



# Defining in Detail and Evaluating Reliability of *DSM-5* Criteria for Autism Spectrum Disorder (ASD) Among Children

C. E. Rice<sup>1,2</sup> · L. A. Carpenter<sup>3</sup> · M. J. Morrier<sup>1</sup> · C. Lord<sup>4</sup> · M. DiRienzo<sup>2</sup> · A. Boan<sup>3</sup> · C. Skowyra<sup>5</sup> · A. Fusco<sup>6</sup> · J. Baio<sup>2</sup> · A. Esler<sup>7</sup> · W. Zahorodny<sup>6</sup> · N. Hobson<sup>8</sup> · A. Mars<sup>9</sup> · A. Thurm<sup>10</sup> · S. Bishop<sup>11</sup> · L. D. Wiggins<sup>2</sup>

Accepted: 13 November 2021

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021

## Abstract

This paper describes a process to define a comprehensive list of exemplars for seven core Diagnostic and Statistical Manual (DSM) diagnostic criteria for autism spectrum disorder (ASD), and report on interrater reliability in applying these exemplars to determine ASD case classification. Clinicians completed an iterative process to map specific exemplars from the CDC Autism and Developmental Disabilities Monitoring (ADDM) Network criteria for ASD surveillance, *DSM-5* text, and diagnostic assessments to each of the core *DSM-5* ASD criteria. Clinicians applied the diagnostic exemplars to child behavioral descriptions in existing evaluation records to establish initial reliability standards and then for blinded clinician review in one site (phase 1) and for two ADDM Network surveillance years (phase 2). Interrater reliability for each of the *DSM-5* diagnostic categories and overall ASD classification was high (defined as very good .60–.79 to excellent  $\geq .80$  Kappa values) across sex, race/ethnicity, and cognitive levels for both phases. Classification of *DSM-5* ASD by mapping specific exemplars from evaluation records by a diverse group of clinician raters is feasible and reliable. This framework provides confidence in the consistency of prevalence classifications of ASD and may be further applied to improve consistency of ASD diagnoses in clinical settings.

**Keywords** Autism · Autism spectrum disorder · Diagnosis · *DSM-5* · Classification · Clinician reliability

## Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder defined by a profile of persistent challenges in social communication and interaction and the presence of restrictive, repetitive patterns of behavior, interests, or activities that are present early in childhood and cause clinically significant functional impairments (APA, 2013). At the individual level, being diagnosed with ASD can inform intervention, supports, and understanding of a person's strengths and challenges. At a community or population-level, diagnoses of autism spectrum conditions can inform policies, services, and supports available and may impact ultimate functioning. Understanding why and how people are diagnosed with behavioral conditions such as ASD can provide insight into the meaning of ascribing such labels for individuals and communities. This paper describes an iterative process of specifying and reliably applying exemplars to endorse the presence/absence of each of the core ASD *DSM-5* criteria and to, ultimately, classify children as meeting an ASD diagnostic classification. This goal is to increase

---

✉ C. E. Rice  
cqr8@cdc.gov

<sup>1</sup> Emory University, Atlanta, GA, USA

<sup>2</sup> Centers for Disease Control and Prevention, Atlanta, GA, USA

<sup>3</sup> Medical University of South Carolina, Charleston, SC, USA

<sup>4</sup> University of California Los Angeles, Los Angeles, CA, USA

<sup>5</sup> Washington University in St. Louis, St. Louis, MO, USA

<sup>6</sup> Rutgers New Jersey Medical School, Newark, NJ, USA

<sup>7</sup> University of Minnesota, Minneapolis, MN, USA

<sup>8</sup> Independent Consultant, Keller, TX, USA

<sup>9</sup> Hunterdon Healthcare System, Flemington, NJ, USA

<sup>10</sup> National Institute of Mental Health, Bethesda, MD, USA

<sup>11</sup> University of California San Francisco, San Francisco, CA, USA

conceptual and practical consistency in the way clinicians consider behavioral criteria for diagnosis ASD.

Autism was first described as a condition by Kanner (1943) and “Infantile Autism” was first included in *the Diagnostic and Statistical Manual (DSM)* in 1980 as one subtype of several “Pervasive Developmental Disorders” (PDDs) (APA, 1980). ASD is now considered in the *DSM-5* as a single categorical spectrum of behaviors with heterogeneous core and co-occurring features (APA, 2013). Across taxonomic changes, diagnosis has relied on behaviorally-defined diagnostic criteria to meet a categorical diagnosis (Volkmar & McPartland, 2014). Guided by current diagnostic criteria, professionals with appropriate training, credentials, and experience (Shulman et al., 2020a) assess a person’s profile of behavior to determine if the evidence supports endorsing persistent deficits in the domains of Social Communication and Interaction (SCI) and Restricted and Repetitive Behaviors or Interests (RRBI) that are present early, cause impairment, and are not better explained by other conditions, specifically intellectual disability or global developmental delay (APA, 2013).

Given the behaviorally-defined and heterogeneous nature of an ASD diagnosis, clinician judgment is an important factor to consider in arriving at an ASD diagnosis. Studies comparing community-based diagnoses of ASD with an independent assessment battery and best-estimate diagnosis found a lack of agreement on ASD diagnoses for 1/4 of the children (Hausman-Kedem et al., 2018; Wiggins et al., 2020). In another study, only 60% of differential diagnoses for ASD were made with high clinician certainty (McDonnell et al., 2018). An analysis of clinical team diagnostic discussions indicated that subjective “feelings of ASD” and professional evaluation of the quality of informant reports were prominent among clinicians when making the diagnosis (Hayes et al., 2020). Given the identification of ASD by affirming the presence of a profile of relevant behaviors, understanding the way different professionals ascribe specific behaviors as evidence of the diagnosis can add specificity to clinical and research efforts to link symptoms with treatment or with risk factors or etiology in research.

Research has focused on developing and evaluating tools to increase the reliability of autism spectrum diagnoses (including “autistic disorder” and other PDD historical subtypes) for both research and clinical purposes compared to “best-estimate clinical judgment” of experienced research clinicians (Bishop et al., 2017; Hausman-Kedem et al., 2018; Kamp-Becker, 2018; Klin et al., 2000; Lord et al., 2012a, 2012b). Specific behaviors that constitute the way these criteria are identified, measured, and applied individually and together are often guided by the specific questions asked in diagnostic instruments (Aiello et al., 2017; Rutherford et al., 2016; Shulman et al., 2020b; Wiggins et al., 2015). Instruments, such as the *Autism Diagnostic Observation Schedule*

(*ADOS-2*) (Lord et al., 2000) and *Autism Diagnostic Interview-Revised (ADI-R)* (Lord et al., 1994), have organized specific autism-relevant behaviors rated from person observation and caregiver interview and provide research-based algorithms that determine whether the person satisfies the instrument criteria for ASD with validation comparing to clinical diagnosis. However, meeting the threshold on a single instrument is not sufficient for diagnosis. Ultimately, ASD diagnosis depends on clinician confirmation of the *DSM-5* criteria from multiple sources of information (Shulman et al., 2020a; Volkmar & McPartland, 2014), and little attention has been given to how clinicians identify and endorse behavioral exemplars that constitute the evidence needed for an ASD diagnosis. Despite robust research on the overall diagnosis and supporting subdomains (Bishop et al., 2016; Frazier et al., 2012; Lord et al., 2012a, 2012b; McPartland et al., 2012; Thurm et al., 2019; Uljarevic et al., 2017; Zheng et al., 2020), research on behavioral exemplars that clinicians consider in endorsing the social and behavioral diagnostic criteria is lacking.

