

Rapid phenotyping of autism spectrum disorders: inclusion of direct observation in feasible paradigms for clinical assessment

John N Constantino^{*1,2}, Yi Zhang², Anna M Abbacchi², Amanda Calhoun², Faith Scofield² & Sarah J Grafeman²

Practice points

- Key parameters of establishing an autism spectrum disorder (ASD) diagnosis (as currently operationalized) include: positive developmental history; presence of current symptoms consistent with ASDs (most reliable when endorsed by multiple informants); absence of alternate developmental or psychiatric diagnosis that may better explain the symptoms; and observational confirmation by a clinician.
- As a test of principle for including direct observation in a rapid phenotyping paradigm, we examined the ability of clinicians to discriminate level of severity of autistic symptomatology among 65 ASD subjects using an adaptation of the Childhood Autism Rating Scale-2. Clinician's quantitative severity ratings reliably captured levels of gradation in severity ascertained independently using scores for current social impairment derived from the Autism Diagnostic Interview-Revised ($r = 0.60$), and exhibited item-level and scale-level inter-rater reliability on the order of 0.85.
- Standardized clinician ratings based on brief observations – without the need for extensive rater training – show tremendous promise for the assessment of ASD, especially when combined with rapidly obtainable information on developmental history and current symptomatology in daily social contexts.
- Until there are published norms for the relationships between social impairment and gradations in severity of specific intellectual, attentional and anxiety-related impairments, individuals whose suspected autistic syndromes are complicated by 'comorbid' clinical-level deficiencies in one or more of these domains will continue to require more intensive structured interviewing and observation to resolve differential diagnosis.

SUMMARY Traditional methods for the diagnostic assessment of autism spectrum disorders (ASDs), which have been increasingly adopted in the USA as prerequisites for both service eligibility and research participation, are expensive and difficult to acquire consistently in public health settings. Based on immediately available methodologies, we propose a cost-effective strategy for the assessment of children affected by ASDs, which would allow a shift in available resources toward treatment and could facilitate the acquisition of repeated-measures data, which is vital to the evaluation of response-to-intervention. As a test of principle, we examined the ability of clinicians to discriminate level of severity of autistic

¹Departments of Psychiatry and Pediatrics, Washington University 660 S. Euclid, Campus Box 8134, St Louis, MO 63110, USA

²Department of Psychiatry, Washington University, 660 S. Euclid, Campus Box 8134, St Louis, MO 63110, USA

*Author for correspondence: Tel.: +1 314 747 6772; Fax: +1 314 747 6777; constantino@wustl.edu

symptomatology among 65 ASD subjects using an adaptation of the Childhood Autism Rating Scale-2. Clinician's quantitative severity ratings reliably captured levels of gradation in severity ascertained independently using scores for current social impairment derived from the Autism Diagnostic Interview – Revised ($r = 0.60$), and exhibited item-level and scale-level inter-rater reliability on the order of 0.85. We conclude that standardized clinician ratings based on brief observation – without the need for extensive rater training – show tremendous promise for the rapid assessment of ASDs, especially when combined with ascertainment of developmental history and current symptomatology in daily social contexts.

The US CDC have recently estimated that autism spectrum disorders (ASDs) affect approximately one out of every 88 children in the general population [1]. Although the diagnosis of most ASDs relies fundamentally on clinical evaluation (an exception is Rett syndrome), advances are being made monthly in understanding the complex genetic and neural structure of this family of conditions, and the boundaries of what does and does not constitute autistic symptomatology are in a steady and ongoing process of revision. For example, large independent studies, each using independent methods have demonstrated that the phenotypic traits of ASD are continuously (not categorically) distributed in nature [2–4]. Recent studies of motor impairment in ASDs have suggested that motor abnormalities (which are not represented in current diagnostic criteria for autism) occur nearly as commonly as language dysfunction (an established diagnostic criterion domain) in affected children. The largest twin study of developmental disorder symptoms to date ($n = 10,895$ pairs) has suggested substantial overlap in the inherited causes of autistic disorders, ADHD, learning disabilities and coordination disorders [5]. Moreover, a selection of large chromosomal rearrangements associated with autism have also been found to be associated with other neuropsychiatric syndromes, including schizophrenia, intellectual disability, ADHD and epilepsy. This suggests a marked diversity in the possible phenotypic manifestations of specific genetic abnormalities that confer susceptibility to autistic syndromes. It is becoming increasingly clear that categorical genetic variations that give rise to autistic syndromes (e.g., 16p11.2 deletions, see [6]) do not honor the arbitrary boundaries for 'caseness' incorporated into the current diagnostic paradigms for ASDs [7].

