New Developments in Autism Clinical Trials

By Eric Hollander, MD, Ricki Robinson, MD, MPH, and Doug Compton, MA

This issue of CNS Spectrums represents a milestone for both those affected by autism and related disorders and for clinicians and researchers who provide help and support to these individuals. In order to improve the lives of a rapidly increasing number of patients with this diagnosis, it is critical for the medical and research community to apply a collaborative and consistent strategy for clinical studies in autism.

The five articles that follow are a synopsis of the five subcommittees of the Autism Clinical Trials Task Force (ACTTF), a collaborative panel of experts convened by the Cure Autism Now (CAN) foundation in the spring of 2002. These subcommittees focused on key issues in clinical trials of autism, including subject selection, outcome measures, study design, biological measures, and governmental issues. CAN brought together this panel of participants from academia, government, and industry for the ACTTF think tank. The panel was organized and chaired by Eric Hollander, MD, and Ricki Robinson, MD, MPH. The overarching goal was to clarify what was known about the state of the art of the field in this area, what key information was still unknown, and to implement specific approaches and suggestions for studies which would provide information to fill in the missing gaps in our knowledge base.

CAN consists of parents, physicians, and researchers dedicated to promoting and funding research with direct clinical implications for treatment of autism and finding a cure. CAN has a three-pronged approach toward promoting advances in the field: funding autism research and open resources, educating and organizing families, and performing political action. CAN was instrumental in the introduction and passage of the Children's Health Act of 2000, which established the coordinated centers of excellence, Studies to Advance Autism Research and Treatment (STAART). CAN also maintains a strong commitment to clinical research and provides ongoing funding to treatment-related studies. CAN has long recognized the need for a consistent core methodology embedded in multicenter autism clinical trials in order to compare and contrast data from multiple sites and multiple therapeutic areas.

The articles in this issue include a compilation of the state of the art for each area, exemplar data, recommendations for future research directions and concepts from the outcome of discussions from the meeting. It is the hope of the foundation that these articles will provide a resource from which to move the field forward and build upon current results through common core components and approaches in future autism trials.

Lawrence Scahill, MSN, PhD, and Catherine Lord, PhD, describe the key issues in subject selection for autism clinical trials and how this variable may influence outcome measure selection and study design. They highlight specific issues from the Research Units on Pediatric Psychopharmacology (RUPP) Autism Network consortium on risperidone versus placebo in disruptive behavior of autism to highlight challenges to subject selection and solutions to these problems.

Michael G. Aman, PhD, and colleagues highlight outcome measures for clinical trials by describing how marked variability in age and levels of functioning make selection of assessment tools a challenge; how there is no definitive tool for assessing core features of autism but there are several promising tools worthy of further study; and how instruments for language and communication are dependent on the participant's developmental level. They identify instruments for challenging behaviors, (ie, repetitive behaviors, irritability, and anxiety), cognitive function, and side-effect assessment.

Eric Hollander, MD, and colleagues examine challenges in the design of pharmacologic trials in autism; discuss the need to stratify the autism population for specific symptom domains in ongoing and future trials; and describe recent trials with various agents on specified symptom domains in autism to illustrate stratification strategies.

George M. Anderson, PhD, and colleagues review biomedical measures that are recommended for inclusion in the selection and screening phase of an autism clinical trial; review and discuss biomedical measures of potential utility in understanding drug response and underlying pathophysiology in autism patients; and consider possible approaches to examining genetic influences.
on drug response in autism clinical trials.

Benedetto Vitiello, MD, and Ann Wagner, PhD, suggest that, given the high public health relevance of autism treatment research and the relatively low interest of the pharmaceutical industry in autism, the role of the National Institutes of Health (NIH) in supporting this research is paramount. The RUPP, STAART Autism Networks, and other government activities in autism clinical trials, are briefly reviewed.

This month’s articles reflect the active examination by a consortium of academic researchers, consumer advocates, industry researchers, and government officials of state-of-the-art and unresolved issues in autism clinical trials. This area of study is rapidly evolving given the increased prevalence of autism, the increased NIH and industry support in autism research and drug development, the increasing number of targets discovered by genetics and neurobiology research for drug development, the increasing number of trials with different classes of medications for different symptoms, and the lack of current Food and Drug Administration-indicated medications for autism.

This issue of CNS Spectrums inspires and serves as a reference to investigators, clinicians, family members, and consumer advocates who treat and study this most fascinating and challenging of neuropsychiatric conditions. CNS

**APPENDIX: THE AUTISM CLINICAL TRIALS TASK FORCE PARTICIPANTS**

Natacha Akshoomoff, PhD; Albert J. Allen, MD, PhD; Michael G. Aman, PhD; George M. Anderson, PhD; David Baskin, MD; Peter Bell, BS, MBA; Matthew Belmonte, PhD; Michael Chez, MD; Diane C. Chugani, PhD; Doug Compton, MA; Therese Finazzo; Lisa Ford, MD; Kenneth D. Gackow, PhD; Donald Gutheinz, PhD; Walter Herlihy, PhD; William Hirinstein, PhD; Deborah Hirtz, MD; Eric Hollander, MD; Porcia Iversen, BS; Bryan H. King, MD; Paul Law, MD; Elizabeth Leonard, PhD; Catherine Lord, PhD; Janet Miller, PhD, JD; Tan Nguyen, MD, PhD; Sherie Novotny, MD; Caralynn Nowinski, BA; Thomas Owley, MD; John Pomeroy, MD; Ricki Robinson, MD, MPH; Carole Samango-Sprouse, EdD; Lawrence Scull, PhD; Sharon Shelton, BS; Jonathan Shestack, BS; Lin Sikich, MD; Jeremy Silverman, PhD; Sarah Spence, MD, PhD; Elizabeth Telegy; Benedetto Vitiello, MD; Ann Wagner, PhD; Serena Wieder, PhD; and Andrew Zimmerman, MD.

**REFERENCE**


We would like to thank the following peer-reviewers who contributed to CNS Spectrums in 2003:

Sheena Aurora, MD  
David Avery, MD  
Richard Balon, MD  
David P. Bernstein, PhD  
John L. Beyer, MD  
Marcelo Bigal, MD  
Martin Bohnus, MD  
Soo Borson, MD  
Stephanie Bouchard, MD  
Charles L. Bowden, MD  
Julie Bower, PhD  
Timothy J. Bruce, PhD  
Richard Bryant, PhD  
Daniel Buyse, MD  
John Capitanio, PhD  
Cheryl Carmin, PhD  
Yael Caspi, ScD, MA  
Kiki D. Chang, MD  
Kathleen Chard, PhD  
Teofilo Lee Chiang, MD  
Dan Connor, MD  
James Couch, MD  
Paul Crist-Christoph, PhD  
CS De Kloet, MD  
Melissa P. DelBello, MD  
Edward DeMice, PhD  
Darin Dougherty, PhD  
Ronald Duman, PhD  
Graham Emslie, MD  
Milton Erman, MD  
John A. Fairbank, PhD  
Steven H. Ferris, PhD  
Robert Findling, MD  
Martin E. Franklin, PhD, MA  
Elber Geuze, MA  
Michael Girdin, MD  
Robert Golden, MD  
Igor Grant, MD  
Madeleine Grigg-Dummerber, MD  
Katherine A. Halmi, MD  
Harry Haroutunian, PhD  
Jack Hirschowitz, MD  
Ralph Hoffman, MD  
Stefan Hofmann, PhD  
Eric Hollander, MD  
Mustafa Husain, MD  
Gail Ironson, PhD, MA  
David Jemison, MD  
David Kennaway, BSc, PhD  
Ehud Klein, MD  
Lorrie M. Koran, MD  
Robert Kowatch, MD  
Lawrence Labbate, MD  
Ruth Lanius, MD, PhD  
John Lauriello, MD  
Helen Lavrersky, MD  
Steve Levine, MD  
Hans-Peter Lipp, MD  
Susan Lugtendorf, PhD  
John Mann, MD  
Randall Marshall, MD  
Alessandro Mathe, MD  
Peter Mattsson, MD, PhD  
Jack Mc Ardle, PhD  
Linda M. Mc Lean, PhD  
John McQuaid, PhD  
Douglas Mennin, MD  
Joost Mertens, MD  
Emanuel Mignot, MD  
Merry Miller, MD  
Jacob Minter, MD  
Marie-Christine Miquel, MD  
Dimos D. Mitsikostas, MD  
Frederick G. Moeller, MD  
Francis M. Mordimore, MD  
Lawrence Newman, MD  
Eric Nofinger, MD  
Mauric M. Ohayon, MD  
Lars-Goran Ost, PhD  
Joel Paris, MD  
David Pauls, PhD  
Mani Pavaluri, MD  
Roger K. Pittman, MD  
Mark Pressman, PhD  
Lawrence H. Price, MD  
Suzan Priddmore, MD  
Jeffrey M. Pyne, MD  
Judith Rabkin, PhD, MPH  
Judith Rapoport, MD  
Ann M. Rasmussen, MD  
Scott L. Rauch, MD  
Sheila Rauch, PhD  
Friedel M. Reischies, MD  
Thomas Rinne, MD, PhD  
Craig W. Ritchie, RCPsych  
Robert Robinson, MD  
Jerry Rosen, MD  
Deborah A. Roth, PhD  
Marwan Sabbagh, MD  
Michael Saita, MD  
Christian Schmid, MD  
Soraya Seadat, FCPsych  
Martin E. Shapiro, MD  
Vansh Sharma, MD  
Marwan Sabbagh, MD  
Robert Robinson, MD  
Jerry Rosen, MD  
Deborah A. Roth, PhD  
Marwan Sabbagh, MD  
Michael Saita, MD  
Christian Schmid, MD  
Soraya Seadat, FCPsych  
Martin E. Shapiro, MD  
Vansh Sharma, MD  
Richard Weiner, MD  
Myron F. Weiner, MD  
Dan Weintraub, MD  
Maureen Whittall, PhD  
Timothy Wilens, MD  
Jeffrey J. Wilson, MD  
Paul Winner, DO  
Rachel Yehuda, PhD  
Vincent Zarcone, MD  
Irina Zhdanova, MD, PhD  
Stephen D. Silberstein, MD, FACP  
Jeremy Silverman, PhD  
Gerald F. Smith, MD  
Brent Solvason, MD  
Jeppe Sturis, MD  
Trisha Suppes, PhD, MD  
John David Sween, PhD  
Martin Teicher, MD, PhD  
David Tolin, PhD  
David Trachtenberg, MD  
Glen Treisman, MD, PhD  
Phebe Tucker, MD  
Oliver Turnbull, PhD  
Eric Vermetten, MD, PhD  
Adelle Viguera, MD  
Herbert E. Ward, MD  
Richard Weiner, MD  
Myron F. Weiner, MD  
Dan Weintraub, MD  
Maureen Whittall, PhD  
Timothy Wilens, MD  
Jeffrey J. Wilson, MD  
Paul Winner, DO  
Rachel Yehuda, PhD  
Vincent Zarcone, MD  
Irina Zhdanova, MD, PhD  
Stefan D. Silberstein, MD, FACP  
Jeremy Silverman, PhD  
Gerald F. Smith, MD  
Brent Solvason, MD  
Jeppe Sturis, MD  
Trisha Suppes, PhD, MD  
John David Sween, PhD  
Martin Teicher, MD, PhD  
David Tolin, PhD  
David Trachtenberg, MD  
Glen Treisman, MD, PhD  
Phebe Tucker, MD  
Oliver Turnbull, PhD  
Eric Vermetten, MD, PhD  
Adelle Viguera, MD  
Herbert E. Ward, MD  
Richard Weiner, MD  
Myron F. Weiner, MD  
Dan Weintraub, MD  
Maureen Whittall, PhD  
Timothy Wilens, MD  
Jeffrey J. Wilson, MD  
Paul Winner, DO  
Rachel Yehuda, PhD  
Vincent Zarcone, MD  
Irina Zhdanova, MD, PhD
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

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

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.

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. 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), the Childhood Autism Rating Scale (CARS), the Social Responsiveness Scale, the Matson Evaluation of Social Skills in Youngsters (MESSY), the Gilliam Autism Rating Scale (GARS), the Ritvo-Freeman Real Life Rating Scale for Autism (RLRS), and the Autism Behavior Rating Scale. 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, 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. 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 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, and as an informant-based rating scale. 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 a 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. 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, 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 was introduced to assess social deficits in ASDs. 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. 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) 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, 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. 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 Checklist25,26 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 stereotypes 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.

Prélinguistic Communication

Forms of pré-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, 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. The Expressive Vocabulary Test is for patients 2–90+ years of age and administration averages 15 minutes. 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. 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. Finally, there is the Clinical Evaluation of Language Fundamentals-Preschool for 3–6 years of age.

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, another is the Scales of Independent Behavior—Revised (SIB–R), and the third is the American Association on Mental Retardation Adaptive Behavior Scale. 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. 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. The Vineland is under revision and, according to the publisher’s website, the Vineland-II is expected to be available in 2005.

The Assessment of Basic Language and Learning Skills (ABLLS) 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 generalizable 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),\textsuperscript{39} Child Symptom Inventory (CSI; 5–12 years of age),\textsuperscript{40} or Adolescent Symptom Inventory (ASI),\textsuperscript{41} all of which include all sections of the Diagnostic and Statistical Manual of Mental Disorder, Fourth Edition (DSM-IV)\textsuperscript{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.\textsuperscript{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.\textsuperscript{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)\textsuperscript{46} and (less so) the Yale-Brown Obsessive-Compulsive Scale and its close relative, the Children’s Yale-Brown Obsessive-Compulsive Scale.\textsuperscript{47,50}

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;\textsuperscript{46} Anxiety, Depression, and Mood Scale;\textsuperscript{48} Behavior Problems Inventory;\textsuperscript{49} the Children’s Yale-Brown Obsessive-Compulsive Scale and the Yale-Brown Obsessive-Compulsive Scale;\textsuperscript{47,50} the adapted Children’s Psychiatric Rating Scale;\textsuperscript{51} Developmental Behaviour Checklist;\textsuperscript{52} Diagnostic Assessment for the Severely Handicapped—Version II;\textsuperscript{53} ECI/ASI;\textsuperscript{19–41} Emotional Disorders Rating Scale;\textsuperscript{54} Fear Survey for Children—Revised;\textsuperscript{55,56} Fear Survey for Children with and without Mental Retardation;\textsuperscript{57} Nisonger Child Behavior Rating Form;\textsuperscript{58,59} Preschool Behavior Questionnaire;\textsuperscript{60,61} Repetitive Behavior Scale—Revised;\textsuperscript{62} Self Injurious Behavior Questionnaire;\textsuperscript{63} and the Stereotypic Behavior Scale.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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. A test, such as the Leiter International Performance Test-Revised, 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.

**Neuropsychological Approach**

Selecting tests from published batteries has the

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

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

* Intensity, 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.

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.

Table 1. **Instruments for Assessing Target Symptoms of Autism Spectrum Disorders**

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

---

**“Experimental” Tests with Roots in Pharmacologic Trials**

These tests often attempt to assess the child while he or she is actually manipulating information “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 and Pearson and colleagues have adapted Continuous Performance Tests with storybook figures as the stimuli. Aman 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.