The *DSM-5* ASD criteria were described broadly to encompass a range of symptom presentations across age and developmental level. Beyond the initial categorical diagnosis (ASD/non-ASD), further indicators of functioning such as language and cognitive levels, adaptive behavior, and co-occurring symptoms and features also must be specified (APA, 2013; Bolte & Diehl, 2013; DiRezze et al., 2016; Gardner et al., 2018). Moreover, some investigators recommend shifting from dependence on overall (categorical) diagnosis to recognition of specific phenotypic features, behaviors, and/or biologic pathways and to intervention research that cuts across behaviorally-defined psychiatric conditions (Constantino & Charman, 2016; Cuthbert & Insel, 2013). Clinical and research trends have emphasized a dimensional approach in overall symptom distribution across the population (Constantino & Todd, 2005; Riglin et al., 2020), subdimensions of primary social and behavioral domains of ASD (Bishop et al., 2016; Uljarevic et al., 2017), and stipulation of specifiers such as cognitive level, language capacities, and the presence of other co-occurring conditions (APA, 2013; Gardner et al., 2018). With the move to “lump” the overall diagnostic category into a single ASD diagnosis without subtypes, and “split” the dimensional aspects across the profile of symptoms and into domains and by co-occurring features, the challenge has become establishing valid and reliable categorical *and* dimensional ways of characterizing behaviors to inform etiologic and treatment research and meaningful supports (Gardner et al., 2018; Lord et al., 2012a, 2012b). Careful specification of the behaviors representing each of the ASD diagnostic criteria may be informative in clarifying a dimensional approach to the spectrum of presentation representing each criterion. For example, the diagnostic criterion A3 “deficits in developing,

maintaining, and understanding” relationships is likely to be influenced by multiple dimensions such as age, verbal and nonverbal communication skills, cognitive profile, intensity of other autism spectrum behaviors, etc. Evidence of the deficits in relationships applied to a toddler with minimal expressive verbal language and limited attention to other people manifests differently than for an adult with above average expressive language who lacks flexible understanding of unstated social rules; however, the end result may still be sufficient criteria to endorse criterion A3. Documenting behaviors across age and developmental levels that represent the expression of each of the ASD social and behavioral criteria is important in evaluating the reliability, validity, and utility of further defining each diagnostic criterion.

Specification of the behaviors within each criterion may also be useful for reliably classifying ASD over time in the population (Arvidsson et al., 2018; Durkin et al., 2015; Miller et al., 2013). According to the CDC’s Autism and Developmental Disabilities (ADDM) Network, ASD prevalence increased from an estimated 1 in 150 in 2000 to 1 in 59 among 8-year-old children in 2014 based on *DSM-IV-TR* (CDC, 2007; Baio et al., 2018) and 1 in 54 children in 2016 based on *DSM-5* criteria (Maenner et al., 2020). ADDM prevalence estimates from surveillance years 2000–2016 were derived from multiple sites across the US using consistent procedures previously described (Bertrand et al., 2001; CDC, 2007; Rice et al., 2007; Yeargin-Allsopp et al., 2003). Across these years, ADDM ASD case confirmation required clinician reviewers to reliably apply a detailed assessment of behavioral exemplars documented in existing evaluation records as meeting evidence of the *DSM-IV-TR* autism/PDD criteria. Establishing and maintaining reliability across multiple sites and clinicians in endorsing each criterion has been dependent on operationalizing the diagnostic criteria.

With the advent of the *DSM-5* criteria, a process was undertaken to update operationalized criteria for ASD in a manner that could be replicated in an effort to specify key behaviors associated with the autism spectrum behind the overall and domain levels in clinical and research diagnoses of ASD. The objective of this paper is to describe an iterative process used to define the spectrum of exemplars for each of the seven core *DSM-5* ASD criteria and associated features to determine ASD case status, and to evaluate inter-rater reliability at the criterion and case classification levels.

## Methods

Identifying and coding *DSM-5* exemplars for each of the diagnostic criteria followed procedures previously advanced by the ADDM Network. ADDM has published population-based estimates for ASD every two years since surveillance year 2000 (CDC, 2007) through 2016 (Maenner et al., 2020).

In these reports, ADDM procedures involved a two phase approach with phase 1: abstracting diagnostic and behavioral descriptions from developmental evaluation reports contained in health and education service records (including professional completing the evaluation, reason for referral, all text descriptions of development and behavior, type and results of ASD assessment tools, adaptive and intellectual assessment results, and diagnostic summary); and phase 2: reviewing the information to identify and code *DSM-IV-TR* (APA, 2000) behaviors and associated characteristics; and using a standardized coding procedure to classify a child as ASD or non-ASD based on clinician review of behaviors described in developmental evaluation reports and by consensus of clinicians when certainty of classification was in question. The standardized coding procedures were outlined in an ASD clinician review manual created to establish and maintain consistency in classification of ASD across a network of clinicians with varied of clinical and research backgrounds.

The standardized ADDM coding procedure used involved line-by-line review of information abstracted from health and education records by a study clinician to determine if text descriptors exemplified the *DSM* criteria, associated features (AFs), or other coding items such as timing and type of early developmental concerns. Based on the exemplars identified and the *DSM* criteria endorsed across all available information in the record, the child was classified as an ASD Case (meeting *DSM-5* ASD criteria), Suspected Case (meeting some, but not all criteria and not counted in final case counts), or Not Case (no evidence of social criteria or previous ASD diagnosis or special education eligibility). Clinicians also provided ratings based on their overall assessment of the quality and consistency of the information to indicate their degree of certainty of the final classification, and levels of impairment. Based on all information in the composite record, including IQ and Adaptive test scores, when available, the clinicians’ overall certainty included their evaluation of the ASD diagnostic criteria of C, D, and E. (APA, 2013), indicating early manifestations of symptoms, impairment, and not being better accounted for by intellectual disability, global delay, or other potential conditions. If the clinician classified the child as an ASD Case based on the coding of exemplars, but was not certain they met criteria for ASD, a second independent clinician review was conducted, and a final consensus case classification was made by the two clinicians.

## Operationalizing DSM-5 ASD Criteria for Case Review

An initial criterion-level *DSM-IV-TR* to *DSM-5* mapping was completed and reported in a study estimating impact of the change in criteria on ASD prevalence (Maenner et al., 2014).

More detailed specification of criterion-based exemplars and assessment of reliability in using these exemplars to endorse *DSM-5* ASD criteria was conducted in two phases.

### Phase 1: Adaptation and Pilot Study

Specification of the detailed criteria was first implemented in the South Carolina ADDM site as part of a supplemental study to compare *DSM-IV-TR* and *DSM-5* ASD case classification through record review and direct screening with in-person clinical evaluation from a subset of the SC 2012 ADDM surveillance year (SC South Carolina Children's Educational Surveillance Study, SUCCESS; Carpenter et al., 2016). An initial criterion-level *DSM-IV-TR* to *DSM-5* mapping was completed (Supplementary Appendix, Fig. 1). To provide greater specification needed for evaluation of record review, a team of clinicians (subset of authors) completed an iterative process to identify specific exemplars for each of the seven core *DSM-5* behavioral diagnostic criteria and associated features (AFs). AFs were defined as behaviors commonly exhibited by individuals with ASD, but not represented in the primary diagnostic criteria such as abnormalities in sleeping, attention deficits, and self-injurious behavior. The operationalization of the *DSM-5* criteria included mapping exemplars to the ASD criteria (Table 1) from the *DSM-5* text (APA, 2013), the ADDM *DSM-IV-TR* PDD clinician review manual exemplars, and *ADOS-2* and *ADI-R* items (Huerta et al., 2012). Additional exemplars were identified from clinicians' examples from clinical experience and examples from other assessment tools or diagnostic evaluation reports. At each phase of mapping, discussions were conducted to clarify discrepancies and to reach consensus on the exemplar-criterion/feature match. The process resulted in a version 1 clinician review manual detailing concepts and specific behavioral exemplars within each of the Seven Social/Communication and Behavioral criteria.