Despite these advances in the understanding of the complexity and diversity of ASD-related syndromes in nature, many clinical, educational and research systems have continued to rely on methods that are perceived as 'gold standards' for categorical designation of affectation. These

procedures can incur substantial costs and may not represent the most appropriate resource allocation if there exist more feasible methods for confirming clinical diagnosis and estimating the severity of a majority of autistic syndromes. By way of example, it is common for diagnostic centers and research programs to routinely acquire a set of iconic psychological diagnostic assessments: The Autism Diagnostic Interview – Revised (ADI-R), a 2.5-h DSM-IV-based interview which ascertains developmental history deemed relevant to ASD, and the Autism Diagnostic Observation Schedule (ADOS), a 60–90 min semi-structured clinical observation and coding schema designed to elicit and ascertain the presence or absence of social communicative behaviors and autistic symptoms [8,9]. There is no question that these instruments advanced the field by offering a first uniform method of diagnostic assessment by which to calibrate the investigation of ASD worldwide. The use of these measures in research settings, however, has perpetuated the notion that they are 'necessary' for accurate characterization of what are currently considered autistic syndromes. The identification of rapid phenotyping methods that offer a comparable level of accuracy of characterization carries the prospect of enormous cost savings (potentially on the order of tens of millions of dollars per year) and a number of other potential advantages over traditional diagnostic methods:

- Minimizing the time required to complete an assessment;
- Minimizing the training required for a clinician to complete an assessment or to publish the results of work performed using the assessment (for the ADOS, e.g., the training alone can require 4–6 months of full-time effort by a doctoral-level clinician);
- Feasible quantitative characterization of severity over time and in response to intervention;
- Quantitative characterization by multiple informants (Kim and Lord recently asserted

the importance of combining information from multiple sources for accurate diagnosis of ASDs in young children [10], concluding that parent-reported developmental history and clinician observation make independent, additive contributions to accurate diagnostic decision-making by clinicians, a conclusion that has been reinforced by other studies [11,12]);

- Facilitating standardization of clinical assessment across clinical, educational and research settings;
- Dissemination/implementation of feasible methods in non-English-speaking environments, developing countries and in public health settings.

Numerous recent studies have demonstrated that very brief rating scales are capable of reliably capturing core constructs ascertained by traditional measures with the added advantage of offering feasible methods for tracking variation over time or in response to intervention [11,13–23]. The importance of feasibility in diagnostic assessment and symptom monitoring is particularly relevant for a condition that is now recognized as a common condition of childhood. The prospect for the accrual of valid data based on brief naturalistic observations of parents and classroom teachers additionally makes it possible to acquire repeated measures by multiple informants at relatively low costs, as is routinely and inexpensively implemented for tracking children's responses to treatment for ADHD, a condition for which extensive diagnostic testing has been made almost obsolete in clinical practice settings given the acceptability and precision of brief measurement methods.

The potential value of rapid phenotyping has been recently challenged in a manner that underscores the contrast between appropriate use and misuse of rapid phenotyping methods. Warren and colleagues studied a highly selected sample of 57 children who had community diagnoses of ASD but turned out not to meet criteria for autism on the ADI-R (out of a total sample of 333) [24]. They examined whether single-informant use of a rapid assessment measure could have differentiated the children who met full diagnostic criteria on the ADI-R and ADOS from those who did not and not surprisingly observed that the positive predictive value for identifying this highly selected group was marginal. Although this is neither the purpose nor

the appropriate implementation of a rapid phenotyping paradigm, the authors conclude from this study design that “clinical assessments and clinical assessors with adequate training are irreplaceable in terms of ultimate accurate diagnostic identification of ASD” and questioned what is gained by rapid phenotyping. We would contend that comparing joint positivity on the ADI-R and ADOS to cutoffs from a single-informant screening measurement has no real bearing on the potential utility of rapid phenotyping measures for the vast majority of children affected by ASDs, for whom it is becoming increasingly clear that such methods accurately capture the core constructs that establish both diagnosis and estimation of the severity of the condition [20] at dramatically reduced cost.

An inherent tautology in research anchoring standardized assessments to clinician diagnosis was recently highlighted by Lord and colleagues who demonstrated that the best-estimate clinical diagnosis of highly trained experts in the USA did not reliably predict the conclusions from the ADI-R and ADOS [25], even when using the information from those measures to formulate their diagnostic classifications, and even though the validation of the ADI-R and ADOS has been based on expert clinician diagnosis in the first place. The authors advocate “a move from existing subgroupings of ASDs to dimensional descriptions of core features of social affect and fixated, repetitive behaviors together with characteristics such as language level and cognitive function.” This conclusion is in agreement with a substantial body of recent literature on the quantitative nature of autistic symptomatology [2]. It is now possible to capture quantitative variation in these constructs by obtaining highly cost-effective ratings by parents, teachers, non-specialty-trained clinicians and other caregivers of children that are standardized by age, gender and informant. The time and cost savings also readily allow for the inclusion of brief validated measurements of other domains of behavioral dysfunction that very commonly complicate autistic syndromes (including motor abnormalities, attentional problems and anxiety-related symptoms), thereby enriching a comprehensive assessment of a developing child with information that is directly relevant to intervention and adaptation, as well as diagnosis.

For ASD, it remains clear that direct observation by a reasonably experienced clinician represents a highly valuable component of

multiple-informant assessment, both from the standpoint of diagnostic confirmation and severity estimation. Fundamentally, it is used to verify that social, communicative and stereotypic behaviors characterized by adult informants are more likely attributable to ASD than to other neuropsychiatric syndromes (e.g., general intellectual disability, anxiety disorders and ADHD), and are manifest at a clinical level of severity. Ideally, a brief observation for these purposes would reasonably standardize the context of the observation, and rapidly ascertain severity ratings for core autistic symptomatology in a manner that could additionally ascertain any appreciable contribution of anxiety, inattention/hyperactivity and cognitive impairment to these ratings.