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. 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. 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)\(^1\); extrapyramidal side effects (eg, Simpson Angus Scale)\(^2\); and tardive dyskinesia (eg, Abnormal Involuntary Movement Scale).\(^3\) The Dyskinesia Identification System Condensed User Scale is a tool for assessing tardive dyskinesia that was developed and normed for people with developmental disabilities.\(^4\) 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


Subject Selection and Characterization in Clinical Trials in Children with Autism

By Lawrence Scahill, MSN, PhD, and Catherine Lord, PhD

FOCUS POINTS

• The goal of clinical research is to provide guidance to clinicians on the risks and benefits of any given treatment.
• Careful subject selection and characterization provides essential information about who was in the trial and for whom the treatment is relevant.
• A fundamental feature of subject selection is identification of the primary outcome measure.
• Setting a threshold for severity on the primary outcome measure can help reduce variability in the sample and potentially improve the signal to noise ratio on the primary outcome measure.
• Effect size is a useful metric for expressing magnitude of change because it accounts for variability in the primary outcome measure.
• Sample size can be determined by estimating the minimum clinically meaningful effect size.

ABSTRACT

The goal of psychopharmacologic research in autism is to provide guidance to clinicians and families on the risks and benefits of any given treatment. Careful subject selection and characterization in clinical trials are necessary for replication, to inform clinicians about the sample, and to elucidate the type of patients who might benefit from the treatment. At minimum, subject characterization includes demographic information, diagnosis (autism, Asperger’s syndrome, or pervasive developmental disorder—not otherwise specified), intellectual functioning, adaptive functioning, symptom severity, general behavioral profile, health status, pertinent clinical laboratory measures, height, weight, current treatments, and educational placements. Subject selection, sample size, and choice of the primary outcome measure are closely interrelated and linked to the study hypothesis. The magnitude of expected improvement on the primary outcome measure, which can be expressed by effect size, has direct implications for sample size. Large sample sizes are required to detect small effect sizes. To facilitate interpretation of study results, research reports should provide descriptive characteristics of the sample as well as the mean change and standard deviation on the primary outcome measure data to permit calculation of the effect size.

CNS Spectr 2004;9(1):22-32

INTRODUCTION

Subject selection and characterization for clinical trials in autism and pervasive developmental disorders (PDDs) share several challenges in common with other serious psychiatric disorders in children, such as trade-offs between generalizability to the wide range of individuals with autistic spectrum disorders (ASDs) and clear interpretation of results. For example, in a placebo-controlled study of attention-deficit/hyperactivity disorder (ADHD), investigators must decide whether to include children with the inattentive type, the hyperactive/impulsive type or both. Other design challenges, such as non-verbal subjects, limited understanding of the underlying pathophysiology, and the overall disappointing track record of medications for the core symptoms of the disorder are more unique to autism.

1 Dr. Scahill is associate professor of nursing and child psychiatry in the Departments of Nursing and Child Psychiatry at the Child Study Center at Yale University in New Haven, Connecticut. Dr. Lord is professor of psychology and psychiatry in the Departments of Psychology and Psychiatry at the University of Michigan in Ann Arbor.

Acknowledgements: The authors would like to thank Andres Martin, MD, George Anderson, PhD, and Kathy Koenig, MSN, and Dierdre Carroll, MSN, for their collaboration, and Dan Karstofsky and Erin Kustan for their help in preparing this manuscript.

Disclosures: This work has been supported by the following United States Public Health Service grants: MH-70009, MH-30929, MH-49351, HD-03008, MO1-RR-06022, MH-066496, HD-35482, by Research Units on Pediatric Psychopharmacology Autism Network of the National Institute of Mental Health. Dr. Scahill has received a grant from Janssen, is on the Speakers’ Bureau of Janssen, and is a consultant for Bristol Myers-Squibb and Janssen. Dr. Lord receives royalties on the sales of the Autism Diagnostic Observation Schedule and the Autism Diagnostic Interview, both published by Western Psychological Services. This paper was submitted on August 11, 2003, and accepted November 20, 2003.

Please direct all correspondence to: Lawrence Scahill, MSN, PhD, Yale Child Study Center, P.O. Box 207900 New Haven, CT 06520; E-mail: lawrence.scahill@yale.edu.
Another factor is the underdeveloped state of outcome measurements in autism. This limitation constrains the questions that can be answered in a clinical trial to those that can be evaluated with available measures or questions for which crude measures of global functioning are acceptable. This paper first reviews the goals of clinical trials with a focus on psychopharmacology, however, many issues may be relevant to psychosocial interventions as well. We then examine the challenges of subject selection and characterization of children and adolescents with autism who participate in clinical trials. We close with a discussion of how to select measures to be used in clinical trials. Reports from recent clinical trials in autism are used to illustrate the salient issues of subject selection.

**Goal of Clinical Trials**

The stated aim of most clinical trials in children and adolescents with developmental disorders is to evaluate the efficacy and safety of a given treatment. Beyond this fundamental aim, the value of a clinical trial can be judged by the degree to which it guides clinical practice. First, if the selection and characterization of subjects is clear, clinicians can easily match the treatment to specific patients with similar characteristics. Second, an informative study also provides estimates about the probability and magnitude of positive response. Third, a useful clinical trial also provides guidance on the dose schedule and time to effect (duration from baseline to signs of therapeutic benefit). Finally, the types and frequency of adverse events associated with the treatment are essential for family education and ongoing clinical management. Due consideration of these goals during the study design phase can help clinicians and families make informed decisions about the risks and benefits of a given treatment.

**ISSUES IN SUBJECT SELECTION**

**Generalizability**

In order for clinicians to judge the relevance of a treatment trial to a particular patient, study subjects must be characterized with respect to demographic characteristics (age, gender, ethnicity, living situation, socioeconomic status), diagnosis (autism, Asperger's disorder, or PDD-not otherwise specified [PDD-NOS]), intellectual functioning, adaptive functioning, symptom severity, general behavioral profile, health status, pertinent laboratory measures, as appropriate (e.g., complete blood count, liver function tests, triglyceride levels, electrocardiogram), height, weight, current and prior treatments, and educational placements. These clinical features enable the clinical community to discern who was in the study and, therefore, the type of patients that might also benefit from the treatment. Thus, in the design phase of a study, decisions concerning the entry requirements (inclusion and exclusion criteria) are critical. The scientific debate usually turns on the issue of heterogeneity versus homogeneity of the sample. From a practical perspective, highly restrictive entry criteria ensure a homogeneous sample, but may make it more difficult to identify and recruit subjects and limit the generalizability of findings.

The concern about accepting too much heterogeneity in the study sample is that it could increase the variability in treatment response and decrease the likelihood of detecting a difference between groups. For example, preliminary evidence suggests that younger children with PDDs are more vulnerable to serotonin reuptake inhibitor-induced behavioral activation. If these age effects on clinical response are true, there are two rational choices: study only one age group to avoid the problem of variability across age groups; or include both older and younger children and compare the response across age groups. The first strategy will yield findings that are only relevant to the selected age group (limiting generalizability). The second strategy may indeed result in a clinical trial that applies to a wider population. To detect the differences in response by age, however, would require a larger sample size so that each age group could be evaluated separately. To achieve an appropriate balance between heterogeneity and homogeneity in a study sample, therefore, investigators should carefully consider the hierarchy of questions to be answered by the clinical trial.

Once the inclusion and exclusion criteria are determined, the study investigators must develop systematic methods of assessment to ensure that only appropriate subjects are enrolled. A valuable aid for judging the generalizability of the study sample is to specify the number of subjects who were screened and did not meet entry criteria or who refused to participate in the study. For example, in the Research Units on Pediatric Psychopharmacology (RUPP) Autism Network risperidone trial of 101 children with autism, 270 children were screened. Of these, 112 children did not meet the entry criteria; the parents of 57 apparently eligible children declined to participate. Tracking the number of subjects who were screened but not eligible, screened and declined, and those ultimately enrolled is now recommended.
Defining a sample by certain clinical features and selecting a sample according to specific entry criteria are somewhat different issues. Incomplete characterization of subjects has been a recurring problem in clinical trials in autism and related disorders.\(^6,7\) Without adequate sample characterization, it is difficult to determine who participated in the trial and to compare results across studies. Providing a clear characterization of subjects may also help to identify subgroups with a higher or lower likelihood of response. Thus, systematic assessment during the screening phase serves three essential purposes: to establish the pretreatment baseline, to match subjects to the entry criteria, and to characterize the sample. Of course, the importance of characterizing the sample must be balanced against cost and subject burden. Table 1 shows the domains of interest in evaluating and characterizing subjects in a clinical trial as well as examples of measures used in the assessment of each domain.\(^1\)

The choice of measures used to characterize the sample depends a great deal on the research questions. There are several instruments that provide information about the core symptoms of ASD (Table 1). Instruments, such as the Autism Diagnostic Interview-Revised (ADI-R)\(^8\) and the Autism Diagnostic Observation Schedule (ADOS),\(^9\) which are considered diagnostic instruments, provide algorithms that map directly onto Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)\(^10\) and International Classifications of Diseases, Tenth

**TABLE 1. SAMPLE DEFINITION AND SELECTION FOR CLINICAL TRIALS IN ASD**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Data Point or Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td></td>
<td>Ethnicity and race</td>
</tr>
<tr>
<td></td>
<td>Parental education and occupation</td>
</tr>
<tr>
<td></td>
<td>Living situation</td>
</tr>
<tr>
<td></td>
<td>Source of subjects (ascertainment)</td>
</tr>
<tr>
<td>PDD Diagnosis</td>
<td>Autism Diagnostic Interview-Revised</td>
</tr>
<tr>
<td></td>
<td>Childhood Autism Rating Scale</td>
</tr>
<tr>
<td></td>
<td>Autism Diagnostic Observation Schedule</td>
</tr>
<tr>
<td>Core Symptoms of PDD</td>
<td>Social Communication Questionnaire</td>
</tr>
<tr>
<td></td>
<td>Social Responsiveness Scale</td>
</tr>
<tr>
<td>Intellectual Functioning</td>
<td>Weschler tests</td>
</tr>
<tr>
<td>Differential Abilities Scale</td>
<td>Mullen Scales of Early Learning</td>
</tr>
<tr>
<td></td>
<td>Leiter International Performance Scales</td>
</tr>
<tr>
<td>Language Skill</td>
<td>Peabody Picture Vocabulary Test</td>
</tr>
<tr>
<td></td>
<td>Vineland Communication Domain (for very low functioning individuals)</td>
</tr>
<tr>
<td>Adaptive Functioning</td>
<td>Vineland Adaptive Behavior Scales</td>
</tr>
<tr>
<td></td>
<td>Adaptive Behavior Language and Learning Skills</td>
</tr>
<tr>
<td>General Behavioral Profile</td>
<td>Aberrant Behavior Checklist (non-core features of PDD)</td>
</tr>
<tr>
<td>Comorbid Psychiatric Disorder</td>
<td>Child Symptom Inventory</td>
</tr>
<tr>
<td></td>
<td>Diagnostic Interview Schedule for Children</td>
</tr>
<tr>
<td>Health Status</td>
<td>Height, weight, head circumference</td>
</tr>
<tr>
<td></td>
<td>Blood pressure, pulse</td>
</tr>
<tr>
<td></td>
<td>Medical history</td>
</tr>
<tr>
<td></td>
<td>Physical examination and review of body systems</td>
</tr>
<tr>
<td></td>
<td>Sleep, appetite, and activity level</td>
</tr>
<tr>
<td>Laboratory Measures</td>
<td>Depends on study medication*</td>
</tr>
<tr>
<td>Current and Prior Treatments</td>
<td>Medication history (psychotropic)</td>
</tr>
<tr>
<td></td>
<td>Medication history (nonpsychotropic)</td>
</tr>
<tr>
<td></td>
<td>Current Educational Placement</td>
</tr>
<tr>
<td></td>
<td>Behavioral therapy</td>
</tr>
</tbody>
</table>

\*eg, antipsychotics and anticonvulsants require more detailed laboratory testing than stimulants.

PDD=pervasive developmental disorder.

In a study of very young children, or when they have, algorithms for diagnosis may be under-inclusive in children <4 years of age.\textsuperscript{14,15} In a study of very young children, diagnostic measures, such as the ADOS, that are appropriate for children <24 months of age mentally are recommended.

Except for the Social Responsiveness Scale (SRS),\textsuperscript{17} which has not been well studied in children with low IQs, most diagnostic and “core symptom” measures yield scores that are correlated either with IQ, mental age equivalents or both. If the purpose is to document behavior deficits in the sample, measures such as the Autism Behavior Checklist or the Gilliam Autism Rating Scale may be sufficient. However, if the intention is to ensure that the sample meets diagnostic criteria for autism, these measures are probably inadequate.\textsuperscript{18,19}

Similarly, the ADI-R gives the most comprehensive information, both historical and current, about the three domains associated with a diagnosis of autism. However, it is reliant on parent report and may, due to parental biases, result in both false positives and false negatives when compared to expert judgment.\textsuperscript{19} It requires considerable training even for experienced clinicians and is expensive because of the length of time it takes to administer (ie, at least 2 hours). How much this is a detriment again needs to be considered for a particular study. If the study treatment will eventually be used in a wide range of patients who may not all conform to diagnostic criteria, this is less of an issue than if the goal is to apply the treatment to autism specifically. If the goal of the study is to address the response of children with different forms of ASD, such as Asperger's syndrome or PDD-NOS, more complex diagnostic algorithms with clinical best estimate diagnoses may be required. Given the level of controversy about these diagnoses, such as study would do well to report interrater reliability and the reliability of the diagnoses over time. However, this level of diagnostic rigor may not be justified if the sample sizes are not sufficient to examine subgroup responses.

Intelligence scores are necessary for characterization. The level of intelligence functioning, may also moderate treatment response. For example, preliminary evidence suggests that children with lower IQs are less likely to show a positive response to methylphenidate.\textsuperscript{2} Ideally, the same IQ measures would be used for the entire sample, but this may not be possible. Another issue is that some measures such as the Wechsler tests of intelligence (Wechsler Abbreviated Scale of Intelligence,\textsuperscript{21} Wechsler Intelligence Scale for Children–Third Edition\textsuperscript{22}) provide more insight into broader areas of skill, but may not be useful.
in low functioning children. Instruments that can be used across a wider range of IQ include Raven’s Progressive Matrices, which measures visual-spatial nonverbal problem-solving and the Differential Abilities Scale (DAS). Which measures general intelligence in children with mental ages between 2 and 16 years of age. It also separates verbal from nonverbal skills more clearly than the Weschler Scales. The Mullen Scales of Early Learning, a developmental test that spans from infancy to 5 years of age and separates verbal from nonverbal tasks, may be especially useful in very young or very handicapped children. One way of handling the selection of IQ measures is to set up a hierarchy of tests. Given that not all subjects can complete the same tests, the hierarchy establishes the order by which the tests will be attempted. The hierarchy for preschool children might be the preschool DAS as the preferred measure, with the Mullen Scales used if the DAS is too difficult for the child. Once established, the hierarchy could also be used, in combination with age criteria, in studies that extract scores from existing datasets or records. Another approach is to use a measure, such as the DAS, the Raven’s Matrices, or the Vineland Adaptive Behavior Scales, that are applicable to almost all ages and then select another test that is more comprehensive for specific subgroups.

Although essential for characterizing a sample of children with developmental disorders, few clinical trials in autism have described subjects’ language level well. The characterization does not need to be a detailed. Indeed, for most studies, clarification of whether subjects communicate using more than a few words on a daily basis, may be sufficient. General measures, such as age equivalents on the Vineland Communication domain or the ADI-R Overall Level of Language code, may be adequate for children or adults with limited language, though neither of these scales provides much differentiation of oral language skills in older or higher-functioning individuals. Other measures, such as the Peabody Picture Vocabulary Test, a measure of receptive single-word vocabulary, or even more detailed measures of expressive or receptive language, may be used to provide additional information. In general, scores on various language tests in samples of children and adults with autism are highly correlated, so for the purposes of comparisons across samples, tests that yield a specific verbal score will often provide comparable information. However, because of the variability and discrepancies in performance scores and verbal measures, it is not possible to infer level of language for individuals with autism from nonverbal IQ scores. Thus, some direct description of language is essential.

