



Published in final edited form as:

JAMA Netw Open. ; 5(4): e229503. doi:10.1001/jamanetworkopen.2022.9503.

Race and Sex Bias in the Autism Diagnostic Observation Schedule (ADOS-2) and Disparities in Autism Diagnoses

Zachary J. Williams, BS

Medical Scientist Training Program, Vanderbilt University School of Medicine, Nashville, Tennessee; Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, Tennessee; Vanderbilt Brain Institute, Vanderbilt University, Nashville, Tennessee.

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental condition that affects an estimated 1 in 44 children in the US, with a male-to-female ratio of approximately 4:1.¹ The prevalence of diagnosed ASD has increased substantially over the past 2 decades, although this increase in prevalence has been greater in certain demographic groups, such as female patients² and those from minoritized racial ethnic groups,³ suggesting the presence of diagnostic disparities by race and sex.⁴ Notably, the prototypical behavioral manifestations of ASD (on which existing diagnostic criteria and standardized diagnostic instruments are based) were derived from samples of children who were predominantly White and male,⁴ and, thus, systematic biases in the diagnostic tools used to evaluate individuals with suspected ASD could theoretically contribute to observed diagnostic disparities.

Kalb et al⁵ sought to investigate this issue by quantifying the magnitude and practical impact of race-based and sex-based bias in the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2), a clinician-administered measure that is widely used in both research and clinical practice to establish or confirm a diagnosis of ASD. Leveraging ADOS-2 data from 6269 youth attending specialist diagnostic evaluations at a university-based ASD clinic, the authors⁵ quantified the degree to which different versions (ie, modules) of the ADOS-2 systematically underestimated the features of ASD in Black/African American children (vs White children) and female children (vs male children). Although their analysis did reveal significant race-based bias in 8 items and sex-based bias in 5 items,⁵ estimated effect size metrics⁶ indicated that for all but 2 items (D4, Repetitive Interest [race-based bias], and D2, Hand Mannerisms [sex-based bias]), these effects were small and unlikely to be of practical significance. Moreover, for all ADOS-2 modules tested, the maximum difference in expected ADOS-2 total scores attributable to measurement bias was less than 1 scale point (range, 0.07–0.91 point).⁵ Overall, these findings indicate that the degree of race

Open Access. This is an open access article distributed under the terms of the [CC-BY License](#).

Corresponding Author: Zachary J. Williams, BS, Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, 1215 21st Ave S, Medical Center East, Rm 8310, Nashville, TN 37232 (zachary.j.williams@vanderbilt.edu).

Conflict of Interest Disclosures: Mr Williams reported receiving consulting fees from Roche and Autism Speaks. He also serves as a member of the autistic researcher review board of the Autism Intervention Network for Physical Health and as a community partner for the Autism Care Network Vanderbilt site. No other disclosures were reported.

and sex bias present in the ADOS-2 is low and unlikely to contribute to the systematic underdiagnosis of ASD in Black or female children.

A major strength of the study by Kalb et al⁵ is its innovative use of item response theory (IRT) models to examine the ADOS-2 items and quantify the degree of bias (or differential item functioning [DIF] in IRT terms) between demographic groups. IRT is a modern, large-sample, psychometric method used to develop, evaluate, and score psychological tests,⁷ and IRT models provide mathematical descriptions of how certain item responses (eg, the endorsement of a symptom on the ADOS-2 as 0, 1, or 2) relate to unmeasured latent variables (ie, ASD severity in the case of the ADOS-2) that are assumed to underlie all items on the scale. Within an IRT framework, researchers are able to test for DIF between groups by examining whether a given item is differentially related to the latent variable in 2 or more samples. For instance, in the study by Kalb et al,⁵ the ADOS-2 item D2 (Hand Mannerisms) was found to be more difficult for female children, meaning that clinicians were less likely, on average, to rate this symptom as present for female children compared with male children with the same underlying level of ASD severity. However, given the large samples often used in IRT research, tests of DIF are often powerful enough to detect trivially small differences between groups that do not translate to meaningful between-group biases in practice. To better contextualize their findings, Kalb and colleagues⁵ additionally reported DIF effect size metrics,⁶ which allowed them to quantify the degree of bias in each item, as well as the expected difference in ADOS-2 total scores between male and female or White and Black children with the same ASD severity levels (known as differential test functioning [DTF]). Observed DIF effect sizes for most items were quite small, and when combining all DIF across items to quantify DTF, total bias was well below the proposed cutoffs for practically significant DTF proposed by the authors (ie, 2 points on the ADOS-2 scale or a standardized difference of 0.2 SD units).⁵ Thus, despite the significant DIF by race and sex observed in the current study,⁵ the degree of bias in ADOS-2 total scores was small, suggesting that measurement bias in this widely used instrument contributes little to the diagnostic disparities reported in the epidemiological literature.