Fortunately, these domains are covered in an existing measure that has a track record and the potential to be adapted to brief clinical observations, the Child Autism Rating Scale-2 (CARS-2), which has relatively modest requirements for time, training and expense. Although the diagnostic validity of the Childhood Autism Rating Scale (CARS) has been established through psychometric analyses involving thousands of children, use of the instrument for rating the severity of symptomatology exclusively ascertained in a time-limited direct observation has not yet been extensively validated. Here we describe an empiric test of principle demonstrating how an instrument such as CARS-2 would function if the ratings were made on the sole basis of a brief clinical observation, standardized in content in a straightforward manner, as would be feasible for implementation in public health, as well as clinical, research and educational settings. Given a diverse sample of clinically affected children with ASD (including a subset who 'outgrew' a full clinical diagnosis) we sought to determine whether brief observational ratings were capable of capturing gradations in severity across this wide clinical range. In addition, we examined the association between these ratings and DSM-IV diagnosis as operationalized on the ADI-R, and tested whether the ratings were reliable between clinicians across disciplines.

Methods

The sample for this test-of-principle consisted of 65 subjects (56 male, nine female) between the ages of 3–16 years (mean age: 7.6 years) from 59 families who gave informed consent to participate in the study, as approved by the Washington

University Human Research Protection Office (MO, USA; HRPO#201105377). All subjects had a history of an ASD based on documented clinical diagnosis in the community and/or a lifetime diagnosis of autistic disorder on the ADI-R, which was performed within weeks of the observation. The subjects were further subcategorized based on current functioning ascertained on the ADI-R. In addition, subjects were categorized as either verbal or nonverbal based on whether or not they were capable of generating meaningful spontaneous or functional phrased speech used on a daily basis.

Subjects were also characterized using standardized quantitative ratings of current autistic symptomatology on the Social Responsiveness Scale (SRS; teacher and/or parent report [n = 65]). For a subset of the patients, data on nonverbal IQ using the Ravens Progressive Matrices (n = 29) and receptive vocabulary, using the Peabody Picture Vocabulary Test (n = 38) were available. A number of the subjects were untestable on these latter two measures, and there were tendencies for these subjects to exhibit higher autism symptom scores on the CARS-2 (p = 0.03). The subjects were consecutive participants in a Washington University research program investigating the genetic origins of autism. There were 47 Caucasian, 15 African-American, and three biracial subjects in the study, generally representing the ethnic composition of the metropolitan St Louis area.

CARS-2 consists of two versions: the Second Edition-Standard Version (CARS-2-ST – equivalent to the original CARS) and a new version for higher-functioning children (CARS-2-HF), defined as verbally fluent with estimated IQ >80 or mental age of 6.0 years or greater [26]. The CARS rating system consists of a total of 15 items, each of which is endorsed on a seven-point response scale that is anchored by detailed descriptions of behavior that characterize gradations in severity for each item. The scale is inexpensive and in widespread use; it capably distinguishes autistic disorders from other developmental and neuropsychiatric conditions [27,28]. The CARS-2-ST and CARS-2-HF are designed to be completed by individuals who have clinical-level experience in the evaluation, education, or intervention of children with ASDs, and requires only minimal additional instrument-specific training (review of a brief manual and practice on a small number of test cases). We implemented a very specific adaptation of the CARS-2 rating system,

which we hereafter refer to as CARS-2^{obs}. The modifications were as follows: first, other than subjects' basic demographic information, the only information made available to clinician raters to score the instrument was a 15-min videotaped observation of the subject being sequentially engaged by a trained examiner in: conversation (verbal subjects) or imitative play (nonverbal subjects – ~5 min); symbolic interactive play using an array of age-appropriate toys (~5 min); and a transition to a sensory activity (e.g., blowing bubbles or allowing the child to interact with highly stimulating sensory toys appropriate for mental age – ~5 min) there was otherwise no consistent script for the three segments of the coded observation, however, individuals performing the examinations that were videotaped for this study generally had experience in the clinical observational assessment of children with developmental disorders. This allowed a straightforward, relative standardization of the context for the observation. The clinicians who coded the videotapes using the CARS-2^{obs} were not certified for research use of the ADOS. Second, autistic severity indices were generated exclusively by the first eight items of the CARS-2 which quantify domains of social impairment, social communication, and stereotypic behavior that define autistic syndromes and that have recently been supported as domains of behavior that differentiate affected from unaffected individuals in taxometric research involving a large national registry sample [18]. The remaining items serve as flags for the presence of symptoms that potentially confound diagnosis (anxiety, inattention/hyperactivity and cognitive impairment) and would trigger more in-depth assessment if a clinician deemed ratings for the severity of autistic symptoms to be more likely attributable to an alternate primary diagnosis than to an ASD.

The 15-min videotaped observations were coded independently by a clinical psychologist (S Grafeman) and a school psychologist (F Scofield) who had completed the standard brief training procedure for the CARS-2 rating system, and who were blind to subjects' clinical status (historic versus current ADI-r diagnosis) when performing CARS-2^{obs} ratings; ratings were obtained from both coders for 58 of the 65 subjects. **Table 1** depicts mean CARS-2^{obs} scores as a function of diagnostic category. CARS-2^{obs} scores were unrelated to age in this sample ($r = 0.08$; $p = 0.50$).