Two clinicians (CER and LAC) piloted the mapping in the SC SUCCESS Study by independently applying the classification guide to sample composite health and educational evaluation records for a random sample of 10 children with either an autism spectrum or potentially related diagnosis, classification, or social behaviors associated with ASD documented in evaluation records. Following the pilot and refinement of the classification manual, composite records were reviewed and classified by two clinicians with a 10% independent reliability sample. Acceptable agreement was defined as at least 90% agreement on Final ASD Case Status and at least 80% agreement on each of the *DSM* criteria with corresponding Kappa values  $\geq 0.80$  were considered excellent; 0.60–0.79 very good; 0.40–0.59 good; 0.20–0.39 questionable; and lower than 0.20 unacceptable (Clarke et al., 2013; Klin et al., 2000).

### Phase 2: ADDM Network Implementation

Next the *DSM-5* ASD clinician review manual for SC ADDM was evaluated by ADDM clinicians. The ADDM clinician review team included over 30 professionals with a diverse range of experience with the developmental evaluation of children with and without ASD (e.g., speech-language therapists, licensed and research psychologists, developmental pediatrician). Based on ADDM clinician review, some behaviors were re-classified as an example of another diagnostic criterion or associated feature and sub-headings were added to improve reliability. These changes were especially important in distinguishing between A1. "Deficits in social emotional reciprocity" and A3. "Deficits in developing, maintaining, and understanding relationships." Ultimately, clinicians were able to distinguish these two criteria by coding A1 as the *observable* exchange of social behaviors and A3 as the *implied* or conceptual awareness of, interest in, and understanding of other people/relationships. A similar process was undertaken for each SCI and RRBI criteria and core concepts summarized (Table 2).

Revisions to the SC ADDM manual were reviewed by an ADDM clinician workgroup who analyzed sample records and resolved mapping and coding questions. Clinician review was then implemented for the 2014 surveillance year (Baio et al., 2018) across 11 ADDM sites. All reviewers met pre-training, pre-specified reliability standards of at least 90% agreement on Final ASD Case Status and at least 80% agreement on each of the *DSM* criteria. A blinded 10% interrater reliability review was then implemented. Following the conclusion of the 2014 surveillance year, the clinician workgroup reviewed and made additional clarifications to the coding guide, established initial reliability, and completed a 10% blinded reliability review for the 2016 surveillance year (Maenner et al., 2020). (Extract of Criterion Behavioral Exemplars from the ADDM SY2016 *DSM-5* ASD Clinician Review Manual, Supplementary Appendix).

### Results

The phase 1 SC SUCCESS study represented a population of 8780 8-year-old children from which 240 children had records abstracted for possible ASD based on a documented classification or suggestion of autism spectrum behaviors in source records. For the pilot reliability sample, composite records detailing developmental evaluations ( $n=79$ ) of 10 children were independently reviewed by two clinicians. Percent and weighted average agreement met or exceeded acceptable thresholds with 90% ( $\kappa=0.78$ ) agreement on final ASD case status. For full phase 1 reliability sample, 2 clinicians completed blinded reviews for 155 developmental evaluations compiled for 20 children (Table 3) with

**Table 1** Example of Mapping Exemplars for *DSM-5* (APA, 2013) Autism Spectrum Disorder (ASD) Criterion A1

<i>DSM-5</i> ASD criterion text	Abnormal social approach and...	Failure of normal back-and-forth conversation	To reduced sharing of interests, emotions, or affect	To failure to initiate... to social interactions	To failure to... respond to social interactions
<p><b>A1. Exemplars from <i>DSM-5</i> Text</b></p> <ul style="list-style-type: none"> <li>Deficit in the ability to engage with others</li> <li>Difficulties processing and responding to complex social cues</li> </ul>	<p>Abnormal social approach and...</p>	<p>Failure of normal back-and-forth conversation</p> <ul style="list-style-type: none"> <li>Language often one-sided, lacking in reciprocity</li> <li>Language used to request or label rather than comment, share feelings, or converse</li> </ul>	<p>To reduced sharing of interests, emotions, or affect</p> <ul style="list-style-type: none"> <li>Deficits in sharing thoughts or feelings</li> <li>No sharing of emotions</li> </ul>	<p>To failure to initiate... to social interactions</p> <ul style="list-style-type: none"> <li>Little or no initiation of social interaction</li> </ul>	<p>To failure to... respond to social interactions</p> <ul style="list-style-type: none"> <li>Reduced or absent imitation of others' behavior</li> </ul>
<p><b><i>DSM-IV-TR</i> Criteria (APA, 2000)</b></p> <p>1d. Lack of social or emotional reciprocity</p>	<p>Abnormal social approach and...</p> <ul style="list-style-type: none"> <li>Touching or acting on others without regard to other's involvement or reaction (must indicate lack of social awareness instead of hyperactive or impulsive touching)</li> </ul>	<p>Failure of normal back-and-forth conversation</p> <p>2b. In individuals with adequate speech, marked impairment in the ability to initiate or sustain a conversation with others</p>	<p>To reduced sharing of interests, emotions, or affect</p> <p>1a. Marked impairment... emotional expression... regulate social interaction 1c. A lack of spontaneous seeking to share enjoyment, interests, or achievements with other people (e.g., by a lack of showing, bringing, or pointing out objects of interest)</p>	<p>To failure to initiate... to social interactions</p> <ul style="list-style-type: none"> <li>Does not seek help when needed</li> <li>Does not seek out others</li> </ul>	<p>To failure to... respond to social interactions</p> <ul style="list-style-type: none"> <li>Does not respond to name</li> <li>Lack of response to people</li> <li>Often appears deaf in response to people's talking or noisemaking that is being directed to the child to get their attention (without hearing loss)</li> </ul>
<p><b><i>DSM-IV-TR</i> Exemplars from ADDM Network Clinician Review*</b></p> <ul style="list-style-type: none"> <li>Cuddle, failure to (infancy)</li> <li>Lack of emotional reciprocity</li> <li>Lack of social reciprocity</li> <li>Does not or limited engagement in simple imitation games</li> </ul>	<p>Abnormal social approach and...</p> <ul style="list-style-type: none"> <li>Marked impairment in the ability to initiate or sustain conversation</li> <li>Deficit in pragmatic language (social use)</li> <li>Monologues, one-sided conversation</li> <li>Can talk, but doesn't</li> <li>Says things that do not make sense without clarification</li> <li>Tangential, associative talk</li> <li>Self-directed communication</li> </ul>	<p>Failure of normal back-and-forth conversation</p>	<p>To reduced sharing of interests, emotions, or affect</p>	<p>To failure to initiate... to social interactions</p>	<p>To failure to... respond to social interactions</p>
<p><b><i>ADOS-2</i>, Module (AM#) and item; and <i>ADI-R</i> Item</b></p> <ul style="list-style-type: none"> <li>AMT: -</li> <li>AM1: B14</li> <li>AM2: B9, B10, B11</li> <li>AM3: B8, B10</li> <li>AM4: B10, B12</li> <li>ADI: -</li> </ul>	<p>Abnormal social approach and...</p> <ul style="list-style-type: none"> <li>AMT: B15</li> <li>AM1: B12</li> <li>AM2: B8</li> <li>AM3: B7</li> <li>AM4: B9</li> <li>ADI: 56</li> </ul>	<p>Failure of normal back-and-forth conversation</p> <ul style="list-style-type: none"> <li>AMT: A2, A9</li> <li>AM1: A2</li> <li>AM2: A5, A6, B10</li> <li>AM3: A6, A8, B9</li> <li>AM4: A6, A7, A8, B12</li> <li>ADI: 34, 35</li> </ul>	<p>To reduced sharing of interests, emotions, or affect</p> <ul style="list-style-type: none"> <li>AMT: B6, B11, B12</li> <li>AM1: A8, A9, B5, B8, B9</li> <li>AM2: B3, B5</li> <li>AM3: A5, B4, B5</li> <li>AM4: A5, B4, B5, B6</li> <li>ADI: 52, 53, 54</li> </ul>	<p>To failure to initiate... to social interactions</p> <ul style="list-style-type: none"> <li>AMT: B8, B9, B10, B13, B16</li> <li>AM1: B7, B10, B12, B13</li> <li>AM2: A2, B6, B8, B9</li> <li>AM3: B7, B8</li> <li>AM4: B9, B10</li> <li>ADI: 55</li> </ul>	<p>To failure to... respond to social interactions</p> <ul style="list-style-type: none"> <li>AMT: B7, B14</li> <li>AM1: B2, B6, B11</li> <li>AM2: B4, B7, B9</li> <li>AM3: B9</li> <li>AM4: B11</li> <li>ADI: 46, 59, 63</li> </ul>

