

Designing Research Studies on Psychosocial Interventions in Autism

Tristram Smith · Lawrence Scahill ·
Geraldine Dawson · Donald Guthrie ·
Catherine Lord · Samuel Odom · Sally Rogers ·
Ann Wagner

© Springer Science+Business Media, Inc. 2006

Abstract To address methodological challenges in research on psychosocial interventions for autism spectrum disorder (ASD), a model was developed for systematically validating and disseminating interventions in a sequence of steps. First, initial efficacy studies are conducted to establish interventions as promising. Next, promising interventions are assembled into a manual, which undergoes pilot-testing. Then, randomized clinical trials test efficacy under controlled conditions. Finally, effectiveness studies evaluate outcomes in community settings. Guidelines for research designs at each step are presented. Based on the model, current priorities in ASD research include (a)

preparation for efficacy and effectiveness trials by developing manuals for interventions that have shown promise and (b) initial efficacy studies on interventions for core features of ASD such as social reciprocity.

Keywords Autism · Psychosocial intervention · Behavioral treatment · Clinical trial · Single-subject design

Introduction

From October, 2002, to May, 2004, a working group supported by the National Institute of Mental Health (NIMH) developed guidelines for designing research studies of psychosocial interventions for individuals with autism spectrum disorders (ASD). The working group built on previous efforts by a committee that

The views expressed in this article are those of the authors and do not necessarily reflect the official position of the National Institute of Mental Health, the National Institutes of Health, or any other part of the U.S. Department of Health and Human Services.

T. Smith (✉)
Department of Pediatrics, Strong Center for Developmental Disabilities, University of Rochester Medical Center, 601 Elmwood, Box 671, Rochester, NY 14642, USA
e-mail: Tristram_Smith@URMC.Rochester.edu

L. Scahill
Yale University School of Nursing
and Child Study Center, Yale University, New Haven, CT, USA

G. Dawson
Department of Psychology, University
of Washington, Seattle, WA, USA

D. Guthrie
Department of Psychiatry and Biobehavioral Sciences,
University of California, Los Angeles, CA, USA

C. Lord
Department of Psychology, University of Michigan, Ann Arbor, MI, USA

S. Odom
Department of Curriculum and Instruction, School of Education, Indiana University, Bloomington, IN, USA

S. Rogers
Department of Psychiatry and Behavioral Sciences,
M.I.N.D. Institute, University of California, Davis, CA, USA

A. Wagner
Division of Pediatric Translational Research and Treatment Development, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

reviewed the state of the science in psychosocial interventions for ASD (Lord et al., 2002; National Research Council, 2001) and a meeting at NIH on how to advance the field (Lord et al., 2005). While this working group focused on design issues, a parallel working group addressed measurement. Members of the working group on design included investigators in the field of autism, a biostatistician, and an NIMH project officer. Outside experts in developing intervention manuals and evaluating psychosocial interventions spoke with the group at a meeting held in May 2004; other investigators and stakeholders in the ASD community, including parents and practitioners, also contributed to this meeting and commented on a draft.

The resulting model and recommendations are intended to serve two purposes: First, they offer guidance to investigators on designing and conducting studies on psychosocial interventions in ASD, based on methodological considerations and identified priorities. Second, these recommendations may help funding agencies identify current areas of need and standardize criteria for evaluating research proposals. Because the mission of the working group was to provide direction to research on psychosocial interventions, the group did not make recommendations for studies of psychopharmacological trials. Guidelines for such studies are presented elsewhere (Hollander, Robinson, & Comp-ton, 2004; Scahill & Lord, 2004).

The working group began by expanding on prior discussions concerning the multiple challenges in developing, testing and disseminating psychosocial interventions for individuals with ASD. Although previous reviews had documented some replicated findings on specific intervention techniques to enhance language, social interaction, and other skills in individuals with ASD (Goldstein, 2002; McConnell, 2002), many methodological challenges and gaps in evidence remain. For example, because of the severity of ASD, a comprehensive intervention program is likely to require many hours per week of service delivery over months or years. In a clinical trial of this kind of program, participants and families understandably may be reluctant to accept random assignment to a no-treatment control group. Also, administering the intervention program, monitoring fidelity of implementation, and evaluating outcomes are expensive and complex, but necessary (Lord et al., 2005). Not surprisingly, therefore, only one randomized clinical trial has examined a comprehensive, intensive treatment program for individuals with ASD (Smith, Groen, & Wynn, 2000b). Hence, there is very little reliable information about long-term outcomes, “active ingre-

dients” (i.e., which components are most responsible for therapeutic effects), and the relative efficacy of less versus more intensive programs (Kasari, 2002; Lord et al., 2005).

Group members concluded that no single research design could possibly deal with all methodological challenges and that no single study could fill in all the gaps in knowledge. Accordingly, the group set out to delineate a model or “road map” for developing and validating interventions in a stepwise manner from identification of new techniques to implementation in real world settings. The model draws upon ideas from engineering, in which a new discovery or invention undergoes a series of tests. Initial tests refine techniques and yield evidence that the discovery or invention has the potential for useful applications; subsequent tests confirm these findings, explore interactions among factors that may affect the utility of the invention, and assess generalizability of applications in different contexts. Another influence was the paradigm put forth by the Medical Research Council (2000) in the United Kingdom to develop complex interventions for chronic medical problems. This paradigm outlines a number of developmental steps for an intervention prior to conducting randomized clinical trials.

In tailoring a model to the particular requirements of research on psychosocial interventions for ASD, the group examined two relevant types of research designs for intervention studies: single-case experiments and experiments with between-group designs. To illustrate applications of the model, the group used it to analyze the state of the science for both (a) comprehensive intervention programs that intensively address multiple skill domains over a relatively long period (e.g., two years) and (b) skills-focused interventions that address a particular aspect of functioning. An example of a comprehensive intervention program is the UCLA Young Autism Project (Lovaas & Smith, 2003), which provides 40 h per week of instruction to preschool children with ASD for 2–3 years. An example of a skills-focused intervention is peer-mediated social skills training (Strain & Schwartz, 2001), in which typically developing peers are coached on how to enhance social interaction in children with ASD.

