational approaches to predict scores on a clinician/researcher-administered measure of adaptive function (VABS) from scores on a parent-reported measure (ABAS).

Overall, our results supported the feasibility of harmonization, although the prediction error was larger than the minimum clinically significant difference in the VABS (2–3.75 points for the VABS (Chatham et al., 2018)). In particular, we achieved median absolute prediction errors of 4.96 ± 0.34 and 4.80 ± 0.47 points for the full and autism-only samples, respectively, using ridge regression. The should be interpreted in the context of the estimated measurement error for the two instruments (VABS: 2.3–3.8; ABAS: 1.5–2.6) [11, 19]) and the minimum clinically significant difference of 2–3.75 points for the VABS (Chatham et al., 2018). In particular, our achieved accuracy was larger than both by just over one point. The computational methods examined in this study provided similar prediction accuracies, consistent with the linear nature of the association between VABS and ABAS [7]. As such, we recommend the use of linear regression given its relative simplicity. From a clinical perspective, we have demonstrated the feasibility of predicting VABS scores relatively accurately from ABAS scores. Given that the ABAS is a parent-reported measure, this may offer advantages in terms of scalability and cost depending on the setting. These results are also encouraging in the context of data harmonization when ABAS or VABS have been used across different datasets.

We investigated the correlates of the estimation error to determine subgroups in our sample for whom the VABS-ABAS conversion may be particularly accurate or erroneous. We found that participant-specific