In North America and an increasing number of countries, determination of mental retardation requires attention to impairment in daily living. Consequently, a measure of adaptive functioning, such as the Vineland Adaptive Behavior Scales, is an important part of sample description. The Vineland Scales also have the added benefit of a single metric across a wide age range.

As discussed earlier, medication studies in autism focus on core symptoms or specific behavioral problems—though the later has been the more common approach. There has been growing concern that, independent of ASD diagnosis, intellectual level, and adaptive functioning, psychiatric symptoms—may influence treatment related to outcome in autism. Thus, measurement of behavior, mood, and anxiety symptoms beyond the developmental disorders is worthy of consideration for sample selection and description. As with other measures, attention to the content and psychometric properties of the measures is warranted in the selection process. For example, the Abberant Behavior Checklist (ABC), which is a commonly used parent and teacher rating of behavioral problems, may be less useful for very young children.

MAGNITUDE OF CHANGE

Results from a clinical trial are typically expressed by reference to a P value, such as P = .05 or P = .01. These conventions for statistical significance only indicate that the difference between groups (eg, active drug and placebo) is unlikely to have occurred by chance. They do not indicate the magnitude of response. Stated differently, a P value of .01 does not imply greater improvement compared with a P value of .05. The difference between a P value of .01 and .05 reflects the level of confidence (99% versus 95%, respectively) that the finding did not occur by chance. The same magnitude of clinical improvement that is significant in a sample of 100 (50 active, 50 placebo) might fall short of statistical significance in a sample of 40 (20 active, 20 placebo).

The magnitude of treatment effect is typically expressed in reference to its impact on the primary outcome measure. Because this measure reflects the central hypothesis of the trial, the primary outcome measure forms the basis of statistical comparison between groups. In a practical sense, the primary outcome measure reflects the proposed match
between treatment and the symptoms of interest.

There are several ways to express the magnitude of response to a treatment in a clinical trial. The mean change in the primary outcome measure from baseline to endpoint in the active treatment group compared with mean change for placebo is a common metric. The mean change in active minus the change in placebo yields the net change score in the primary outcome measure attributable to the study medication. This method of expressing the magnitude of change has intuitive appeal, but requires familiarity with the scale to interpret the results. A mean change of 5 points may reflect clinically meaningful improvement on one measure, but virtually no change in another. In the RUPP Autism Network study of risperidone the change score was 14 points on the Irritability subscale of the ABC and 5 points on the Stereotypy subscale. This apparent difference in magnitude is due to the differences in the number of items in these scales, differences in the severity at baseline, as well as differences in treatment effect. Thus, reporting only change scores is often unsatisfactory for the practicing clinician. In a recent report by Kern and colleagues, the investigators evaluated whether secretin would be effective in a subgroup of children with PDD accompanied by chronic gastrointestinal (GI) problems. Nineteen children were roughly evenly divided by positive and negative history of GI problems were enrolled in the crossover study. No primary outcome measure was stated, but results focused on the various subscales of the ABC. Figures presented in the report suggest that secretin was more effective in the subgroup of subjects with a history of GI problems compared with those without GI problems. Unfortunately, mean scores and standard deviations at baseline were not reported making it difficult to interpret the clinical implications of the change scores.

Improvement can also be expressed as the percent change from baseline. This is computed by subtracting the endpoint score from baseline divided by the baseline score and multiplied by 100. The net change can be computed by subtracting the percent change in the placebo group from the percent change in the active group. Although the percent change enables easier comparison across studies and measures, neither the mean change nor the percent change take into account variability on the measure. The most common metric that incorporates variability is called effect size. Effect size brings variability into the equation by dividing the net change (change in active minus change in placebo) by the standard deviation of the measure observed in the trial. In other words, effect size expresses change from baseline on the primary outcome measure in standard deviation units. Effect sizes in the range of 0.5 to 0.7 are considered moderate; ≥0.8 is considered large. Data from the RUPP Autism Network trial can be used to illustrate approaches to expressing magnitude of change. As shown in Table 2, there is a 57% improvement on the Irritability subscale of the ABC. With correction for the 14% in the placebo group, the net change is 43%; the effect size is estimated at 1.5. By contrast, there was a 4.8 point (45%) change on Stereotypy subscale for an estimated effect size of 0.67. Because it incorporates variability in expression of change in a clinical trial, effect size also permits straightforward comparison of benefit across studies—even when different measures are used.

In the design phase, the estimated difference in the primary outcome measure in the active treatment compared to the control group is the basis for calculating sample size. There are several methods for calculating sample size and available computer programs and reference books may be helpful in this regard. Perhaps the most straightforward approach is to determine the minimum effect size that would be accepted as meaningful. This is essentially a clinical judgment with statistical implications. As suggested above, effect sizes between 0.5 (or one-half of a standard deviation unit on the outcome measure) and 0.7 is a moderate effect size. Effect sizes of 0.8–1.0 (improvement by a standard deviation) is considered large. Depending on the nature of the problem and the risks of the treatment, the answer to the question of a meaningful effect size may vary. Smaller effect sizes require larger samples to detect a difference between groups.

**INCLUSION AND EXCLUSION CRITERIA**

With the discussion of generalizability and effect size as background, the entry criteria (inclusion and exclusion) become the practical application of these overarching principles. Ideally, as noted above, the methods used to characterize subjects also serve to confirm the eligibility of subjects for the trial. The primary outcome measure should be clearly identified in the study design phase and used to estimate the sample size. Thus, clarity about the entry criteria is essential for the operation of the study and also bears on the scientific question to be addressed by the trial. A recent study by King and colleagues enrolled 39 subjects with PDD, 5–15 years of age, in a placebo-controlled study of amanta-
The entry criteria accepted children with high ratings on the ABC Irritability subscale or the ABC Hyperactivity subscale. If these two subscales are highly correlated, the either/or criterion will have little impact on the sample or the study results. On the other hand, if the subscales are moderately correlated or uncorrelated, the dual criterion is likely to enroll subjects with quite different characteristics. Given that these two subscales were derived by factor analysis, they are presumably measuring different constructs and only moderately correlated.\(^{39,40}\) Simply stated, subjects that are high on one rating may be low on the other. Thus, it is not surprising that the range of baseline scores were 3–38 for the Irritability subscale and 16–46 for the Hyperactivity subscale. The study showed no difference between amantadine and placebo on either outcome. An alternative approach in this situation might have been to test two separate hypotheses: amantadine versus placebo for Irritability and amantadine versus placebo for Hyperactivity. This approach implies either one larger trial or two separate trials.

In addition to the scientific importance of developing clear inclusion and exclusion criteria, clear elucidation of the entry criteria in published reports is essential for providing guidance to practicing clinicians. As noted above, treatment may be directed at specific symptoms (e.g., hyperactivity or aggression) or the core features of autism. In either case, in the design phase and in the eventual report of the study results, the primary outcome measure should be clearly stated. To avoid excessive variability on that symptom cluster, it is often wise to set a specific threshold on the primary outcome measure for entry into the study. If the threshold is set too low, two consequences are likely to follow. First, there may be “floor” effects in which it is difficult for the mean score to move much lower—even if the treatment is effective. Second, the resulting wide range will result in a large standard deviation. Recalling that standard deviation (i.e., variability) goes into the denominator in the calculation of effect size, accepting a wide range on primary outcome measure faces the fundamental problem of signal detection (clinical benefit) amidst too much noise (variability). Setting the threshold too high may reduce the signal to noise problem, but may make it difficult to recruit subjects. Furthermore, in addition to floor effects, there can also be ceiling effects. Here the concern is that the treatment may...

### TABLE 2. CHANGE ON IRRITABILITY AND STEREOTYPY SUBSCALES OF THE ABC IN THE RUPP AUTISM NETWORK STUDY OF RISPERIDONE IN CHILDREN WITH AUTISM AND SERIOUS BEHAVIORAL PROBLEMS

<table>
<thead>
<tr>
<th></th>
<th>Irritability Subscale</th>
<th>Stereotypy Subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risperidone</td>
<td>Placebo</td>
</tr>
<tr>
<td>Baseline Mean (SD)</td>
<td>26.2 (7.9)</td>
<td>25.5 (6.6)</td>
</tr>
<tr>
<td>Endpoint Mean (SD)</td>
<td>11.3 (7.4)</td>
<td>21.9 (9.5)</td>
</tr>
<tr>
<td>Mean Net Change</td>
<td>14.9</td>
<td>3.6</td>
</tr>
<tr>
<td>% Change</td>
<td>57%</td>
<td>14%</td>
</tr>
<tr>
<td>% Net Change</td>
<td>43%</td>
<td>45%</td>
</tr>
<tr>
<td>ES*</td>
<td>1.5</td>
<td>.67</td>
</tr>
<tr>
<td>P</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

|                  | Risperidone           | Placebo             |
| Baseline Mean (SD) | 10.6 (4.9)            | 9.0 (4.4)           |
| Endpoint Mean (SD) | 5.8 (4.6)             | 7.3 (4.8)           |
| Mean Net Change   | 4.8                   | 1.7                 |
| % Change          | —                     | 19%                 |
| % Net Change      | 45%                   | 26%                 |
| ES*               | —                     | .67                 |
| P                 | —                     | <.001               |

* Effect Size: Irritability: 14.9–3.6/7.3; for Stereotypy 4.8–1.7/4.65 (average of SD across groups at baseline). An even better estimate of SD (though more complicated) is the so-called pooled SD which is a weighted average of the SDs at baseline and endpoint. When pooled SD was used the ES was 1.2 for Irritability and 0.8 for Stereotypy subscales, respectively.

produce change that is not detected because the true score was artificially constrained by the maximum score at baseline. Table 3 shows the inclusion and exclusion criteria used in the RUPP Autism Network risperidone trial. These criteria include relatively high threshold for the primary outcome measure, homogeneity with respect to diagnosis, but accepts a wide range of age and intelligence.

The entry criterion of 18 on the Irritability subscale represents a score that is ~1.3 standard deviations (88th percentile [Michael G. Aman, PhD, written communication, November 2003] above the mean for this scale in developmentally disabled populations (the distribution of scores varies slightly by age and gender). This cut score assured at least moderate problems in the domain of aggression, tantrums and self-injury, but was not too restrictive. Setting the moderate score of 18 as the minimum score on the Irritability subscale all but guaranteed that the mean score would be >18. Indeed, as shown in Table 2, the mean score at baseline was in the mid twenties at baseline. Establishing the minimum score of 18 at baseline leaves room for improvement even for those subjects at the minimum entry score. Furthermore, it does not limit the study to only the most severe cases, but it does limit variability on the primary outcome measure.

**SELECTION OF THE PRIMARY OUTCOME MEASURE**

**Core Symptoms Versus Target Symptoms**

Among the first questions to be answered when organizing a clinical trial in children and adolescents with PDDs is to decide whether the treatment will be aimed at the core symptoms (delayed communication, repetitive behavior, impaired social interaction) or selected behavioral problems. To date, few medications have been shown to reduce the core symptoms of ASD in controlled studies. One drawback of directing a treatment at core symptoms is that most measures that assess core symptoms were developed for diagnostic purposes. Measures built for classification incorporate history as well as current symptoms. The inclusion of historical features with the current diagnostic features limits the capacity of these instruments to show change. This difficulty could be avoided by developing measures of “current” functioning or by using observational measures that are capable of detecting change. Many medications may be more effective at modifying relatively non-specific behaviors (eg, activity level, irritability) and through these effects may also change core aspects of ASD, such as communication and social behaviors. Tracking the particular effects of a medication on individuals with ASD, even for non-core symptoms, is worthy of study because of the varied trajectories and qualitatively different patterns of behavior seen in ASD.

**Types of Outcome Measures**

Following the decision about the focus of the clinical trial (ie, autism as a syndrome or a more specific symptom cluster), the study must settle on a primary outcome measure. Several factors enter into the selection of the primary outcome measure including the type of measure, informant, subject burden, rater training, and, of course, availability of an appropriate measure. In the best of all circumstances, the measure would be simple to administer and rated by a clinician based on a focused, semi-structured interview.

In the field of developmental disabilities, valid, reliable, clinician-rated measures that are sensitive to change are few in number. The chief advantage of clinician-rated measures is that they make use of an expert to sort through the report of the parent

<table>
<thead>
<tr>
<th>TABLE 3. INCLUSION AND EXCLUSION CRITERIA IN THE RUPP AUTISM STUDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion criteria</td>
</tr>
<tr>
<td>Autism</td>
</tr>
<tr>
<td>5–17 years of age</td>
</tr>
<tr>
<td>Irritability subscale score &gt;18</td>
</tr>
<tr>
<td>CGI-Severity &gt;4</td>
</tr>
<tr>
<td>Mental age &gt;18 months</td>
</tr>
<tr>
<td>Medication free (14–28 days depending on drug, except anticonvulsants for seizure disorder)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of any other form of PDD</td>
</tr>
<tr>
<td>Presence of another psychiatric disorder in need treatment</td>
</tr>
<tr>
<td>Weight &lt;15 kg</td>
</tr>
<tr>
<td>Presence of a serious medical condition</td>
</tr>
<tr>
<td>Unstable seizure disorder (seizure within the last 6 months, need for medication change in the last month)</td>
</tr>
<tr>
<td>Prior adequate trial with risperidone (1 mg for at least 2 weeks)</td>
</tr>
</tbody>
</table>

RUPP=Research Units on Pediatric Psychopharmacology; CGI=Severity=Clinical Global Impressions-Severity Scale; PDD=pervasive developmental disorder.

and/or the child. The chief disadvantage is that clinician-rated measures may require meticulous training to achieve reliability and are relatively expensive because they are usually administered by an expert clinician. To advance clinical research in autism and related conditions practical, reliable, and valid clinician-rated instruments are needed.

Another approach to evaluating treatment effects relies on informant-based measures. In children, the most common informants are parents and teachers. Informant-based measures have been used effectively in stimulant trials in ADHD for over 30 years. They have also been used in clinical studies in children with developmental disabilities. The chief advantages of informant-based measures are that they are inexpensive and provide a relatively easy way to collect data from multiple informants. Depending on the measure, population data may be available as a way to pinpoint the severity threshold at baseline and to estimate a meaningful and plausible change over time. The primary disadvantage of parent and teacher measures is that the informant may not understand the language or the intent of items on the rating. These measures are also vulnerable to expectations on the part of the informant, which may contribute to placebo effects in some trials.

Several studies, particularly behavioral interventions and recent studies of secretin have also employed observational measures. These measures may be highly individualized, such as counts of specific behaviors, severity checklists done during an office visit, or coded observations during a semi-structured interaction, such as the ADOS. The purported advantage of observational measures is that they are objective and less vulnerable to misunderstanding or filtering effects of informant-based and ratings based on clinician interview.

There are several disadvantages to observational measures. Many involve the coding of videotaped sessions. Therefore, the rating is only as good as quality of videotape. Moreover, subsequent coding of videotapes is a time-consuming process. When the time required to code the tapes is added to the time required to collect the data, videotaped observational measures become expensive. Although coding observations in real time is less expensive than videotaping and later coding, it may be difficult to achieve reliability across raters in a multisite study. Even with measures such as the ADOS, which has established reliability, it is not yet known whether it is sensitive to change (Table 4). Given that an observation session is a small slice of time with a potential for practice effects, validity is an important empirical question that needs to be addressed.