The group recognized that the model of phases of research is an ideal and that, in reality, investigators may have reason to deviate from this model. For example, evaluating the effectiveness of a treatment in community settings is the last phase of the model; however, if a treatment is already in wide use, community effectiveness studies may be warranted even in

the absence of studies at previous phases. Also, preliminary outcome studies are not included in the model at all but, as a practical matter, may be useful to investigators in securing support for randomized clinical trials. More generally, valuable findings may emerge from studies outside the model. Nevertheless, the model offers clear direction on research that systematically advances the state of the science and fosters dissemination of effective interventions.

Phases of Research

The working group identified four phases in research on psychosocial interventions for individuals with ASD: (a) formulation and systematic application of a new intervention technique, (b) developing a manual and research plan for evaluation of the intervention across sites, (c) randomized clinical trials (RCTs), and (d) community effectiveness studies. The first stage establishes an intervention as efficacious in highly selected patients and indicates that the intervention is worthy of evaluation in large-scale clinical trials. The second stage standardizes the intervention in preparation for clinical trials. The third stage tests the efficacy of the intervention in controlled settings, and the final stage assesses outcomes achieved when the intervention is disseminated to competent practitioners in community settings. Table 1 contains a summary of the main goals and design considerations at each phase.

Formulation and Systematic Application of a New Intervention Technique

In the first phase of research, the goals are to refine the intervention technique and conduct initial efficacy studies to provide “proof of concept” (i.e., to show that the technique may have therapeutic benefit). The investigator may begin by informally applying the new technique to a series of cases and using clinical judgment to evaluate the results. However, it is also useful to test the technique with an experimental research design and to obtain evidence that the intervention may be efficacious for some individuals with ASD. Because the effects of new interventions are unknown, oversight by an institutional review board is essential to monitor the safety and ethics of research at this preliminary phase, as well as at all subsequent phases.

One appropriate strategy to begin testing an intervention is to use single-case experimental designs. These designs compare behavior during a baseline

period when the individual receives no intervention to one or more intervention phases. Data are collected continuously by directly observing the target behavior throughout baseline and treatment conditions. If the behavior improves in an intervention period relative to baseline, one may conclude that the individual benefited from the intervention.

One single-case methodology, the reversal design, involves introducing and withdrawing a treatment two or more times to examine its impact on a particular behavior that an individual displays. This design is useful for assessing whether an intervention changes an already acquired behavior (e.g., whether a procedure alters the frequency of a behavior problem). It is less useful for determining whether an intervention promotes the development of a new skill because, if the skill is well-chosen, the individual should continue to display it after the intervention is withdrawn. Another methodology is the multiple baseline design. This approach involves having two or more baseline periods that vary in length and applying the intervention after each baseline. For example, three individuals could be in baseline and then begin intervention at different times (a procedure called multiple baseline across participants). Alternatively, three different behaviors displayed by the same individual could be in baseline and targeted for intervention beginning at different times (multiple baseline across behaviors). Because the multiple baseline design does not require withdrawing treatment, it is often appropriate for evaluating whether an intervention establishes new skills.

Single-case designs are a useful starting point for establishing efficacy because they yield evidence that the technique has a clear, replicable effect on a specific behavior. These designs also involve repeated observations, which provide close monitoring of the effects and opportunities to refine the technique. In addition, due to the small number of participants, they are likely to require less time and fewer resources than studies that compare a treatment group to a control group. As such, they can be performed by independent practitioners and small teams, enabling many investigators to contribute to the identification of new intervention techniques (Hayes, Barlow, & Nelson-Gray, 1999). A number of references describe how to plan, conduct, and analyze studies with single-case designs (Bailey & Burch, 2002; Barlow & Hersen, 1984; Kazdin, 1982). In addition, Horner et al. (2005) and the National Academy of Sciences committee (Lord et al., 2002) identified quality indicators, and Wolery (2004) reviewed procedures for monitoring fidelity of intervention implementation. Table 2 summarizes indicators that

Table 1 Phases of research on psychosocial interventions in autism

Phase	Goals and activities	Research designs	Comments
1. Formulation and systematic application of a new intervention	Conduct initial efficacy studies to refine techniques and document clinical significance of effects	Single-case experimental designs such as reversal or multiple baseline (Barlow & Hersen, 1984; Kazdin, 1982), OR Between-group designs (Kazdin, 2002)	Advantages of single-case designs for initial studies: (1) examine whether a specific intervention has a clear, replicable effect on a focal target behavior; (2) allow for individualization of intervention across subjects; (3) measure progress at multiple time points, enabling careful analysis of intervention; (4) require fewer resources than clinical trials Limitations of single-case designs for later phases of research: (1) Focus on short-term rather than long-term effects; (2) Have small sample sizes, reducing generalizability of results; (3) Yield little information on comparative outcomes of alternate interventions Advantages and disadvantages of group designs: (a) potentially avoid limitations of single-case designs but (b) pose other challenges (e.g., limited resources may necessitate a small sample size, resulting in low statistical power to evaluate intervention effects)
2. Manualization and protocol development	Assemble efficacious interventions into manual Put together a manual for comparison group Devise treatment fidelity measures Test feasibility of implementing manuals across sites Assess acceptability of interventions to clinicians and families Examine sustainability of interventions in community settings Consult with biostatistician to estimate sample size for RCT	Small, multisite pilot studies for feasibility testing (e.g., implementing manual and assessment protocol with 2–3 subjects at each treatment site; Hibbs et al., 1997) Surveys and focus groups with clinicians and families at each site Focus groups with consumers and providers in the community	Challenges: Develop manual that standardizes intervention yet allows for individualization Use modules that may be selected for individual participants Employ decision rules for modifying intervention plan based on subjects' response to intervention Specify procedures for addressing family concerns (e.g., meeting goals identified by families) and site-specific issues (e.g., available school placements) Ensure that manual is complete and current Revise manual based on feedback from surveys and focus groups Formation of multisite network to conduct pilot testing Promotion of collaboration and dissemination of information across sites Secure adequate resources to support clinicians at each site Consider partnerships with community funders Identifying fidelity measures Use both objective measures of fidelity and regular case conferences to maintain adherence to the protocol
3. Efficacy studies	Evaluate efficacy of an intervention in a large-scale trial Demonstrate consistent effects across sites, as a step toward disseminating the intervention	Randomized clinical trials (RCTs; Meinert, 1986)	Advantages of RCTs: Definitive test of whether intervention is efficacious under controlled conditions Limitations of RCTs: (1) Yield little information on “active ingredients” of intervention; (2) Use carefully selected subjects and clinicians in research settings, restricting generalizability to community practice