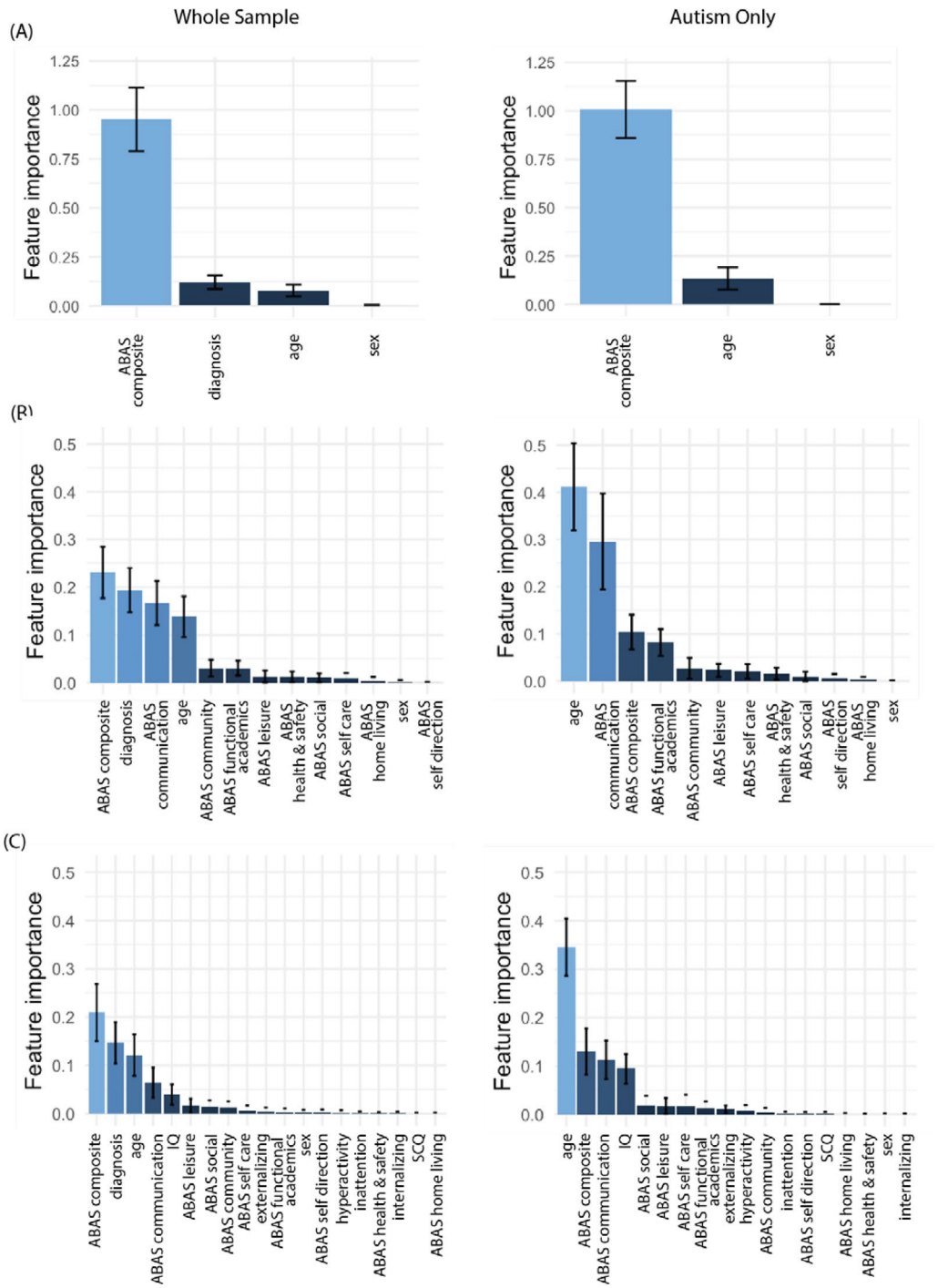

Fig. 2 Feature importance values for linear regression averaged over 10 cross validation folds for **A** feature set 2, **B** features set 3, and **C** feature set 4. Bars represent standard deviation

features available in our dataset, including age, sex, diagnosis, or any of the symptom scores were not significant sources of prediction error. Although diagnosis was not a significant predictor of regression error, the results of the present study replicated the systematic differences between ABAS and VABS scores in the autistic, but not neurotypical, subgroups observed in previous studies [7, 15]. In the context of harmonization, this suggests that conversion of scores should occur on population-specific models.

Measure-specific features may also contribute to differences between the VABS and the ABAS. These, for example, may include differences in the administration method for each instrument (VABS clinical interview versus ABAS parent completed questionnaire measure) and the phrasing of the questions.

In this study, we examined four sets of variables for prediction of VABS scores, each set containing additional features to aid the prediction task. The first set contained only the ABAS composite whereas the second second set consisted of the ABAS composite scores, age, sex, and diagnosis. The third set further included raw ABAS scores to account for ceiling and floor effects. Finally, the fourth set additionally included phenotypic variables characterizing core and co-occurring features (autism-like features: SCQ; ADHD-like features: SWAN; IQ; internalizing and externalizing behaviours: CBCL). While set 4 led to the lowest prediction error, the differences between the results obtained from the four feature sets were smaller than a clinically meaningful difference in the VABS score. As the phenotypic measures may not be available in all datasets, we suggest that a model trained on a feature set consisting of participant demographic variables and ABAS composite and raw scores may be sufficient for most applications.

We quantified the contribution of each variable in the feature set to the predictive model using permutation feature importance. When considering only the ABAS composite, diagnosis, age, and sex (feature set 2), the ABAS composite was the most prominent variable for prediction of VABS scores for both the pooled and autism-only samples. Once ABAS domain scores were added to the feature set, diagnosis, age, and the ABAS communication raw score additionally contributed to the model; However, the ABAS composite remained the most prominent contributor in the pooled sample. This finding is consistent with previous work which found the highest correlation between ABAS and VABS scales to be between the composite scores [7]. Interestingly, for the autism-only subsample, age and the ABAS communication domain scores contributed most to the model. In our results, age emerged as an important feature in the prediction of VABS, especially in the autistic subsample. While ABAS and VABS are both age-normed scores, we also found a significant negative association between age and the VABS-ABAS difference.

Other than IQ, the remaining phenotypic variable did not contribute significantly to the final models, further confirming that feature set 3 is sufficient for prediction of VABS scores in practice. This finding may also reflect the high correlation among these variables, with the ABAS ABAS composite score likely summarizing the information needed for prediction.