Global measures, such as the Clinical Global Impression for Severity (CGI-S) and Improvement (CGI-I), have also been used as an outcome measures in recent studies in autism. These measures have stood the test of time, but surprisingly little work has been done on the reliability and validity performance-based measures. The CGI-S can be incorporated into the inclusion criteria as a part of establishing the overall severity of the sample. The CGI-I is, by definition, a measure of change ranging from very much improved (score of 1) through no change (score of 4) and very much worse (score of 7). By convention, scores of 1 and 2 (very much or much improved, respectively) are used to define responders and subjects with scores of 3–7 are regarded as nonresponders. In a placebo-controlled study, the CGI-I can make a simple and powerful statement about the rate of positive response in the active treatment group compared with the placebo. For example, in the RUPP Autism Network study, 75% of subjects in the risperidone group were rated as responders compared with 11% in the placebo group. On the other hand, the behavioral domain of improvement on the CGI-I may not always be clear—leaving practicing clinicians uncertain about which symptoms actually improved in the study and what to expect from the treatment. Planning for secondary analyses that spotlight specific areas of improvement may be helpful in this regard.

TABLE 4. EFFECT SIZE OBSERVED IN SECRETIN TRIAL ON THE IRRITABILITY SUBSCALE OF THE ABC AND COMBINED SOCIAL AND COMMUNICATION SCORE OF THE ADOS

<table>
<thead>
<tr>
<th>Measure</th>
<th>Secretin Baseline Mean (SD)</th>
<th>Secretin Endpoint Mean (SD)</th>
<th>Placebo Baseline Mean (SD)</th>
<th>Placebo Endpoint Mean (SD)</th>
<th>ES*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Interaction</td>
<td>18.7 (3.7)</td>
<td>18.1 (2.5)</td>
<td>17.3 (3.5)</td>
<td>16.4 (3.5)</td>
<td>.03</td>
</tr>
<tr>
<td>Irritability</td>
<td>11.6 (7.5)</td>
<td>10.1 (10.2)</td>
<td>10.1 (7.1)</td>
<td>10.9 (8.1)</td>
<td>.3</td>
</tr>
</tbody>
</table>

* Effect size: (change in active-change in placebo/average SD at baseline across groups.

ABC=Aberrant Behavior Checklist, ADOS=Autism Diagnostic Observational Scale; ES=effect size.

CONCLUSION

The goal of clinical research in autism is to provide guidance to clinicians and families on the risks and benefits of a given treatment. Subject selection and subject characterization are fundamental features of a well-designed study because they inform clinicians and families who was in the trial and the type of patients that might benefit from the treatment. Subject selection is intimately tied to the study hypothesis and the choice of the primary outcome measure. The primary outcome measure is a critical choice in study design as is setting the threshold for study entry. Failure to set a threshold or setting the threshold too low will almost certainly increase the variability in the sample and may make it difficult to detect a treatment effect. Whenever possible, a dimensional, clinician-rated measure is preferred, but this is an underdeveloped area in autism clinical research. An important related concern is the expected magnitude of improvement on the primary outcome measure. The magnitude of improvement, which can be expressed by effect size, has direct implications for sample size. In the design phase, investigators first estimate the minimum clinically meaningful effect size. The sample size then becomes the number of subjects needed to detect that difference between groups. To ensure that a study can provide guidance for the clinical community, research reports should describe the characteristics of the study sample, track who did and did not enter the trial, and provide the necessary raw data to calculate effect size. These components of a study report also permit comparison with past and future studies.

REFERENCES

Impact of Recent Findings on Study Design of Future Autism Clinical Trials

By Eric Hollander, MD, Ann Phillips, PhD, Bryan H. King, MD, Donald Guthrie, PhD, Michael G. Aman, PhD, Paul Law, MD, Thomas Owley, MD, and Ricki Robinson, MD, MPH

FOCUS POINTS
- There are many challenges in the study design of pharmacologic trials in autism.
- To discuss the need to stratify the autism population for specific symptom domains in ongoing and future clinical trials.
- To describe recent trials with various agents on specified symptom domains in autism and to illustrate stratification strategies.

ABSTRACT
There are specific challenges to studying the design of pharmacologic trials in child/adolescent and adult autism, such as subject stratification and parallel versus crossover designs. This article describes how optimal study design is influenced by subject selection and outcome measures chosen. Lessons learned in study design from the Research Units on Pediatric Psychopharmacology Autism Network trial with risperidone, Seaver Center trials with fluoxetine and valproate, Dartmouth trials with amantadine, and National Institutes of Health secretin trials are highlighted. The Internet System for Assessing Autistic Children system for managing multicenter clinical trials in autism and statistical issues in autism research are also described.

CNS Spectr 2004;9(1):49-56

INTRODUCTION
Autism is a pervasive neurodevelopmental disorder characterized by social and communicative deficits and compulsive/repetitive behaviors. It has a major impact on the lives of individuals with the disorder, their families, and society. While there is a great need to find medications that can be used to help with core symptoms and/or to help manage secondary symptoms, to date there are no medication treatments approved by the Food and Drug Administration for an indication in autism, and no proven efficacious pharmacologic treatment for autism, (with the exception of risperidone for treating aggression in the subgroup of autistic children with aggressive behavioral problems). There are a number of medications being routinely prescribed for children with autism for which there is inadequate evidence of efficacy. Studies of potentially useful medications are needed, and they must be carried out using methods that yield the most information for both research and clinical practice and have the greatest scientific validity.

Scientifically ideal study designs are limited by practical factors, including but not limited to level of funding and sample size. Optimal study design must take into account subject selection (ie, age,
intelligence quotient, language function, comorbidity, concomitant medications) and targets for outcome measures (ie, global response, core symptoms, target domains). In addition, there are ethical concerns, such as whether or not to take children off current medications and whether or not to include placebo-control groups. There are also scientific validity concerns, such as self-selection of participants, such that families of children with more severe forms of the disorder or more disruptive symptomatology may conceivably be more willing to participate in pharmacologic trials than families of less symptomatic children. These factors must be balanced against one another in order to design studies that can offer scientific and clinical insight, maximize subject benefit, and reducing negative impact on the subject population. Recent improvements in study design, including the use of double-blind, placebo controls, the use of more focused diagnostic groups stratified for specific symptoms, and the use of reliable and validated outcome measures, have resulted in improved methodology in the study of medications in the treatment of developmental disorders.

Remaining challenges include identifying target symptoms, defining inclusion criteria that are pertinent to clinical practice, developing better outcome measures, and creating study designs that evaluate both short- and long-term effects. Below we discuss these issues in more detail and outline some of the ways in which these concerns have been met in recent autism research. There is still the need, however, for creative solutions to many of these issues.

**Length of Study, Dosage, and Titration**

In determining the length of a study, there are several considerations, which will partially depend upon the nature of the drug being studied. Medication must typically be started at lower than recommended dose in this population in order to avoid adverse effects. The half-life of the study medication (among other factors) will influence how long it takes to reach a steady-state, which will then influence optimal length of study. Another issue is the optimal number of visits; more frequent visits result in the more finely tuned and sensitive data but place more demand on subjects. This can result in a higher rate of non-compliance and greater drop-out rate. Also, some medications have a therapeutic lag-time before onset of action (eg, selective serotonin reuptake inhibitors [SSRIs]), whereas others have an immediate effect (eg, stimulants). This influences optimal length of study. In addition, it is recommended that subjects remain on a stable dose of medication for a period of time (ie, 4 weeks), although dosage may still require adjustment over time to get to a therapeutic dose.

Dosage can either be linked to subject's weight, be standardized, or be titrated. If dosage titration is too slow, while there may be greater tolerability, the patient will have a shorter time on an effective dose, or the study length will need to be increased. A more gradual titration will result in less drop-out due to adverse events, but could also result in a greater drop-out due to lack of efficacy. To get a good sense of whether or not a drug is effective, subjects often need to be at the optimal dose and steady-state for at least 4 weeks, except for stimulants.

Optimizing therapeutic response is also an issue that needs to be addressed. Some medications lend themselves well to forced titration, with side effects limiting further increases dosage. Others have defined therapeutic windows. Most medications, however, display pharmacokinetic and pharmacodynamic variability, suggesting we should also be investigating the effect of liver enzyme systems and variability in the genes regulating the production of the proteins responsible for molecular targets of the medication.

**Blinding**

It is especially necessary for treating and evaluating study clinicians to be blind to group placement (ie, drug or placebo) of the study subjects. This is because inadvertent differential treatment by clinicians can influence the outcome of the study and because knowledge of study group can influence clinical assessments. Because knowledge of side effects can reveal group membership, it is optimal to have a second clinical evaluator, (who has no access to side-effect information), make clinical assessments. These evaluations may either be compared with primary evaluations to make sure that there is no bias, or be used as primary outcome measures for data analyses.

**Parallel Versus Crossover Design**

In a crossover design, all subjects receive both placebo and study drug in separate phases of the study. There are several advantages to this design. Subjects act as their own controls, which reduces error because random differences between subjects is automatically controlled, substantially increasing the power of statistical analyses. Another advantage is that experimenters can guarantee that all subjects will get medication in one phase of the trial, which aids enrollment and provides useful clinical data for eventual treatment. Drawbacks with this design are that there can be carryover, such that subjects who move from medication to placebo may still be
experiencing benefits of medication after discontinuation. Alternatively, subjects may experience discontinuation effects in coming off the medication in Phase I that adversely influence ratings early on in Phase II. Adding a washout phase helps eliminate carryover. However, it adds to overall length of the study and increases the proportion of the time that subjects do not receive medication. There are also concerns about how to interpret order effects. The experience of medication and placebo may be different for subjects depending on which treatment they receive first. The parallel design is simpler to interpret, and can be followed by open-label treatment for placebo nonresponders, so that everyone who might benefit is offered the opportunity for treatment, but without order and carryover effect of the crossover design. However, the open-label data cannot be included along with the double-blind data for analysis, and is less informative.

**Placebo Response**

Placebo controls are appropriate for autism research because there are no established medications for this disorder. Therefore, it is ethical to keep a group on placebo for the duration of a study. Placebo controls are also important in this population due to a significant placebo response that has been found across a number of studies. The behavior of individuals with autism varies over time, and often improves with time. These factors can only be separated from the effects of the treatment medication by the inclusion of a placebo-control group. The Table summarizes recent studies that have highlighted the placebo response in autism clinical trials. These range from 12% to 37%, and are clearly influenced by subject selection, outcome measures, study design issues, and claims made for the study medication. The use of a placebo control or other control group is essential in order to control for these effects. In order to attempt to minimize placebo effects, placebo run-ins have been used for several weeks before beginning the medication trial, however they have not in fact proven to decrease placebo response. Another option is to have the study duration be long enough for placebo effects to run their course, minimizing the likelihood of significant numbers of placebo responders.

**Subject Selection**

Intervention studies generally employ common diagnostic tools, primarily the Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule-Generic (ADOS-G), which give subscores for the three core diagnostic features of autism. In addition to the core symptom domains, autistic patients are often impaired by comorbid conditions (ie, seizures, mental retardation) and associated features (eg, impulsivity/aggression, affective instability). Subjects may also vary on level of speech and language or other developmental domains. Thus treatment studies must also take into account these comorbid conditions and associated features, which clearly influence treatment response and side effects.

**Core Versus Associated Features**

While difficulties with social interaction, and speech/communication, and the presence of compulsive/repetitive behaviors are the core features of autism, other associated problem behaviors occur within the autistic syndrome (eg, attentional difficulties, self-injurious behavior, mental retardation, self-stimulation, affective instability, seizures). These symptoms themselves cut across a number of disorders. Thus, it is possible we will make more progress in understanding autism by investigating commonalities between core symptoms in autism and similar symptoms in other disorders. This would involve investigating, for instance, whether they have distinct neurobiological/genetic mechanisms, and whether individuals with different levels of symptom severity will have differential treatment response. A number of studies of autism have benefited from

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>N</th>
<th>Intervention</th>
<th>Measure</th>
<th>% Placebo Responders</th>
<th>% Drug Responders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sandler et al</td>
<td>1999</td>
<td>25</td>
<td>Secretin</td>
<td>CGI</td>
<td>28</td>
<td>33</td>
</tr>
<tr>
<td>Dunn-Geier et al</td>
<td>2000</td>
<td>48</td>
<td>Secretin</td>
<td>PLS-3 (&gt;4 pts)</td>
<td>21</td>
<td>34</td>
</tr>
<tr>
<td>Buitelaar et al</td>
<td>1998</td>
<td>18</td>
<td>ORG 266</td>
<td>ABC (withdrawal)</td>
<td>22</td>
<td>34</td>
</tr>
<tr>
<td>King et al</td>
<td>2001</td>
<td>19</td>
<td>Amantadine</td>
<td>ABC (hyper/irrit)</td>
<td>37</td>
<td>47</td>
</tr>
<tr>
<td>King et al</td>
<td>2001</td>
<td>19</td>
<td>Amantadine</td>
<td>CGI</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>McCraken et al</td>
<td>2002</td>
<td>52</td>
<td>Risperidone</td>
<td>ABC (irrit)</td>
<td>12</td>
<td>69</td>
</tr>
</tbody>
</table>

CGI=Clinical Global Impressions Scale; PLS-3=The Preschool Language Scale, Third Edition; pts=points; ORG 266=an andrenocorticotropic hormone-(4-9) analog; ABC=Aberant Behavior Checklist; hyper/irrit=hyperactivity/irritability subscales.

restricting the population studies based on core symptom domains. For example, genetic studies have applied this approach to identify potential chromosomal regions that may contribute to autism.8,9 This approach to subject selection was recommended by the Autism Clinical Trials Task Force Meeting held in April 2002.

Treatments that address one symptom domain may also modulate other symptom domains either directly or indirectly. SSRIs, which improve repetitive behaviors, may also indirectly improve social deficits because patients less impaired by severe compulsiveness, craving for sameness, and rigid adherence to routines, may have greater opportunities for improving social functioning.

Several investigators have highlighted the fact that their sample sizes were either too heterogeneous or too small to allow for the study of the effects of an intervention among subgroups of children with autism.5-7,10,11 Here, there is a trade-off between generalizability (by making the target group too narrowly specified, you lose the ability to generalize the results to the diagnostic group in general) versus including too many subgroups, losing an effect for group, while still having too small a sample to analyze subgroups individually. Multicenter collaborations allow for the increase in power necessary to allow for these kinds of analyses.

One strategy for dealing with these issues is to start with a focal subgroup for whom there is preliminary evidence and/or theoretical justification for the trial. Results from such a trial can then be used to guide further research in order to test for the generalizability of the results. On a cautionary note, it is important to aim to strike a balance with respect to symptom severity at the time of enrollment such that response to treatment is not assured merely as a regression toward behavioral means. It may also be helpful to limit the cohort to a relatively narrow age group as developmental effects may influence treatment response as suggested by Eckman and colleagues.12 This latter issue may be particularly important for some interventions, such as serotonergic drugs, given the evidence cited above that suggests serotonergic measures change with maturation.

**Outcome Measures**

Outcome measures for intervention studies usually rely on the subjective assessment of parents and/or clinicians. No instrument currently in use, with the possible exception of the most global of measures of change, will capture all aspects of improvement that are meaningful to children and their families. Little is known about the factors that most influence observer ratings in drug studies of children with autism. One approach to this problem is to cast a broad net, anchored to clinically meaningful change, and supported by attempts to identify areas of specific change through more specific measures. Whenever possible, researchers should define target behaviors that they theorize will be affected by the medication being tested. The development of improved outcome assessment tools will expand our ability to design studies that can speak to clinically relevant questions.

**Data Management**

Data management for multisite, multiple-investigator studies requires uniform coding schemes and data structures. Studies involving more than two sites should appoint a single data monitor, often called a “gatekeeper,” who will have authority and responsibility for data integrity. The structure of the recorded data should be kept as simple as is consistent with ease of access. Relational databases consist of separate tables for each measurement field, while large rectangular files contain all data items for a subject on a single record. Investigators involved in multisite studies will have to consider whether to have data centrally located or resident at each of the sites.