Table 1 continued

Phase	Goals and activities	Research designs	Comments
	Conduct hierarchical analyses of mediators and moderators		
4. Community effectiveness studies	Assess whether competent clinicians in community can implement treatment	Randomized clinical trials or other between-group designs	<p>Direct test of whether intervention is effective as typically administered in community settings</p> <p>Difficult to maintain the level of experimental control that is characteristic of randomized clinical trials</p> <p>Clinicians are independent of the study and may use an eclectic approach rather than adhering to one manual</p> <p>Community agencies may not have specific inclusion and exclusion criteria</p> <p>Primary unit of analysis in the study might be each provider or agency, rather than each individual with autism</p>

Table 2 Quality indicators in single-case research on psychosocial interventions for individuals with ASD

1. Use of a single-case experimental design such as reversal or multiple baseline
2. Specific inclusion and exclusion criteria for enrollment in the study along with documentation of drop-outs and intervention failures
3. Well-defined samples of participants in the study (i.e., standardized diagnostic tests to confirm diagnosis, standardized tests of intelligence and adaptive behavior to document developmental level)
4. Replication of intervention effects across three or more participants
5. Assessment of generalization of intervention effects to at least one other setting or maintenance of effects over time
6. Measurement of outcome conducted blind to the purpose of the study
7. Fidelity of intervention implementation monitored through direct observation

are particularly relevant to research on individuals with ASD. An example of a single-case study with these indicators is Whalen and Schreibman (2003). Due to the repeated observations that occur in single-case studies, the most challenging design feature to incorporate may be measurement by raters blind to the purpose of the study. Strategies such as videotaping observations for later scoring by independent raters can be used (Barlow & Hersen, 1984). Implementing such strategies requires time and resources but is vital for objective data collection.

Although well-suited to initial studies of an intervention, single-case designs have limitations. Because they involve a small number of participants, they yield little information on how the intervention compares with other interventions and what proportion of individuals with ASD would benefit. Multiple single-case studies by independent investigators are necessary to confirm that the intervention may help individuals with ASD beyond those in the original study. Also, these designs test whether the introduction of an intervention is associated with an immediate change in behavior but produce much less information on long-term outcomes. In addition, most studies with single-

case designs focus on a particular intervention technique; as such, they provide a valuable test of that technique but do not document the results obtained from combining different techniques into an intervention package. Moreover, the focus on measurement of individual behaviors may miss generalized effects of the intervention on other behaviors. Finally, interpretation of the results may be complex. Results are usually interpreted by visually inspecting graphs of the data (Barlow & Hersen, 1984), but this process may be subjective. Statistical methods of interpretation are also available, but selection of the appropriate statistical model may be difficult (Huitema & McKean, 1998).

For some skills-focused interventions (e.g., procedures for teaching a particular self-help skill), these limitations may be relatively unimportant, and single-case experiments may be sufficient to document that the intervention is clinically useful. Horner et al. (2005) suggested that an intervention could be considered validated based on a minimum of five well-designed single-case studies, conducted by at least three independent investigators with a total of 20 or more subjects across studies. For other interventions, however,

the limitations may be significant. For example, single-case experiments may show that a psychosocial intervention helps reduce behavior problems in some individuals, but may not help clinicians choose between the intervention and an alternate one such as a psychopharmacological intervention. Also, single-case experiments may indicate that an intervention helps some individuals acquire a new social skill, but it may be unclear whether and how the intervention should be integrated into a comprehensive intervention program. To address these situations in the model presented here, investigators would proceed to develop manuals for the interventions in preparation for randomized clinical trials.

As an alternative to single-case designs, between-group designs in which participants are assigned to an intervention or control group may be appropriate for some initial efficacy studies. For example, Kasari, Freeman, and Paparella (2006) compared groups that received interventions to enhance joint attention or pretend play to a control group that received no intervention. The findings are not definitive, but do provide information for future study. Some advantages of between-groups designs include the opportunity to assign subjects randomly to intervention conditions, the inclusion of a large number of subjects (allowing for analysis of individual differences in response to treatment and increasing generalizability of the results), and the ability to analyze long-term outcomes on global measures of functioning.

Between-group designs have potential disadvantages: Because results are aggregated across groups of individuals, it can be unclear whether individuals within the groups made gains that substantially improve their functioning or quality of life. Analyses of effect size and secondary outcome measures may help address this issue. Also, because of resource limitations, studies may have a small number of participants, especially if the research is conducted at a single site. With a small sample, a failure to show a difference between groups may be due to an ineffective intervention or inadequate statistical power. Resource limitations may also give rise to other methodological problems (e.g., insufficient monitoring of fidelity of intervention), making a finding of differences between groups difficult to interpret. Thus, in the absence of resources to conduct a large, well-designed between-group study, it is often advantageous to evaluate and refine the component parts of a proposed intervention package through single-case designs.

