

Supplemental Online Content

Barbaresi W, Cacia J, Friedman S, et al. Clinician diagnostic certainty and the role of the Autism Diagnostic Observation Schedule in autism spectrum disorder diagnosis in young children. *JAMA Pediatr*. Published online October 17, 2022.
doi:10.1001/jamapediatrics.2022.3605

eMethods. Statistical Analyses.

eTable. Predictions of Diagnostic Stability by Use of Receiver Operating Curve Analyses

eFigure. ROC Curves of Individual Predictors With at Least Fair Effect Sizes.

This supplemental material has been provided by the authors to give readers additional information about their work.

eMethods. Statistical Analyses

The difference between two deviance statistics is distributed as a chi-square and improvement in model fit is estimated as the difference in deviance for the difference in the number of estimated parameters.

A variety of normed, standardized measures of cognitive, language and adaptive function were used at the 8 sites, consistent with usual clinical practice at each site. Given the limitation resulting from combining standardized scores from different measures which had not been previously equated, a categorical score of the merged variable was created using quartiles.

The model estimates the binomial outcome using a logit link function in which the log odds of “consistency of diagnosis” is estimated as follows:

$$\eta_{ij} = \log \left(\frac{\varphi_{ij}}{1-\varphi_{ij}} \right) \quad (\text{Equation 1})$$

With φ being the probability of obtaining a value of 1 versus a zero. These predicted log-odds η_{ij} can be converted to a predicted probability using the following formula:

$$\varphi_{ij}(1|0) = \frac{1}{1+\exp(-\eta_{ij})} \quad (\text{Equation 2})$$

With the probability being bounded between 0-1. Initially, a null model was fit to the data to identify the magnitude of stability versus ASD diagnosis change using the following model:

Level-1 Model

Diagnostic Stability_{ij} = β_{0j}

Level-2 Model

$\beta_{0j} = \gamma_{00} + u_{0j}$

with β_{0j} being equal to η_{ij} as in Equation 1, i denoting a participating child nested within a clinician j ¹. The terms β_{0j} reflects the stability of diagnosis in logits and can then be transformed onto a predicted probability using Equation 2. The term u_{0j} reflects random residual variations around clinicians.

Building multivariate models in the presence of missing data would result in a listwise deletion as valid cases would involve full data only. Consequently, only a subset of predictors with available data was utilized in the multivariate model.

Diagnoses Among Child Participants Not Receiving a Reference Standard Diagnosis of ASD

Other diagnoses among children not receiving a Reference Standard Diagnosis of ASD included the following: language/communication disorders (n=62), global developmental delay (n=30), motor disorders (n=18), anxiety (n=10), attention-deficit/hyperactivity disorder (n=9), stereotypic movement (n=3), social/social language disorders (n=7), delayed developmental milestones (n=7), and behavior problems (n=14).

¹ Clinicians were not nested within sites as the number of sites was too small to satisfy the modeling of an additional level of analysis (i.e., 3-level model).

eTable. Predictions of Diagnostic Stability by Use of Receiver Operating Curve

Analyses

Variable	n	AUC	S.E.	C.I.95	E.S. ^a	Sensitivity %	Specificity %	Cutoff
Child age	349	0.513	0.058	0.459-0.567	Poor	75.80	37.14	-
Child gender (Boys)	349	0.516	0.034	0.462-0.570	Poor	20.38	82.86	-
Child race	348	0.511	0.047	0.449-0.608	Poor	77.00	28.57	-
Child ethnicity	343	0.527	0.033	0.472-0.581	Poor	20.06	85.29	-
Severity of DSM A Criteria	267	0.732 ^b	0.056	0.675-0.785	Fair	71.95	76.19	2
Severity of DSM B Criteria	258	0.656 ^b	0.072	0.595-0.714	Poor	70.42	66.67	-
Clinician Certainty at Time 1	349	0.860 ^b	0.032	0.819-0.894	Good	76.43	80.00	7
ADOS Classification	238	0.595 ^c	0.054	0.530-0.658	Poor	63.33	53.57	-
ADOS Module Scores	220	0.664 ^b	0.053	0.598-0.726	Poor	42.71	85.71	-
Cognition	145	0.523	0.082	0.439-0.607	Poor	51.52	61.54	83
Vocabulary	76	0.612	0.148	0.493-0.722	Poor	75.71	66.67	-
Language	188	0.707 ^b	0.075	0.637-0.771	Fair	66.86	76.92	68
Adaptive Behavior	218	0.593	0.060	0.525-0.659	Poor	34.01	90.48	-
Time child directly observed	325	0.623 ^b	0.046	0.518-0.674	Poor	46.53	74.29	-
Evaluation by Trainee	347	0.541	0.036	0.487-0.594	Poor	88.14	20.00	-
Availability of Measures at time of Index Diagnosis								
Cognitive	349	0.585 ^c	0.044	0.531-0.637	Poor	57.01	60.00	-
Language	348	0.563	0.042	0.509-0.616	Poor	44.09	68.57	-
Social Behavior	332	0.530	0.045	0.475-0.585	Poor	54.55	51.43	-
Adaptive Behavior	308	0.536	0.044	0.478-0.593	Poor	44.32	62.86	-
Behavioral Problems								