As described above, the ADI-R [9] is a standardized, semi-structured interview in which a parent provides information on key aspects of the lifetime developmental history and current functioning of his/her child relevant to a DSM-IV diagnosis of autistic disorder. In addition to the level of symptomatology present at age 4–5 years or 'ever' (for children who have reached or exceeded that age), which is used in the ADI-R to establish lifetime diagnosis, the scale also generates scores for current symptoms (the 3-month period preceding the interview). Due to the fact that cutoffs have not been established for 'current' (as opposed to lifetime) diagnosis, for the purposes of this study, we adapted cutoff scores for lifetime symptoms using conventions established by the Autism Genetic Resource Exchange (AGRE) and adapted these for current symptoms (see [101] for detailed description of the AGRE categories). Subjects whose scores met AGRE criteria for autistic disorder or 'not quite autism' were deemed as meeting the threshold for a current diagnosis of autism, while subjects who scored in the range for 'broad spectrum' or 'criteria not met' were categorized as not meeting current diagnosis of autism. For all subjects in this study, the ADI-R was performed by raters certified for research use of the scale.

Results

Inter-rater intraclass correlations for the CARS-2^{obs} are presented in **Table 2**. They were consistently in the range of 0.73 to 0.97 considering either the item or total score level, for both the higher function and standard versions of the CARS-2^{obs}. The correlation between

Table 1. Mean Childhood Autism Rating Scale-2 scores as a function of diagnostic categories derived from current symptomatology ascertained by the Autism Diagnostic Interview – Revised.

	HF ASD (CARS-2-HF)		LF ASD (CARS-2-ST)	
	A [†]	B [‡]	C [§]	D [¶]
CARS-2 ^{obs} Total of items 1–8 (mean ± SD)	15.6 ± 3.0	14.6 ± 3.5	20.8 ± 5.2	12.4 ± 2.2

When comparing the two groups within the HF ASD group, $t = 0.796$, $df = 25$, $p < 0.434$; when comparing the two groups within the LF ASD group, $t = 3.642$, $df = 36$, $p < 0.001$. When combining HF and LF subjects, the comparison between those with versus without criteria for a current diagnosis was $t = 4.310$, $df = 56$, $p < 0.0001$.
[†]Meeting ADI-R criteria for current diagnosis of autistic disorder ($n = 14$).
[‡]ADI-R lifetime or historic community diagnosis positive, but ADI-R current score falls below threshold for clinical diagnosis ($n = 13$).
[§]Meeting ADI-R criteria for current diagnosis of autistic disorder ($n = 29$).
[¶]ADI-R lifetime or historic community diagnosis positive, but ADI-R current score falls below threshold for clinical diagnosis ($n = 9$).
 ADI-R: Autism Diagnostic Interview-Revised; ASD: Autism spectrum disorder; CARS-2: Childhood Autism Rating Scale-2; HF: Higher functioning; LF: Lower functioning; SD: Standard deviation; ST: Standard.

Table 2. Inter-rater reliability (expressed as intra-class correlation coefficient with 95% CI) for item scores and summary score on the CARS-2^{obs}, segregated by version of the instrument (higher functioning vs standard); selected characteristics of the children observed in each subgroup are provided to contextualize the results.

Item number	Reliability (95% CI)	
	CARS-2-HF (n = 21)	CARS-2-ST (n = 37)
1. HF: social–emotional understanding; ST: relating to people	0.93 (0.82–0.97)	0.92 (0.85–0.96)
2. HF: emotional expression and regulation of emotions; ST: imitation	0.76 (0.40–0.90)	0.89 (0.79–0.94)
3. HF: relating to people; ST: emotional response	0.77 (0.44–0.91)	0.85 (0.70–0.92)
4. Body use	0.83 (0.58–0.93)	0.87 (0.75–0.94)
5. HF: object use in play; ST: object use	0.76 (0.41–0.90)	0.83 (0.66–0.91)
6. HF: adaptation to change/restricted interests; ST: adaptation to change	0.87 (0.68–0.95)	0.91 (0.82–0.95)
7. Visual response	0.73 (0.33–0.89)	0.90 (0.81–0.95)
8. Listening response	0.79 (0.49–0.92)	0.82 (0.66–0.91)
Sum of items 1–8	0.93 (0.84–0.97)	0.97 (0.94–0.99)
Mean age	9.6 ± 2.8	5.4 ± 3.1
Proportion male	85.20%	86.80%
Proportion caucasian	74.10%	71.10%
Mean SRS parent report raw score	106.8 ± 31.6 (n = 27)	98.3 ± 30.3 (n = 38)
Mean SRS teacher report raw score	85.9 ± 34.0 (n = 22)	105.7 ± 29.7 (n = 34)
Mean PPVT standard score	94.4 ± 25.3 (n = 18)	68.7 ± 22.5 (n = 20)
Mean Raven estimated nonverbal IQ	99.9 ± 17.9 (n = 25)	82.0 ± 12.8 (n = 4)

CARS-2: Childhood Autism Rating Scale-2; HF: Higher functioning; PPVT: Peabody Picture Vocabulary Test; SRS: Social Responsiveness Scale; ST: Standard.