**Statistical Analysis**

The principle behind the concept of intent-to-treat is that data can become biased and, therefore, misleading due to subject self-selection. Subjects may differentially drop-out of a study because of nonresponse, side effects, or may self-select based on severity of the disorder or of secondary symptoms, and participants may do so differentially between placebo and medication groups. An intent-to-treat analysis attempts as far as possible to include all subjects who have begun the study. On the other hand, since subjects do end their participation, there is the problem of how to deal with missing data. If missing data are not included, the intent-to-treat principle is not obeyed.

There are two ways to compensate for this bias. One is to carry the last observation forward. This is a conservative approach, such that the chance of finding an effect is decreased proportional to the amount of missing data. Another approach is to use a mixed model, which looks at both individual and group data. This random regression model will estimate change values given at least two data points for an individual subject. This minimizes bias while taking advantage of as much information inherent in the data as possible to make the analyses as powerful as possible. Another advantage of mixed models is that
they have the ability to compensate for uneven time spacing. Although longitudinal study designs assume the even spacing of data points across time, because of less than perfect subject compliance, it is difficult to completely avoid some data that is unevenly spaced. Therefore, compensation for uneven spacing is another advantage of the mixed model.

**Statistical Power**

Power refers to the ability of a design to detect effects that may be present. Based on the “Null Hypothesis Testing” framework of statistical inference, power is simply the probability that a statistically significant result will be found. This probability depends on a variety of factors, including for example the nature of random variations (errors), the relationship among measured variables, and the impact of an experimental intervention. These factors can be difficult to estimate, however, for drugs that have not been previously tested. Since there are so few studies conducted in autism to date, it may be necessary to make these assumptions based on pervious studies with other populations. With simplifying assumptions it is possible to obtain reasonable estimates of power and apply those to the experimental design. The usual result of a power analysis is to determine the sample size needed to address the research question. In the context of experimental design, power is the index quantifying the trade-off among risk of false findings, likelihood of detecting a real effect, and cost of the study as reflected in the size of the sample. A power analysis should be based on assumptions about how big an effect is scientifically expected or clinically interesting. The rate of clinically significant improvement may be lower for individuals with autism than for other groups. These assumptions are then balanced against the needed sample size and the risk of failing to detect an effect if it is present. See Hintze for a computer program that will do power analysis.

The following sections illustrate how study design issues have been dealt with studies conducted by the authors. These examples illustrate the principles of study design implementation discussed above.

**RECENT TRIALS IMPACTING ON STUDY DESIGN**

This section reviews some specific clinical trials in autism that utilize very different agents (atypical antipsychotics, SSRIs, anticonvulsants, dopamine agonists, and secretin). These trials were presented at the Autism Clinical Trials Taskforce Meeting and are used as examples to illustrate different study design issues. For example, parallel versus crossover design, stratification issues, dosing issues, length of trial and blinding issues are highlighted. The initial Research Units on Pediatric Psychopharmacology (RUPP) trial with risperidone is described in detail to illustrate several design points, and the other trials are described more briefly but highlight specific design issues as well.

**Research Units on Pediatric Psychopharmacology Trial**

**Atypicals for Disruptive Behavior**

This study found that children with the target symptom of disruptive behavior showed improvement in those behaviors when treated with risperidone, but not placebo. Inclusion criteria stratified for subjects on the high end of the spectrum for disruptive behaviors. Defining inclusion criteria in this way was considered most relevant to clinical practice. This multisite study included an 8-week, double-blind, placebo-controlled trial of risperidone, followed by open-label treatment for 4 months and then a 2-month discontinuation. This design allowed researchers to assess both the short-term and long-term effects of the medication. Phase I was an 8-week, double-blind, randomized, parallel-groups comparison of risperidone and placebo. Subjects assigned to placebo who did not improve were offered an 8-week open-label trial of risperidone. This helped in subject recruitment and provided some additional data, although the data could not be merged with double-blinded data. Phase II included a 4-month open-label treatment with risperidone for patients who showed improvement in the 8-week acute trial. Phase III was a randomized, double-blind, placebo-controlled discontinuation study. Open-label treatment was offered to children in the placebo group who did not show improvement in their behavior, and children in the risperidone group who had a positive response to treatment.

According to the intent-to-treat model, every subject who was randomized was included in the study. The study used a mixed-model analysis. Children were selected based on a narrow diagnosis of autism (rather than including children with pervasive developmental disorder or Asperger's syndrome), as well as the presence of serious behavioral disturbances with tantrums, aggression and/or self-injury. Children were included if they had a score of ≥18 on the Irritability subscale of the Aberrant Behavior Checklist (ABC), (1.3 to 1.5 SD above the mean of the population of developmentally disabled children). In order to compensate for a perceived lack of variability of core symptoms of autism in the Clinical Global Impressions (CGI) score, a CGI score of 3 was assigned for autism symptoms alone, and higher scores were based on
comorbid symptoms (ie, aggression, self-injury).

Medication schedule was based partially based on weight. Children weighing at least 20 kg began with a dose of 0.5 mg, which was increased to 1 mg/day. After that, the dose was increased at the discretion of the treating clinician, to no >2.5 mg/day. Children weighing ≥45 kg or more were increased to no >3.5 mg/day, with no increases in dosage for any children after day 29. Dosage could be lowered due to adverse effects, subject again to clinical judgment.

In order to maintain at least one blind clinician who was not exposed to side-effect information, a clinical evaluator who did not get information about side effects or dosage saw the subjects. The decision about whether subjects were responders or nonresponders was made while both clinicians were still blind.

Positive treatment response was defined by a decrease of 25% on the Irritability subscale of the ABC and a rating of Much Improved or Very Much Improved on the Global Improvement item on the CGI at endpoint. Relapse was defined as the return of target symptoms—a ≥25% increase on the Irritability subscale of the ABC and two consecutive ratings of “much worse” or “very much worse” on the Global Improvement item of the CGI by the blinded clinical evaluator.

This study documented significant improvement in disruptive behavior with risperidone. Advances in study design included the use of a blind, independent evaluator, dosage titration based on weight, the use of maintenance and discontinuation phases to evaluate long-term efficacy and side effects, the use of intent-to-treat analysis that included in the database all subjects randomized, and especially the use of a stratification strategy to include subjects with a high score on a particular target symptom at baseline.

Seaver and New York Autism Center of Excellence Trials

Anticonvulsants for Impulsive/Aggression and Affective Instability

A recent study and an ongoing study are described that illustrate different study designs. The first study assessed the effect of divalproex sodium on aggressive behaviors (both self- and other-directed) and impulsivity. Valproate has been used in the past for patients with epilepsy, migraine, and manic episodes associated with bipolar disorder. As a mood stabilizer, divalproex sodium may be an effective treatment for the affective instability seen accompanying autistic disorders, and as an anticonvulsant, valproate might have a favorable impact on electroencephalographic (EEG) abnormalities often seen in autism.

In the preliminary pilot study,14 divalproex sodium was effective in treating core dimensions and associated features of autism. Of the 14 patients in the study, 12 had comorbid diagnoses, including anxiety disorders (n=7), impulsive control disorders (n=6), mood disorders (n=4), attention-deficit/hyperactivity disorder (n=3) and psychotic disorder (n=1). Ten (71%) were rated as sustained-treatment responders with final CGI-Improvement scores of “much improved” (n=2) to “very much improved” (n=1). Overall, the mean CGI-Improvement score was 2.36±1.48 (much improved). Six patients manifested improvement in affective instability, five patients became less impulsive, and four patients were less aggressive. Responder and nonresponder groups did not marginally or significantly differ in endpoint divalproex sodium dose (887.50±619.28 mg/day versus 468.75±386.96 mg/day) or in maximum valproate levels (75.61±13.43 mcg/mL versus 76.50±12.02 mcg/mL).

Our ongoing study, funded by the National Institute of Neurological Disorders and Stroke, uses a double-blind, placebo-controlled, parallel 12-week design. A longer length of treatment (12 weeks) allows for medication titration and a longer period on a therapeutic and stable dosage of valproate. Inclusion criteria were used to narrow subject characteristics and establish a more homogeneous group. The Overt Aggression Scale-Modified (OAS-M)15 and the ABC-Community Version are used to assess aggression. The following cut-off criteria are employed: OAS-M ≥13 or ABC-Irritability subscale ≥18 (raw scores). Subjects with stable seizure disorder or EEG abnormalities are included. An independent evaluator is utilized to assess outcome independent of side effect assessment. Dosage is titrated based on therapeutic effect, blood levels, and side effect data utilizing a third independent blood level checker. The outcome measure for evaluating aggressive behavior is the OAS-M. To assess for change in affective instability, two scales are used: the Young Mania Rating Scale16 and the Children’s Depression Rating Scale.17

This is the first placebo-controlled, parallel-design trial of an anticonvulsant in autism, and investigates the impact of valproate on impulsive aggression, affective instability, and impact of history of seizures and EEG abnormalities on treatment response.

Fluoxetine in Children and Adults

Early studies with fluoxetine, in both adult and child subjects, were limited by several factors. First, the trials included a heterogeneous population consisting of subjects over a wide age range with diagnoses ranging from high-functioning autism and...
Asperger's syndrome to low-functioning patients with profound mental retardation and other comorbid symptoms. Subjects were not selected for severity or for specific symptoms (ie, higher level of repetitive behavior). Early studies employed a short-term crossover design (8 weeks for each phase), which may have prevented subjects from reaching steady-state levels of fluoxetine, may have been complicated by carry-over or washout effects, and limited response detection. Finally, this study did not evaluate for treatment relapse or symptom regression during maintenance and following discontinuation.

Because of these limitations, a follow-up study that addresses many of these limitations and addresses important gaps in our knowledge is currently being conducted, funded by the Orphan Products Division of the FDA. Notable design changes include the addition of a 4 week washout phase in between the two 8-week crossover phases and a change in the age range which limits subjects to children and adolescents (between 5 and 17 years of age). We have also added a series of new outcome measures to evaluate the severity of symptoms in a range of specific symptom domains. We have included the use of an independent examiner to assess outcome blind to side effect information. We have replaced cumbersome coding of videotape with parent and clinician ratings of more targeted primary and secondary outcome measures. This trial utilizes a liquid form of fluoxetine to initiate at very low doses to minimize early activation. The titration is influenced by the subject's weight.

**Dartmouth Trials with Dopaminergic Agonists for Hyperactivity**

This multisite study utilized the dopamine agonist amantadine versus placebo in 40 children and adolescents with autism 5–15 years of age. It used inclusion criteria that stratified for irritability and hyperactivity by including only those subjects at or above the top 25th percentile on the ABC. The study design was randomized, double-blind, parallel-group, placebo-controlled, with a 1-week placebo run-in and a 4-week treatment phase. The short-term trial was appropriate given the rapid onset of the dopamine agonist. Dosage was titrated up on a dosage adjusted for weight. Outcome measures were the ABC and CGI change score as rated by clinicians and parents. This study found a positive effect of amantadine hydrochloride for the ratings of clinicians, but not for parents. A significant percentage of the parents chose to keep their children on the medication, indicating that they found it to be beneficial, but that the ratings scales were not able to capture the relevant dimensions of improvement.

**Design of the National Institutes of Health Secretin Trial**

The objective of this multisite study was to examine the efficacy of intravenous porcine secretin for the treatment of autistic disorder. This study used a randomized, double-blind, placebo-controlled, crossover design. Subjects on stable doses of other psychotropic medication were included and asked not to change dose during the study or 8 weeks before the study. Fifty-six children between 3 and 12 years of age with autism received either a secretin or placebo infusion at baseline and the other substance at week 4. The primary outcome measure was the ADOS-G, specifically the social-communication subscore, which was hypothesized to capture the dimension of autism that might have been responsive to secretin in case reports. The Gilliam Autism Rating Scale (GARS) was also completed by the parents or caregivers at 2-week intervals. At week 4 there was no change in the measure of social-communication score from week 0 to 4, and there was no statistically significant difference between secretin and placebo. No significant differences were found with the GARS. Because the study ended after 4 weeks, it is possible that effects could be found later on, however, case reports had found improvement shortly after infusion. The results of this study highlight the need for controlled trials. In this case a medication for which there was only anecdotal evidence for treatment efficacy turned out not to be efficacious when tested in a more rigorous manner. This study tested the impact of a single dose of secretin versus placebo, and additional evaluation of the impact of subgroups (ie, gastrointestinal status) and treatment response was not undertaken.

**THE INTERNET SYSTEM FOR ASSESSING AUTISTIC CHILDREN**

Due to the increasing need for autism research and the number of collaborative efforts increasingly involved in treatment efficacy studies, the need was seen for a consistent database management system that can be shared. The Internet System for Assessing Autistic Children database addresses this need, and is dedicated to aiding researchers in the development of effective treatments and a cure for autism. The Internet System for Assessing Autistic Children's Web-based services promote collaboration among researchers while increasing efficiencies and eliminating redundancies in the collection, management, and analysis of autism research data. The purpose of
this system is to provide an easy-to-use system that eliminates or reduces the data management concerns that typically face health researchers. In addition, the hope is that it will lead to a Web-based community for sharing and discussing autism research data and analysis. Another aspiration is to foster collaboration and to reduce redundancies in autism research.

The development of this system was envisioned and funded by Cure Autism Now, and the web-based application was used and tested by AGRE during its transition from hard-coded forms to the database of questions and answers. It is now available to all autism researchers and provides a database of questions and answers, rigorous security, data entry, data sharing at multiple levels, and a flexible and adaptable design. The following forms are available for entry and download (including agreements with publishers): ADOS-G, ADI-R, CARS, AGRE family medical history, and Raven’s Progressive Matrixes. In addition, there is automated scoring on ADOS-G, ADI-R, and the Raven Progressive Matrixes. The Vineland Adaptive Behavior Scales, Peabody Picture Vocabulary Test, Version III, and ABC will be added (pending publisher’s approval).

CONCLUSION

This article reflects some of the presentations and discussion regarding study design of autism clinical trials presented at the Autism Clinical Trials Task Force Meeting. Optimal study design needs to take into account multiple factors, including, but not limited to level of funding, sample size, subject selection (ie, age, IQ, level of language functioning, comorbidity, and concomitant medications) and targets for outcome measures (ie, global severity, target symptoms, or development over time). Several specific trials conducted in autism are highlighted, and study design features contrasted. This addresses what we know, what we do not know, and studies needed to address these gaps and help obtain this missing knowledge.

REFERENCES

Autism Clinical Trials: Biological and Medical Issues in Patient Selection and Treatment Response

By George M. Anderson, PhD, Andrew W. Zimmerman, MD, Natacha Akshoomoff, PhD, and Diane C. Chugani, PhD

FOCUS POINTS
- Biomedical measures recommended for inclusion in the selection and screening phase of an autism clinical trial are discussed.
- There are several biomedical measures of potential utility in understanding drug response and underlying pathophysiology in patients with autism.
- Possible approaches to examining genetic influences on drug response in autism clinical trials are considered.

ABSTRACT
Biomedical measures are critical in the initial patient-screening and selection phases of a clinical trial in autism and related disorders. These measures can also play an important role in the assessment and characterization of response. In addition, clinical trials offer an opportunity to employ research-oriented measures in order to study underlying etiologic and pathophysiologic processes. In the patient-selection phase, biomedical measures are used to screen for potential safety-related problems, to decrease biological and genetic heterogeneity, and to define subgroups. Neurobiological research measures can be employed to study mechanisms and extent of drug action, to perform baseline investigations of possible pathophysiologic alterations and to carry out longitudinal studies in the active agent and placebo-treated groups.