Self-Injurious	312	0.555	0.018	0.498-0.611	Poor	13.82	97.14	-
Aggressive	309	0.508	0.038	0.451-0.565	Poor	78.83	22.86	-
Hyperactive	328	0.516	0.044	0.460-0.571	Poor	63.14	40.00	-
Inattentive/Distractible	332	0.527	0.045	0.471-0.581	Poor	59.60	45.71	-

Abbreviations: AUC, area under the curve; S.E., standard error; C.I.95, 95th percentile confidence interval; E.S., effect size; DSM A Criteria, Diagnostic and Statistical Manual of Mental Disorders category A autism spectrum disorder symptoms of deficits in social communication and social interaction; DSM B Criteria, Diagnostic and Statistical Manual of Mental Disorders category B autism spectrum disorder symptoms of restricted, repetitive patterns of behavior, interests, or activities; DSM, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; ADOS, Autism Diagnostic Observation Schedule

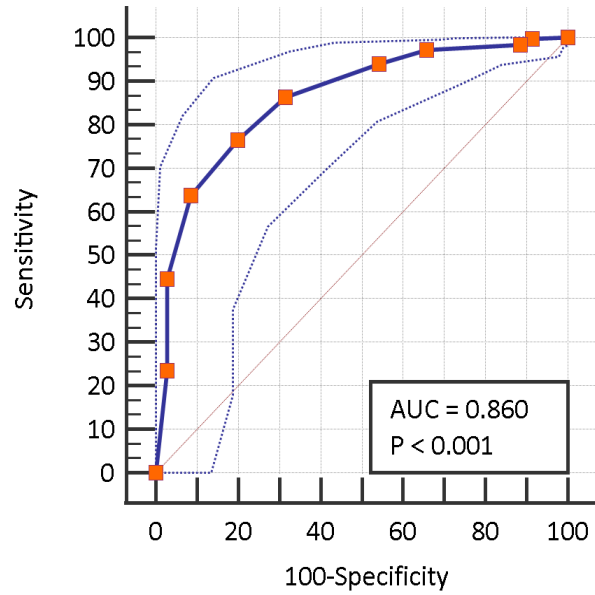
^aConventions on effect size are: 90%–99% = excellent; 80%–89% = good; 70%–79% = fair; 60%–69% = poor. Cutoff values are not shown in the presence of non-significant AUCs, poor prediction by use of effect size conventions, or nominal predictors. (-) dashed lines indicate the absence of estimates as poor prediction renders those estimates meaningless.

^bp<.05, two-tailed test

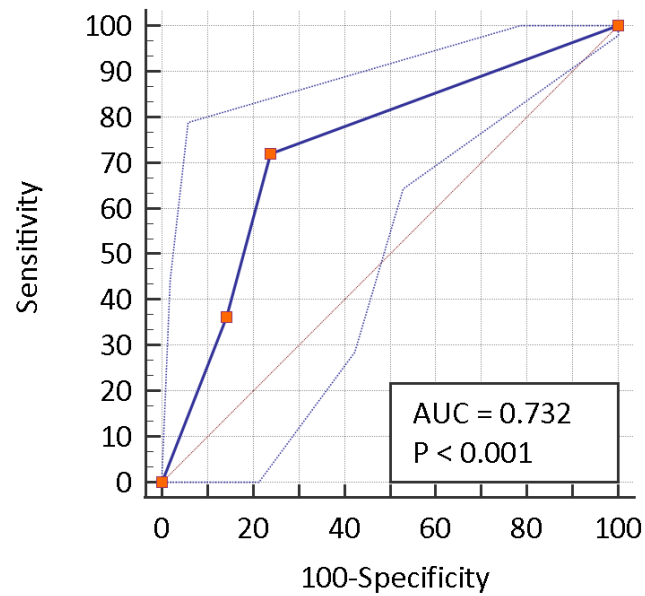
^cp<.05, one-tailed significance.

eFigure. ROC Curves of Individual Predictors With at Least Fair Effect Sizes.

**Prediction of Consistency Between Index and Reference Diagnosis
from Clinician Diagnostic Certainty at Time of Index Diagnosis**



Severity of Social Communication Behaviors



Prediction of Stability of Diagnosis from Language Scores

