Outcome Measures for Clinical Drug Trials in Autism

By Michael G. Aman, PhD, Sherie Novotny, MD, Carole Samango-Sprouse, EdD, Luc Lecavalier, PhD, Elizabeth Leonard, PhD, Kenneth D. Gadow, PhD, Bryan H. King, MD, Deborah A. Pearson, PhD, Morton Ann Gernsbacher, PhD, and Michael Chez, MD

FOCUS POINTS

- Marked variability in age and levels of adaptive functioning often make selection of assessment tools for randomized clinical trials a challenging task.
- At our current level of refinement, no definitive instrument has been identified for assessing core features of autism, but five tools worthy of consideration were identified for future work.
- Choice of assessment instruments for language and communication should be largely dependent on the participant's developmental level.
- Common types of comorbid challenging behaviors include irritability; hyperactivity; compulsive, ritualistic, and perseverative behavior; excessive anxiety; and self-injury. Suitable instruments for assessing these are identified.
- For assessing cognitive function, the major challenge usually is finding tasks that are attractive to participants and which participants can successfully perform.
- Given that psychotropic agents are often prescribed for extended periods, frequently in children and adolescents, it is best to err on the side of safety and to probe specifically for side effects in randomized clinical trials.

ABSTRACT

This paper identifies instruments and measures that may be appropriate for randomized clinical trials in participants with autism spectrum disorders (ASDs). The Clinical Global Impressions scale was recommended for all randomized clinical trials. At this point, however, there is no "perfect" choice of outcome measure for core features of autism, although we will discuss five measures of potential utility. Several communication instruments are recommended, based in part on suitability across the age range. In trials where the intention is to alter core features of ASDs, adaptive behavior scales are also worthy of consideration. Several "behavior complexes" common to ASDs are identified, and instruments are recommended for assessment of these. Given the prevalence of cognitive impairment in ASDs, it is important to assess any cognitive effects, although cognitive data from ASD randomized clinical trials, thus far, are minimal. Guidance from trials in related pharmacologic areas and behavioral pharmacology may be helpful. We recommend routine elicitation of side effects, height and weight, vital signs, and (in the case of antipsychotics) extrapyramidal side-effects assessment. It is often appropriate to include laboratory tests and assessments for continence and sleep pattern.

CNS Spectr *2004;9(1):36-47*

Dr. Aman is professor of psychology and psychiatry in the Departments of Psychology and Psychiatry at the Nisonger Center at Ohio University in Columbus. Dr. Novotny is assistant professor of psychiatry in the Department of Psychiatry at the University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Campus. Dr. Samango-Sprouse is associate clinical professor in the Department of Pediatrics at the George Washington University in Washington, D.C. Dr. Lecavalier is assistant professor of psychology in the Department of Psychology at Ohio State University. Dr. Leonard is lecturer on psychology in the Department of Psychiatry at Harvard Medical School in Boston, Massachusetts. Dr. Gadow is professor of psychiatry in the Department of Psychiatry at the State University of New York at Stony Brook. Dr. King is the director of Child and Adolescent Psychiatry in the Department of Psychiatry at Dartmouth Medical School in Hanover, New Hampshire. Dr. Pearson is associate professor of psychiatry and behavioral sciences in the Department of Psychiatry and Behavioral Sciences at the University of Texas Medical School in Houston. Dr. Gernsbacher is Sir Frederic C. Bartlett Professor of Psychology in the Department of Psychology at the University of Wisconsin, Madison. Dr. Chez is assistant professor of Neurology and Pediatrics at the Rush Medical School and Lecturer at the Chicago Medical School, both in Chicago, IL.

Disclosure: Dr. King has received honoraria from Janssen and grant support from CAN and the National Institutes of Health. This paper was submitted on June 20, 2003, and accepted on September 14, 2003.

Please direct all correspondence to: Michael G. Aman, PhD, The Nisonger Center, Ohio State University, 175 McCampbell Hall, 1581 Dodd Dr. Columbus, Ohio 43210-1257; E-mail: aman.1@osu.edu.

INTRODUCTION

In this paper, we briefly review representative measures of treatment response for possible use in randomized clinical trials. Our goal is not to conduct an exhaustive review of the existing literature, but rather to provide a starting point for continuing discussion and development. Although we have made a sincere effort to identify measures that may be sensitive indicators of treatment effects (and to be consistent with other reviews where present), our efforts inevitably also reflect our personal experiences and biases. The authors of this article were recruited in two groups. The first group comprised a number of researchers and clinicians approached by Cure Autism Now (CAN) based on recommendations from clinical researchers, their known clinical experience, and/or publication track record. This "Outcomes Committee" met with several other committees, who collectively comprised CAN's Autism Clinical Trials Task Force, in Santa Monica, California, in March, 2002. The second group was co-opted by the Outcomes Committee to expand the committee and reduce possible areas of deficiency within the committee. This resulted in the 10 individuals who are the authors of this article. Besides the face-to-face meeting in Santa Monica, the committee met via conference calls and through a series of E-mail communications.

OBJECTIVES OF THE TRIAL

A primary consideration in designing randomized clinical trials in autism spectrum disorders (ASDs) is to determine the objective of the trial. One aim could be to alter the very course of the disorder. The other, far more common aim is to modify impairing behaviors associated with ASDs. Although psychopharmacologists have been studying therapies for autism for over 50 years, it is important to note that there are currently no Food and Drug Administration-approved indications for the treatment of autism or associated behavior problems for any agent. In this paper, we discuss both approaches in the sections that follow. Regardless of the objective, one measure that should be universal in all ASD clinical trials is the Clinical Global Impressions scale (CGI), which has two key domains: the Severity and Improvement subscales.[1] It is common to obtain CGI-Severity scores at the beginning and end-point of a trial, whereas the CGI-Improvement scale should be used to measure change during the trial and at the endpoint. Raters should use the CGI to assess all behavior of the participants (in as many contexts as possible) so that the score is truly a reflection of the participant's global functioning. The CGI can be used to reflect both changes in core autism symptoms and in comorbid behaviors or specific symptom clusters (eg, aggression) as well.

OUTCOME MEASURES FOR CORE AUTISTIC SYMPTOMS

Core autistic symptoms include qualitative deficits in social interaction; restricted, repetitive, and stereotyped patterns of behavior, interests, or activities; and deficits in communication and language. There are no universally accepted outcome measures developed for measuring changes in core symptoms from treatment, and there are no drug products currently approved for the treatment of this disorder. Nevertheless, investigators have made attempts at quantifying such changes by using some of the following measures.

Global Symptoms of Autism

These measures include the Autism Diagnostic Observation Scale-G (ADOS-G),[2] the Childhood Autism Rating Scale (CARS),[3,4] the Social Responsiveness Scale,[5] the Matson Evaluation of Social Skills in Youngsters (MESSY),[6] the Gilliam Autism Rating Scale (GARS),[7] the Ritvo-Freeman Real Life Rating Scale for Autism (RLRS),[8] and the Autism Behavior Rating Scale.[9] Besides the MESSY and the Ritvo-Freeman Scale, all were primarily designed for diagnostic purposes rather than for assessing effects of therapeutic agents. We employed an A to C ranking procedure to score the available instruments for assessing core features of autism in randomized clinical trials. We only assigned a grade of B or lower for all of these instruments because all appear to have at least some deficiencies. This decision is not intended to denigrate any of the instruments available; it merely reflects the fact that tools developed for other reasons are not ideally suited for pharmacologic purposes. Although this article addresses measures for the range of ASDs, in the interest of brevity, we only discuss autism scales. These instruments are discussed in the next section.

Deficits of Social Interaction

Recently, there has been some attempts at developing or adapting scales to rate change in social behavior over the course of a randomized clinical trial. ADOS-G,[2] is a standardized protocol for observing social and communicative behavior in children, adolescents, and adults suspected of having an ASD. The ADOS-G consists of standard activities that allow the examiner to observe the occurrence or absence of

behavior relevant to the diagnosis of ASDs across developmental levels. Ratings are based on a series of structured and semi-structured "presses" (or prompts) for social interaction and communication, and cut-off criteria are ascertained using a diagnostic algorithm. Administration time is ~45 minutes.

There are four modules to the ADOS-G: Module 1 is for children who are nonverbal or do not consistently use three-word phrases; Module 2 is for children with expressive language skills between 30- and 47-months of age; Modules 3 and 4 is for individuals with expressive language skills at 48 months of age. Module 3 has a greater emphasis on the use of toys, whereas Module 4 focuses more on interview questions. Although primarily a diagnostic scale, the ADOS-G has been used in a few randomized clinical trials to assess social and communication behavior over a relatively short interval.[10-12] At this point it is not clear if the ADOS-G would provide a sensitive assessment of change if a truly therapeutic agent were being tested because these trials had negative outcomes. As the ADOS-G was developed primarily for diagnosis, it is probable that scores will tend to be stable over time. It is difficult to integrate the four modules when using the scale to determine outcome, as the modules are not completely compatible with one another when determining the extent of a child's impairment. Reliability on this measure is difficult and time consuming to establish, but the strict reliability standards required by the instrument's developers is also a strength.

The CARS[3,4] has been translated and validated in several languages and used in numerous published studies. It was designed as a diagnostic instrument for young children and was intended to be completed by clinicians following behavioral observations. It has also been used with adolescents and adults[13] and as an informant-based rating scale.[14] The CARS contains 15 areas rated on a seven-point Likert scale. The first 14 areas represent different domains of child functioning, while the last item is a global rating of autism. Ratings are done on a four-point scale (normal to severely abnormal). Midpoints are used when the child's behavior falls between the descriptors used as anchor points. Individuals are designated as Not Autistic to Severely Autistic, depending on the total score and number of items scored as severely abnormal. The CARS can be a difficult scale on which to achieve interrater reliability, it does not have a standardized series of prompts, and each rater is on his/her own to create a mental picture of the "normal child of equivalent age," which is the basis of the rating. Reliability standards and coding criteria are less well developed than those of the ADOS-G. Also, some items (eg, Taste) are hard to ascertain in the short assessment period and carry equal weight to core symptoms in the scoring. The CARS has no subscales, so it is not optimal for an randomized clinical trial in which the drug is expected to target a single core area or symptom.

In general, reliability, criterion validity, and construct validity appear to be good.[3,4,15-18] The CARS is widely used for screening and diagnostic purposes, but it was not designed to measure behavior change. Although some studies have reported that certain areas may be sensitive to change,[13,15] the subjective nature of the ratings, broadly defined categories, and (perhaps) lack of normative data may reduce the scale's appeal.

The Social Responsiveness Scale[5] was introduced to assess social deficits in ASDs.[5,19,20] This scale is a 65-item informant-rated assessment scale that requires ~15–20 minutes to complete. The instrument measures both specific and observable items for social behavior and social language use, as well as characteristics of ASDs. Some psychometric properties have been reported; intraclass correlation coefficients for reliability were about 0.80, and interrater reliability was ~0.75 for various informants.[5] This scale is relatively new and time will tell whether it is useful for assessing changes in social interaction in randomized clinical trials.

The Gilliam Autism Rating Scale (GARS)[7,21] consists of 56 items divided into four subscales (Social Interaction, Communication, Stereotyped Behaviors, and Developmental Disturbances). The items are rated on a four-point scale (0=never to 3=frequently observed) based on a 6-hour observation period. The scores are then added together for each subscale, and across all subscales and rated as to probability of having an ASD. Some workers have challenged the accuracy of GARS for diagnosis of children with milder presentation of autism,[22] but the GARS does have a method for grading intermediate changes. The GARS could be sensitive to subtle changes, but given its uncertain sensitivity for diagnosis, it is unclear whether it would miss important autistic symptoms. The design of the scale is appropriate for repeated use, but some items do not appear to be appropriately subgrouped (eg, play and repetitive behavior appear in Social Interactions subscale). Overall, this scale has some potential, but additional sensitivity and further psychometric data are needed.

Another of the informant-rated scales for the social dimension is the MESSY.[6] The MESSY was designed to obtain information regarding an array of

social skills. It has been used to assess acquisition of social skills for deaf children. The scale was evaluated in one comparison study of typical and autistic children and found to distinguish between these two groups.[21] Another version (Matson Evaluation of Social Skills in the Severely Retarded [MESSIER]) was used to evaluate adults with profound mental retardation with and without ASDs; the MESSIER distinguished between the groups with and without ASDs.[23] It has not been used in any published autism randomized clinical trials to date, and its drug sensitivity is unknown. Also, we are not aware of reliability and validity data. Its administration time is ~10 minutes. As the MESSIER focuses on social skills, it warrants further attention in ASD trials.

We did not feel that the Autism Behavior Checklist[9,24] should be recommended as an outcome measure for randomized clinical trials. Each of its items is simply endorsed or not (0 or 1), leaving little room for intermediate change over time and for detecting subtle changes. We also declined to recommend the RLRS.[8] Difficulties that have emerged include problems in obtaining interrater reliability, administering the scale consistently over time, and structuring the observation period to probe for all included behaviors. The RLRS does seem to assess a good range of social skills.

Restricted Interests and Repetitive Behavior

Although this is one of the core domains, it is also considered a comorbid target behavior in many treatment trials. Restricted interests are covered to some extent by the ADOS-G. Scales for assessing rituals, compulsive behavior, and stereotypies are addressed in the section on comorbid target behaviors in the following sections.

Communication Impairment

Because communication deficits comprise a core aspect of ASDs, their assessment warrants consideration for most clinical trials with ASDs. Communication is not addressed again in the section (below) on cognitive assessment, but it should be noted that some measures of communication may be good ways to assess cognitive outcomes as well. Care should be taken to select an instrument that is sensitive to the small changes that typically occur over the short time-frame of most clinical trials (usually weeks or a few months), so that if communication gains are noted during the trial, the instrument will be able to detect them. As with IQ tests (discussed below), if a particular communication measure can only assess developmental changes greater than a few months, it probably will not be treatment sensitive. Because of the interface between cognition and communication, it is important to separate the influence of communication deficits on cognitive abilities among children with ASD.

Assessing Communication

An excellent review of the areas of communication that can be reliably assessed and the common measures for such assessment is provided by an authoritative chapter that we will draw heavily upon.[25] We shall discuss the structure of communication and provide cautions relevant to assessing communication in ASDs.

Prelinguistic Communication

Forms of pre-linguistic communication include babbling during the first year of life,[25,26] joint attention,[27] and pointing. Protoimperative pointing is the action typically used to request objects, whereas protodeclarative pointing is an action used to draw attention to an object or comment on the object. The motor planning or motor development for pointing might be impaired in children with autism, a finding relevant to assessment techniques requiring children to point in order to respond.

Linguistic Communication

Linguistic communication traditionally comprises phonology (the sounds of a language), prosody (rhythm and intonation), morphology (the combination of morphemes, which are the smallest units of meaning in a language), syntax (rules for combining words into phrases and sentence), semantics (meanings associated with words), pragmatics (the situational contexts within which utterances are made, including the knowledge and beliefs of the speaker and the relation between speaker and listener), and discourse (the combination of words into sentences, sentences into paragraphs, and paragraphs into narratives). The vast majority of standardized assessments of linguistic communication assess morphology and syntax (eg, word inflection, sentence and phrase construction) and semantics (eg, vocabulary). Fewer assess phonology, prosody, pragmatics, or discourse.

When selecting communication assessments for use in randomized clinical trials we provide two general recommendations. First, assessment instruments that are comprehensive across ages (eg, scales that range from early childhood across adolescence or beyond) might be preferable to those that are limited to a smaller age range. Secondly, a careful assessment of the presence and the extent of any motor planning dysfunction should be made prior to selection of any communication assessment, as motor planning

dysfunction can complicate assessment of many behaviors, communication, and cognition.

Recommended Tests: Prelinguistic Communication

There are two scales that may be useful in assessing prelinguistic communications skills. These tools are the Early Social Communication Scales,[27] observation schedule, which might be inappropriate for older children, and Rosetti Infant Toddler Language Scale, a care-provider interview.

Recommended Tests: Linguistic Communication

For linguistic communication assessment, there are several options. The Peabody Picture Vocabulary Test, Version III is for 2–90+ years of age with an average 15-minutes administration time.[28] The Expressive Vocabulary Test is for patients 2–90+ years of age and administration averages 15 minutes.[29] For patients 3–21 years of age, there is the Comprehensive Assessment of Spoken Language, which has an average administration time of 30–45 minutes.[30] The Clinical Evaluation of Language Fundamentals—Revised and Clinical Evaluation of Language Fundamentals Preschool is useful for assessing individuals 5–21 years of age; it has an average 30–45 minutes administration time.[31] Finally, there is the Clinical Evaluation of Language Fundamentals-Preschool for 3–6 years of age.[32]

Whereas these are good clinical assessment tools, it is not yet known whether these are sensitive to treatment effects. On some of these instruments, changes on a few items corresponds to months of developmental change, which may compromise their sensitivity in brief clinical trials.

Adaptive Behavior

Subaverage adaptive behavior is not listed as a core symptom of autism, but adaptive behavior deficits are seen in the vast majority of individuals with ASDs. There are three well-established adaptive behavior scales commonly used in this population. One is the Vineland Adaptive Behavior Scales,[33] another is the Scales of Independent Behavior–Revised (SIB–R),[34] and the third is the American Association on Mental Retardation Adaptive Behavior Scale.[35] These are semi-structured informant interviews that assess an individual's daily functioning. They can be administered to a caretaker/family member or teachers. The Vineland has been recently normed on individuals with autism.[36] These scales are primarily designed for diagnostic and prognostic purposes, and they are unlikely to show change in short-term treatment studies. The Vineland was used in two autism pharmacological studies, but the results did not show changes with the agents assessed.[10,37] The Vineland is under revision and, according to the publisher's website, the Vineland-II is expected to be available in 2005 (http://www.agsnet.com/assessments/vineland2.asp).

The Assessment of Basic Language and Learning Skills (ABLLS)[38] is a criterion referenced skill-tracking system designed to assess a variety of language and daily living skills. It was also designed to account for a child's motivation to respond, ability to attend to a variety of environmental stimuli and generalize skills, and tendency to use those skills spontaneously. Most of the ABLLS items were designed for children functioning at or below that of a typical child 5 years of age. For this reason, the ABLLS may be a reasonably good assessment for children with moderate to severe symptoms of autism, as its items appear to assess relatively fine steps in development. The new National Institute of Mental Health (NIMH) Research Units in Pediatric Psychopharmacology and Psychosocial Intervention (RUPP-PI) Autism Network is attempting to assess adaptive behavior as one outcome measure in a study of atypical antipsychotic medicine and parent management training. The RUPP-PI has adopted the ABLLS as one outcome measure. Because the ABLLS addresses behavior usually seen in quite young children, the RUPP-PI decided to create an upward extension so that it will be relevant to older participants as well.

Standardized Rating Instruments for Comorbid Maladaptive Behavior

Besides the use of therapies for altering the course of core features of ASDs, the more common type of pharmacological trial is to manage disruptive and/or emotional behaviors. Extreme irritability, hyperactivity, perseverative behaviors, and anxiety are some key areas. Although there is a growing literature on conventional psychiatric syndromes in individuals with ASDs, in the authors' view, it is seldom possible to make simple comorbid diagnoses for most individuals with ASDs (even when there are significant comorbid emotional or problem behavior). For this reason, we focused on "behavior complexes" and tried to find the best assessments for each. The complexes chosen include the following: irritability; hyperactivity; compulsive, ritualistic, and perseverative behavior; anxiety; and self-injury (a subset of perseverative behavior). Each of these terms is defined in Table 1, where various instruments are summarized.

We encountered at least three challenges in trying to identify suitable instruments. First, there is a small data base in autism from which to make recommendations. Second, children with ASDs are being much more commonly diagnosed today, with greater variability in their intellectual abilities. Third, ASDs occur over the life span, so that covering all possible combinations of behavior complex, functional level, and age is a tall order. For children and adolescents with normal/near-normal IQ, it may be sensible to employ relevant portions of the Early Childhood Inventory (ECI; preschool ages),[39] Child Symptom Inventory (CSI; 5–12 years of age),[40] or Adolescent Symptom Inventory (ASI),[41] all of which include all sections of the *Diagnostic and Statistical Manual of Mental Disorder, Fourth Edition* (*DSM-IV*)[42] relevant to children. For participants with mental retardation, it makes more sense to adopt instruments created for people with developmental disabilities (particularly as the severity of intellectual handicap increases).

Most of the instruments discussed here have been reviewed in greater detail elsewhere.[43-45] Table 1 contains our best attempt to identify suitable instruments for autism clinical trials, where we have made some attempt to determine whether the instrument has been assessed psychometrically. In the "Reliability" columns, I-R refers to the existence of interrater reliability data, IC refers to internal consistency, and T-T refers to test-retest data. In the "Validity" columns, Cn refers to presence of construct validity, and Cr means that criterion group validity data exist. Construct validity was used to refer to the presence of factor analytic derivation or a compelling link to an accepted nosological system, such as the *DSM-IV*. Criterion group validity refers to whether the instrument has been found successfully to discriminate between clinical groups. These ratings were based on several authors' (M.G.A., L.L., K.G.) familiarity with the field and reflect our best scientific judgment, based on our familiarity with psychometric work and on previous published reviews.[43-45] Such a process is nevertheless somewhat subjective. All psychometric data pertain to individuals with ASD or mental retardation. A notation that an instrument was assessed for reliability or validity does not indicate that the outcome was necessarily good; it merely means that the psychometric work was done. The "Limitations" column was used to signify concerns about a given scale. The "Outcome Studies" column is presented to indicate whether any outcome work has been done with ASD participants (usually in clinical trials). This is obviously a minimal standard. The only instruments known to us to have been used repeatedly in the ASD field are the Aberrant Behavior Checklist (ABC)[46] and (less so) the Yale-Brown Obsessive-Compulsive Scale and its close relative, the Children's Yale-Brown Obsessive-Compulsive Scale.[47]

In this section, we suggest several instruments for consideration in autism randomized clinical trials. In the interest of brevity, they are listed here. When convenient, we refer to them by abbreviations in Table 1. The instruments are: ABC;[46] Anxiety, Depression, and Mood Scale;[48] Behavior Problems Inventory;[49] the Children's Yale-Brown Obsessive-Compulsive Scale and the Yale-Brown Obsessive-Compulsive Scale;[47,50] the adapted Children's Psychiatric Rating Scale;[51] Developmental Behaviour Checklist;[52] Diagnostic Assessment for the Severely Handicapped—Version II;[53] ECI/CSI/ASI;[39-41] Emotional Disorders Rating Scale;[54] Fear Survey for Children—Revised[55,56] [and Fear Survey for Children with and without Mental Retardation];[57] Nisonger Child Behavior Rating Form;[58,59] Preschool Behavior Questionnaire;[60,61] Repetitive Behavior Scale—Revised;[62] Self Injurious Behavior Questionnaire;[63] and the Stereotypic Behavior Scale.[64]

In general, an "A" indicates that the measure is highly recommended for the target behavior, whereas a "B" signifies less enthusiasm. Hence, the ABC and Developmental Behaviour Checklist were most strongly recommended for assessing irritability. The ABC, the Nisonger Child Behavior Rating Form, and (contingent on IQ) ECI/CSI-4 were recommended for assessing hyperactivity. We did not have a top-line recommendation for compulsive/perseverative behavior or for assessing anxiety. The Behavior Problems Inventory was recommended for assessing self-injury. These ratings have to be viewed as somewhat tentative, as the objectives of the study may alter the investigator's choice.

Assessing Cognition in Autism

Evaluating cognition in children with ASDs is challenging for several reasons. First, these children may vary markedly in IQ, with previous reports (which have been disputed in some sectors) of up to 75% of children with autism having mental retardation.[42] Second, the core deficits in ASDs often prevent adherence to standard test protocols. Lack of language, noncompliant behavior, and impaired object use may depress performance.[65]

IQ tests can be a foundation for identifying a population for study, but it would not be expected that IQ tests would be effective outcome measures because developmental processes underlying the growth of cognition evolve slowly over time and are resistant to rapid change. One criticism of many popular IQ tests is that

their basal scores is too high to establish reliable measurements in children with IQs <50.[66] A test, such as the Leiter International Performance Test-Revised,[67] can be useful for identifying children who can be stratified into subgroups for study rather than aggregating children (ie, assuming equal ability across the sample). There are really two traditions for testing in children with developmental disabilities. One might be described as a "neuropsychological" approach, whereas the other might be called the "pharmacological tradition." The former has its roots in assessing for strengths and weaknesses and usually employs tests having norms, whereas the latter often uses automated equipment and often does not have norms.

TABLE 1. INSTRUMENTS FOR ASSESSING TARGET SYMPTOMS OFTEN ASSOCIATED WITH ASDS

Constellation/Scale Name (Subscale) [Age Groups]	Recommendation	Reliability
Irritability*		
Aberrant Behavior Checklist (irritability) [C, TA, Ad]	A	I-R, IC, T-T
Developmental Behaviour Checklist (disruptive/antisocial) [C, TA]	A	I-R, IC, T-T
ECI/CSI/ASI (oppositional defiant disorder; conduct disorder) [P, C, TA]	B	IC
Children's Psychiatric Rating Scale (anger/uncooperativeness) [P, C, TA]	B	I-R
NCBRF (conduct problem) [C, TA]	B	I-R, IC, T-T
DASH-II (impulse control) [TA, Ad]	B	I-R, IC, T-T
Preschool Behavior Questionnaire (hostile-aggressive) [P]	B	I-R, T-T
***Hyperactivity/Inattention/Impulsiveness**†*		
Aberrant Behavior Checklist (hyperactivity) [C, TA, Ad]	A	I-R, IC, T-T
ECI/CSI/ASI (ADHD) [P, C, TA]	B	IC
Children's Psychiatric Rating Scale (hyperactivity factor) [P, C, TA]	B	I-R
NCBRF (hyperactive) [C, TA]	A	I-R, IC, T-T
Preschool Behavior Questionnaire (hyperactive-distractible) [P]	B	I-R, T-T
***Compulsive, Ritualistic, Perseverative**‡*		
Repetitive Behavior Scale-Revised (stereotyped; self-injurious; compulsive; ritualistic; sameness; restricted behavior) [C, TA, Ad]	B	I-R, IC, T-T
C-YBOCS/YBOCS (obsessions, compulsions) [C, TA, Ad]	B	I-R, T-T
Stereotypic Behavior Scale [C, TA, Ad]	B	I-R, IC, T-T
Aberrant Behavior Checklist (stereotypic behavior) [C, TA, A]	B	IR, IC, T-T
***Anxiety and Fears**§*		
ADAMS (general anxiety) [TA, Ad]	B	I-R, IC, T-T
Emotional Disorders Rating Scale (anxiety) [C, TA]	B	I-R, T-T
Fear Survey for Children-Revised [C, TA]	B	I-R, T-T
Preschool Behavior Questionnaire (anxious/fearful) [P]	B	I-R,TT
ECI/CSI/ASI (generalized anxiety disorder, separation anxiety disorder) [P, C, TA]	B	IT
Self-Injury**		
Behavior Problems Inventory (self-injury) [C, TA, Ad]	A	I-R, IC, T-T
DASH-II (self-injurious behaviors) [TA, Ad]	B	I-R, IC
Repetitive Behavior Scale-Revised [C, TA, Ad]	B	I-R, IC, T-T
Self Injurious Behavior Questionnaire [TA, Ad]	B	Unknown
Other		
ECI/CSI/ASI (major depressive disorder, dysthymia, bipolar disorder) [P, C, TA]	B	I-R, IC

* Volatile, emotional, sometimes explosive behavior. The following behaviors may be evident: temper tantrums, aggression, mood swings, self-injury, destructiveness, outbursts, and/or screaming.
† Physical overactivity, marked difficulty sustaining attention, and impulsiveness. Such patients may be at risk in parking lots or near roads. They may be exceptionally difficult to manage in stimulating environments, such as supermarkets and stores. In some patients, only one or two of the elements may be present.
‡ Preoccupation with repetitive behaviors, expectation of repetitions from others, repeated speech (immediate or delayed), insistence on sameness within environment or routine, excessive preoccupation with narrow interests, and physically stereotyped movements.
§ Excessive worrying, nervousness, avoidance, and/or phobic responses in relation to events or stimuli.
** Repetitive mechanical acts, done voluntarily, that cause tissue damage to the person.

Aman MG, Novotny S, Samango-Sprouse C, Lecavalier L, Leonard E, Gadow KD, King BH, Pearson DA, Gernsbacher MA, Chez M. *CNS Spectr*. Vol 9, No 1. 2004.

Neuropsychological Approach

Selecting tests from published batteries has the advantage of standardized administration as well as test stimuli that can be used in multisite trials. There are tests of attention, motor speed, executive function, visual and auditory memory, and visual-spatial and visual-constructive reasoning that can be used on children as young as 2 years of age. These tests can tap selected neurocognitive abilities underlying cognition.

"Experimental" Tests with Roots in Pharmacologic Trials

These tests often attempt to assess the child while he or she is actually manipulating information

Validity	Outcome Studies	Raters	Limitations
Cn, Cr	Y	Pr, T, O	Few aggression items
Cn, Cr	Y	Pr, T, O	Some tangential items
Cn, Cr	N	Pr, T	Devised for typically developing children
Cn	Y	CL	Only 4 items
Cn	Y	Pr, T	Emphasizes disruptive behaviors
Cn	N	O	Normed with severe/profound MR group. *DSM-IV*-derived
Cn	N	T	Teacher only
Cn, Cr	Y	Pr, T, O	N/A
Cn, Cr	N	Pr, T	Devised for typically developing children
Cn	Y	CL	Only 3 items
Cn	Y	Pr, T	N/A
Cn	N	T	Only 4 items; teacher only
Cn, Cr	N	O	Psychometric data presented in conference; peer-reviewed data not yet available
	Y	CL	Derived from normal-ability patients; compulsions only in non-verbal patients; sensitivity to mild change may be limited
Cn	N	O	Confined to stereotypies
Cn, Cr	Y	Pr, T, O	Confined to stereotypies
Cn, Cr	N	Pr, CL,O	Only 7 items
Cr	N	O	Only 6 items; mediocre T-T reliability; difficult to obtain?
Cn, Cr	N	S	N/A
Cn	N	T	Designated for preschoolers
Cn, Cr	N	Pr, T	Devised for typically developing children
Cn	N	O, Pr	N/A
Cr	N	O	Primarily developed for subjects with severe mental retardation
Cn, Cr	N	O	Peer-reviewed data not available yet
Unknown	Y	O	Mixes aggressive and self-injurious behaviors
Cn, Cr	N	Pr, T	Devised for typically developing children

ASDs=autism spectrum disorders; C=children (6–12 years of age); TA=teens/adolescents; Ad=Adults; A=highly recommended for the target behavior; I-R=interrater reliability; IC=internal consistency; T-T=test-retest reliability; Cn=construct validity; Cr=criterion group validity Y=yes; Pr=parent; T=teacher; O=other; ECI=Early Childhood Inventory; CSI=Child Symptom Inventory; ASI=Adolescent Symptom Inventory; P=preschoolers; B=less enthusiasm that measure would be appropriate for the target behavior; N=no; CL=clinician; NCBRF=Nisonger Child Behavior Rating Form; DASH-II=Diagnostic Assessment for the Severely Handicapped-Version II; MR=mentally retarded; *DSM-IV=Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*; N/A=not applicable; ADHD=attention-deficit/hyperactivity disorder; C-YBOCS=Children's Yale-Brown Obsessive-Compulsive Scale; YBOCS=Yale-Brown Obsessive-Compulsive Scale; ADAMS=Anxiety, Depression, and Mood Scale; S=self.

"on the run." Hence, static information, such as might be assessed on many IQ tests, would not qualify. Norms are often not available. It can be challenging to find materials that a youngster with autism will work with, but investigators have been able to do so in the past. The Continuous Performance Test (a vigilance task) is probably the most commonly used measure in pediatric psychopharmacology. Aman[68] and Pearson and colleagues[69] have adapted Continuous Performance Tests with storybook figures as the stimuli. Aman[68] has used a Matching-to-Sample task (memory for colors) in which the child is shown a color (red, yellow, or blue). The child presses the screen on which the stimulus appears and three colors appear shortly thereafter. The subject has to select the color that is the same as the preceding color. There is an algorithm that raises or reduces the time of delay between the stimulus and test colors in relation to the subject's accuracy to adjust to subject differences in ability. The Matching-to-Sample paradigm was originally used to test pigeons, which raises the question of whether the behavioral pharmacology literature should be examined to see if there are useful tasks that could be adapted for participants who experience difficulty with standard tests. One laboratory has had success using a discrimination-learning task for assessing treatment effects in children with autism.[70]

At time of this writing, the RUPP is employing the Discrete Trials Trainer, a computer-controlled program for teaching language to children with ASDs and other disabilities. This program entails the presentation of common objects followed by a computer-generated voice asking the child to touch one of the stimuli. The RUPP Autism Network had to constrain the program so that it could not change in difficulty level (ie, we have deliberately programmed it to be constant for each child). The Discrete Trials Trainer certainly measures motivation and may also assess elements of distractibility. Of the various cognitive "tests" used by the RUPP, this has been the easiest for children to perform (a large majority able to do). Another test that has been used with children having mental retardation is the Classroom Analogue task.[69] This simply involves presenting the child with paper-and-pencil materials (usually mathematics) and determining how many problems the child can solve in some standard time. Although cognitive processing tasks usually do not have norms, this should not be considered a fatal flaw, because all controlled trials should have an appropriate comparison group or condition. All of the cognitive processing tasks mentioned here have been used successfully with developmentally disabled children.

Cognitive Functioning: Conclusion

In general, the greatest challenge is finding learning tasks that the participant can successfully perform, are not overly motor dependent, and are intriguing to the children of various cognitive levels. To do so, the task may have to be seductive (as in the case of the Discrete Trials Trainer), or adaptations may have to be made. For example, to make the Purdue Pegboard Test less frustrating for children with autism and highly irritable behavior, the NIMH Autism RUPP group permitted the tester to place masking tape over one column of holes to avoid confusion for the child. As long as such adaptations are done in a consistent manner with both treatment and control groups, this is a reasonable approach. However, such adaptations would invalidate the use of norms.

Side-Effects Assessment

We have relatively few recommendations for assessing side effects within randomized clinical trials. They are as follows. Interested readers are referred to an excellent discussion of side effects assessments in the developmental disabilities.[70] Our recommendations are based on measures previously used in ASDs or mental retardation as well as issues (eg, presence of enuresis, seizures) that often emerge when working with patients having developmental disabilities. To take a concrete example, degree of continence may wax and wane in a subset of patients. The investigator will naturally want to know if this is related to the agent under study or if it merely is random "noise."

Eliciting Side Effects

A drug that is highly effective, but unsafe, will have little clinical utility. Thus, the importance of measuring side effects in randomized clinical trials, particularly in children, cannot be over-emphasized. There are different views on whether specifically to probe for side effects or to rely upon spontaneous caregiver/participant report. Our opinion is that there is more risk in underreporting side effects than in over-reporting them. Hence we prefer that investigators elicit side effects by direct inquiry. Some researchers may be concerned that active probing will result in "too many" side effects. If the power of suggestion is responsible for such reports, then they should appear equally often in the placebo group as in the drug group.

Height, Weight, and Laboratory Measures

Most psychotropic medicines have the capability of causing weight gain or weight loss. In children, concerns about growth suppression may also be important to address. Moreover, if the agent concerned has no demonstrable effect on appetite and/or weight, such knowledge may be very important in shaping clinical decisions. All else being equal, a particular side-effect profile may well determine medication choice. Given the ease with which height and weight are obtained, it is hard to argue against measuring them. It is rather standard to expect to see relevant laboratory studies before and after treatment in a randomized clinical trial. Some psychotropics may cause alterations in cardiac conduction, liver, or kidney function, and so on. In many situations for drugs involving the population with ASDs, it is likely that medications will be used for months or years rather than days or weeks. Few studies look at the emergence of side effects associated with chronic treatment, but this is an area of investigation that should be encouraged.

What Side Effect Scale To Use?

No current side effect scale is fully satisfactory for every randomized clinical trial. Almost all current instruments leave potential adverse events out that could affect organ systems. Most drug researchers develop lists of potential side effects based on the current literature, the package insert (if available), and theoretical mechanisms that might occur. For example, there is a substantial literature indicating that psychostimulants may cause stereotyped behavior in rodents and other animals. This led the NIMH Autism RUPP Network to include stereotypic behavior on its list of potential side effects, despite the fact that it is seldom seen with stimulant treatment.

Increased hyperactivity, tics, or ritualistic/stereotypic behavior have been observed in prior trials using selective serotonin reuptake inhibitors, secretin, and stimulants. Questionnaires for eliciting these clinical or behavioral side effects should be considered for trials with subjects having ASDs. Sleep difficulty is another area of potential change, as many children with ASDs already have disrupted sleep cycles. It is important to elicit information about sleep before the trial begins. With participants having mental retardation, some measure of incontinence (urinary and fecal) should be considered as it often does wax and wane over the course of a trial.

Extrapyramidal Side Effects

Most antipsychotic drugs are capable of causing a range of neurological side effects, including tardive dyskinesia, although the newer atypical antipsychotics appear less prone to do so. For this reason, trials of antipsychotic agents should include indices of akathisia (eg, Barnes scale)[73]; extrapyramidal side effects (eg, Simpson Angus Scale)[74]; and tardive dyskinesia (eg, Abnormal Involuntary Movement Scale).[75] The Dyskinesia Identification System Condensed User Scale is a tool for assessing tardive dyskinesia that was developed and normed for people with developmental disabilities.[72] Tics are another potential side effect in subjects with ASDs, and tic scales may be of special relevance in trials of stimulants and related agents.

CONCLUSION

We are at a fairly early stage in developing a "psychopharmacology of ASDs." At this point, partly because we have no proven intervention for core autism symptoms, we do not know which instruments are best for assessing these features, although there are several candidates. As they are related to a core feature of autism, language and communication outcomes warrant assessment in randomized clinical trials; they may also be useful for assessing cognitive effects of drug interventions. Several key comorbid problem areas were identified. Depending on the area, potentially promising instruments are available, although further psychometric assessment and evaluation in clinical trials are clearly needed. The biggest challenge with cognitive measures at this stage is finding tasks that are attractive and flexible enough that individuals with a wide range of abilities will be able to perform them. Some tests that have been used in pharmacological trials with subjects having mental retardation may be applicable in this field. Finally, side-effects assessments should probe for adverse events known to occur with the agent concerned, while addressing issues that may be of particular relevance to the ASDs (eg, hyperactivity, ritualistic behaviors, sleep pattern, incontinence).

REFERENCES

1. Guy W. *ECDEU Assessment Manuel for Psychopharmacology, Revised*. Rockville, MD: National Institute of Mental Health. US Department Health Education and Welfare Pub (ADM). 1976;76-338.
2. Lord C, Rutter M, DiLavre PC. *Autism Diagnostic Observation Schedule-Generic (ADOS-G)*. San Antonio, Tex: Psychological Corp; 1998.
3. Schopler E, Reichler RJ, Devellis RF, Daly K. Toward objective classification of childhood autism. *J Autism Dev Disord*. 1980;10:91-103.
4. Schopler E, Reichler RJ, Renner BR. *The Childhood Autism Rating Scale*. Los Angeles, Calif: Western Psychological Services; 1988.

5. Constantino JN, Davis SA, Reich W, et al. Validation of a brief quantitative measure of autistic traits: comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *J Autism Dev Disord*. In press.
6. Matson JL. *Matson Evaluation of Social Skills with Youngsters*. Worthington, Ohio: IDS Publishing; 1990.
7. Gilliam JE. *Gilliam Autism Rating Scale*. Austin, Tex: Pro-Ed; 1995.
8. Freeman BJ, Ritvo ER, Yokota A, Ritvo A. A scale for rating symptoms of patients with the syndrome of autism in real life settings. *J Am Acad Child Adolesc Psychiatry*. 1986;25:130-136.
9. Krug DA, Arick J, Almond P. *Autism Screening Instrument for Educational Planning*. Portland, Or: ASIEP Educational Co.; 1980.
10. Belsito KM, Law PA, Kirk KS, Landa RJ, Zimmerman AW. Lamotrigine therapy for autistic disorder: a randomized, double-blind, placebo controlled trial. *J Autism Dev Disord*. 2001;31:175-181.
11. Owley T, Steele E, Corsello C, et al. A double-blind, placebo-controlled trial of secretin for the treatment of autistic disorder. *Med Gen Med*. 1999;E2.
12. Owley T, McMahon W, Cook EH, et al. Multisite, double-blind, placebo-controlled trial of porcine secretin in autism. *J Am Acad Child Adolesc Psychiatry*. 2001;40:1293-1299.
13. Mesibov GB, Schopler E, Schaffer B, Michal N. Use of the Childhood Autism Rating Scale with autistic adolescents and adults. *J Am Acad Child Adolesc Psychiatry*. 1989;28:538-541.
14. Tobin LE, Glenwick DS. Relation of the Childhood Autism Rating Scale-Parent version to diagnosis, stress, and age. *Res Dev Disabil*. 2002;23:211-223.
15. DiLalla DL, Rodgers SJ. Domains of the Childhood Autism Rating Scale: relevance for diagnosis and treatment. *J Aut Dev Disord*. 1994;24:115-128.
16. Eaves RC, Milner B. The criterion-related validity of the Childhood Autism Rating Scale and the Autism Behavior Checklist. *J Abnorm Child Psychol*. 1993;21:481-491.
17. Pilowsky T, Yirmiya N, Shulman C, Ronit D. The Autism Diagnostic Interview-Revised and the Childhood Autism Rating Scale: differences between diagnostic systems and comparison between genders. *J Aut Dev Disord*. 1998;28:143-151.
18. Sevin JA, Matson JL, Coe DA, Fee VE, Sevin BM. A comparison and evaluation of three commonly used autism scales. *J Aut Dev Disord*. 1991;21:417-432.
19. Constantino JN, Hudziak JJ, Todd RD. Deficits in reciprocal social behavior in male twins: evidence for a genetically independent domain of psychopathology. *J Am Acad Child Adolesc Psychiatry*. 2003;42:458-467.
20. Constantino JN, Przybeck T, Friesen D, Todd RD. Reciprocal social behavior in children with and without pervasive developmental disorders. *J Dev Behav Pediatr*. 2000;21:2-11.
21. South M, Williams BJ, McMahon WM, et al. Utility of the Gilliam Autism Rating Scale in research and clinical populations. *J Autism Dev Disord*. 2002;32:593-599.
22. Matson JL, Compton LS, Sevin JA. Comparison and item analysis of the MESSY for autistic and normal children. *Res Dev Disabil*. 1991;12:361-369.
23. Njardvik U, Matson JL, Cherry KE. A comparison of social skills in adults with autistic disorder, pervasive developmental disorder not otherwise specified, and mental retardation. *J Autism Dev Disord*. 1999;29:287-295.
24. Krug DA, Arick J, Almond P. Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. *J Child Psychol Psychiatry*. 1980;21:221-229.
25. Kelley E, Jones G, Fein D. Language assessment in children. In: Goldstein G, Beers, SR, eds. *The Comprehensive Handbook of Psychological Assessment, Volume 1: Intellectual and Neuropsychological Assessment*. New York, NY: Wiley & Sons; 2003.
26. Filipek PA, Accardo PJ, Baranek GT, et al. The screening and diagnosis of autistic spectrum disorders. *J Aut Dev Disord*. 1999;29:439-484.
27. Mundy P, Hogan A. *A preliminary manual for the abridged Early Social Communication Scales (ESCS)*. Coral Gables, Fla: University of Miami Press; 1996.
28. Dunn LM, Dunn LA. *Peabody Picture Vocabulary Test*. 3rd ed. Circle Pines, Minn: American Guidance Service; 1997.
29. Williams KT. *Expressive Vocabulary Test*. Circle Pines, Minn: American Guidance Service; 1997.
30. Carrow-Woolfolk E. *Comprehensive Assessment of Spoken Language*. Circle Pines, Minn: American Guidance Service; 1999.
31. Semmel E, Wiig EH, Secord WA. *Manual for the Clinical Evaluation of Language Fundamentals*. 3rd ed. San Antonio, Tex: The Psychological Corporation; 1995.
32. Secord WA & Semel E. *Manual for the Preschool Clinical Evaluation of Language Fundamentals*. 3rd ed. San Antonio, Tex: The Psychological Corporation; 1995.
33. Sparrow S, Balla D, Ciccetti D. *Vineland Adaptive Behavior Scales*. Circle Pines, Minn: American Guidance Service; 1984.
34. Bruininks RH, Woodcock RW, Weatherman RF, Hill BK. *Scales of Independent Behavior*. rev ed. Manual. Boston, Mass: Riverside Publishing Company; 1996.
35. Nihira K, Leland H, Lambert N. *AAMR Adaptive Behavior Scale-Residential Community:* Second Edition. Austin, Tex: PRO-ED; 1993.
36. Carter AS, Volkmar FR, Sparrow SS, et al. The Vineland Adaptive Behavior Scales: supplementary norms for individuals with autism. *J Aut Dev Disord*. 1998;28:287-302.
37. Kern JK, Miller VS, Cauller PL, Kendall PR, Mehta PJ, Dodd M. Effectiveness of N,N-dimethylglycine in autism and pervasive developmental disorder. *J Child Neurol*. 2001;16:169-173.
38. Partington JW, Sundberg ML. *Assessment of Basic Language and Learning Skills*. Pleasant Hill, Calif: Behavior Analysts, Inc; 1998.
39. Gadow KD, Sprafkin J. *Early Childhood Symptom Inventory Manual*. Stony Brook, NY: Checkmate Plus; 1997.
40. Gadow KD, Sprafkin J. *Child Symptom Inventories Manual*. Stony Brook, NY: Checkmate Plus; 1994.
41. Gadow KD, Sprafkin J. *Adolescent Symptom Inventory Manual*. Stony Brook, NY: Checkmate Plus; 1997.
42. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Press; 1994.
43. Aman MG, *Assessing Psychopathology and Behavior Problems in Persons with Mental Retardation: A Review of Available Instruments*. DHHS Publication No. [ADM] 91-1712. Rockville, MD: U.S. Department of Health and Human Services; 1991.
44. Aman MG. Review and evaluation of instruments for assessing emotional and behavioural disorders. *Australia and New Zealand Journal of Developmental Disabilities*. 1991;17:127-145.
45. Lecavalier L, Aman MG. Rating instruments. In: Matson JL, Matson ML, eds. *Behavior Modification for Persons with Developmental Disabilities: Empirically Supported Treatments*. Kingston, NY: National Association for the Dually Diagnosed Press; In press.
46. Aman MG, Singh NN. *Aberrant Behavior Checklist-Community. Supplementary Manual*. East Aurora, NY: Slosson Educational Publications; 1994.

47. Scahill L, Riddle MA, McSwiggin-Hardin M, et al. Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. *J Am Acad Child Adolesc Psychiatry*. 1997;36:844-852.
48. Esbensen AJ, Rojahn J, Aman MG, Ruedrich S. The reliability and validity of an assessment instrument for anxiety, depression and mood among individuals with mental retardation. *J Aut Dev Disord*. 2003;33:617-629.
49. Rojahn J, Matson JL, Lott D, Esbensen AJ, Smalls Y. The Behavior Problems Inventory: an instrument for the assessment of self-injury, stereotyped behavior, and aggression/destruction in individuals with developmental disabilities. *J Aut Dev Disord*. 2001;31:577-588.
50. Goodman WK, Price LH, Rasmussen SA, et al. The Yale-Brown Obsessive Compulsive Scale (Y-BOCS): Part I. Development, use, and reliability. *Arch Gen Psychiatry*. 1989;46:1006-1011.
51. Overall JE. Campbell M. Behavioral assessment of psychopathology in children: infantile autism. *J Clin Psychol*. 1988;44:708-716.
52. Einfeld SL, Tonge BJ. *Manual for the Developmental Behavior Checklist*. Sydney, Australia: Monash University Center for Developmental Psychiatry and School of Psychiatry, University of New South Wales; 2002.
53. Matson JL. *Manual for the Diagnostic Assessment for the Severely Handicapped-II*. Baton Rouge, La: Louisiana State University Press; 1995.
54. Feinstein C, Kaminer Y, Barrett RP, Tylenda B. The assessment of mood and affect in developmentally disabled children and adolescents: The Emotional Disorders Rating Scale. *Res Dev Disabil*. 1988;9:109-121.
55. Ollendick TH. Reliability and validity of the Revised Fear Survey Schedule for Children (FSSC-R). *Behav Res Ther*. 1983;21:685-692.
56. Sarphare G, Aman MG. Parent- and self-ratings of anxiety in children with mental retardation: agreement levels and test-retest reliability. *Res Dev Disabil*. 1996;17:27-39.
57. Ramirez SZ, Kratochwill TR. Development of the Fear Survey Schedule for Children with and without mental retardation. *Behav Assess*. 1990;12:457-470.
58. Aman MG, Tassé MJ, Rojahn J, Hammer D. The Nisonger CBRF: a child behavior rating form for children with developmental disabilities. *Res Dev Disabil*. 1996;17:41-57.
59. Tassé MD, Aman MG, Hammer D, Rojahn J. The Nisonger Child Behavior Rating Form: age and gender effects and norms. *Res Dev Disord*. 1996;17:59-75.
60. Behar LB, Stringfield SA. A behavior rating scale for the preschool child. *Dev Psychol*. 1974;10:601-610.
61. Aman MG, Rojahn J. The psychometric characteristics of the Preschool Behavior Questionnaire in preschoolers with developmental handicaps. *Journal of Developmental and Physical Disabilities*. 1994;6:311-325.
62. Bodfish JW, Symons FJ, Parker DE, Lewis MH. Varieties of repetitive behavior in autism. *J Autism Dev Disord*. 2000;30:237-243.
63. Schroeder SR, Rojahn J, Reese RM. Reliability and validity of instruments for assessing psychotropic medication effects on self-injurious behavior in mental retardation. *J Autism Dev Disord*. 1997;27:89-102.
64. Rojahn J, Matlock ST, Tassé MJ. The Stereotyped Behavior Scale: Psychometric properties and norms. *Res Dev Disabil*. 2000;21:437-454.
65. Zelazo PR. Infant-toddler information processing assessment for children with pervasive developmental disorder and autism: Part I. *Infants Young Child*. 1997;10:1-14.
66. Arnold LE, Aman MG, Martin A, et al. Assessment in multisite randomized clinical trials of patients with autistic disorder: the Autism RUPP Network. *J Autism Dev Disord*. 2000;30:99-111.
67. Roid GH, Miller LJ. *Leiter International Performance Scale*. rev. Wood Dale, Ill: Stoelting; 1997.
68. Aman MG. Applications of computerized cognitive-motor measures to the assessment of psychoactive drugs. In: Dodson WE, Kinsbourne M, Hiltbrunner B, eds. *The Assessment of Cognitive Function in Epilepsy*. New York, NY: Demos Publications; 1991.
69. Pearson DA, Yaffee LS, Loveland KA, Lewis KR. A comparison of sustained and selective attention in children who have mental retardation with and without attention deficit hyperactivity disorder. *Am J Ment Retard*. 1996;100:92-607.
70. Anderson TT, Campbell M, Grega DM, Perry R, Small AM, Green WH. Haloperidol in the treatment of infantile autism: Effects on learning and behavioral symptoms. *Am J Psychiatry*. 1984;141:1195-1202.
71. Handen BL, Breaux AM, Gosling A, Ploof DL, Feldman H. Efficacy of Ritalin among mentally retarded children with ADHD. *Pediatrics*. 1990;86:922-930.
72. Kalachnik JE. Measuring side effects of psychopharmacologic medication in individuals with mental retardation and developmental disabilities. *Ment Retard Dev Disabil Res Rev*. 1999;5:348-359.
73. Barnes TR. A rating scale for drug-induced akathisia. *Br J Psychiatry*. 1989;154:672-676
74. Simpson GM, Angus JW. A rating scale for extrapyramidal side effects. *Acta Psychiatr Scand* 1970;212:11-19.
75. National Institute of Mental Health. Abnormal Involuntary Movement Scale (AIMS). *Psychopharmacol Bull*. 1985;21:1077-1080.