Assembly of Manuals and Development of Research Plan

After initial efficacy studies, the next phase is to assemble intervention techniques into a manual and put together protocols for implementing and evaluating the manualized intervention in subsequent clinical trials. An intervention may be viewed as a candidate for inclusion in a manual if three or more well-designed single-case studies from independent investigators or at least one well-designed between-group study indicate that the intervention is possibly efficacious (Chambless et al., 1996).

For research purposes, a manual serves to standardize an intervention and make it available to a wider audience, for both clinical use and continued research. Thus, it must give enough guidance to ensure uniformity of intervention implementation and assessment at different centers. The manual explains how the intervention addresses key features of the disorder or target behaviors and outlines step-by-step instructions for each stage of the intervention, as well as common problems and problem-solving strategies (Hibbs et al., 1997). The level of specificity may vary across manuals. For example, a manual that emphasizes highly structured intervention techniques such as discrete trial training might detail a large number of particular intervention programs, whereas a manual that emphasizes less structured interventions may offer less specific guidelines.

A major challenge in developing an intervention manual is to balance uniformity with the need to individualize the intervention. McMahon (2004) proposed “constrained flexibility” such that limited variation in implementation is acceptable. For example, the intervention may be divided into components with decision rules for selecting which components to deliver and under what circumstances. The manual may describe acceptable variations in delivering an intervention (e.g., alternate instructions and prompts in a discrete trial program) and courses of action if the initial implementation of the intervention is unsuccessful (e.g., possible modifications to a discrete trial program or other intervention techniques that can be introduced). The subject’s own interests could be incorporated into the manual (e.g., using preferred objects as therapy materials; Kendall & Chu, 2000). For activities such as parent training, the manual may give an outline of what to cover rather than a complete script for sessions (Frankel, 2004). Also, procedures for collaboratively setting

intervention goals with caregivers may be included (cf. Kendall & Chu, 2000; Kendall, Chu, Gifford, Hayes, & Nauta, 1998).

In addition to an intervention manual for the intervention group, a manual for the control or comparison group is often desirable. In the absence of this manual and evidence of adherence to it, results from a control group may be ambiguous. For example, finding that an intervention group outperformed a control group may not mean that the intervention was beneficial, but that the control intervention was harmful or not relevant for participants.

A manual for a comprehensive intervention package to implement for individuals with ASD is likely to be more complex than intervention manuals developed for other populations. Customarily, such manuals were written for circumscribed problems (e.g., oppositional behavior in otherwise typically developing children), with one or two hours per week of intervention provided over a period of three to six months. However, these manuals may be useful for illustrating how to develop a manual for treating individuals with ASD (cf. Woody & Sanderson, 1998). Also, although not developed specifically for clinical trials, there are several existing manuals on ASD intervention, notably Lovaas et al. (1981), Lovaas (2003), Taylor and McDonough (1996), Leaf and McEachin (1999), and McClannahan and Krantz (1999). For research purposes, it would be useful to examine procedures in these manuals for selecting particular intervention components and adapting these components based on the needs of individual participants. It would also be important to conduct systematic pilot or feasibility testing, as discussed below.

Feasibility testing involves delivering the manualized intervention to a small number of subjects as a prerequisite for a multisite clinical trial (perhaps 3–5 subjects at each participating site). Because the feasibility test involves only a few subjects, it does not provide reliable information about outcome or allow for meaningful comparisons between an intervention group and control groups. However, it does serve two important purposes: First, it shows that the intervention can in fact be delivered as planned across sites. Second, it yields information about the acceptability of the manual to clinicians and subjects (e.g., whether they view the manual as relevant, helpful, complete and user-friendly, or burdensome). The assessment may take the form of surveys, focus groups, or other subjective reports of the clinicians' and families' impressions of the intervention. Based on the feasibility test, investigators can revise the manual by adding, removing, or modifying intervention activities

and clarifying how to administer intervention procedures (Albano, 2004; DeRosier, 2004).

The feasibility test also affords an opportunity to work out measures of intervention fidelity. The two main elements of fidelity are clinician competence and adherence to the protocol (Waltz, Addis, Koerner, & Jacobson, 1993). Competence may be demonstrated by establishing minimum training criteria or by conducting tests of skill at implementing intervention techniques or formulating a subject (e.g., developing an intervention plan for an individual with ASD based on a manual). Adherence may be documented by observing a sample of intervention sessions and recording occurrence or nonoccurrence of relevant intervention techniques. Collecting information on the extent to which participants and significant others (e.g., parents) follow through on assignments also may be important. On-site supervision, frequent case reviews (e.g., weekly teleconferences with clinicians from each site), and periodic re-training are necessary to maintain adherence throughout a study and handle issues that were not anticipated in the manual (Albano, 2004; Hibbs et al., 1997).

In conjunction with feasibility testing, another prerequisite for a successful multisite RCT is to formulate a plan for data collection and analysis. Consultation with a biostatistician is often invaluable for this purpose. Table 3 summarizes key aspects of formulating a plan. One central task is calculating a sample size needed for adequate statistical power in the RCT. For this purpose, it is necessary to identify the primary unit of analysis. In some RCTs, study personnel may provide the intervention directly to individuals with ASD. Accordingly, individuals with ASD are the primary unit of analysis; they are randomly assigned to groups, and power analyses are based on how many subjects are needed in the intervention and control groups to detect a difference between groups. However, other RCTs involve training caregivers or clinicians to deliver the intervention and comparing the group that received training to a control group. In these studies, each caregiver or clinician is the primary unit of analysis, as opposed to the individual with autism. Each is randomly assigned to groups, and power analyses focus on how many parents or clinicians are needed to examine the impact of training.