**Table 1** (continued)

<i>DSM-5</i> ASD criterion text	A1. Deficits in social-emotional reciprocity, ranging, for example, from...	Abnormal social approach and...	Failure of normal back-and-forth conversation	To reduced sharing of interests, emotions, or affect	To failure to initiate... to social interactions	To failure to... respond to social interactions
Other Exemplars from clinician observations, reports, tests, etc. *	<ul style="list-style-type: none"> <li>One-sided interactions (impairment in reciprocal interaction; for example, child only hugs when he or she initiates)</li> </ul>		<ul style="list-style-type: none"> <li>Talking or verbalizing to oneself in the presence of others (when talking to self would be considered socially unacceptable or odd)</li> <li>Disturbance in the pragmatic (social use) of language</li> </ul>	<ul style="list-style-type: none"> <li>Very few attempts to get, maintain, or direct others' attention</li> </ul>	<ul style="list-style-type: none"> <li>Gets things for self when it would be "easier" to ask for help; walks away rather than requesting help; does not request assistance in general or indicate need in situation where it would be expected)</li> </ul>	<ul style="list-style-type: none"> <li>Does not respond to others or the social environment, when there is clear social opportunity</li> <li>Often appears deaf in response to people's talking or noisemaking that is being directed to the child to get their attention (without hearing loss)</li> </ul>

\*Sample of exemplars included, complete mapping not included for space reasons (See Supplementary Appendix for more complete Exemplar List)

90% agreement ( $\kappa = 0.70$ ) on the *DSM-5* criteria (range 85%–95%;  $\kappa = 0.57$ –0.85); 92% ( $\kappa = 0.83$ ) on early developmental concerns; 85% ( $\kappa = 0.69$ ) on AFs; 90% ( $\kappa = 0.69$ ) on the presence of an earlier PDD/ASD diagnosis; and 90% ( $\kappa = 0.62$ ) on overall ASD Case Classification ( $n = 17$  ASD Case;  $n = 3$  not ASD Case) (Table 3). Manual refinements were undertaken to increase clarity, particularly for any item with less-than-optimal reliability (B2, AF Cognitive Scatter) prior to the next phase including all ADDM sites.

In phase 2, 11 ADDM sites for the 2014 surveillance year with a base population of 263,775 8-year-old children had 9304 records abstracted for possible ASD. For the blinded 10% reliability review, 34 clinicians from 11 sites completed review of 4319 evaluations for 924 children. Agreement standards were at or well above prescribed levels with the exception of "good" agreement on criterion A1. The percent agreement was between 82% (B3) to 89% (A2 and B4) on *DSM-5* criteria (range:  $\kappa = 0.58$ –0.78) and 92% ( $\kappa = 0.85$ ) on overall ASD Case Classification ( $n = 449$  ASD Case;  $n = 475$  not ASD Case) (Table 3). There was consistently high interrater agreement on ASD Case Status across sites (MN, 87% to AR, 97%) and by sex (male, 92% and female, 95%), race/ethnicity (Asian or Pacific Islander, 90% to Hispanic, 100%), cognitive level (Average or Above Average, 91% to No cognitive data, 95%), and previous classification of ASD (ASD Suspected, 85% to ASD classification on record, 97%) (Table 4). Further, interrater agreement was consistently high for endorsement of all seven ASD diagnostic criteria across sex, race/ethnicity, and cognitive level (Fig. 1).

Clinician review procedures were repeated for the ADDM Surveillance Year 2016 in 11 sites ( $n = 952$  reliability pairs) with excellent overall agreement on Case/Non-Case Classifications (0.95;  $K = 0.89$ ) [range 89%, GA to 99%, AR], and a range of overall percent agreement on the *DSM-5* criteria from a low of 87% (A3 and B3) to a high of 92% (B2 and B4).

## Discussion

Meaningful clinical diagnosis of ASD involves an appreciation of the complex presentation and functional impact and clinical support of the caregivers and identified individual (Craddock & Mynors-Wallis, 2014). Better specification of the clinical decision process may increase the reliability and validity of the ASD diagnosis for clinical and research purposes. While there is evidence on the stability of early research-based autism spectrum diagnoses, there is conflicting evidence of community clinician agreement on these diagnoses (Daniels et al., 2011; Hausman-Kedem et al., 2018; van Daalen et al., 2009; Wiggins et al., 2020). At the same time, little is known about how clinicians apply

**Table 2** Overview of core *DSM-5* autism spectrum disorder (ASD) criterion-level concepts for the Social Communication and Interaction (SCI) and Restricted and Repetitive Behavior and Interest (RRBI) Domains

A1: Deficits in social emotional reciprocity	A2: Deficits in nonverbal communicative behaviors	A3: Deficits in developing, maintaining, and understanding relationships	
Involves the exchange of social behaviors (responding, initiating, and reciprocating with others)	Used to communicate affect, regulate social interactions, or supplement language	Involves an awareness of, interest in, and/or understanding (insight) of other people/relationships and playing with children	
Impairments in: <ul style="list-style-type: none"> <li>• Initiating social interactions</li> <li>• Responding to social interactions</li> <li>• Interacting with others (initiating or responding) given social opportunity</li> <li>• Conversing with others</li> <li>• Sharing enjoyment, interests, or objects</li> <li>• Imitating others</li> <li>• Cuddling with familiar persons</li> <li>• Reciprocating social interactions</li> </ul>	Impairments in: <ul style="list-style-type: none"> <li>• Using facial expressions (including smiles)</li> <li>• Using eye contact</li> <li>• Using gestures</li> <li>• Integrating verbal and nonverbal communication</li> <li>• Using other people to communicate</li> <li>• Having appropriate voice characteristics (e.g., intonation)</li> </ul>	Impairments in: <ul style="list-style-type: none"> <li>• Being aware of others</li> <li>• Being interested in others</li> <li>• Preferring to be with others</li> <li>• Differentiating adults and others, and self from others</li> <li>• Adjusting behavior to suit various social contexts</li> <li>• Understanding social conventions</li> <li>• Having an unusual quality of social overtures or responses</li> <li>• Playing with children</li> </ul>	
B1: Stereotyped or repetitive motor movements, use of objects or speech	B2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behavior	B3. Highly restricted interests that are abnormal in intensity or focus	B4. Hyper- or hypo-reactivity to sensory input or unusual interest in sensory aspects of the environment
<ul style="list-style-type: none"> <li>• Demonstrating motor stereotypies</li> <li>• Demonstrating repetitive use of objects</li> <li>• Demonstrating stereotyped or repetitive use of speech</li> </ul>	<ul style="list-style-type: none"> <li>• Insisting on sameness</li> <li>• Showing inflexible adherence to routines or restricted patterns of behavior</li> <li>• Having ritualized patterns of verbal behavior</li> <li>• Having rigid thinking patterns</li> </ul>	<ul style="list-style-type: none"> <li>• Having highly restricted interests or obsessions</li> <li>• Having unusual interests</li> <li>• Focusing on parts of objects</li> </ul>	<ul style="list-style-type: none"> <li>• Showing unusual and general sensory reactions or interests</li> <li>• Showing unusual and specific sensory reactions or interests (i.e., sound, smell, texture or touch, visual, vestibular, pain or temperature, or food-based reactions)</li> </ul>

the specific diagnostic criteria to arrive at those diagnoses. In addition, although practice guidance emphasizes the importance of using evidence-based tools in making a diagnosis, these tools rarely provide explicit instruction on how the overall score or the information obtained relates to the ultimate endorsement of *DSM* criteria, and diagnosis of behaviorally-defined conditions such as ASD cannot be reduced to a clinically-agnostic formula of applying criteria as has been attempted in some machine learning efforts (Bone et al., 2015). This paper reports an iterative process to operationalize the *DSM-5* criteria for ASD with specific behavioral exemplars and guidance to classify children as an “ASD Case” for multi-site surveillance. These results indicated that a diverse group of clinicians can agree upon more nuanced definitions of each of the ASD *DSM-5* criteria and reliably evaluate the presence/absence of these criteria based on the same set of information.

Overall, our findings indicate that a diverse group of clinicians from multiple sites across the US can reliably categorize specific behavioral exemplars documented in evaluation records to classify presence/absence of *DSM-5* criteria, associated features, ASD case status, and other overall diagnostic specifiers. The process and product of operationalizing the *DSM-5* ASD criteria resulted in a detailed guide which can facilitate a transparent and consistent process for evaluating

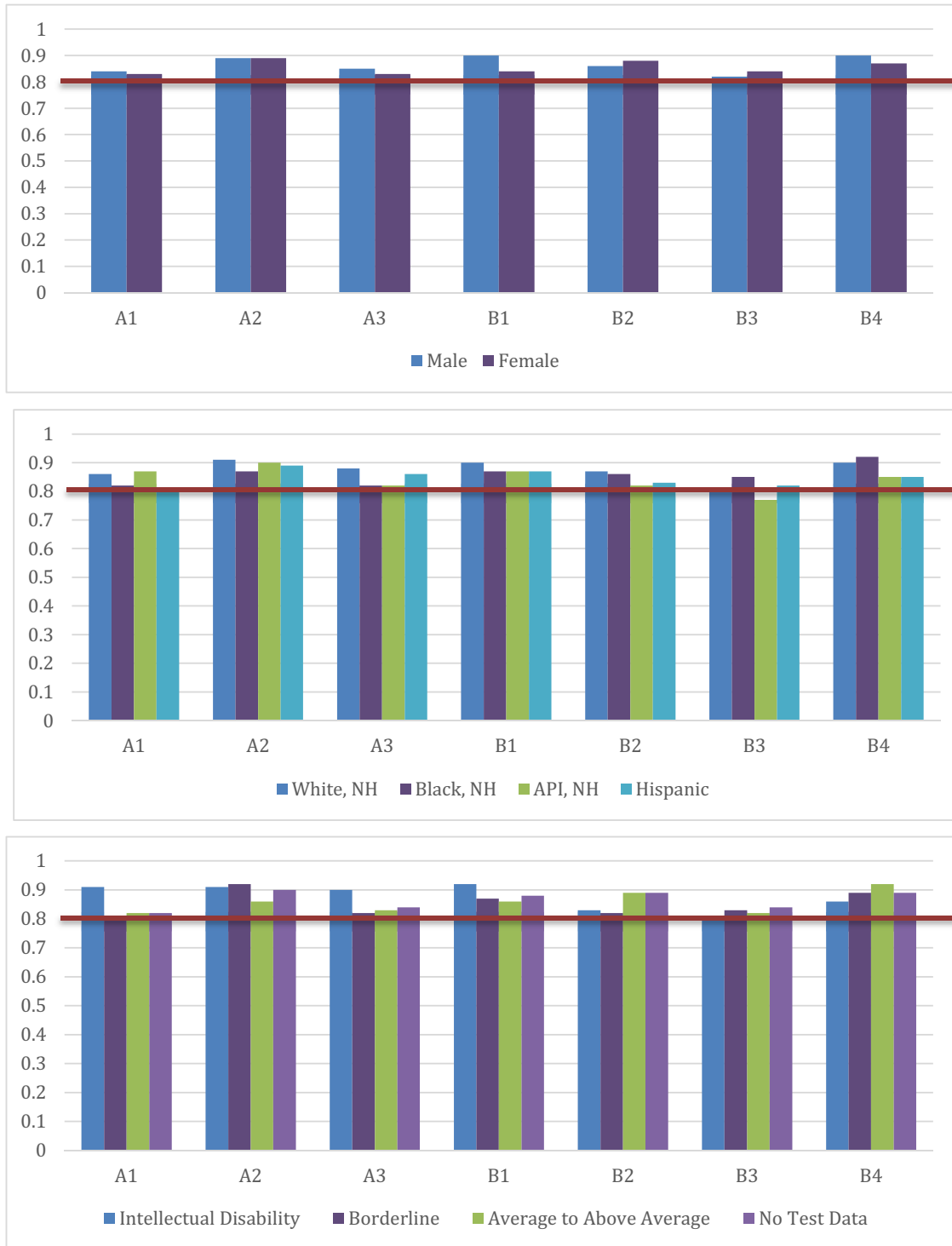
**Table 3** Reliability samples, South Carolina (SC) SUCCESS and the autism and developmental disabilities monitoring (ADDM) network, surveillance year 2014

<i>DSM-5</i> criterion	SC SUCCESS n = 20 pairs (children) 155 evaluations 2 reviewers		ADDM network 2014 n = 924 pairs (children) 4329 evaluations 34 reviewers	
	% Agree	Kappa	% Agree	Kappa
A1	.90	.62	.84	.58
A2	.90	.62	.89	.73
A3	.90	.62	.85	.61
B1	.90	.62	.88	.71
B2	.85	.57	.86	.72
B3	.90	.79	.82	.64
B4	.95	.86	.89	.78
Final ASD case classification	.90	.62	.92	.85

both the specific behavioral descriptions and the *DSM* criteria, as well as the overall clinical decision on the presence of ASD. These findings can have implications for research and clinical practice.

We found high reliability across the initial study in SC and across ADDM for 2 surveillance years in application of exemplars to criteria from records-based descriptions and in the overall case classification. In addition, we found

no significant variation based on site location, the child's sex, or race/ethnicity. Robust agreement on the presence of the *DSM* criteria and final classification was also found across cognitive levels, despite the varied functional and



**Fig. 1** ASD clinician review percent agreement (y-axis) on each *DSM*-5 criterion (x-axis) for autism spectrum disorder by (a) sex, (b) race/ethnicity, and (c) cognitive level; ADDM Network, Surveillance Year 2014



**Table 4** ASD clinician review interrater reliability by site, sex and race/ethnicity, cognitive level, and ASD classification, ADDM network surveillance year 2014

	ADDM 2014 (n=924 pairs; 34 reviewers)	
	% Agree	Kappa
ADDM network site*		
Arkansas	.97	.93
Arizona	.89	.78
Colorado	.91	.81
Georgia	.91	.82
Maryland	.89	.77
Minnesota	.87	.75
Missouri	.92	.82
North Carolina	.96	.92
New Jersey	.92	.84
Tennessee	.89	.75
Wisconsin	.93	.86
Sex		
Male (N = 710)	.92	.83
Female (N = 214)	.95	.90
Race/ethnicity*		
White, Non-Hispanic (N = 433)	.93	.85
Black, Non-Hispanic (N = 249)	.92	.85
Asian or Pacific Islander, Non-Hispanic (N = 39)	.90	.80
Other Race or Multiracial, Non-Hispanic (N = 28)	.93	.86
Race Missing/Not Stated, Non-Hispanic (N = 16)	.91	1.0
Hispanic (N = 155)	1.00	.82
Cognitive level from most recent test in records		
Intellectually disabled (IQ < = 70) (N = 211)	.92	.81
Borderline (IQ 71–85) (N = 160)	.92	.83
Average or above average (IQ > = 85) (N = 315)	.91	.83
No cognitive test data (N = 238)	.95	.89
ASD classification from source records		
ASD diagnosis or eligibility on record (N = 430)	.97	.77
ASD suspected, but no documented diagnosis or eligibility on record (N = 199)	.85	.70
No mention of ASD on record (N = 295)	.90	.68

clinical presentations across intellectual ability (Thurm et al., 2019). Given concerns about differential identification of ASD in community practice, it is encouraging that a basic level of consistency in evaluating behavioral descriptions was achieved across the demographic categories, thereby providing evidence in support of a process to more consistently determine if ASD characteristics are present. There is evidence that improved identification across different sociodemographic groups have contributed to the increases in ASD prevalence (Durkin et al., 2017), and that overall autism spectrum diagnoses reflect lower symptom

intensity over time (Arvidsson et al., 2018). For the first time, the most recent ADDM report indicated similar ASD prevalence among children identified as non-Hispanic white, non-Hispanic black, and Asian/Pacific Islander (Maenner et al., 2020). ASD prevalence among Hispanic children has increased from earlier reports but was still below other groups. Continued periodic monitoring of ASD prevalence using replicable and consistent methods for clinician review in a subset of the population could help inform trends in prevalence and to inform targeted community identification efforts (Rice et al., 2012).

While the ADDM method to date is dependent on a child coming to the attention of a health or educational professional and receiving a developmental evaluation, that evaluation does not have to be for an “autism spectrum disorder” per se to be identified for clinician review. The presence of the evaluation is an indication that the child is experiencing some functional impairment; however, the characteristics of children who receive evaluations and the ways behaviors are documented in evaluations may change over time (Arvidsson et al., 2018). Not surprisingly, agreement on case status was highest when a child was already classified as having an ASD by a community clinician or for special education, but there was also high agreement when children were not previously classified with ASD as well, indicating the robustness of the methodology described here.

These ADDM clinician review guidelines were developed for identifying children with ASD for the purpose of monitoring population prevalence; however, the operationalization of the *DSM-5* criteria also has additional utility for research and clinical practice. The standardized methods employed resulted in high inter-rater agreement, and if utilized in clinical settings, could also improve the consistency and the precision of clinical diagnoses. This consistency may be particularly useful for efforts to disentangle racial/ethnic- or sex-associated bias from differential service and diagnostic patterns (Imm et al., 2019). Consistency in applying the diagnostic criteria across these demographic factors was an encouraging finding in the present studies.

While it is assumed that clinicians interpret diagnostic criteria in consistent ways, the multiple rounds of clinician discussions of the concepts behind each criterion indicated that this is not necessarily the case. Most challenging was developing guidance to distinguish (A1) Deficits in social-emotional reciprocity, and (A3) Deficits in developing, maintaining, and understanding relationships. For example, the *DSM-5* text indicates that (A1) includes “little or no initiation of social interaction” and (A3) includes “reduced social interest... preference for solitary activities.” When clinicians consider descriptions of behaviors written in diagnostic evaluations such as “child rarely played with other children” or “child stood quietly in the circle as the other students engaged in a game”, additional guidance was needed

to increase consistency in deciding if these statements better represent a deficit in social emotional reciprocity (A1) or in relationships (A3). As such, the authors established guidelines to specify that (A1) “involves the exchange of social behaviors” such as initiating, responding, conversing, sharing, reciprocating and is represented by observable behaviors, while (A3) represents the “awareness, interest, or insight into other people” and is more about less observable concepts.

In addition, clarifying the concepts represented in the diagnostic criteria (e.g., A3 further detailed into “awareness”, “interest”, and “insight” of other people) provides a framework to further delineation of the phenotypic spectrum represented by these criteria. For example, when engaging in communication with other professionals, parents, or an individual diagnosed, it is useful to have a schema for mapping the person’s profile to specific criteria in a way that specifies current needs on different dimensions relevant to the spectrum (Di Rezze et al., 2016). Additional information on developmental history; age, place, and professional for evaluations; language level, rating of degree of impairment etc. was also classified but not reported in this paper. However, the framework used for characterizing exemplars within each criterion reflects an iterative process taking research from existing diagnostic classifications, tools, and clinical wisdom into account. This framework could be utilized to develop a more structured clinical and research tools and training to quantify and qualify “best estimate expert diagnosis” that considers both categorical and dimensional aspects of ASD. Future research could further evaluate the validity and utility of the behavioral exemplars and concepts representing the expression of the full spectrum within each domain and may serve as a platform for establishing research domain criteria (rdoc) within criteria and across the ASD phenotype.

## Limitations

This paper represents a schema and process for increasing consistency in identification of *DSM-5* ASD criteria and in establishing overall ASD diagnoses. While the very good to excellent agreement was established across multiple samples, the emphasis was on operationalizing the current diagnostic framework of the *DSM-5* and not on the validation of the domains or diagnosis itself. However, consistency across clinicians and sites reflecting the current standard is important. Validation of the ASD definition is a much more complicated issue and some have questioned the validity and utility of behavioral diagnoses (Craddock & Mynors-Wallis, 2014). At the very least, the ADDM clinician review method has shown high agreement between ASD case classification for surveillance and in person assessment based on *DSM-IV-TR* criteria (Avchen et al., 2011). However, in-person assessments

were not conducted in these specific projects and not all existing diagnostic information for each child may have been obtained from source records. Another limitation is that study data reflect information contained in documented evaluations of children up through age 8 years. The ADDM method has been adapted to younger (4 years) and older (15 years) children (Christensen et al., 2019; Nicholas et al., 2012). Further refinements are likely necessary for application across ages, the spectrum of clinical presentations, presence of co-occurring conditions, and across multiple contexts. Expansion of the framework to characterize ASD across ages and levels of support could also be done by further documenting key features of ASD across the entire expression of the autism spectrum. The information evaluated by the clinician reviewers is constrained by the community practices of identifying, evaluating, and documenting the behaviors and histories of children with ASD or other associated developmental challenges. While the content of the information may change over samples, a process that can consistently be replicated and further evaluated is an important strength of this work. In addition to using this framework in ADDM, adaptation and use for improving training and consistency of ASD diagnoses among community clinicians is ongoing. Finally, it is important to note that agreement on definitions and reliable application of diagnostic criteria does not necessarily result in useful application of this approach. Continued evaluation of the impact and utility of classification processes on the people involved is an essential part of our evolving clinical science.

## Conclusion

This paper reports a large-scale effort to detail exemplars and define a process for characterizing for endorsing ASD *DSM-5* criteria. Consistency in the way those criteria are defined and endorsed is an important research and clinical need. The data presented here indicate that a diverse group of clinicians across multiple sites can reliably apply an operationalized guide to endorsing the *DSM-5* ASD criteria regardless of child sex, race/ethnicity, or cognitive level. The described methods and procedures can be used to inform the development and evaluation of tools and training for clinical, research, and surveillance efforts.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10803-021-05377-y>.

**Author Contributions** CER and LAC led the first phase of the *DSM-5* ASD criteria mapping and all co-authors contributed to subsequent refinements. LAC led the SC SUCCESS Project. AB analyzed the SC data and JB the ADDM Network data. The first draft of the manuscript

was written by CER. All authors commented drafts of the manuscript and approved the final manuscript.

**Funding** Funding for the SC SUCCESS project was provided by Autism Speaks (Grant No. 7793 and 8408), Foundation for the National Institutes of Health (Grant No. UL1TR001450), and the Centers for Disease Control and Prevention (Grant No. DD10-1002). Funding for co-author (Thurm), Intramural Research Program of the NIMH (1ZICMH002961). The ADDM Network is supported by the Department of Health and Human Services, Centers for Disease Control and Prevention, National Center on Birth Defects and Developmental Disabilities.

**Conflict of interest** C. Lord acknowledges royalties from diagnostic instruments from Western Psychological Services. *The findings and conclusions in this report are those of the authors and do not necessarily represent the views of CDC, NIH, or Autism Speaks.*

## References

- Aiello, R., Ruble, L., & Esler, A. (2017). National study of school psychologists' use of evidence-based assessment in autism spectrum disorder. *Journal of Applied School Psychology, 33*(1), 67–88. <https://doi.org/10.1080/15377903.2016.1236307>
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). APA.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed, text revision). APA.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.978089425596>
- Arvidsson, O., Gillberg, C., Lichtenstein, P., & Lundström, S. (2018). Secular changes in the symptom level of clinically diagnosed autism. *Journal of Child Psychology and Psychiatry, 59*(7), 744–751. <https://doi.org/10.1111/jcpp.12864>
- Avchen, R. N., Wiggins, L. D., Devine, O., Braun, K. V. N., Rice, C., Hobson, N. C., Schendel, D., & Yeargin-Allsopp, M. (2011). Evaluation of a records-review surveillance system used to determine the prevalence of autism spectrum disorders. *Journal of Autism and Developmental Disorders, 41*(2), 227–236. <https://doi.org/10.1007/s10803-010-1050-7>
- Baio J., Wiggins L., Christensen D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Rosenberg, C. R., White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L. C., Harrington, R., Lopez, M., Fitzgerald, R. T., Hewitt, A., ... Dowling, N. F. (2018). Prevalence of autism spectrum disorder among children aged 8 years – Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2014. *Morbidity and Mortality Weekly Report, Surveillance Summaries, 67*(SS-6), 1–23.
- Bertrand, J., Mars, A., Boyle, C., Bove, F., Yeargin-Allsopp, M., & Decoufle, P. (2001). Prevalence of autism in a United States population: The Brick Township, New Jersey, investigation. *Pediatrics, 108*(5), 1155–1161. <https://doi.org/10.1542/peds.108.5.1155>
- Bishop, S. L., Havdahl, K. A., Huerta, M., & Lord, C. (2016). Subdimensions of social-communication impairment in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 57*(8), 909–916. <https://doi.org/10.1111/jcpp.12510>
- Bishop, S. L., Huerta, M., Gotham, K., Havdahl, K. A., Pickles, A., Duncan, A., Bal, V. H., & Lord, C. (2017). The autism symptom interview, school-age: A brief telephone interview to identify autism spectrum disorders in 5-to-12-year-old children. *Autism Research, 10*(1), 78–88. <https://doi.org/10.1002/aur.1645>
- Bolte, E. E., & Diehl, J. J. (2013). Measurement tools and target symptoms/skills used to assess treatment response for individuals with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(11), 2491–2501. <https://doi.org/10.1007/s10803-013-1798-7>
- Bone, D., Goodwin, M. S., Black, M. P., Lee, C. C., Audhkhasi, K., & Narayanan, S. (2015). Applying machine learning to facilitate autism diagnostics: Pitfalls and promises. *Journal of Autism and Developmental Disorders, 45*(5), 1121–1136. <https://doi.org/10.1007/s10803-014-2268-6>
- Carpenter, L. A., Boan, A. D., Wahlquist, A. E., Cohen, A., Charles, J., Jenner, W., & Bradley, C. C. (2016). Screening and direct assessment methodology to determine the prevalence of autism spectrum disorders. *Annals of Epidemiology, 26*(6), 395–400. <https://doi.org/10.1016/j.annepidem.2016.04.003>
- Centers for Disease Control and Prevention. (2007). Prevalence of autism spectrum disorders—Autism and developmental disabilities monitoring network, six sites, United States, 2000. *Morbidity and Mortality Weekly Report, Surveill Summ, 56*(SS-1), 1–11.
- Christensen, D. L., Maenner, M. J., Bilder, D., Constantino, J. N., Daniels, J., Durkin, M. S., ... & Shenouda, J. (2019). Prevalence and characteristics of autism spectrum disorder among children aged 4 years—Early Autism and Developmental Disabilities Monitoring Network, seven sites, United States, 2010, 2012, and 2014. *Morbidity and Mortality Weekly Report, Surveillance Summaries, 68*(2), 1–19
- Clarke, D. E., Narrow, W. E., Regier, D. A., Kuramoto, S. J., Kupfer, D. J., Kuhl, E. A., Greiner, L., & Kraemer, H. C. (2013). DSM-5 field trials in the United States and Canada. I. Study design, sampling strategy, implementation, and analytic approaches. *American Journal of Psychiatry, 170*(1), 43–58. <https://doi.org/10.1176/appi.ajp.2012.12070998>
- Constantino, J. N., & Charman, T. (2016). Diagnosis of autism spectrum disorder: Reconciling the syndrome, its diverse origins, and variation in expression. *The Lancet Neurology, 15*(3), 279–291. [https://doi.org/10.1016/S1474-4422\(15\)00151-9](https://doi.org/10.1016/S1474-4422(15)00151-9)
- Constantino, J. N., & Todd, R. D. (2005). Intergenerational transmission of subthreshold autistic traits in the general population. *Biological Psychiatry, 57*(6), 655–660. <https://doi.org/10.1016/j.biopsych.2004.12.014>
- Craddock, N., & Mynors-Wallis, L. (2014). Psychiatric diagnosis: Impersonal, imperfect and important. *The British Journal of Psychiatry, 204*(2), 93–95. <https://doi.org/10.1192/bjp.bp.113.133090>
- Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine, 11*(1), 126. <https://doi.org/10.1186/1741-7015-11-126>
- Daniels, A. M., Rosenberg, R. E., Law, J. K., Lord, C., Kaufmann, W. E., & Law, P. A. (2011). Stability of initial autism spectrum disorder diagnoses in community settings. *Journal of Autism and Developmental Disorders, 41*(1), 110–121. <https://doi.org/10.1007/s10803-010-1031-x>
- Di Rezze, B., Rosenbaum, P., Zwaigenbaum, L., Hidecker, M. J., Stratford, P., Cousins, M., Camden, C., & Law, M. (2016). Developing a classification system of social communication functioning of preschool children with autism spectrum disorder. *Developmental Medicine & Child Neurology, 58*(9), 942–948. <https://doi.org/10.1111/dmcn.13152>
- Durkin, M. S., Bilder, D. A., Pettygrove, S., & Zahorodny, W. (2015). The validity and usefulness of public health surveillance of autism spectrum disorder. *Autism, 19*(1), 118–119. <https://doi.org/10.1177/1362361314548732>
- Durkin, M. S., Maenner, M. J., Baio, J., Christensen, D., Daniels, J., Fitzgerald, R., Imm, P., Lee, L. C., Schieve, L. A., Braun, K. V. N., Wingate, M. S., & Yeargin-Allsopp, M. (2017). Autism

- spectrum disorder among US children (2002–2010): Socioeconomic, racial, and ethnic disparities. *American Journal of Public Health*, 107(11), 1818–1826. <https://doi.org/10.2105/AJPH.2017.304032>
- Frazier, T. W., Youngstrom, E. A., Speer, L., Embacher, R., Law, P., Constantino, J., Findling, R. L., Hardan, A. Y., & Eng, C. (2012). Validation of proposed DSM-5 criteria for autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(1), 28–40. <https://doi.org/10.1016/j.jaac.2011.09.021>
- Gardner, L. M., Campbell, J. M., Keisling, B., & Murphy, L. (2018). Correlates of DSM-5 autism spectrum disorder levels of support ratings in a clinical sample. *Journal of Autism and Developmental Disorders*, 29, 1–11. <https://doi.org/10.1007/s10803-018-3620-z>
- Hausman-Kedem, M., Kosofsky, B. E., Ross, G., Yohay, K., Forrest, E., Dennin, M. H., Patel, R., Bennett, K., Holahan, J. P., & Ward, M. J. (2018). Accuracy of reported community diagnosis of autism spectrum disorder. *Journal of Psychopathology and Behavioral Assessment*, 40(3), 367–375. <https://doi.org/10.1007/s10862-018-9642-1>
- Hayes, J., McCabe, R., Ford, T., & Russell, G. (2020). Drawing a line in the sand: Affect and testimony in autism assessment teams in the UK. *Sociology of Health & Illness*, 42(4), 825–843. <https://doi.org/10.1111/1467-9566.13063>
- Huerta, M., Bishop, S. L., Duncan, A., Hus, V., & Lord, C. (2012). Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. *American Journal of Psychiatry*, 169(10), 1056–1064. <https://doi.org/10.1176/appi.ajp.2012.12020276>
- Imm, P., White, T., & Durkin, M. S. (2019). Assessment of racial and ethnic bias in autism spectrum disorder prevalence estimates from a US surveillance system. *Autism*, 23(8), 1927–1935. <https://doi.org/10.1177/1362361319827510>
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2(3), 217–250.
- Klin, A., Lang, J., Cicchetti, D. V., & Volkmar, F. R. (2000). Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV autism field trial. *Journal of Autism and Developmental Disorders*, 30(2), 163–167.
- Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D. M., Ousley, O., Guy, L., Bernier, R., Gerdtts, J., Algermissen, M., Whitaker, A., Sutcliffe, J. S., Warren, Z., Klin, A., Saulnier, C., Hanson, E., Hundley, R., Piggot, P., et al. (2012a). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*, 69(3), 306–313. <https://doi.org/10.1001/archgenpsychiatry.2011.148>
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The autism diagnostic observation schedule – generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule* (2nd Edn.) (ADOS-2) Manual (Part I): Modules 1–4. Western Psychological Services.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659–685.
- Maenner, M. J., Shaw, K. A., Baio, J., Washington, A., Patrick, M., DiRienzo, M., Christensen, D. L., Wiggins, L. D., Pettygrove, S., Andrews, J. G., Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C. R., Lee, L. C., Harrington, R. A., Huston, M., et al. (2020). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *Morbidity and Mortality Weekly Report, Surveillance Summaries*, 69(4), 1–12.
- Maenner, M. J., Rice, C. E., Arneson, C. L., Cunniff, C., Schieve, L. A., Carpenter, L. A., Van Naarden Braun, K., Kirby, R. S., Bakian, A. V., & Durkin, M. S. (2014). Potential impact of DSM-5 criteria on autism spectrum disorder prevalence estimates. *JAMA Psychiatry*, 71(3), 292–300. <https://doi.org/10.1001/jamapsychiatry.2013.3893>
- McDonnell, C. G., Bradley, C. C., Kanne, S. M., Lajonchere, C., Warren, Z., & Carpenter, L. A. (2018). When are we sure? Predictors of clinician certainty in the diagnosis of autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 49(4), 1391–1401. <https://doi.org/10.1007/s10803-018-3831-3>
- McPartland, J. C., Reichow, B., & Volkmar, F. R. (2012). Sensitivity and specificity of proposed DSM-5 diagnostic criteria for autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(4), 368–383.
- Miller, J. S., Bilder, D., Farley, M., Coon, H., Pinborough-Zimmerman, J., Jenson, W., Rice, C. E., Fombonne, E., Pingree, C. B., Ritvo, E., Ritvo, R. A., & McMahon, W. M. (2013). Autism spectrum disorder reclassified: A second look at the 1980s Utah/UCLA autism epidemiologic study. *Journal of Autism and Developmental Disorders*, 43(1), 200–210. <https://doi.org/10.1007/s10803-012-1566-0>
- Nicholas, J. S., Carpenter, L. A., King, L. B., Jenner, W., Wahlquist, A., Logan, S., & Charles, J. M. (2012). Completeness of case ascertainment for surveillance of autism spectrum disorders using the autism developmental disabilities monitoring network methodology. *Disability and Health Journal*, 5(3), 185–189.
- Rice, C. E., Baio, J., Van Naarden Braun, K., Doernberg, N., Meaney, F. J., Kirby, R. S., & Network, A. D. D. M. (2007). A public health collaboration for the surveillance of autism spectrum disorders. *Paediatric and Perinatal Epidemiology*, 21(2), 179–190. <https://doi.org/10.1111/j.1365-3016.2007.00801>
- Rice, C. E., Rosanoff, M., Dawson, G., Durkin, M. S., Croen, L. A., Singer, A., & Yeargin-Allsopp, M. (2012). Evaluating changes in the prevalence of the autism spectrum disorders (ASDs). *Public Health Reviews*, 34(2), 1–22. <https://doi.org/10.1007/BF03391685>
- Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism diagnostic interview-revised*. Los Angeles, CA: Western Psychological Services.
- Rutherford, M., McKenzie, K., McClure, I., Forsyth, K., O'Hare, A., McCartney, D., & Finlayson, I. (2016). A national study to investigate the clinical use of standardized instruments in autism spectrum disorder assessment of children and adults in Scotland. *Research in Autism Spectrum Disorders*, 29, 93–100. <https://doi.org/10.1016/j.rasd.2016.05.003>
- Shulman, C., Esler, A., Morrier, M. J., & Rice, C. E. (2020a). Diagnosis of autism spectrum disorder across the lifespan. *Child and Adolescent Psychiatric Clinics*, 29(2), 253–273. <https://doi.org/10.1016/j.chc.2020.01.001>
- Shulman, C., Rice, C. E., Morrier, M. J., & Esler, A. (2020b). The role of diagnostic instruments in dual and differential diagnosis in autism spectrum disorder across the lifespan. *Child and Adolescent Psychiatric Clinics*, 29(2), 275–299. <https://doi.org/10.1016/j.chc.2020.01.002>
- Thurm, A., Farmer, C., Salzman, E., Lord, C., & Bishop, S. (2019). State of the field: Differentiating intellectual disability from autism spectrum disorder. *Frontiers in Psychiatry*, 10, 526.
- Uljarević, M., Arnott, B., Carrington, S. J., Meins, E., Fernyhough, C., McConachie, H., Le Couteur, A., & Leekam, S. R. (2017). Development of restricted and repetitive behaviors from 15 to 77

- months: Stability of two distinct subtypes?. *Developmental Psychology*, 53(10), 1859–1868. <https://psycnet.apa.org/doi/https://doi.org/10.1037/dev000032>.
- Van Daalen, E., Kemner, C., Dietz, C., Swinkels, S. H., Buitelaar, J. K., & Van Engeland, H. (2009). Inter-rater reliability and stability of diagnoses of autism spectrum disorder in children identified through screening at a very young age. *European Child & Adolescent Psychiatry*, 18(11), 663–674. <https://doi.org/10.1007/s00787-009-0025-8>
- Volkmar, F. R., & McPartland, J. C. (2014). From Kanner to DSM-5: Autism as an evolving diagnostic concept. *Annual Review of Clinical Psychology*, 10, 193–212. <https://doi.org/10.1146/annurev-clinpsy-032813-153710>
- Wiggins, L. D., Durkin, M., Esler, A., Lee, L. C., Zahorodny, W., Rice, C., Yeargin-Allsopp, M., Dowling, N. F., Hall-Lande, J., Morrier, M. J., Christensen, D., Shenouda, J., & Baio, J. (2020). Disparities in documented diagnoses of autism spectrum disorder based on demographic, individual, and service factors. *Autism Research*, 13(3), 464–473. <https://doi.org/10.1002/aur.2255>
- Wiggins, L. D., Reynolds, A., Rice, C. E., Moody, E. J., Bernal, P., Blaskey, L., Rosenberg, S., & A., Lee, L. C., & Levy, S. E. (2015). Using standardized diagnostic instruments to classify children with autism in the study to explore early development. *Journal of Autism and Developmental Disorders*, 45(5), 1271–1280. <https://doi.org/10.1007/s10803-014-2287-3>
- Yeargin-Allsopp, M., Rice, C., Karapurkar, T., Doernberg, N., Boyle, C., & Murphy, C. (2003). Prevalence of autism in a US metropolitan area. *Journal of the American Medical Association*, 289(1), 49–55. <https://doi.org/10.1001/jama.289.1.49>
- Zheng, S., Kaat, A., Farmer, C., Kanne, S., Georgiades, S., Lord, C., Esler, A., & Bishop, S. L. (2020). Extracting latent subdimensions of social communication: A cross-measure factor analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 60, 768–782.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.