CNS Spectr 2004;9(1):57-64

INTRODUCTION
Biomedical measures are critical in the initial patient selection and various screening processes necessary for clinical trials in autism and other pervasive developmental disorders. These measures can also play an important role in the assessment and characterization of response. In addition, clinical trials offer an opportunity to employ research-oriented measures in order to study underlying etiologic and pathophysiologic processes. In the patient-selection phase, biomedical measures are used to screen for potential safety-related problems, to decrease biological and genetic heterogeneity, and to ascertain subgroups of special interest. In the response realm, the measures can be examined as possible predictors and modifiers of therapeutic response and adverse effects. At times the measures can even be seriously considered as possible surrogates of response. Although grouped separately, the boundary between the selection and response measures is certainly not hard and fast. Biomedical research measures can be employed to study mechanisms and extent of drug action, to perform baseline investigations of possible pathophysiologic alterations and to carry out longitudinal studies in the active agent and placebo-treated groups.
A published set of guidelines put forth by the Cure Autism Now Consensus Group in 1998 for the screening and diagnostic evaluation of individuals with autism is very relevant to the issues involved in clinical characterization of individuals participating in clinical trials. The physical examination, medical history, and basic laboratory tests are essential, and a family history is highly recommended. Other measures, including chromosomal and metabolic screening and electroencephalography (EEG), are often desirable depending upon patient characteristics and the agent under study. A range of more research-oriented measures, including neuropsychological, neurophysiological, neuroimaging, neurochemical and neuroendocrine indices, plasma drug levels, and pharmacogenetic analyses, may be of particular utility in specific studies.

**SELECTION AND SCREENING**

Therapeutic responses and adverse effects may relate to findings that are relevant to subgroups within otherwise heterogeneous and overlapping diagnostic categories within the autism spectrum. Thus, the traditional medical history and physical examination are especially important for the basic evaluation of all subjects in autism clinical trials. Subgroups might be defined in children with histories of regression or a plateau in language development (typically, from 18–24 months of age); recurrent infections, such as otitis media, or allergies, including drugs; gastrointestinal symptoms, diarrhea or esophageal reflux; seizures; hypotonia or signs of autonomic dysfunction. Information from each subject should build a database for that individual, as well as the study group, that can be correlated to behavioral, cognitive, and other outcome measures.

**Physical Examinations**

The physical examination of subjects is a necessity and should include plotting of growth parameters on standard charts: height, weight, and head circumference in percentiles for age. Many young subjects have macrocephaly and accurate head measurements are best obtained using narrow cloth tapes pulled tightly from the inion of the occiput to the mid-frontal bone. Dysmorphic features occur commonly in autism, and may also reflect underlying chromosomal or other genetic disorders. Among other findings, these dysmorphologies may include unusual facial features, position or size of the ears, size of the hands and feet, or syndactyly of the toes. The skin should be carefully examined for signs of café-au-lait spots (neurofibromatosis-type 1 [NF-1]) or depigmented spots (tuberous sclerosis), and an ultraviolet lamp can be useful for highlighting these. A neurodevelopmental evaluation, with emphasis on speech and language and social domains, is important to assess appropriate brain development for age. The neurological examination should include cranial nerve function (e.g., Moebius syndrome), muscle bulk, strength and tone, deep tendon reflexes, coordination, and movement disorders, including stereotypes. Asymmetries as well as handedness should be noted.

**Medical and Family History**

Obtaining a complete medical history is essential, and a family psychiatric history is highly desirable. Important elements of the medical history should include the mother's pregnancy history, including drug and environmental exposures, premature labor, infections, and obstetrical complications. The child's neonatal, infancy, and early childhood histories should be recorded, along with immunizations, food sensitivities, surgeries, and hospitalizations. The child's fever response may be of special interest, in that some children with autism frequently seem to show little rise in temperature, while others may have atypical responses. Important basic elements of the complete history and examination have been organized in the Autism Genetic Resource Exchange (AGRE) database.

The family history should track maternal and paternal lines separately because imprinted genes may be important for the behavioral expression of autism and may influence drug metabolism (pharmacokinetics) and drug response (pharmacodynamics). For autism, the family history should include questions about traits of reclusiveness, aloofness, and social ineptitude, as well as language disorders and delays in development of speech. Learning disorders, school failure, and mental retardation or intellectual disability in relatives may indicate genetic factors, such as Fragile X syndrome or tuberous sclerosis. Neuropsychiatric disorders should be recorded. This would include depression, bipolar disorder, obsessive-compulsive disorder, and schizophrenia. Stereotypic behaviors, tics, and Tourette's syndrome may be genetically linked to autism in some families. Autoimmune disorders may occur with increased frequency in close relatives of children with autism in some families, especially rheumatoid arthritis, lupus, and autoimmune thyroid disease. Medical disorders of note may include gout (with possible relevance to “purine autism”), neurocutaneous disorders (including NF-1), and mitochondrial disorders, including unexplained hearing loss and myopathies.

**Chromosomal Analysis and Metabolic Screening**

Basic laboratory testing should include a complete
blood count and serum chemistries, lead level, a chromosomal analysis with high resolution banding, and DNA for Fragile X syndrome. Additional genetic testing that may be indicated by dysmorphic features on the physical examination include fluorescent in situ hybridization (FISH) for specific disorders, such as Williams and Angelman syndromes, telomere deletion screening, and testing for carbohydrate deficient glycoprotein and 7-dehydrocholesterol. Metabolic screening, such as quantitative plasma amino acids and urinary organic acids, is useful to screen for mitochondrial disorders, as well as to rule out rare metabolic disorders that may be associated with autism.

**ELECTROENCEPHALOGRAPHY**

EEG can be used to detect abnormal brain electrical activity associated with seizures. The EEG may demonstrate electrical seizure activity in autistic subjects, even when the subject with autism has not had clinical seizures. Such abnormal “spike and wave” discharges occur most commonly in deep sleep, which is obtained optimally on overnight studies. Up to 40% of children with autism will develop clinical seizure activity over time, usually by adolescence. A possible relationship between brain electrical seizure activity and language regression suggests that, in some children with autism, loss of language may be a form of epileptic aphasia or Landau-Kleffner syndrome. Although abnormal EEGs are important for the diagnosis and treatment of epilepsy, their usefulness for treatment and prognosis of autism have not been established. Routine EEGs are probably not warranted. However, EEGs are strongly recommended if seizures are suspected and EEGs may be useful in studies targeting seizures or if anticonvulsants are being studied. Preliminary results from a retrospective study suggest that EEG abnormalities might be predictive of overall response in individuals treated with the anticonvulsant valproic acid. Likewise, sleep studies (ie, polysomnography) are useful for the evaluation of sleep disorders, which occur commonly in autism and may be associated with rapid eye movement sleep behavior disorder. However, sleep studies are difficult to obtain in uncooperative patients and are not indicated routinely in the evaluation of autism.

**RESEARCH-ORIENTED MEASURES: RESPONSE AND PATHOPHYSIOLOGY**

**Neuropsychological and Neurophysiologic Assessment**

Visual fixation patterns can now be measured using infrared and other methods for tracking eye movements. These techniques show patterns of eye movements during visual pursuit that are distinctive, even in high-functioning autistic individuals, for tracking faces and emotional expressions. The duration of visual focus on mouths, as opposed to objects, correlates with the social functioning abilities of individuals. Such measures might be useful outcome measures in clinical trials that seek to affect social competence. New techniques for measuring central auditory processing in autism spectrum disorders (ASDs) using standardized testing and long latency event-related potentials (ERPs) may also provide reliable clinical and physiological measures of drug and training effects. In high-functioning children with autism, Ceponiene and colleagues found that involuntary orienting to changes in auditory stimuli (as indexed by the P3a ERP response) was severely affected in children with autism when the change occurred in a speech sound (vowel) but was normal when it occurred in acoustically matched non-speech stimuli. Experimental studies of this type may be of use in monitoring improvement in auditory sensitivity and therapy-related improvements in language processing. More generally, methods that examine basic aspects of sensory processing may prove especially useful in defining autistic subgroups and in revealing critical dimensions of response.

A need for autonomic nervous system (ANS) evaluations is suggested by clinical observations that children with autism frequently show dilated pupils, unusual cold tolerance, decreased shivering and sweating, and sympathetic hyperresponsiveness to everyday stimuli. Basic measurements of heart and respiratory rates, blood pressure, electrodermal responses, and pupillography, when integrated over time, can provide a dynamic picture of ANS functions, both sympathetic and parasympathetic. The balance of these functions depends on developmentally determined interactions within brainstem nuclei, from sensory inputs, as well as the cerebral cortex and hypothalamus. Measurements of these functions are becoming more relevant as measures of responses to drug and other therapies, and are now increasingly practical with new tools available for ANS evaluations.

**Functional Neuroimaging**

There are a number of potential applications of functional neuroimaging in pharmacological treatment studies of autism. Positron emission tomography (PET) or single photon emission computed tomography (SPECT) can be used to assess drug kinetics and receptor occupancy by imaging radiolabeled drugs.
over time following administration. PET, SPECT, and functional magnetic resonance imaging (fMRI) can be used to measure changes in blood flow in the brain before and after drug treatment. PET, SPECT, and magnetic resonances spectroscopy can be used to monitor changes in brain biochemistry with treatment. Various imaging parameters in the future may be used to predict which patients may be most likely to respond to a drug treatment based upon their biochemical phenotype. These imaging modalities also may be used to understand developmental changes in biochemistry and thereby lead to more rational design of drug treatments of children of different ages.

For example, global brain values for serotonin (5-HT) synthesis capacity, measured with PET using the tryptophan analogue alpha[C-11]methyl-L-tryptophan, were >200% of adult values until 5 years of age and then declined toward adult values. In autistic children, 5-HT synthesis capacity increased gradually between 2 and 15 years of age to values 1–1.5 times adult normal values. Although there is some disagreement concerning the interpretation of alpha[C-11]methyl-L-tryptophan data, the results suggest that humans may undergo a period of high brain 5-HT synthesis capacity during childhood, and that this developmental process may be disrupted in autistic children. Similarly, the N-methyl-D-aspartic acid A (GABA_A) receptor complex can be measured with PET using the tracer [C-11]flumazenil (FMZ), a ligand that binds the GABA_A receptor at the benzodiazepine site. In non-autistic subjects with epilepsy, global brain values for FMZ volume of distribution were highest in the youngest children (2 years of age, 50% higher than adults) and declined exponentially with age. Analysis of FMZ PET scans in nine children and young adults with autism showed volume of distribution values were lower in four of the autistic subjects, whereas the remaining autistic subjects showed values that did not differ from the non-autistic age-matched children. This suggests that the GABA_A receptor is affected in some children with autism but not in others. Information about developmental changes in receptors and neurotransmitters can provide guidance in determining doses to be used in children of different ages.

Task-related studies using fMRI have found baseline differences that might either predict or change with drug response. Thus, the reduced level of fusiform face area activation seen in individuals with autism, compared with normal controls when processing human faces might distinguish subgroups suitable for specific interventions. Measures of regional activation might also be used to assess the effects of interventions, such as training, to increase familiarity with faces and administration of agents, such as selective serotonin reuptake inhibitors (SSRIs). Due to the cost and the difficulty of performing functional imaging studies in children with autism, the application of these methods would not be expected to be a routine component of drug treatment trials. These techniques may be instrumental in trials designed to study a specific subset of subjects.

Structural Neuroimaging

Data from structural brain imaging studies have provided a great deal of useful information to autism researchers. While most researchers agree that clinical magnetic resonance imaging (MRI) scans from children and adults with autism rarely indicate any specific abnormalities, studies that have relied on high resolution imaging protocols and quantification of neuroanatomic structures have identified abnormalities in the brain in autism.

A number of quantitative MRI studies in autism have reported that the cerebellar hemispheres and subregions within the cerebellar vermis are abnormally reduced in size. The majority of these studies have included adolescents and adults with autism, many of them higher-functioning. Some results have suggested that the degree of abnormality may be related to level of functioning. MRI studies of high-functioning older children and adults with autism have concluded that the size of the anterior and posterior cerebellar vermis does not differ from normal, while studies that include both low- and high-functioning individuals consistently report decreased size of the posterior vermis.

The cross-sectional area of the posterior corpus callosum has also been found reduced in size compared with normal. There is also some evidence for reduced size of the amygdala and hippocampus. Among older children and adults with autism, brain volume has been found to be normal or slightly below normal.

In order to examine early brain development in autism, children need to be studied as young as possible. In a recent study of 60 children with autism and 52 normal controls (2–16 years of age), whole brain volume was significantly larger that normal in the 2- to 4-year-olds with autism. Cerebral white and gray matter were significantly larger than normal in the youngest children but not the older children and adolescents with autism. The most dramatic finding was the degree to which cerebellar white matter volumes were larger than normal in the youngest children with autism. The posterior cerebellar ver-
mis was smaller across all ages examined. Excessive cerebral growth appeared to be largest in the frontal lobes in the youngest children, there were different patterns within frontal lobe regions. Abnormally large whole brain volume in preschoolers with an autism spectrum diagnosis was also reported in a subsequent study by another laboratory. The abnormalities in whole brain volumetric measures for young children with autism appear to be quite robust. In an extension to the Courchesne and colleagues study, Akshoomoff and colleagues found that MRI brain measures correctly classified 95% of the ASD cases and 92.3% of the control cases (head-to-head comparison). This set of variables also correctly classified 85% of the ASD cases as lower functioning and 68% of the ASD cases as higher functioning.

It is possible that structural MRI could be useful in clinical treatment trials for autism. Potential participants could be selected or defined based on certain neuroanatomic characteristics. For example, there may be reason to select children who have whole brain volume measures that are closer to the normal mean than those who have excessively large whole brain volume. On the other hand, there may be reason to use variations in neuroanatomic abnormality to assess outcome or symptom change following a treatment trial. For example, a certain type of treatment may be less effective in individuals with abnormally reduced hippocampal volumes than in those with normal volumes. The data from MRI and postmortem studies make it clear that widespread areas of the brain are affected in autism, and affect the developmental process in many systems. These factors also need to be considered before attempting to apply simple brain-behavior models to this disorder.

There are additional caveats to be considered. The data briefly reviewed here indicate that age-related differences in brain measures and patient characteristics need to be taken into consideration when evaluating quantitative MRI data. Imaging protocols and analysis techniques vary, making it difficult to compare absolute measurements across studies and laboratories. These techniques can be quite costly, making them impractical for large-scale studies. However, these techniques may be useful to identify potential explanations for study results and the development of animal models.

**Neurochemical and Neuroendocrine Measures**

The neurochemical and neuroendocrine measures that may be of use in autism clinical trials include those related to mode of drug action and those assessing possible underlying pathophysiologic alterations. There is an extensive and natural overlap between these areas, as drug effects in autism and on autism-related behaviors are often used to suggest etiological hypotheses. Conversely, hypotheses regarding underlying neurobiology have served to argue for the testing of certain classes of agents. The principle central systems of interest have included the serotonergic, dopaminergic, noradrenergic, cholinergic, glutamatergic and gaba-ergic, as well as the hypothalamic-pituitary-adrenal axis and the sympathoadrenomedullary system. The more commonly employed agents, including the neuroleptics, the serotonin reuptake inhibitors and SSRIs, the stimulants and anticonvulsants, all act on one or more of the aforementioned systems.

The longstanding and well-replicated finding of platelet hyperserotonemia in autism has generated considerable interest in the serotonergic system, and the SSRIs are now one of the most widely used classes of agents in autism. Early studies of the serotonergic agent fenfluramine used measures of platelet 5-HT to assess the biochemical effects of the drug. The diminutions in platelet 5-HT following SSR1 administration are much more easily interpreted and have been widely used as a reflection of the extent of uptake inhibition. In a recent treatment study of fluvoxamine in children and adolescents with pervasive developmental disorder, it was clear that response and nonresponse, as well a possible gender difference in response, was not due to under- or overmedication, as similar reductions in platelet 5-HT were seen across the various response groups. This approach to the pharmacodynamics and bioavailability of the SSRIs adds a valuable perspective to the treatment study. In a related but less well-established approach, changes in plasma levels of the neurohormone prolactin have been used to assess central response to the nonspecific serotonergic enhancer fenfluramine. Similar neuroendocrine challenge strategies have been used in autism after administration of more receptor-specific serotonergic agents.