Also necessary for the power analyses is the selection of a primary outcome measure and a small number of secondary outcome measures to be used in the RCT (Scahill & Lord, 2004). The choice of measures follows from having a clear rationale for why the measure is a valid index of change in areas targeted by the intervention. In addition, measures in an RCT of a

Table 3 Planning data collection and analysis

1. Conduct a power analysis
 - a. Identify the unit of analysis (e.g., individuals with ASD, families, or service providers)
 - b. Select one primary outcome measure and a small number of secondary measures
 - c. Determine the smallest difference on outcome measures that would be considered clinically meaningful
2. Work out a reliable system for managing data (e.g., developing a database and procedures for accurate data entry)
3. Outline a schedule for repeated administrations of outcome measures
4. Develop an analysis plan
 - a. Specify what data to use from outcome measures (e.g., raw scores, scaled scores, or age equivalents)
 - b. Identify optimal time points of assessment for use in specific types of analyses
 - c. Select appropriate statistical tests for evaluating differences between groups (e.g., analyses of variance, linear mixed-model regression, or non-parametric tests).
 - d. Evaluate the feasibility of analyzing variables that may influence favorable or unfavorable outcomes (i.e., mediators and moderators); if this analysis is feasible, determine what methods will be used for this analysis
 - e. Select appropriate tests of effect size and clinical significance

comprehensive intervention should be sensitive to changes in core features of autism (difficulties in reciprocal social interaction and pragmatic communication, repetitive behaviors or circumscribed interests). Investigators should identify the smallest difference on the primary outcome measure that is considered clinically meaningful. Power analyses are then performed to determine sample sizes that would be needed to detect this difference. Table 3 outlines additional tasks involved in planning data collection and analysis.

Randomized Clinical Trials

Once an intervention manual has been assembled and feasibility testing has been conducted, the multisite randomized clinical trial (RCT) authoritatively tests the efficacy of the intervention. If the trial is successful, it also confirms the replicability of the intervention across sites. Large-scale RCTs enable investigators to analyze variables such as subject characteristics that may be associated with favorable or unfavorable responses to intervention, though the statistical power for such analyses may be limited.

RCTs have made a major contribution to medicine and have become increasingly important in research on psychosocial interventions. Many references provide information on running RCTs (e.g., Kazdin, 2002;

Kendall, Flannery-Schroeder, & Ford, 1999; Meinert, 1986). Table 4 presents quality indicators such as a recruitment plan to obtain a representative sample of individuals with autism, clear criteria for including or excluding participants, careful assessment of participants prior to intervention, and, especially in studies of comprehensive interventions, outcome measures that assess core features of autism. (See Lord et al., 2005; Scahill & Lord, 2004, for further discussion.)

An important issue is the identification of variables that influence outcomes (mediators and moderators; Kraemer, Wilson, Fairburn, & Agras, 2002). A mediator is a variable that is influenced by the intervention, which, in turn, influences the outcome variable. It can be thought of as the mechanism by or through which the independent variable operates. For example, intervention may increase communication skills, which may help reduce disruptive behavior. A moderator is a factor that alters the impact of the intervention on the outcome variable. It specifies for whom or under what circumstances an intervention works. For example, language ability at entry into the study may influence how much communication skill an individual acquires during intervention, and the amount of change in communication skill may affect the extent to which the intervention reduces disruptive behavior.

Table 4 Quality indicators in clinical trials on psychosocial interventions for individuals with ASD

1. Random assignment of participants to intervention and control groups
2. Manuals for all groups
3. A recruitment plan to obtain a representative sample
4. Clearly stated inclusion and exclusion criteria
5. Careful characterization of participants at entry into the study (e.g., diagnosis, symptom severity, and level of functioning)
6. Systematic monitoring of intervention fidelity
7. Clear rationale for the choice of outcome measures and, especially in studies of comprehensive intervention packages, inclusion of measures that assess core features of autism such as reciprocal social interaction
8. Use of outcome measures collected blind to intervention group
9. Appropriate statistical analyses of differences between groups after intervention, effect size and clinical significance of differences, and variables that may influence outcomes (i.e., mediators and moderators)

As discussed by Lord et al. (2005), investigators frequently propose mediators and moderators to account for the tremendous variability in intervention outcomes among individuals with ASD. However, large-scale RCTs with appropriate analyses of mediators and moderators have not yet been conducted on psychosocial interventions for this population. RCTs on other childhood disorders, notably the Multimodal Treatment Study of Children with Attention Deficit Hyperactivity Disorder (Owens et al., 2003), illustrate suitable analytic methods.

Community Effectiveness Studies

Following the demonstration of efficacy in a large-scale RCT, effectiveness studies examine whether similar outcomes can be achieved when the intervention is applied outside specialized research centers, in real world settings by community practitioners. Thus, effectiveness studies are the ultimate test of whether an intervention is helpful in clinical practice. As such, they tend to be more consumer-driven than are efficacy studies, with a focus on whether an intervention should be widely disseminated.

RCTs in research centers differ in many important respects from clinical practice in community settings (Weisz, Weiss, & Donenberg, 1992). For example, in RCTs, specifically trained clinicians systematically deliver a well-defined intervention. By contrast, effectiveness studies may be carried out by community clinicians with eclectic theoretical perspectives and, perhaps, less training on the intervention techniques. Clinicians' professional backgrounds also may differ between RCTs and community settings. Furthermore, in RCTs, investigators specify exclusion and inclusion criteria for participants; effectiveness studies tend to use less restrictive entry criteria. For reasons such as these, intervention outcomes in community settings frequently differ from those in RCTs (Weisz et al., 1992).