Limitations

This study has several strengths, including the large sample assessed using both the VABS and ABAS scores. Additionally, the deep phenotypic characterization of our sample allowed us to investigate the contribution of various domains to the VABS-ABAS relation. We did not examine the conversion of domain scores across the two measures of adaptive functioning and suggest this as a future area of investigation.

Conclusion

In this study, we examined the feasibility of mapping scores on two measures of adaptive functioning namely, the VABS and the ABAS. Our results suggest that a linear regression model trained on the ABAS composite score, the ABAS raw domain scores, and age, sex, and diagnosis would provide an acceptable trade-off between accuracy, parsimony, and measurement effort/cost.

Abbreviations

ABAS	Adaptive behavior assessment system
ABIDE	Autism brain imaging data exchange
ADI-R	Autism diagnostic interview–revised
ADOS	Autism diagnostic observation schedule
AIMS	Autism innovative medicine studies
CBCL	Child behavior checklist
GAC	General adaptive composite
HBN	Healthy brain network
LEAP	Longitudinal European autism project
MAE	Mean absolute error
OLS	Ordinary least squares
POND	Province of Ontario neurodevelopmental disorders
SCQ	Social communication questionnaire
SWAN	Strengths and weaknesses of attention-deficit/hyperactivity disorder symptoms and normal behavior scale
VABS	Vineland adaptive behavior scales

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13229-024-00630-4.

Additional file 1.

Acknowledgements

This research was conducted with the support of the Province of Ontario Neurodevelopmental Network, as part of the Ontario Brain Institute. The opinions, results, and conclusions are those of these authors and no endorsement by the Ontario Brain Institute is intended or should be inferred.

Author contributions

CS and AK substantially contributed to the drafting of the manuscript, concept and design of the study, analytic plan, interpretation of results, and critical review and revision of the manuscript. JC, EK, MA, AI, RS, SG, RN, PA, JP, EA and JJ contributed to the acquisition of data and the critical review and revision of the manuscript. AL and TC contributed to design of study, interpretation of results, and critical review and revision of the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Funding
This work was supported by research funds from the Natural Sciences and Engineering Research Council of Canada and the New Frontiers in Research Fund.

Availability of data and materials
Participants were drawn from the Province of Ontario Neurodevelopmental Disorders (POND) network (exported July 2021). Data from the POND network is available via a controlled data release through Ontario Brain Institute's Brain-CODE: https://www.braincode.ca/). The code for the models is available at https://github.com/azadeh-kushki/ABAS_VABS_conversion.

Declarations

Ethics approval and consent to participate
The POND study was approved by each participating institution's research ethics boards; written informed consent or assent was obtained from the primary caregiver or participant where appropriate.

Consent for publication
Not applicable.

Competing interests
All co-authors meet the requirements for authorship and have reviewed and approved this manuscript. Dr Kushki is the inventor of a software called the "holly". She is involved in commercialising holly (patents US 9,844,332 B2 and US 16/276,208 (pending)) and will financially benefit from its sales. Dr Kushki served on the board of advisors for Shaftesbury, a media company developing virtual reality products for autistic children, from February 2020 to February 2021, and was compensated financially for this role. She has also received consulting fees from a company, DNAStack. Dr Kushki has received donations of hardware for her research program from Samsung Canada. Dr Anagnostou reports grants from Roche, grants from Anavex, personal fees from Roche, personal fees from Quadrant, personal fees from Impel, personal fees from Ono, personal fees from Wiley, book royalties from Springer, book royalties from APPI, and non-financial support from AMO Pharma outside the submitted work. In addition, Dr Anagnostou holds a patent on the software called the "Anxiety Meter"; the Anxiety Meter is being commercialised and she will financially benefit from its sales. The other authors report no potential conflicts of interest.

Author details
[1]Autism Research Centre, Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, 150 Kilgour Road, Toronto, ON M4G 1R8, Canada. [2]Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. [3]Department of Psychiatry, University of Toronto, Toronto, Canada. [4]Department of Psychiatry, The Hospital for Sick Children, Toronto, ON, Canada. [5]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Canada. [6]Department of Psychiatry, Western University, London, Canada. [7]Department of Psychology, Queen's University, Kingston, Canada. [8]Centre for Neuroscience Studies, Queen's University, Kingston, Canada. [9]Department of Psychiatry, Queen's University, Kingston, Canada. [10]The Mathison Centre for Mental Health Research & Education, Cumming School of Medicine, University of Calgary, Calgary, Canada. [11]Program in Neurosciences & Mental Health, The Hospital for Sick Children, Toronto, Canada. [12]Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, FMRIB, University of Oxford, Oxford, UK. [13]Department of Medical Biophysics, University of Toronto, Toronto, Canada. [14]Institute of Medical Science, University of Toronto, Toronto, Canada. [15]Institute of Biomedical Engineering, University of Toronto, Toronto, Canada.

Received: 12 January 2024 Accepted: 4 November 2024
Published online: 03 December 2024

References

1. Achenbach TM, Rescorla LA. Manual for the ASEBA preschool forms and profiles: an integrated system pf multi-informant assessment. University of Vermont, Research Center for Children, Youth & Families; 2000.
2. Achenbach TM, Rescorla LA. Manual for the ASEBA school-age forms and profiles: an integrated system of mult-informant assessment. University of Vermont, Research Center for Children, Youth & Families; 2001.
3. American Psychiatric Association. Neurodevelopmental disorders. In: Diagnostic and statistical manual of mental disorders (DSM-5). American Psychiatric Association; 2013. https://doi.org/10.1176/APPI.BOOKS.9780890425596.DSM01
4. Arashi M, Roozbeh M, Hamzah NA, Gasparini M. Ridge regression and its applications in genetic studies. PLOS ONE. 2021;16(4):4. https://doi.org/10.1371/journal.pone.0245376.
5. Bedford SA, Park MTM, Devenyi GA, Tullo S, Germann J, Patel R, Anagnostou E, Baron-Cohen S, Bullmore ET, Chura LR, Craig MC, Ecker C, Floris DL, Holt RJ, Lenroot R, Lerch JP, Lombardo MV, Murphy DGM, Raznahan A, Chakravarty MM. Large-scale analyses of the relationship between sex, age and intelligence quotient heterogeneity and cortical morphometry in autism spectrum disorder. Mol Psychiatr. 2020;25(3):614–28. https://doi.org/10.1038/s41380-019-0420-6.
6. Charman T, Loth E, Tillmann J, Crawley D, Wooldridge C, Goyard D, Ahmad J, Auyeung B, Ambrosino S, Banaschewski T, Baron-Cohen S, Baumeister S, Beckmann C, Bölte S, Bourgeron T, Bours C, Brammer M, Brandeis D, Brogna C, Buitelaar JK. The EU-AIMS Longitudinal European Autism Project (LEAP): Clinical characterisation. Mol Autism. 2017;8(1):1–21. https://doi.org/10.1186/s13229-017-0145-9.
7. Dupuis A, Moon MJ, Brian J, Georgiades S, Levy T, Anagnostou E, Nicolson R, Schachar R, Crosbie J. Concurrent validity of the ABAS-II questionnaire with the vineland II interview for adaptive behavior in a pediatric ASD sample: high correspondence despite systematically lower scores. J Autism Dev Disord. 2021;51(5):1417–27. https://doi.org/10.1007/s10803-020-04597-y.
8. Ellegood J, Anagnostou E, Babineau BA, Crawley JN, Lin L, Genestine M, DiCicco-Bloom E, Lai JKY, Foster JA, Peñagarikano O, Geschwind DH, Pacey LK, Hampson DR, Laliberté CL, Mills AA, Tam E, Osborne LR, Kouser M, Espinosa-Becerra F, Lerch JP. Clustering autism: Using neuroanatomical differences in 26 mouse models to gain insight into the heterogeneity. Molec Psychiat. 2015;20(1):1. https://doi.org/10.1038/mp.2014.98.
9. Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA. The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. Trends in Cognitive Sciences. 2019;23(7):7. https://doi.org/10.1016/j.tics.2019.03.009.
10. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci. 1997;55(1):1. https://doi.org/10.1006/jcss.1997.1504.
11. Harrison PL, Oakland T. Adaptive behavior assessment system® Second Edition ABAS®-II. Harcourt; 2003.
12. Kushki A, Anagnostou E, Hammill C, Duez P, Brian J, Iaboni A, Schachar R, Crosbie J, Arnold P, Lerch JP. Examining overlap and homogeneity in ASD, ADHD, and OCD: a data-driven, diagnosis-agnostic approach. Transl Psychiatr. 2019. https://doi.org/10.1038/s41398-019-0631-2.
13. Lombardi A, Amoroso N, Diacono D, Monaco A, Tangaro S, Bellotti R. Extensive evaluation of morphological statistical harmonization for brain age prediction. Brain Sci. 2020;10(6):1–12. https://doi.org/10.3390/brainsci10060364.
14. Lombardo MV, Lai MC, Baron-Cohen S. Big data approaches to decomposing heterogeneity across the autism spectrum. Mol Psychiatry. 2019;24(10):1435–50. https://doi.org/10.1038/s41380-018-0321-0.
15. Lopata C, Smith RA, Volker MA, Thomeer ML, Lee GK, McDonald CA. Comparison of adaptive behavior measures for children with HFASDs. Autism Res Treat. 2013;2013:1–10. https://doi.org/10.1155/2013/415989.
16. Loth E, Ahmad J, Chatham C, López B, Carter B, Crawley D, Oakley B, Hayward H, Cooke J, San José Cáceres A, Bzdok D, Jones E, Charman T, Beckmann C, Bourgeron T, Toro R, Buitelaar J, Murphy D, Dumas G. The meaning of significant mean group differences for biomarker discovery. PLoS Comput Biol 2021;17(11):11. https://doi.org/10.1371/journal.pcbi.1009477
17. Rutter M, Bailey A. The social communication questionnaire (SCQ). Western Psychological Services; 2003.

18. Sparrow SS, Balla DA, Cicchetti DV, Harrison PL Vineland adaptive behavior scales. American Guidance Service, Circle Pines. 1984
19. Sparow SS, Cicchetti DV, Balla DA. Vineland adaptive behavior scales, 2nd ed (VinelandTM-II). American Guidance Service; 2005.
20. Swanson JM, Schuck S, Porter MM, Hartman CA, Sergeant JA. Categorical and dimensional definitions and evaluations of symptoms of ADHD: history of the SNAP and the SWAN rating scales. Int J Educ Psychol Assess. 2017;176(12):139–48.
21. Thorogood A, Rehm HL, Goodhand P, Page AJH, Joly Y, Baudis M, Rambla J, Navarro A, Nyronen TH, Linden M, Dove ES, Fiume M, Brudno M, Cline MS, Birney E. International federation of genomic medicine databases using GA4GH standards. Cell Genomics. 2021;1(2):2. https://doi.org/10.1016/j.xgen.2021.100032.
22. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J Royal Stat Soc: Ser B (Methodol). 1996;58(1):1. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
23. Trost B, Engchuan W, Nguyen CM, Thiruvahindrapuram B, Dolzhenko E, Backstrom I, Mirceta M, Mojarad BA, Yin Y, Dov A, Chandrakumar I, Prasolava T, Shum N, Hamdan O, Pellecchia G, Howe JL, Whitney J, Klee EW, Baheti S, Yuen RKC. Genome-wide detection of tandem DNA repeats that are expanded in autism. Nature. 2020;586(7827):80–6. https://doi.org/10.1038/s41586-020-2579-z.
24. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. J Roy Stat Soc Ser B: Stat Methodol. 2005;67(2):2. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Publisher's Note