In sum, the study by Kalb et al⁵ leverages a large clinical data set to show that biases inherent in the ADOS-2 algorithm are not the primary reason that ASD diagnoses may be missed or delayed more often in certain groups. Importantly, the results of this study do not mean that race-based or sex-based disparities in ASD diagnoses do not exist or are not significant; rather, the findings suggest that these disparities are associated with factors other than bias in the clinical evaluation, which may include disparities in access to health care, differential patterns of specialist referrals, or different levels of parental concern in response to early signs of ASD. Moreover, because the study was limited to individuals younger than 18 years and only evaluated DIF in 1 racial minority group (Black children), substantial bias in the ADOS-2 could still be present when the tool is used to evaluate adults or different racial or ethnic minority groups than the one tested (eg, Asian, Pacific Islander, Native American, or Hispanic/Latinx children). Despite not being able to definitively pinpoint the specific factors associated with the race-based or sex-based disparities in ASD diagnosis, Kalb et al⁵ provide considerable evidence to suggest that these disparities are not due to inherent biases in the specialist diagnostic evaluation itself. Additional research on this topic is necessary to further explore the primary factors associated with race-based and sex-based

diagnostic disparities, laying the groundwork for targeted public health interventions that seek to promote earlier and more equitable ASD diagnoses for individuals from multiple minoritized groups.

Funding/Support:

Mr Williams is supported by the National Institute on Deafness and Other Communication Disorders (grant F30-DC019510), National Institute of General Medical Sciences (grant T32-GM007347), the Nancy Lurie Marks Family Foundation, and the Misophonia Research Fund.

Role of the Funder/Sponsor: The funders had no role in the analysis or interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Maenner MJ, Shaw KA, Bakian AV, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years: Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2018. *MMWR Surveill Summ.* 2021;70(11):1–16. doi:10.15585/mmwr.ss7011a1
2. Russell G, Stapley S, Newlove-Delgado T, et al. Time trends in autism diagnosis over 20 years: a UK population-based cohort study. *J Child Psychol Psychiatry.* Published online August 19, 2021. doi:10.1111/jcpp.13505
3. Nevison C, Zahorodny W. Race/ethnicity-resolved time trends in United States ASD prevalence estimates from IDEA and ADDM. *J Autism Dev Disord.* 2019;49(12):4721–4730. doi:10.1007/s10803-019-04188-6 [PubMed: 31435818]
4. Lai MC, Lin HY, Ameis SH. Towards equitable diagnoses for autism and attention-deficit/hyperactivity disorder across sexes and genders. *Curr Opin Psychiatry.* 2022;35(2):90–100. doi:10.1097/YCO.0000000000000770 [PubMed: 35084380]
5. Kalb LG, Singh V, Hong JS, et al. Analysis of race and sex bias in the Autism Diagnostic Observation Schedule (ADOS-2). *JAMA Netw Open.* 2022;5(4):e229498. doi:10.1001/jamanetworkopen.2022.9498 [PubMed: 35471566]
6. Tay L, Meade AW, Cao M. An overview and practical guide to IRT measurement equivalence analysis. *Organ Res Methods.* 2015;18(1):3–46. doi:10.1177/1094428114553062
7. Thomas ML. Advances in applications of item response theory to clinical assessment. *Psychol Assess.* 2019;31(12):1442–1455. doi:10.1037/pas0000597 [PubMed: 30869966]