The wide use and efficacy of the dopamine receptor blocking neuroleptics in pervasive developmental disorder have served to draw attention to the dopaminergic system. Unfortunately, there are few good available measures for assessing alterations or drug effects on central dopamine and reported basal alterations in central dopamine metabolism in autism have not been substantiated. Hyperprolactinemia stands out as a consistent and potentially damaging adverse effect of treatment with typical and the newer atypical neuroleptics. Hyperprolactinemia is clearly a result of reduced dopaminergic inhibitory control in...
the tuberoinfundibular pathway. Measurement of prolactin can be recommended on safety grounds, and it may also provide an approach to predicting treatment response and adverse effects. Adverse effects of potential interest in this context include excess salivation, motoric side effects, and weight gain.

Frequent serious problems with hyperactivity and related behaviors in individuals with autism have led to the use of stimulants, such as methylphenidate, and adrenergic compounds, like guanfacine. The relationship of the hyperactivity, impulsivity, restlessness, and apparent attention problems to reported hyperreactivity of stress response systems is not clear, however assessment of relevant sympathetic, adrenal medullary and hypothalamic-pituitary-adrenal axis measures may of particular utility when targeting behaviors in this area. Included as possible measures are urine, salivary and plasma cortisol, plasma adrenocorticotropic hormone and β-endorphin and urinary and plasma catecholamines (norepinephrine and epinephrine), and related metabolites. In general, protocols examining urinary measures or plasma compounds with long half-lives assess basal functioning, while those examining patients in clinical situations (testing, blood-drawing) using neurophysiologic (heart rate, blood pressure, galvanic skin response) or short half-life plasma species (eg, adrenocorticotropic hormone, catecholamines) obtain a reflection of reactivity. A review of the studies examining various neurochemical and neuroendocrine measures, as well as related neurophysiologic measures, in autism strongly suggested, that, while basal functioning of the stress response systems are not altered, there is hyperreactivity to acute stress. It is of potential interest to examine the effects of stimulants and other agents on both basal functioning and on reactivity.

Assessment of pre or post-drug functioning of central GABA and glutamate systems is more difficult and would require either lumbar puncture for cerebrospinal fluid measurements or intensive studies of central receptor imaging. Measurement of levels of the pineal indole, melatonin, in plasma and urine may be useful given suggested alterations in pineal function in some individuals with autism and melatonin’s role in sleep, an area of frequent disturbance in autism. Another area of great potential is “omic” analyses. Recently developed methodologies and software permit the analysis of a great number of messenger RNA species, proteins, or small molecules. This approach been labeled transcriptomics, proteomics, and metabolomics, respectively. The approaches have the potential to reveal etiologic and pathophysiologic alterations that may well be undetectable when examining single analytes. The “omic” approach may also be an especially efficient way to search for predictors and correlates of drug response and, in this context, would probably use plasma or, perhaps, urine specimens.

PHARMACOLOGY AND PHARMACOGENETICS

Plasma drug monitoring should be seriously considered as part of any clinical trial, and the autistic individual's often limited capacity to express health complaints may make this window on drug metabolism and disposition of special utility in when treating patients with autism. Guarding against toxicity and avoiding underdosing are critical, and examining dose-response relationships in the clinical range can be crucial to establishing treatment guidelines. The pharmacokinetic approach can be augmented by phenotyping relevant cytochrome P450 enzymes and genotyping functional variants. The choice of enzymes to examine depends upon the class of agent and the specific drug administered.

Possible genetic influences on the pharmacodynamic effects of the SSRIs, neuroleptics, stimulants, and other agents can be examined by typing variants of genes encoding for receptors, transporters, and metabolic and synthetic enzymes related to the neurobiology of drug action. Examples include typing of 5-HT transporter and receptor variants in trials of SSRIs, examining dopamine and 5-HT transporter and receptor variants in trials of atypical neuroleptics, and examining dopamine- and noradrenaline-related variants when studying effects of the stimulants.

CONCLUSION

There is obviously a wide range of possible biomedical measures that can be employed in autism clinical trials. The inclusion of a number of the measures can be strongly recommended as having to do with patient safety or with the characterization necessary for proper patient selection. However, many of the measures can only be recommended if they do not over-burden a study in terms of time and possible aversiveness. Nevertheless, there is an imperative to obtain as much information as possible from any study, both with respect to the drug’s therapeutic and adverse effects and regarding its biological actions. It should be emphasized that patient cohorts offer a well-characterized group of affected individuals that will be available over a defined time-span. Thus, it is useful to consider what parallel studies can be undertaken. Considerations may be focused on, but should
REFERENCES

9. Tuchman R. Treatment of seizure disorders and EEG abnormalities should also be obtained and analyzed with a view toward compatibility with past studies and with an eye toward future application and utility.

not be limited to, measures related to the symptoms or agent of interest. Whatever measures are incorporated should be undertaken in a manner that optimizes their subsequent utility. Just as behavioral assessment should be done in a manner as consistent as possible with future meta-analytic utilization, biomedical measures should also be obtained and analyzed with a view toward compatibility with past studies and with an eye toward future application and utility.

CNS
Government Initiatives in Autism Clinical Trials

By Benedetto Vitiello, MD, and Ann Wagner, PhD

ABSTRACT
Randomized clinical trials remain the most valid method of testing the efficacy and safety of treatments. While efforts to elucidate the genetic and neurodevelopmental bases of autism are underway, clinicians and families are in need of scientifically valid information on how to best treat patients with autism. The effectiveness of many interventions currently used in communities has not been adequately tested. Given the high public health relevance of autism treatment research and the low interest of the pharmaceutical industry in autism, the role of government agencies in supporting clinical trials in autism is critically important.

FOCUS POINTS
• Autism research is a high priority at the National Institutes of Health and its funding has substantially increased in recent years.
• Controlled clinical trials of interventions for individuals with autism are a necessary step toward developing a rational, evidence-based approach to the treatment of these patients.
• Because of the limited involvement of the pharmaceutical industry in autism, the role of government agencies in supporting clinical trials in autism is critically important.

INTRODUCTION
Autism is a major public health concern and research on autism is a high priority at the National Institutes of Health (NIH), as reflected in the substantial increase in research funding over recent years. Autism funding at NIH increased from $22.2 million in 1997 to $73.9 million in 2002, which constitutes a 3.3-fold increase, during a period of time when the total NIH fund appropriation increased 1.8 fold, from $12.7 billion in 1997 to $23.5 billion in 2002. The majority of NIH autism-related research is sponsored by the institutes that comprise the NIH Autism Coordinating Committee (NIH/ACC): the National Institute of Mental Health (NIMH), the National Institute of Child Health and Human Development (NICHD), the National Institute of Neurological Disorders and Stroke (NINDS), the National Institute of Deafness and Other Communication Disorders (NIDCD), and the National Institute of Environmental Health (NIEHS).

The Children’s Health Act of 2000 mandated the establishment of an Interagency Autism Coordinating Committee (IACC) to coordinate autism research and other efforts within the Department of Health and Human Service. In April 2001, Secretary of Health and Human Services Tommy Thompson delegated the authority to establish the IACC to the NIH. Within the NIH, the NIMH has been designated to lead this work. IACC membership includes representatives from seven agencies of the Department of Health and Human Services agencies: the NIH, Dr. Vitiello is chief of the Child and Adolescent Treatment and Preventive Intervention Research Branch, and Dr. Wagner is chief of the Autism and Pervasive Developmental Disorders Intervention Research Program, both in the Division of Services and Intervention Research at the National Institute of Mental Health in Bethesda, Maryland.

Acknowledgements: The authors would like to acknowledge the contribution of the following speakers at the Autism Clinical Trials Task Force Conference held March 2002 in Santa Monica, California: Deborah Hirtz, MD (National Institute of Neurological Disorders and Stroke), Tan Nguyen, MD (Food and Drug Administration), and Peter Bell (Janssen).

Disclosure: The authors do not have an affiliation with or financial interest in a commercial organization that might pose a conflict of interest. The opinions and assertions contained in this report are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of Health and Human Services, the National Institutes of Health, or the National Institute of Mental Health. This paper was submitted on June 2, 2003, and accepted on July 3, 2003.

Please direct all correspondence to: Benedetto Vitiello, MD, National Institute of Mental Health, Room 7147, 6001 Executive Blvd., MSC 9633, Bethesda, MD 20892-9633; E-mail: bvitiell@mail.nih.gov.
Centers for Disease Control and Prevention, Administration for Children and Families, Centers for Medicare and Medicaid Services, Food and Drug Administration, Health Resources and Services Administration, Substance Abuse and Mental Health Services Administration, and the Department of Education’s Office of Special Education. There are several public members, who are patient and family advocates. The primary mission of the IACC is to facilitate the efficient and effective exchange of information on autism activities among the member agencies, and to coordinate autism-related programs and initiatives. The IACC also serves as a forum to assist in increasing public understanding of the member agencies’ activities, programs, policies, and research and in bringing important matters of interest forward for discussion. The IACC meets twice yearly and meetings are open to the public. Information about the IACC and summaries of its meetings can be found at http://www.nimh.nih.gov/autismacc/index.cfm.

NIH-sponsored research on autism spectrum disorders covers domains such as genetics, neurobiology, diagnosis, development, interventions, and services. Most of the NIH’s autism funding is devoted to research to elucidate the pathogenesis of the disorder and, in particular, its genetic and neurobiological substrate. Understanding the basic mechanisms responsible for the clinical manifestations of autism is a critical step towards developing specific preventive and treatment interventions. While these research efforts continue, subjects suffering from autism, their families, and treating clinicians face immediate decisions as to which interventions to use in order to improve behavior and counteract the disabling effects of the disorder.

The level of evidence for the efficacy of many treatment interventions used in autism is less than ideal, being too often based on small and uncontrolled studies or merely on case reports. Only few treatment interventions have received controlled clinical investigation. In this situation, treatment choices for subjects with autism lie on less firm ground than for patients suffering from other disorders, such as attention-deficit/hyperactivity disorder (ADHD), epilepsy, or depression. Controlled clinical trials remain the most valid way of testing the efficacy and safety of potential treatments, and, though a small part of the overall research effort in autism, they constitute an essential component of this effort because of the potential for generating knowledge with immediate and direct impact on clinical practice.

RECENT NATIONAL INSTITUTES OF HEALTH INITIATIVES RELEVANT TO CLINICAL TRIALS IN AUTISM

The table summarizes some recent NIH funded multisite clinical trials in autism. In 1997, the NIMH launched the Research Units on Pediatric Psychopharmacology (RUPP) Autism Network with the purpose of conducting clinical trials of medications commonly used in the community to treat children with autism. The network was funded through NIMH contracts to Indiana University, University of California Los Angeles, Yale University, and Ohio State University, with a subcontract to the Kennedy Krieger Institute. Risperidone was identified as one of the most commonly used medications in the management of severe behavioral disturbances, such as aggression, self-injury, and tantrums, in children with autism, in spite of the absence of evidence from well controlled studies to support the efficacy of this practice. Consequently, a double-blind, placebo-controlled, randomized trial of risperidone in children with autism with severe behavioral problems was conducted and recently reported. Risperidone was found to be superior to placebo in improving behavior and functioning in the short-term (8 weeks). Drowsiness, fatigue, and increased appetite (average weight gain of 2.7 kg) were more common with risperidone than with placebo (average weight gain of 0.8 kg). Children who improved with risperidone were followed-up for up to 6 months of continuous treatment to assess persistence of response and tolerability. A report on these data is forthcoming.

Another trial currently in progress in the RUPP Autism Network, is testing the efficacy and tolerability of methylphenidate in children with autism or other pervasive developmental disorders (PDDs) who also have impairing symptoms of ADHD. Although in the prevailing current psychiatric classification the presence of autism or another PDD is an exclusion criterion for a formal diagnosis of ADHD, symptoms of hyperactivity, impulsiveness, and inattention are common among these children and frequently cause significant impairment. Small studies and case reports have suggested that stimulants can be helpful in some children with autism, but side effects are more common and severe than among non-autistic children. The current RUPP trial aims to establish the degree of efficacy and of intolerable side effects of methylphenidate in autism and to examine possible predictors of treatment effects.

In 2001, the scope of the RUPP network was expanded to include also trials of psychosocial interventions. In this network, the autism group is funded through NIMH cooperative agreement grants...
to Indiana University, Ohio State University, and Yale University. A randomized controlled trial to test the efficacy of an intervention that combines both medication and behavioral therapy in the treatment of severe behavioral disturbances in the context of autism and other PDDs is being planned. This study is expected to address an important question: What is the benefit of adding a behavioral therapy component to a purely pharmacologic management of children with autism-related severe behavioral symptoms? In clinical practice, medications are often not used in isolation, but integrated with nonpharmacologic interventions. In fact, clinical trials that test just one intervention do not necessarily provide all the information that clinicians and families need. Studies that directly compare alternative treatment modalities with each other, and more comprehensive interventions with simpler ones are needed. These types of clinical trials are unlikely to be supported by the pharmaceutical industry, and are therefore especially relevant to the NIMH mission.

In 1997, the NICHD started the Collaborative Programs of Excellence in Autism network, whose primary focus has been on the pathogenesis and genetics of autism, but which has also conducted clinical trials of high public health relevance. An example is the double-blind, placebo-controlled study of secretin in autism by Sandler and colleagues. In the late 1990s, anecdotal reports of dramatic improvements in children who have received intravenous infusions of secretin for diagnostic purposes generated much interest in this peptide among families and practitioners. The results of the trial by Sandler and colleagues, and of other subsequent studies, did not support the efficacy of secretin in autism.

In response to growing public concern about what appears to be an increasing prevalence of individuals diagnosed with autism and related disorders, congress included several items related to autism in the Children’s Health Act of 2000. This legis-

<table>
<thead>
<tr>
<th>Study</th>
<th>Main Aim</th>
<th>Design</th>
<th>Site</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risperidone in Children with Autism and Serious Behavioral Problems</td>
<td>To test the efficacy and tolerability of risperidone in autism</td>
<td>8-week randomized, placebo-controlled, double-blind parallel groups</td>
<td>RUPP Autism Network</td>
<td>Completed</td>
</tr>
<tr>
<td>Longer Term Risperidone Treatment of Autistic Disorder</td>
<td>To examine extended maintenance of tolerability and efficacy of risperidone</td>
<td>Open-label risperidone for 4 months, then placebo-controlled blinded discontinuation</td>
<td>RUPP Autism Network</td>
<td>Completed (report in progress)</td>
</tr>
<tr>
<td>Methylphenidate in the Treatment of Hyperactivity and Impulsiveness in Children with PDD</td>
<td>To test the efficacy and tolerability of methylphenidate in PDD</td>
<td>Randomized, placebo-controlled, double-blind, Within subjects 5 weeks plus extended follow-up</td>
<td>RUPP Autism Network</td>
<td>Completed (report in progress)</td>
</tr>
<tr>
<td>SSRI Treatment of Repetitive Behavior in Children with PDD</td>
<td>To test the efficacy and tolerability of an SSRI medication in reducing repetitive behavior associated with PDDs in children</td>
<td>In preparation</td>
<td>STAART Clinical Trial Network</td>
<td>In preparation</td>
</tr>
<tr>
<td>Risperidone and Behavior Therapy in Children with PDD</td>
<td>To compare the relative efficacy of combined pharmacological and behavioral therapy versus pharmacological treatment alone</td>
<td>In preparation</td>
<td>RUPP-PI Autism Network</td>
<td>In preparation</td>
</tr>
</tbody>
</table>

NIH=National Institutes of Health; RUPP=Research Units on Pediatric Psychopharmacology; PDD=pervasive developmental disorder; SSRI=selective serotonin reuptake inhibitor; STAART=Studies to Advance Autism Research and Treatment; RUPP-PI=Research Units on Pediatric Psychopharmacology and Psychosocial Interventions.

lution mandated the establishment of a new autism research network—the Centers of Excellence in Autism Research—and specified that each center must conduct intervention research projects. In response, the NIMH, NICHD, NINDS, NIDCD, and NIEHS have launched the network of centers for Studies to Advance Autism Research and Treatment (STAART).6 In 2002, two STAART centers were funded at the University of North Carolina, Chapel Hill and Yale University. In 2003, six additional centers (at the University of Washington; University of California, Los Angeles; Mount Sinai Medical School; Kennedy Krieger Institute; Boston University; and the University of Rochester) have been funded. Each center will contribute to the autism research base in the areas of causes, diagnosis, early detection, prevention, control, and treatment. Planned treatment studies cover a wide variety of topics, including development of novel treatments and efficacy trials of commonly used interventions. Both pharmacologic and psychosocial treatment multisite studies will be conducted.