Design considerations are similar for effectiveness and efficacy studies, but outcome variables in effectiveness studies are often more global and fewer in number. For example, variables may center on consumer satisfaction instead of standardized measures of symptom change. Effectiveness studies may be carried out in multiple centers with large sample sizes. As a result, subject characterization is often less detailed than in RCTs, hindering analyses of moderators. Also, checks on fidelity of intervention may be less rigorous. Given these practical realities, community clinicians may deviate from a manual to a greater extent than clinicians in RCTs. Therefore, a key aspect of

effectiveness studies is to define which elements of the intervention program must be retained across clinicians or agencies and which may be permitted to vary. For example, in early intensive behavioral intervention, community providers might diverge in their level of training and supervision, number of service hours they provide, setting where services take place (e.g., home versus school), and typical sequence in which instructional programs are introduced. For this reason, specifying minimal levels of training, supervision, and service hours, along with acceptable variations in setting and sequence of instruction, would be essential. Prior research in the single-case or RCT phase may inform development of these criteria.

The unit of analysis may differ in effectiveness and efficacy studies of the same intervention. For example, efficacy studies may focus on individuals with ASD or caregivers, whereas an effectiveness study may involve training agencies such as school districts to implement an intervention and comparing trained to untrained agencies. In this case, agencies would be randomly assigned to groups, and power analyses would ascertain the number of agencies required for statistical analyses of differences between groups.

Using the Phase Model to Evaluate the State of the Science in ASD Research

The working group applied the phase model to one comprehensive intervention (the UCLA model) and one skills-focused intervention (Peer-Mediated Social Skills Training). These interventions were chosen because they were judged to have heuristic value for evaluating the state of the science, not because they were necessarily the most promising or best-studied interventions.

Beginning in the 1960s (e.g., Lovaas, Berberich, Perloff, & Schaeffer, 1966), Lovaas and colleagues conducted single-case experiments on discrete trial training (DTT), which is a primary intervention procedure in the UCLA model (e.g., Lovaas, 1977; Lovaas, Koegel, & Schreibman, 1979). Many other investigators have also performed single-case experiments on this procedure (Goldstein, 2002). Thus, DTT has completed the first phase of intervention development (initial efficacy studies).

After their first studies, Lovaas et al. (1981) and Lovaas (2003) put together a manual, which has been used in one RCT (Smith et al., 2000b) and several effectiveness studies (e.g., Smith, Buch, & Evslin, 2000a). However, the manual has not undergone systematic pilot testing. Also, although these studies provide preliminary evidence of efficacy and effectiveness, they have limitations such as few measures of core features of autism small sample sizes, provision of

intervention at a single site rather than across sites, assessments conducted by examiners familiar with subjects' intervention status, and little evaluation of intervention fidelity. Thus, the current priority would be to do additional pilot testing on manuals, measurement of fidelity, and outcome evaluation, and then to conduct RCTs that can provide better evidence for efficacy and data regarding mediators and moderators. It would then be appropriate to proceed to community effectiveness studies.

Several comprehensive intervention approaches besides the UCLA model also have support from initial efficacy studies. Examples include the Denver model (Rogers, Hall, Osaki, Reaven, & Herbison, 2001) and intervention programs that emphasize incidental teaching (e.g., Koegel & Koegel, 1995). These interventions may be appropriate candidates for manual development, pilot testing to show feasibility, and multisite RCTs.

Because research continues on components of the UCLA model and other interventions that could be included in a comprehensive intervention program, a challenge is to ensure that manuals are complete and current. Other challenges include improving procedures for monitoring adherence to manuals and obtaining resources to recruit and retain clinicians across sites. For example, in its original form, the UCLA program involves 40 h per week of intervention for 2–3 years. If this intervention were to be undertaken in a large-scale RCT, maintaining the continuity of therapists within and across sites could be a difficult task. Addressing these challenges would largely involve coordination between investigators and funding agencies to identify mechanisms to support developing manuals and conducting feasibility tests. An example of a possible mechanism is the NIMH R34 program (From Intervention Development to Services: Exploratory Research Grants; www.nimh.gov). New funding mechanisms also may need to be developed to deal with the challenges posed by a 2–3 year intervention.

Peer-mediated social skills training is a skills-focused intervention with support from numerous single-case experiments. The first study demonstrated that the intervention might be applicable to children with autism (Strain, Shores, & Timm, 1977). Several subsequent studies tested the efficacy of the intervention in children with autism across different levels of functioning, isolated active ingredients of the intervention (e.g., comparing different kinds of peer communication such as offering to share versus giving assistance), and refined techniques (e.g., adding self-monitoring; Strain & Schwartz, 2001). Hence, this intervention has

completed the phase of initial efficacy studies, though single-case experiments continue to be conducted for further refinement of the procedures. Extending these studies, Odom et al. (1999) developed a manual and conducted a community effectiveness study on peer-mediated social skills training for children with special needs, including but not limited to children with autism. Their study of 98 children is a valuable example of moving from single-case studies to an effectiveness study. This strategy is appropriate when the single-case results are clear about the appropriate subjects for the intervention and the intervention is relatively straightforward. In the stepwise model presented here, there might be additional value in conducting pilot studies with the manual in children with autism and proceeding to a large-scale RCT to prove efficacy for this population. A mechanism such as the R34 program may work well for supporting these activities.

Other examples of skills-focused interventions with extensive support from initial efficacy studies include the Natural Language Paradigm/Pivotal Response Training (Koegel & Koegel, 1995), Functional Communication Training (Carr & Durand, 1985), and video modeling of social skills (Charlop & Milstein, 1989). These interventions, perhaps as components of a more comprehensive program, appear ready for manual development and pilot testing in preparation for RCTs and wider dissemination through effectiveness studies. In most studies of these interventions, outcome measures have assessed selected behavior problems rather than core features of autism. If the interventions are combined into a comprehensive program, appropriate measures of core features would be needed.

Because skills-focused interventions are less complex than comprehensive interventions, development of manuals may be a more straightforward process. Nevertheless, one challenge is likely to be to identify inclusion and exclusion criteria. For example, most studies on peer-mediated social skills have focused on preschoolers, but a question might be whether or not to include school-age children as well. Another question might be whether there are prerequisite skills that participants should have in order to benefit from the intervention (e.g., some proficiency at following instructions or imitating a model). Conclusions about such prerequisite skills may be especially informative for establishing entry criteria for future studies. Also, given the history of individualizing interventions in single-case experiments on the intervention, an additional challenge in manual development is to balance the importance of standardization and individualization.

The foregoing discussion shows that several comprehensive and skills-focused interventions have support from initial efficacy studies and may merit systematic manual development, followed by RCTs with analysis of mediators and moderators, and then effectiveness studies. The discussion also reveals that few studies on either comprehensive or skills-focused interventions included outcome measures assessing core features of autism. Thus, the development of interventions specifically designed to address core features is a priority.

Discussion

This working group extended previous efforts sponsored by the National Research Council and by the NIMH meeting. The overarching goals of the current and previous efforts were to characterize the state of the science and identify methodological challenges in studies of psychosocial interventions for children with autism. In this paper, the working group presents a discursive, step-wise model for developing, validating, and disseminating interventions for children with autism. According to the model, the first phase is to conduct initial efficacy studies such as investigations with single-case designs. Once an intervention has support from these studies, the next step is to assemble intervention manuals and conduct pilot testing to assess the feasibility of implementing the intervention. When possible, it is advantageous to conduct the pilot testing across multiple sites to demonstrate that the manual can be delivered uniformly and is acceptable to clinicians and families. The next important step is the large-scale RCT in well-characterized samples under controlled conditions, with clear entry criteria and outcome measurement. Structured interventions with demonstrated feasibility and efficacy are ready for wider dissemination in community effectiveness studies to show that the intervention can be administered by competent clinicians in typical practice settings.

When the model is applied to the current psychosocial research literature for children with autism, it is clear that no intervention has completed all of these steps. Nonetheless, several interventions, including both comprehensive and more skills-focused programs, have shown encouraging results and appear appropriate for further study. Two recommendations are made:

First, support manual development and testing by dedicating resources to the following activities: (a) assembly of intervention manuals based on inter-

ventions with empirical support, (b) multisite pilot testing of intervention manuals to demonstrate feasibility, (c) large-scale RCTs to assess the efficacy of manualized interventions, demonstrate exportability to different research sites, and obtain information on mediators and moderators, and (e) effectiveness studies of manualized intervention interventions that have shown efficacy.

Second, develop and test new interventions for core features of autism by dedicating resources to the following activities: (a) initial testing of interventions specifically designed to address core features and (b) supporting these efforts by developing and validating outcome measures that are sensitive to change in core features of autism and practical to administer.

References

- Albano, A. M. (2004). *Cognitive behavioral group treatment for social phobia in adolescents (ages 13–17)*. Invited address given at the meeting of the NIH Working Group on Methodological Challenges in Autism Treatment Research, Sacramento, CA.
- Bailey, J. B., & Burch, M. R. (2002). *Research methods in applied behavior analysis*. Thousand Oaks, CA: Sage.
- Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for discussing behavior change* (2nd ed.). New York: Pergamon.
- Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18*, 111–126.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Cristoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L., Williams, D. A., & McCurry, S. (1996). An update on empirically validated therapies. *The Clinical Psychologist, 49*(2), 5–18.
- Charlop, M. H., & Milstein, J. P. (1989). Teaching autistic children conversational speech using video modeling. *Journal of Applied Behavior Analysis, 22*, 275–286.
- DeRosier, M. E. (2004). *Social skills group intervention*. Invited address given at the meeting of the NIH Working Group on Methodological Challenges in Autism Treatment Research, Sacramento, CA.
- Frankel, F. (2004). *Children's friendship training*. Invited address given at the meeting of the NIH Working Group on Methodological Challenges in Autism Treatment Research, Sacramento, CA.
- Goldstein, H. (2002). Communication intervention for children with autism: A review of treatment efficacy. *Journal of Autism and Developmental Disorders, 32*, 373–396.
- Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist-practitioner: Research and accountability in the age of managed care* (2nd ed.). Boston: Allyn & Bacon.
- Hibbs, E. D., Clarke, G., Hechtman, L., Abikoff, H. B., Greenhill, L. L., & Jensen, P. S. (1997). Manual development for the treatment of child and adolescent disorders. *Psychopharmacology Bulletin, 33*, 619–629.
- Hollander, E., Robinson, R., & Compton, D. (2004). New developments in autism clinical trials. *CNS Spectrums, 9*, 20–21.

- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Council for Exceptional Children, 71*, 165–179.
- Huitema, B. E., & McKean, J. W. (1998). Irrelevant autocorrelation in least-squares intervention models. *Psychological Methods, 3*, 104–116.
- Kasari, C. (2002). Assessing change in early intervention programs for children with autism. *Journal of Autism and Developmental Disorders, 32*, 447–462.
- Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology and Psychiatry, 47*, 611–620.
- Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.
- Kazdin, A. E. (2002). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.
- Kendall, P. C., & Chu, B. C. (2000). Retrospective self-reports of therapist flexibility in a manual-based treatment for youths with anxiety disorders. *Journal of Clinical Child Psychology, 29*, 209–220.
- Kendall, P. C., Chu, B., Gifford, A., Hayes, C., & Nauta, M. (1998). Breathing life into a manual: Flexibility and creativity with manual-based treatments. *Cognitive and Behavioral Practice, 5*, 177–198.
- Kendall, P. C., Flannery-Schroeder, E. C., & Ford, J. D. (1999). Therapy outcome research methods. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck, (Eds.) *Handbook of research methods in clinical psychology*, 2nd ed. (pp. 330–363). New York: Wiley.
- Koegel, R. L., & Koegel, L. K. (1995). *Teaching children with autism: Strategies for initiating positive interactions and improving learning opportunities*. Baltimore: Paul H. Brookes.
- Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59*, 877–883.
- Leaf, R., & McEachin, J. (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism*. New York: DRL Books.
- Lord, C., Bristol-Power, M., Cafiero, J. M., Filipek, P. A., Gallagher, J. J., Harris, S. L., Leslie, A. M., McGee, G. M., Odom, S. L., Rogers, S. J., Volkmar, F. R., & Wetherby, A. M. (Eds.). (2002). JADD special issue: NAS workshop papers. *Journal of Autism and Developmental Disorders, 32*, 349–508.
- Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., Dawson, G., Durand, V. M., Grossman, L., Guthrie, D., Harris, S., Kasari, C., Marcus, L., Murphy, S., Odon, S., Pickles, A., Scahill, L., Shaw, E., Siegel, B., Sigman, M., Stone, W., Smith, T., Yoder, P. (2005). Challenges in evaluating psychosocial interventions for autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 35*, 695–708.
- Lovaas, O. I. (1977). *The autistic child: Language training through behavior modification*. New York: Irvington.
- Lovaas, O. I. (2003). *Teaching individuals with developmental delays: Basic intervention techniques*. Austin, TX: Pro-ed.
- Lovaas, O. I., & Smith, T. (2003). Early and intensive behavioral intervention in autism. In A. E. Kazdin, & J. Weisz, (Eds.), *Evidence-based psychotherapies for children and youth* (pp. 325–340). New York: Guilford.
- Lovaas, O., Koegel, R., Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin, 86*, 1236–1254.
- Lovaas, O. I., Ackerman, A., Alexander, D., Firestone, P., Perkins, M., Young, D. B., Carr, E. G., & Newsom, C. (1981). *Teaching developmentally disabled children: The me book*. Baltimore, MD: University Park Press.
- Lovaas, O. I., Berberich, J. P., Perloff, B. F., & Schaeffer, B. (1966). Acquisition of imitative speech in schizophrenic children. *Science, 151*, 705–707.
- MacMahon, R. J. (2004). *The Fast Track Project*. Invited address given at the meeting of the NIH Working Group on Methodological Challenges in Autism Treatment Research, Sacramento, CA.
- McClannahan, L. E., & Krantz, P. J. (1999). *Activity schedules for children with autism: Teaching independent behavior*. New Jersey: Woodbine House.
- McConnell, S. (2002). Interventions to facilitate social interaction for young children with autism: Review of available research and recommendations for educational intervention and future research. *Journal of Autism and Developmental Disorders, 32*, 351–372.
- Medical Research Council. (2000). *A framework for development and evaluation of RCTs for complex interventions to improve health*. London, UK: Author.
- Meinert, C. L. (1986). *Clinical trials: design, conduct, and analysis*. New York, NY: Oxford University.
- National Research Council. (2001). *Educating children with autism*. Lord, C., & McGee, J. P. (Eds.), Washington DC: National Academy Press.
- Odom, S. L., McConnell, S. R., McEvoy, M. A., Peterson, C., Ostrosky, M., Chandler, L. K., Spicuzza, R. J., Skellinger, A., Creighton, M., & Favazza, P. C. (1999). Relative effects of interventions supporting the social competence of young children with disabilities. *Topics in Early Childhood Special Education, 19*, 75–91.
- Owens, E. B., Hinshaw, S. P., Kraemer, H. C., Arnold, L. E., Abikoff, H. B., Cantwell, D. P., Conners, C. K., Elliott, G., Greenhill, L. L., Hechtman, L., Hoza, B., Jensen, P. S., March, J. S., Newcorn, J. H., Pelham, W. E., Severe, J. B., Swanson, J. M., Vitiello, B., Wells, K. C., & Wigal, T. (2003). Which treatment for whom for ADHD? Moderators of treatment response in the MTA. *Journal of Consulting & Clinical Psychology, 71*, 540–552.
- Rogers, S. J., Hall, T., Osaki, D., Reaven, J., & Herbison, J. (2001). The Denver model: A comprehensive, integrated educational approach to young children with autism and their families. In J. S. Handleman, & S. L. Harris, (Eds.), *Preschool education programs for children with autism*, 2nd ed. (pp. 95–134). Austin, TX: Pro-ed.
- Scahill, L., & Lord, C. (2004). Subject selection and characterization in clinical trials in children with autism. *CNS Spectrums, 9*, 22–32.
- Smith, T., Buch, G. A., & Evslin, T. (2000a). Parent-directed, intensive early intervention for children with pervasive developmental disorder. *Research in Developmental Disabilities, 21*, 297–309.
- Smith, T., Groen, A. D., & Wynn, J. W. (2000b). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation, 105*, 269–285.
- Strain, P. S., & Schwartz, I. (2001). ABA and the development of meaningful social relations for young children with autism. *Focus on Autism and Related Disorders, 16*, 120–128.
- Strain, P. S., Shores, R. E., & Timm, M. A. (1977). Effects of peer social initiations on the behavior of withdrawn preschool children. *Journal of Applied Behavior Analysis, 10*, 289–298.

- Taylor, B., & McDonough, K. A. (1996). Selecting teaching programs. In C. Maurice (Ed.), *Behavioral interventions for young children with autism: A manual for parents and professionals* (pp. 63–177). Austin, TX: Pro-ed.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*, 620–630.
- Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist, 47*, 1578–1585.
- Whalen, C., & Schreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 44*, 456–468.
- Wolery, M. (2004). *Procedural fidelity: Lessons from single-subject experimental research*. Invited address given at the meeting of the NIH Working Group on Methodological Challenges in Autism Treatment Research, Sacramento, CA.
- Woody, S. R., & Sanderson, W. C. (1998). *Manuals for empirically supported treatments: 1998 update*. Washington, DC: Division 12, American Psychological Association.