CARS-2^{obs} scores (sum of items 1–8) and ADI-R current score for social impairment was 0.60 ($p < 0.000001$) and the correlation with teacher-report SRS was 0.41 ($p < 0.002$). By contrast, the correlation with parent-report SRS did not reach statistical significance in this clinically ascertained sample. Such contrasts underscore potentially meaningful cross-informant contrasts in quantitative ratings based on accumulated observations by parents and teachers versus DSM-IV-based assessment of autistic impairment from the disparate vantage point of a one-time clinician observation [17]. Here it is critical to note that all subjects had clinical histories of ASD and that the distribution for quantitative scores on the SRS was therefore substantially constrained (to the pathological range), such that the correlation coefficients reported here are highly conservative. Moreover, the relatively small size of this clinical sample limits statistical power to identify all but the most robust associations between variables.

Summary ratings for ASD-specific symptomatology (items 1–8) generated from the brief observations were, on average, highly discrepant between CARS-2-ST-rated subjects with current versus historical (lifetime)-only diagnosis of

autistic disorder, as shown in **Table 1** ($p < 0.001$). For higher functioning subjects (i.e., those for whom the CARS-2-HF was employed), mean differences in these indices were in the same direction but did not reach statistical significance.

Discussion

In this test of principle of the implementation of rapid observational ratings of autistic severity – feasible for use in clinical practice and public health settings – an eight-item index exhibited strong correlations with ADI-R scores for current DSM-IV symptomatology and very high inter-rater reliability among both lower- and higher-functioning subjects. These findings preliminarily, and in principle, support the addition of observational assessment to measurement paradigms that can be rapidly obtained in children with ASD as part of a feasible multi-informant assessment approach [17,20,29].

Limitations of this exploratory case series included the fact that it involved a relatively small number of subjects (with very few females) and that it involved the use of videotaped

recordings of single clinical observations on which ratings were made by independent clinicians reviewing the videotapes. Subsequent research would need to involve larger numbers of subjects, to confirm that the ratings are stable in test–retest analyses and that they would be equally reliable when rendered in real time by a clinician interacting with the patient. Future research should also further explore comparisons of rapid observational ratings with those obtained through lengthier in-depth clinical measurement paradigms such as the ADOS, with the caveat that current understanding of the boundaries of what does and does not constitute autism are rapidly evolving. This is particularly relevant given the prospect for changes in the diagnostic criteria for ASDs proposed for DSM-V, and given that recent research has elucidated the quantitative nature of autistic impairment [3,4,30] as well as causal overlap between the genetic origins of autistic syndromes and those of other neurodevelopmental disorders of childhood [2,5].

Nevertheless, the findings of extremely high inter-rater reliability and differentiation of lower-functioning ASD subjects who had improved versus those who had not improved over the course of development represent an important demonstration-of-principle that brief observational ratings in clinical settings can generate reliable ratings of severity without the requirement for extensive rater training or prolonged observation. A distinct advantage of the rapid assessment strategy proposed here is that the CARS-2 is immediately available to clinicians for use in medical, educational and research settings.

■ Toward a feasible multi-informant assessment strategy

When synthesizing these findings with the results of recent research demonstrating the ability of brief parent- and teacher-report questionnaires to accurately ascertain autistic symptomatology [13–23,29], the possibility of deriving methods for rapid, cost-efficient characterization of most cases of ASD becomes apparent. One such strategy which we present here as a testable model, synthesized from prior research on rapid phenotyping (summarized in this report) and from these new results, is illustrated schematically in **Figure 1**.

It is important to emphasize that key parameters of establishing an ASD diagnosis (as

currently operationalized) include: positive developmental history; presence of current symptoms consistent with ASD (most reliable when endorsed by multiple informants); absence of alternate developmental or psychiatric diagnosis that could better explain the symptoms; and observational confirmation by a clinician. Based on the observations in this report and in previous research by our group and others, we anticipate that up to four out of every five children suspected of autistic syndromes and who screen positive by questionnaire screening for developmental history could be confirmed by clinicians using a rapid phenotyping system such as the one proposed here. In contrast to historic cohorts of ASD-affected children, we note that estimates for the prevalence of co-occurring intellectual disability have fallen into the range of 20–40% [1,29]. For those children whose ratings do not fall within the usual range of severity for autistic impairment (i.e., clinical level impairment in social communication on the CARS-2^{obs}), or whose symptoms are judged better attributed to competing diagnoses, more in-depth testing would be fully warranted. An additional advantage of a protocol comprised of brief quantitative ratings is that it can feasibly be repeated over time to clinically ascertain response to intervention.

This proposed algorithm represents only a testable starting point. The method would require further testing, refinement and receiver-operating curve analysis. We caution researchers, however, to avoid the earlier-described tautology of anchoring novel measurements to arbitrary standards related to what may or may not constitute an autism diagnosis. We and others have shown that autistic impairment manifests itself in a very wide range of severity in nature [3,4,30,31]. Categorical cut-offs for case diagnosis fail to account for the affected status of many children (especially girls) whose social, communicative or cognitive impairments are influenced by the same factors that result in traditionally defined autism, but who nevertheless remain undiagnosed because of traditional criteria for case assignment that invoke arbitrary (nonstandardized) thresholds for case designation [29]. The system that we have proposed in this report differs from an exclusive ADI-R/ADOS phenotyping strategy in the following ways: it requires approximately a fifth of the time and a tenth of the expense; it does not require specially trained

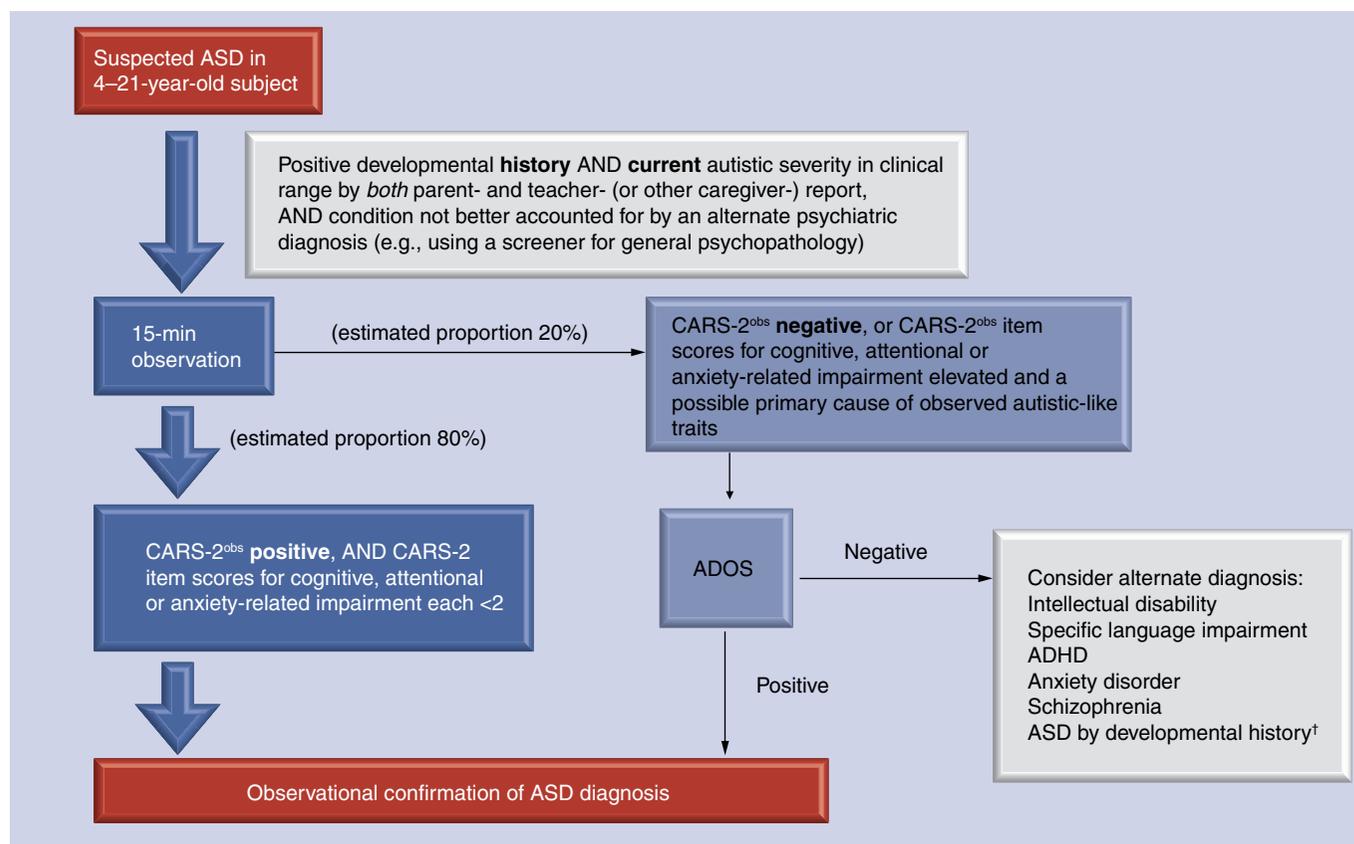


Figure 1. A testable model for feasible diagnostic confirmation of autistic syndrome in public health settings.

[†]Unconfirmed by current clinical observation.

ADOS: Autism Diagnostic Observation Schedule; ASD: Autism spectrum disorder; CARS-2: Childhood Autism Rating Scale-2.

raters (except for those cases in which more in-depth diagnostic testing is required); it allows incorporation of ratings by classroom teachers, who routinely observe children in their natural social contexts; it invokes standardized severity ratings that account for gender differences in the population distribution of autistic traits [2] and in most cases can feasibly be obtained repeatedly over time and in response to intervention. Furthermore, collections of this kind of data allow for the ongoing empiric derivation of profiles of disability that incorporate measured variation in multiple domains of developmental competency rather than doing so on the basis of preconceived notions about the boundaries of specific developmental disabilities.

Until there are published norms for rapid quantitative ratings of social impairment in non-ASD populations reflecting successive gradations in severity of specific intellectual, attentional, and anxiety-related impairments, individuals whose suspected autistic syndromes

are complicated by ‘comorbid’ clinical-level deficiencies in one or more of these domains will continue to require more intensive structured interviewing and observation to resolve differential diagnosis. For example, children with a given level of intellectual disability (e.g., IQ 60) would be expected to have consequent social impairments and possibly some degree of perseveration/repetitive behavior that would register on any autistic symptom scale. In these cases, a diagnosis of ASD may be invoked only in the presence of social impairments or repetitive behaviors whose severity is out of proportion to what would be expected in a child with an IQ of 60. The norms for such individuals (or individuals with an IQ of 70, 50, 40 and so on) have not yet been established but remain an important priority for the next generation of developmental research [32]. Once that is achieved, it will pave the way for a precise quantitative operationalization of autistic impairment akin to what has been used for learning disability (academic achievement 1.5 standard

deviations lower than expected from IQ), and for height versus weight norms as used in standard pediatric practice.

We emphasize that further research is needed to test and refine any phenotyping algorithm in primary care, educational, epidemiologic and tertiary care settings, and to clarify patient profiles which may require more extensive behavioral assessment for diagnostic confirmation. A distinct advantage of the method described in this report is that the observational assessment proposed here is readily available and already in widespread use in clinical and educational settings in the USA (the CARS rating system). What remains is to study and devise optimal methods for harnessing multi-informant rapid phenotyping data in the assessment, education and clinical care of affected children and families. We await the results of ongoing genetic and neuroimaging research that will continue to elucidate the boundaries and diversity of autistic syndromes, and anticipate that quantitative measurement systems that link behavior, genotype and underlying neural mechanisms will continue to facilitate scientific discovery of the causes and potential interventions for this disabling group of conditions of childhood.

Conclusion

Based on immediately available methodologies, we propose a cost-effective strategy for the assessment of children affected by ASD, which would allow a shift in available resources toward treatment and could facilitate the acquisition of repeated measures data, which is vital to the evaluation of response-to-intervention. As a test of principle, we examined the ability of clinicians to discriminate level of severity of autistic symptomatology among 65 ASD subjects using an adaptation of the CARS-2. Clinician's quantitative severity ratings reliably captured levels of gradation in severity ascertained independently using scores for current social impairment derived from the ADI-R ($r = 0.60$), and exhibited item-level and scale-level inter-rater reliability on the order of 0.85. We conclude that standardized clinician ratings based on brief observations – without the need for extensive rater training – show tremendous promise for the diagnostic confirmation of ASD, especially when combined with rapidly obtainable information on developmental history and current symptomatology in daily social contexts.

Future perspective

Traditional methods for the diagnostic assessment of ASD, which have been increasingly adopted in the USA as prerequisites for both service eligibility and research participation, are expensive and difficult to acquire consistently in public health settings. Here, we have proposed a cost-effective strategy for the assessment of children affected by ASD, which would allow a shift in available resources toward treatment and could facilitate the acquisition of repeated measures data, which is vital to the evaluation of response to intervention. This proposed algorithm represents only a testable starting point. The method would require further testing, refinement and receiver-operating curve analysis.

Collections of data of this nature in large populations will allow for the ongoing empiric derivation of profiles of disability that incorporate measured variation in multiple domains of developmental competency, rather than doing so on the basis of preconceived notions about the boundaries of specific developmental disabilities. Standardized clinician ratings based on brief observations – without the need for extensive rater training – show tremendous promise for the assessment of ASD, especially when combined with rapidly obtainable information on developmental history and current symptomatology in daily social contexts.

Financial & competing interests disclosure

JN Constantino receives royalties for the Social Responsiveness Scale, the copyright for which is held by Western Psychological Services, the publisher of the Childhood Autism Rating Scale-2. No royalties to the author were generated by any of the assessments described in this research report. This work was supported by grants from the NIH (HD42541) and from Autism Speaks, Sponsor Award No 5667. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

References

Papers of special note have been highlighted as:

■ of interest

- 1 CDC. Prevalence of autism spectrum disorders – autism and developmental disabilities monitoring network, 14 sites, United States, 2008. *MMWR* 61(3), 1–19 (2012).
- 2 Constantino JN. The quantitative nature of autistic social impairment. *Pediatr. Res.* 69(5 Pt 2), 55R–62R (2011).
- **For further reading on this topic, the above reference provides in-depth coverage of key concepts.**
- 3 Ronald A, Happe F, Price TS, Baron-Cohen S, Plomin R. Phenotypic and genetic overlap between autistic traits at the extremes of the general population. *J. Am. Acad. Child Adolesc. Psychiatr.* 45(10), 1206–1214 (2006).
- 4 Skuse DH, Mandy W, Steer C *et al.* Social communication competence and functional adaptation in a general population of children: preliminary evidence for sex-by-verbal IQ differential risk. *J. Am. Acad. Child Adolesc. Psychiatr.* 48(2), 128–137 (2009).
- 5 Lichtenstein P, Carlstrom E, Rastam M, Gillberg C, Anckarsater H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *Am. J. Psychiatr.* 167(11), 1357–1363 (2010).
- **For further reading on recent scientific advances in understanding the boundaries of autism and its causal overlap with other developmental disorders, the above reference provides in-depth coverage of key concepts.**
- 6 Hanson E, Nasir RH, Fong A *et al.* Cognitive and behavioral characterization of 16p11.2 deletion syndrome. *J. Dev. Behav. Pediatr.* 31(8), 649–657 (2010).
- 7 Kendler KS. Advances in our understanding of genetic risk factors for autism spectrum disorders. *Am. J. Psychiatr.* 167(11), 1291–1293 (2010).
- 8 Lord C, Rutter M, Dilavore PC, Risi S. *Autism Diagnostic Observation Schedule-WPS (WPS Edition)*. Western Psychological Services, Los Angeles, CA, USA (1999).
- 9 Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism. Dev. Disord.* 24(5), 659–685 (1994).
- 10 Kim SH, Lord C. Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age. *J. Child. Psychol. Psychiatr.* 53(2), 143–151 (2012).
- 11 Kanne SM, Abbacchi AM, Constantino JN. Multi-informant ratings of psychiatric symptom severity in children with autism spectrum disorders: the importance of environmental context. *J. Autism. Dev. Disord.* 39(6), 856–864 (2009).
- 12 Risi S, Lord C, Gotham K *et al.* Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J. Am. Acad. Child Adolesc. Psychiatr.* 45(9), 1094–1103 (2006).
- 13 Bolte S, Westerwald E, Holtmann M, Freitag CM, Poustka A. Autistic traits and autism spectrum disorders: the clinical validity of two measures presuming a continuum of social communication skills. *J. Autism. Dev. Disord.* 41(1), 66–72 (2010).
- 14 Charman T, Baird G, Simonoff E *et al.* Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *Br. J. Psychiatr.* 191, 554–559 (2007).
- 15 Cohen IL, Schmidt-Lackner S, Romanczyk R, Sudhalter V. The PDD behavior inventory: a rating scale for assessing response to intervention in children with pervasive developmental disorder. *J. Autism Dev. Disord.* 33(1), 31–45 (2003).
- 16 Constantino JN, Abbacchi AM, Lavesser PD *et al.* Developmental course of autistic social impairment in males. *Dev. Psychopathol.* 21(1), 127–138 (2009).
- 17 Constantino JN, Lavesser PD, Zhang Y *et al.* Rapid quantitative assessment of autistic social impairment by classroom teachers. *J. Am. Acad. Child Adolesc. Psychiatr.* 46(12), 1668–1676 (2007).
- **For further reading on the factor structure of autism, this reference provides in-depth coverage of key concepts.**
- 18 Frazier TW, Youngstrom EA, Speer L *et al.* Validation of proposed DSM-5 criteria for autism spectrum disorder. *J. Am. Acad. Child Adolesc. Psychiatr.* 51(1), 28–40.e3 (2012).
- 19 Kogan MD, Blumberg SJ, Schieve LA *et al.* Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics* 124(5), 1395–1403 (2009).
- 20 Lee H, Marvin AR, Watson T *et al.* Accuracy of phenotyping of autistic children based on Internet implemented parent report. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 153B(6), 1119–1126 (2010).
- **For further reading on rapid phenotyping of autistic syndromes, the above reference provides in-depth coverage of key concepts.**
- 21 Ozonoff S, Goodlin-Jones BL, Solomon M. Evidence-based assessment of autism spectrum disorders in children and adolescents. *J. Clin. Child. Adolesc. Psychol.* 34(3), 523–540 (2005).
- 22 Robins DL, Fein D, Barton ML, Green JA. The modified checklist for autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J. Autism Dev. Disord.* 31(2), 131–144 (2001).
- 23 Wetherby AM, Allen L, Cleary J, Kublin K, Goldstein H. Validity and reliability of the communication and symbolic behavior scales developmental profile with very young children. *J. Speech Lang. Hear. Res.* 45(6), 1202–1218 (2002).
- 24 Warren Z, Vehorn A, Dohrmann E *et al.* Accuracy of phenotyping children with autism based on parent report: what specifically do we gain phenotyping ‘rapidly’? *Autism Res.* 5(1), 31–38 (2012).
- 25 Lord C, Petkova E, Hus V *et al.* A multisite study of the clinical diagnosis of different autism spectrum disorders. *Arch. Gen. Psychiatr.* 69(3), 306–313 (2012).
- 26 Schopler E, Van Bourgondien ME, Wellman GJ, Love SR. *Childhood Autism Rating Scale. (2nd Edition)*. Western Psychological Services, Los Angeles, CA, USA (2010).
- 27 Chlebowski C, Green JA, Barton ML, Fein D. Using the childhood autism rating scale to diagnose autism spectrum disorders. *J. Autism Dev. Disord.* 40(7), 787–799 (2010).
- 28 Perry A, Condillac RA, Freeman NL, Dunn-Geier J, Belair J. Multi-site study of the Childhood autism rating scale (CARS) in five clinical groups of young children. *J. Autism Dev. Disord.* 35(5), 625–634 (2005).
- 29 Constantino JN, Zhang Y, Frazier T, Abbacchi AM, Law P. Sibling recurrence and the genetic epidemiology of autism. *Am. J. Psychiatr.* 167(11), 1349–1356 (2010).
- 30 Constantino JN, Todd RD. Autistic traits in the general population: a twin study. *Arch. Gen. Psychiatr.* 60(5), 524–530 (2003).
- 31 Frazier TW, Youngstrom EA, Sinclair L *et al.* Autism spectrum disorders as a qualitatively distinct category from typical behavior in a large, clinically ascertained sample. *Assessment* 17(3), 308–320 (2010).
- 32 Constantino JN. How continua converge in nature: cognition, social competence, and autistic syndromes. *J. Am. Acad. Child Adolesc. Psychiatr.* 48(2), 97–98 (2009).
- **Website**
- 101 Autism Genetic Resource Exchange. AGRE affected status categories. <http://research.agre.org/agrecatalog/algorithm.cfm>