In addition to supporting research infrastructure and specific clinical trials, the NIH organizes conferences and workshops for researchers in the field of autism with the purpose of advancing the development of novel interventions and improving the methodology for assessing treatment effects. In April 1999, the NIH/ACC and the Department of Education sponsored a meeting on “Treatments for People with Autism and Other Pervasive Developmental Disorders: Research Perspectives.” That meeting made it clear that there was a need for treatment development and resulted in a Request for Applications for treatment development projects. Several exploratory studies were funded from the responses to that application and are in process now, including projects on pharmacologic and behavioral treatments, and the development of an animal model to explore pharmacologic interventions related to repetitive self-injurious behaviors.

As a follow-up, the workshop “Research on Psychosocial and Behavioral Interventions in Autism: Confronting Methodological Challenges” was convened in September 2002.6 The aim of this meeting was for a more detailed discussion of the state of the science with regard to psychosocial and behavioral interventions, and to discuss potential strategies for addressing challenges confronted in conducting this type of research. Although the focus was on psychosocial/behavioral treatment, many of the research design challenges apply to medication trials as well. It was clear from the meeting that there is a continuing need for treatment development, but there is also a need to test the efficacy of commonly used but untested treatments. Some treatments with efficacy evidence, such as early intensive behavioral intervention, need further testing in order to evaluate relative efficacy when compared with other treatments, identification of moderators and mediators of outcome, determining optimum intensity, and other factors related to service delivery. There are clearly special challenges related to research design and outcome measurement due to the wide heterogeneity of the population, the lack of specificity of cause and underlying pathophysiology, and the variability of symptomatology across development.

Specifically focused on treatment development for subjects with Fragile X syndrome, a condition that can present with the clinical features of autism, was the workshop “Mental Health Aspects of Fragile X Syndrome: Treatment Research Perspectives,” which was held in November 2001. A summary of the meeting proceedings is available at http://www.nimh.nih.gov/research/fragilex.cfm.10

A significant impediment to expanding the research effort in autism clinical trials is the limited pool of investigators experienced in both the clinical management of subjects with autism and the research methodology for testing efficacy and safety of treatment interventions. A number of training grant mechanisms exist at NIH, and especially the Mentored Patient-Oriented Research Career Development Award or K23, which can be utilized by prospective autism clinical investigators.

THE FOOD AND DRUG ADMINISTRATION ORPHAN PRODUCTS DEVELOPMENT PROGRAM

The Food and Drug Administration Orphan Drug Act provides for granting special status to a drug product that is being studied as a possible treatment of a “rare disease or condition.”11 This status is referred to as “orphan designation,” which qualifies the developer of the product for the tax credit and marketing exclusivity incentives. The term “rare disease or condition” means any disease or condition that affects ≤200,000 persons in the United States, or affects ≥200,000 individuals in the US and for which there is no reasonable expectation that the cost of developing and making available in the US a drug for such disease or condition will be recovered from sales in the US of such drug. The goal of the Office of Orphan Products Development Grant Program is to encourage clinical development of products for use in rare diseases or conditions. The products studied can be drugs, biologics, medical devices, or medical foods.
the Office of Orphan Products Development Grant Program also administers a clinical research grants program, whereby researchers may compete for funding to conduct clinical trials to support the approval of drugs for rare diseases. Typically, trials are awarded grants from $100,000–$200,000/year in direct costs for up to 3 years. Information on the program is available at http://www.fda.gov/orphan/.12

Over the years, a few drugs received orphan product designation for use in autism under the Orphan Drug Act (eg, fluoxetine, human gammaglobulin for autism with regression), and several grants relevant to autism have been awarded (eg, to study fenfluramine, fluoxetine, and naltrexone). Recently, however, due to reports of a higher prevalence rate of autism in the US population than previously estimated,13 it appears that the prevalence of autism has exceeded the Orphan Drug Act criteria for “rare disease or condition.” Thus, the relevance of this program for autism clinical trials in general is at this time doubtful.

PHARMACEUTICAL INDUSTRY AND AUTISM CLINICAL TRIALS

Industry has played a critical role in facilitating NIH-funded trials, often donating study medication and matched placebo, but its role in launching clinical trials or research programs for developing and testing treatments of autism has been rather limited. It is likely that previous prevalence estimates that autism was a rare disorder had discouraged research and development investments in an area without marketing appeal. In addition, the lack of a clear understanding of the pathogenesis of autism has prevented the identification of potential treatment targets for drug development. All this may change vis-à-vis the recent marketing incentives provided by the 1997 FDA Modernization Act (eg, fluoxetine, human gammaglobulin for autism with regression) and the 2002 Best Pharmaceuticals for Children Act criteria for “rare disease or condition.” Thus, the relevance of this program for autism clinical trials in general is at this time doubtful.

CONCLUSION

Clinical trials are an essential part of the public health research effort in autism. In the last few years, major advances have been made in establishing research networks for multisite clinical trials and successfully completing several studies. New studies are being launched aimed at testing the effects of promising interventions (pharmacologic, psychosocial, or combined) already available to practitioners that may benefit children with autism by improving behavioral disturbances commonly associated with this disorder. In addition to providing clinically useful information, these initiatives have offered an opportunity to build a research infrastructure of multisite centers with expertise in running clinical trials in autism. This infrastructure will remain available for testing novel treatment approaches that may be developed based on a better understanding of autism pathogenesis. While the current investment in the neuroscience of autism seems to be still far from yielding concrete therapeutic tools, it will be crucial to facilitate the prompt translation of basic neuroscience findings into drug discovery and clinical research.

REFERENCES

Comorbidity in Compulsive Hoarding: A Case Report

By Alicia Kaplan, MD, and Eric Hollander, MD

FOCUS POINTS
- Compulsive hoarding is a difficult-to-treat condition, with high rates of treatment resistance, and may be a unique subtype of obsessive-compulsive disorder.
- Comorbid conditions contribute to the refractory nature of hoarding behaviors.
- Stimulants may provide certain benefits to the hoarding picture, especially if comorbid attention-deficit/hyperactivity disorder is present.

ABSTRACT
A 56-year-old male presented with compulsive hoarding along with attention-deficit/hyperactivity disorder and schizotypal personality disorder. Hoarding has been described as difficult to treat both pharmacologically and behaviorally, and this patient's comorbid conditions also contributed to his overall impairment. The patient's treatment regimen of fluvoxamine, amphetamine salts, and risperidone, along with behavioral therapy, has helped with hoarding behaviors, motivation, procrastination, and increased socialization. Hoarding may be a unique subtype of obsessive-compulsive disorder with poorer prognosis and distinct neuroanatomic dysfunction. Augmentation with stimulants may provide benefits in aspects of hoarding such as procrastination, especially if comorbid attention-deficit/hyperactivity disorder is present.

INTRODUCTION
Obsessive-compulsive disorder (OCD) is characterized by obsessions and/or compulsions which cause distress, are excessive, and/or lead to significant impairment in functioning.1 The illness is heterogeneous in that obsessions and compulsions span a wide range of symptomatology. In recent years, some investigations have centered on different collections of symptom factors or clusters in OCD patients that may define unique subgroups in OCD.

Baer2 initially described the following three different symptom factors in OCD patients: pure obsessions, contamination and checking, and symmetry and hoarding symptoms. Leckman and colleagues3 described four symptom clusters: (1) aggressive, sexual, and religious obsessions with checking compulsions; (2) symmetry obsessions with counting, ordering, arranging, and repeating compulsions; (3) contamination obsessions with washing and cleaning compulsions; and (4) hoarding, saving, and collecting symptoms. Hoarding, which was found as a distinct factor in the above two studies, has been shown to be associated with poorer response to pharmacologic and behavioral treatments.4,5 Hoarding is defined as “the collection of and failure to discard possessions which appear to be worthless or of no value to others.” Saved items often include newspapers, magazines, clothing, bottles, and/or other household items. Compulsive hoarders may have obsessional fears of losing items that might be needed later, or describe future need for or sentimental attachments to the saved or collected items.6 Estimates of hoarding were found to be up to 18% of 560 OCD patients in one study,7 and up to 31% of OCD patients in another.8 Hoarders with OCD often have other obsessive-compulsive symptoms as well, such as symmetry, or sexual, aggressive, or religious obsessions. Pathological hoarding is also reported in other psychiatric disorders besides OCD, including schizophrenia,9 obsessive-compulsive personality disorder,1 dementia,10 and autism.11

Dr. Kaplan recently completed a research fellowship at the Compulsive/Impulsive and Anxiety Disorders Program in the Department of Psychiatry at Mount Sinai School of Medicine in New York City. Dr. Hollander is professor of psychiatry in the Department of Psychiatry and the director of The Seaver and New York Autism Center of Excellence at Mount Sinai School of Medicine.

Disclosures: This work was supported in part by National Institute of Mental Health grant KO1 MH64539-01. Dr. Hollander has received research grants from the National Institute of Neurological Disorders and Stroke (R21 NS43979-01), the National Institute of Mental Health (1U54MH66673-01), Abbott’s Orphan Products Division, Bristol-Myers Squibb, Eli Lilly, Food and Drug Administration (FD-R-002026-01), Pfizer, Solvay, and UCB Pharma. Dr. Kaplan received National Institute on Drug Abuse grant T32DA07135-22. This case report was submitted on June 25, 2003, and accepted on July 11, 2003.

Please direct all correspondence to: Eric Hollander, MD, Department of Psychiatry, Mount Sinai School of Medicine, Anneberg Building, Suite 22-66, One Gustave Levy Place, Box 1230, New York, NY 10029-6574; E-mail: eric.hollander@mssm.edu.
We describe a treatment-resistant OCD patient with compulsive hoarding who also has comorbid conditions of attention-deficit/hyperactivity disorder (ADHD) and schizotypal personality disorder.

**CASE REPORT**

A single male, 56 years of age, presented 7 years ago for psychiatric treatment of his compulsive hoarding behavior. He described a longstanding history of hoarding of newspapers, milk bottles, paper towels, and unfinished food items, with great difficulty throwing items away. Other hoarding characteristics included perfectionism and procrastination. His apartment was piled to the ceiling with these objects. He reported distress when others would touch or move his possessions. He had been evicted from two apartments. The first eviction occurred because the landlord determined his “rat pack” apartment was a firetrap, while the second was after a fire occurred in his paper-filled apartment.

Additional symptoms of OCD included compulsive hand-washing and aggressive obsessions, such as violent images that were upsetting to the patient and found to be ego-dystonic. The patient also displayed symptoms of ADHD with current attentional difficulties and a past history of hyperactivity during his school years. His primary social contact was with his sibling. He showed reduced capacity for other interpersonal relationships, although he did attend a weekly group-therapy session that he found helpful. He displayed other schizotypal characteristics, such as excessive social anxiety, ideas of reference, and odd thinking. His past psychiatric history included past hospitalization for depressive symptoms and anxiety with consideration of a schizophrenia diagnosis. The patient had electroconvulsive therapy 12 years ago without any reported effect and had a history of fluctuating depressive symptoms. The patient's past medical history included hypertension, for which he was taking a calcium channel blocker, and arthritis.

The patient was started on fluvoxamine 50 mg/day, which was titrated up to 250 mg/day. He enrolled in cognitive-behavioral therapy with home visits. Methylphenidate was initiated for ADHD symptomatology up to 5 mg BID. This allowed the patient to focus better, be better able to handle social situations, have increased ability to get started better in the morning, and to have less procrastination. The methylphenidate was eventually discontinued because of agitation. A brief trial of buspirone 15 mg BID was given for anxiety with mild benefit. Risperidone 1 mg daily was initiated and the patient showed better insight into his hoarding behaviors. Amphetamine salts 2.5 mg BID was then instituted, and the patient reported it easier to do things, to get up in the morning, have less procrastination, and be better able to keep appointments. Compared with when he began treatment, the combination of fluvoxamine, amphetamine salts, and risperidone, as well as behavioral therapy, helped the patient be better motivated, and he is able to take courses at a local recreational center. Overall, he is calmer, more rational, less upset when throwing things away, but continues to have waxing and waning severity in his hoarding behaviors. His OCD group successfully made home visits to clear a path from the door to the window, so that an air conditioner installer could insert an air conditioner that had been unused for 2 years.

**DISCUSSION**

As described above, hoarding has been reported to be quite difficult to treat, with high rates of treatment resistance. Our patient’s treatment course is an example of this, with overall impairment stemming not only from his hoarding but also from the comorbid conditions of schizotypal personality disorder and ADHD. Frost and colleagues found that OCD hoarding subjects had more schizotypal personality disorder symptoms than any other personality types. Likewise, obsessive-compulsive patients with concomitant schizotypal personality disorder have been found to have high rates of treatment failure. This is of interest as the impairment from associated personality disorders in compulsive hoarding may be part of the refractory picture. More studies of personality in hoarding are critical, as the schizophrenia spectrum may provide a subtype of hoarders with poor prognosis.

Our patient displayed substantial perfectionism and procrastination. Frost and Show described hoarding as part of a discrete clinical syndrome that also includes pervasive indecisiveness, perfectionism, procrastination, and behavioral avoidance. In light of this, it is evident that the stimulants methylphenidate and amphetamine salts prescribed for ADHD also aided in elements of the patient’s hoarding, such as procrastination. This suggests that stimulants may be a viable augmentation treatment option in patients with comorbid ADHD and hoarding OCD.

As hoarding may be a unique subtype of OCD with poorer prognosis, it has been suggested that there may be a distinct neuroanatomic circuitry involved in hoarding. Saxena and colleagues described diminished activity in the cingulate gyrus in compulsive hoarding patients. This differs from earlier neuroimaging positron emission tomography studies in OCD.
patients where elevated metabolism in the cingulate gyrus occurs in addition to that found in the orbital frontal cortex and basal ganglia. Mataix-Cols and colleagues examined neural correlates of anxiety in normal volunteers and found that hoarding relevant anxiety predominantly activated ventral prefrontal regions and the left amygdala.

Comorbid conditions occurring with compulsive hoarding and the distinct neuroanatomic dysfunction in hoarding may account for the treatment-resistant nature of the condition, which is demonstrated by this case report. Augmentation with stimulants may provide benefits in aspects of hoarding, such as procrastination, which may improve level of functioning, especially if comorbid ADHD or severe depression exists. Saxena and colleagues have reported that intensive, multimodal treatment consisting of medication, cognitive-behavioral therapy, and psychosocial rehabilitation is helpful in treating this disabling population. Further research is needed in this not uncommon, quite disabling, and difficult to treat condition.

REFERENCES: