

Current Progress and Future Plans for Developing Effective Treatments: Comments and Perspectives

Alan E. Kazdin

Department of Psychology, Yale University

Philip C. Kendall

Department of Psychology, Temple University

Examined the conclusions and implications of articles in this special issue. Treatments can be differentiated on their empirical basis and, for the problems reviewed, one can identify treatments of choice. To build on the documented advances, we provide a blueprint for progress in treatment research. The blueprint focuses on a series of steps that involve conceptualization of clinical dysfunction and treatment, supportive research on these conceptual views, careful specification of treatment, evaluation of treatment outcome, and tests of mediators and moderators of treatment. To improve research, we recommend expanding assessment, addressing a broad range of questions about treatment, attending to measures of clinical significance, replicating key findings, and investigating the transportability of the findings to service-oriented clinical settings.

Methods of reviewing the literature on the effects of psychological treatments have evolved along with the empirical studies they encompass. First, the dominant review method had been the narrative or qualitative review in which efforts are made to sift through individual studies and to summarize what can and cannot be supported. This method relies on the author's evaluation of the strength and merit of the studies and, of course, was the first to be used for reviewing the evidence for therapies for children and adolescents alone (e.g., Levitt, 1957, 1963) and in combination with treatments for adults (e.g., Eysenck, 1952). Second, more recently quantitative reviews have dominated. Meta-analyses (Smith & Glass, 1977) and, to a lesser extent, box-score methods (Luborsky, Singer, & Luborsky, 1975) were applied to systematically codify and combine studies. By coding characteristics of the studies (e.g., type of treatment, problem, quality of the methodology), summary conclusions were drawn and novel hypotheses were tested. Early meta-analyses focused on adult and child treatments (Smith & Glass, 1977; Smith, Glass, & Miller, 1980) and child treatments by themselves (Casey & Berman, 1985; Weisz, Weiss, Alicke, & Klotz, 1987). Third, the focus on validated or empirically supported treatment (EST) can be considered as another way to review the literature. In this type of review, the evidence is examined from the perspec-

tive of a common set of criteria to select individual studies. The criteria are intended to classify treatments based on the extent to which they are empirically supported (i.e., as having shown treatment produces systematic change in controlled trials). The review asks the general question, What is known to be effective (i.e., empirically supported) for a particular clinical problem based on the extent to which studies are available that meet specific criteria? How the studies are reviewed (i.e., by qualitative or quantitative methods; Chambless et al., 1996; Nathan & Gorman, 1998; Roth & Fonagy, 1996; Task Force on Promotion and Dissemination of Psychological Procedures, 1995) is not as critical as the focus or purpose of the review.

The reviews in this special issue identify ESTs and place treatments into one of three mutually exclusive categories: well established; probably efficacious; or, by default, neither. The criteria for this classification system, specified at the outset of the series (Lonigan, Elbert, & Johnson, this issue), have evolved (Chambless & Hollon, 1998; Chambless et al., 1996; Task Force on Promotion and Dissemination of Psychological Procedures, 1995) and may continue to evolve. The authors of the articles have carefully sifted through the child and adolescent intervention literatures on depression, anxiety, autism, conduct problems, and attention-deficit/hyperactivity disorder and applied the criteria to identify what treatments might be probably efficacious or well established. Each article presents a valuable analysis of treatments in the respective problem areas—a clear statement about the current status of treatment—and identifies key issues or qualifiers.

Requests for reprints should be sent to Alan E. Kazdin, Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520-8205; or Philip C. Kendall, Department of Psychology, Weisz Hall, Temple University, Philadelphia, PA 19122.

From the set of articles as a whole, one can identify salient issues beyond the conclusions for specific treatments. First, one can differentiate among many treatments in use based on the status of their evidence. Clearly, the criteria that were invoked for deciding whether a treatment is well established or probably efficacious can be questioned and multiple issues affecting the classification were identified (Weisz & Hawley, this issue). Nevertheless, differentiating among treatments with reasonable criteria is extremely important. Among techniques that have been studied, not all have been shown to work or to work equally well (i.e., not everyone has won or has all earned prizes; Luborsky et al., 1975). Some treatments emerge as more established than do others.

Second, based on the articles, when children come to treatment for depression, anxiety, conduct problems, or attention deficit hyperactivity disorder, there are specific treatments that would be interventions of choice. Not applying one of the treatments that the authors in this series have identified as probably efficacious or well established would seem to require a very strong justification. The conclusions reached in the articles could change as (a) new or other treatments are investigated, (b) moderators of treatment outcome become clarified, (c) comparative studies of well-established treatments are conducted, and (d) criteria for efficacious and established treatments evolve (e.g., using clinical significance or magnitude of effects as criteria in addition to, or in place of, statistically significant group differences). Yet at this point in time, specific treatments are available with evidence in their behalf and should be considered as the treatments of choice when the problems are presented clinically.

Third, in focusing on probably efficacious and well-established treatments, the articles unwittingly draw attention to a third category—namely, “neither” or “none of the above.” *Neither* only means the criteria for the other categories were not met, but there may be good reason to make distinctions here. For example, as Rogers (this issue) noted in her article on interventions for autism, there are still promising treatments with empirical evidence in their behalf that fall into the “neither” category. There may be value in addressing this category further because treatments that are presently neither well established nor probably efficacious may still have value. Some treatments have not been investigated; others have been investigated but not shown to influence outcome in a few studies; and yet others have been shown to be effective, but below the threshold of probably efficacious (Rogers, this issue). There would be value in delineating the status of such treatments because of the implications for next steps in research and for clinical practice. As an illustration, one could say that nutritional counseling is neither probably efficacious nor well established for the treatment of hyperactivity. But we would want to go farther in light of sev-

eral studies showing that diet does not materially affect key symptoms. Several failed trials may foster stronger statements than merely noting this is in the “neither” category.

Finally, the articles in this issue convey that although we know something about preferred treatments, we do not have complete knowledge about “exactly what works,” leaving aside why it works. There are several hundred outcome studies of child and adolescent therapy (see Durlak, Wells, Cotten, & Johnson, 1995). It is irresistible to raise the question, Did we need hundreds of studies to reach the conclusions that these articles have reached? Apparently, the answer is yes. But the question makes more salient the concerns about the next hundred studies or perhaps the next several hundred studies. What is needed to move the field forward in a more focused fashion? Can we achieve the objectives of developing and identifying well-established treatments? Key issues to improve the quality of studies and reporting of the data, nicely outlined by Weisz and Hawley (this issue) are important areas with which we concur. Addressing deficits in existing research (e.g., expanding assessments, focusing on magnitude of effects), although valuable, does not ensure that we will obtain the knowledge required for developing effective treatments. The purpose of our comments is to move to a somewhat different plane in light of all of the areas that were reviewed and to identify how we can make further progress to improve the scope and strength of conclusions and how to establish that treatments are efficacious and effective.

A Blueprint for Further Progress

Elaborating the Nature and Characteristics of Treatment

The current focus on ESTs emphasizes outcome studies that show that treatment is more effective than a control or other treatment condition. This first step is obviously important, and none of the comments we make in any way gainsays the essential nature of rigorous and replicated outcome evidence. Yet, we seek a more complete picture of what constitutes effective treatment, how and why it operates, and for whom it is optimally effective. To achieve this, we propose a blueprint or broad plan for research. This blueprint includes a portfolio of research on treatment including, but extending beyond methodologically rigorous outcome studies.

Table 1 identifies key steps to evaluate different facets of treatment, how treatment relates to what is known about the disorder or clinical problem, and how and to whom treatment can be applied to achieve optimal gains. The notion of “steps” is useful to convey a progression of research, but the progression does not nec-

Table 1. *Steps Toward Developing Effective Treatment*

1.	Conceptualization of the dysfunction: Propose key areas, processes, and mechanisms that relate to the development, onset, and escalation of dysfunction.
2.	Research on processes related to dysfunction: Efforts to test proposed processes in relation to the dysfunction.
3.	Conceptualization of treatment: Propose key areas, processes, and mechanisms through which treatment may achieve its effects and how the procedures relate to these procedures.
4.	Specification of treatment: Operationalize the procedures, preferably in manual form, that identify how one changes the key processes. Provide material to codify the procedures and the permit of treatment integrity and to replicate in research and practice.
5.	Tests of treatment outcome: Direct tests of the impact of treatment drawing on diverse designs (e.g., open [uncontrolled] studies, single-case designs, and full-fledged clinical trials) and types of studies (e.g., dismantling, parametric studies, and comparative outcome studies) are relevant.
6.	Tests of treatment processes: Studies to identify whether the intervention techniques, methods, and procedures within treatment actually affect those processes that are critical to the model.
7.	Tests of the boundary conditions and moderators: Examination of the child, parent, family, therapist, and other factors with which treatment interacts. The boundary conditions or limits of application are identified through interactions of Treatment $\times$ Diverse Attributes within empirical tests.

essarily require a fixed sequence of studies. Rather, over a period of years of research, the steps refer to the areas that should be covered so that the foundation of knowledge is systematic and builds in a cohesive way. We argue for interventions that employ strategies that are linked to knowledge about the nature of the disorder. The sections that follow address the steps noted in Table 1.

Conceptualization of the clinical dysfunction.

Conceptualization of the clinical dysfunction refers to factors that lead or contribute to the pattern of functioning we wish to change, what processes are involved, and how these processes emerge or operate. A conceptual model of the dysfunction may encompass concurrent correlates, risk, protective, and causal factors. The model may be quite circumscribed to address some facet of the problem or a subtype of many individuals who show the problem. Considering what is known about factors related to onset, maintenance, termination, and recurrence of the problem, we are likely to learn more about what might serve as effective interventions (prevention and treatment). As a beginning, a conceptualization (preferably data based) of onset and course of clinical dysfunction or the specific problem domain is a critical step.

Research on processes related to dysfunction.

This step consists of testing directly the processes hypothesized to be implicated in the clinical problem (i.e.,

tests of the conceptual model). For example, if cognition is proposed to play a pivotal role in the onset or maintenance of a disorder or pattern of functioning, direct tests are needed and should be part of the foundation leading toward treatment development. Treatment studies can test conceptual models insofar as they provide evidence that key processes change over the course of therapy and that these changes mediate or at least relate to outcome. However, with some exceptions (e.g., Treadwell & Kendall, 1996), this is an infrequent focus of child and adolescent therapy research.

Tests of the conceptual model of the clinical problem need not rely on treatment studies. For example, if family processes are proposed to account for or contribute to the problem, direct tests of these processes in cross-sectional and longitudinal studies of the family and of the onset, course, and subtypes of the disorder or clinical problem are very valuable. Research on the nature of the clinical problem is likely to identify various subtypes, multiple paths leading to a similar onset, and key moderators that alter the profile of risk factors leading to the disorder. These findings may also have implications for providing varied forms of treatment. For some treatments (e.g., learning-based therapies, medications), animal studies may be a viable option to evaluate key processes (e.g., neurotransmitter uptake). As a general rule, research on the nature (e.g., characteristics, course, mediators) of clinical problems should be more highly integrated with or used as a resource for generating psychosocial treatments.

Conceptualization of treatment. Treatment should have conceptual underpinnings (i.e., explicit views about what treatment is designed to accomplish and through what processes). The guiding question is, How does this treatment achieve change? Answering this question is pivotal to development and optimization of treatment effects. In many cases, the answer may involve basic psychological processes (e.g., memory, learning, information processing, and conflict resolution). It is no longer sufficient to provide global conceptual views (orientation) about what to do in the sessions. Rather, to ensure progress, those processes (e.g., psychodynamic, cognitive, or familial) that are considered to be responsible for therapeutic change ought to be assessed directly. Moreover, when seeking information about change processes, it is wise for more than one hypothesized process to be evaluated within a single study, in comparative form (Kendall & Flannery-Schroeder, 1998).

It may be that treatment directly addresses those processes considered to be involved in the development of the problem. For example, inept discipline practices influence the development of aggressive behavior in children (Dodge, Pettit, & Bates, 1994; Patterson, 1982; Patterson, Reid, & Dishion, 1992) and parent

management training directly alters these practices and alters aggressive child behavior. In other treatments, the focus may draw on change processes that may or may not address those processes involved in development of the problem. For example, learning models may generate treatments that provide special learning experiences but not necessarily have any implications for the etiology of the problem.

Specification of treatment. Treatments should be operationalized, preferably in manual form, so that the integrity of treatment can be evaluated, the material learned from treatment trials can be codified, and the treatment procedures can be replicated (Wilson, in press). Manuals specify guidelines, session foci, and content; the progression of treatment; and when and how to continue particular practices, tasks, sessions, and themes. The application can be individualized (e.g., Kendall, Chu, Gifford, Hayes, & Nauta, in press), but the essentials of generic treatment can be in the manuals.

Manuals can encompass all aspects of treatment that can be documented. Some of the information may seem relatively trivial, such as the materials used in a session, instructions to explain treatment, and forms used to document the sessions. The effective application of treatment may not require complete adherence to each of the points that is specified. At the same time, replication in research and clinical practice will require knowing precisely what was done. Evaluation of the extent to which any particular practice is essential or optimal then can be researched (Kendall, in press) if one doubts the importance of that facet of treatment.

The value of treatment manuals is not unanimously embraced among researchers (e.g., Gaston & Gagnon, 1996; Strupp & Anderson, 1997). At this time, there is no alternative to manualization of treatment, although the meaning of *manualization* and how and to what extent treatment is codified have a range of options. Yet manualization in some form is essential in light of the likely alternative—namely, clinical practice in which there is improvisation and vast individual therapist differences in training, preferences for treatment, and judgment as a guide. A balance of manuals, improvisation, and individual therapist differences may prove to be important. However, research and clinical practice begin with specification and documentation of the critical facets of the intervention.

Tests of treatment outcome. To develop and advance treatment, direct tests of clinical outcomes are central. A wide range of treatment tests (e.g., open [uncontrolled] studies, single-case designs, randomized clinical trials) can provide evidence that change is pro-

duced and that treatment is responsible for the change. Direct tests of treatment are the most common forms of research of child and adolescent therapy. The articles in this issue by and large consist of reviews of outcome evidence. Clearly, of all steps in the research progression, there has been an emphasis of treatment outcome studies of the type represented in this step. There are, of course, several different types of outcome questions based on whether, for example, variations of parameters of treatment, components of treatment, or combinations of treatment influence outcome, as discussed later.

Tests of treatment processes. Do the intervention techniques, methods, and procedures within treatment sessions actually affect those processes that are considered to be critical to the treatment model? The processes refer to those facets in the conceptualization of treatment that are considered to produce, facilitate, or mediate change (see Holmbeck, 1997). It may be that parenting practices are critical to change in the oppositional child. If so, these practices should change during treatment. Also, the extent to which these practices were problematic to begin with and to which they changed over treatment should relate specifically to child improvement. Among varied treatment models, relationship, alliance, and bonding between the child (or parent) and therapist may be viewed as a critical process to achieve change. Once an intervention is found to be efficacious, tests of such processes are essential to ensure we understand why and how treatment works.

Tests of boundary conditions and moderators. Tests of the conditions on which effective interventions depend are critical. What are the boundary conditions for effective application of treatment or the variables (moderators) that influence efficacy and effectiveness? In child and adolescent treatment, it is conceivable that child, parent, family, and contextual influences affect development of clinical dysfunction and the likelihood that critical treatment processes will be altered (see Holmbeck, 1997). Although moderators of therapy are not studied very often in child and adolescent therapy, one can point to a number of demonstrations in which the effectiveness of treatment depended on such other factors as severity of child dysfunction, parent stress and psychopathology, and socioeconomic disadvantage, to mention a few (Kazdin, 1995; Kazdin & Crowley, 1997; Webster-Stratton & Hammond, 1990). What treatment works for whom and under what conditions has been a phrase that has guided discussions of research (Kiesler, 1966) but has not served as a basis for very many studies in child and adolescent therapy (Kazdin, Bass, Ayers, & Rodgers, 1990).

From Concepts to Research

The blueprint we propose is to organize and steer investigations, but it may be sufficiently abstract and hence not very helpful for designing investigations. Yet, several researchable questions are encompassed by the model, as presented in Table 2. Progress can be made by ensuring that for a given treatment and clinical problem these questions are addressed systematically.

The primary focus of child and adolescent treatment research has been on basic outcome questions, specifically, Questions 1 and 4 in Table 2. Reviewers have consistently lamented the paucity of research on other aspects of treatment, with especially scant research on mediators and moderators (Durlak et al., 1995; Kazdin, et al., 1990). As the previous articles in this issue have carefully documented, important conclusions can be reached already by the available outcome studies.

Although each of the questions in Table 2 is important, we are arguing more broadly for systematic efforts to address the range of questions for a given technique. Presumably, if research proceeds systematically, the priority of the questions that should be considered will change over time. For example, following replicated studies on the impact of treatment, as emphasized in the effort (this special issue) to identify probably efficacious and well-established treatments, we want the emphasis to shift to studies that address other questions that are likely to be neglected. This shift is not an idle or academic exercise. Clinical impact of treatment will derive not only from outcome studies, but also from understanding the processes through which treatment works and the individuals for whom treatment is and is not likely to work.

General Comments

Because much of contemporary outcome research has focused on a narrow set of questions, there are not many exemplars in child and adolescent therapy that illustrate the blueprint in full bloom. Parent management training is one of the better examples. Conceptualization of family processes (inept discipline practices) that promote aggressive behavior has been followed with supportive research that charts the nature and progression of parent-child interaction patterns (Patterson, 1982; Patterson et al., 1992). Also, changing inept parenting practices alters child behavior and the extent to which parents change in these practices influences outcome (i.e., dose-response relation). These findings suggest, but do not prove, that parent behavior mediates therapeutic change. Moreover, child, parent, and family characteristics have been shown to moderate treatment effects (see Dishion, Patterson, & Kavanagh, 1992; Forgatch, 1991; Kazdin, 1997; Patterson et al., 1992; Webster-Stratton, 1996). Many of the techniques

Table 2. *Range of Questions to Guide Treatment Research*

1. What is the impact of treatment relative to no-treatment?
2. What components contribute to change?
3. What parameters can be varied to improve outcome?
4. How effective is this treatment relative to other treatments for this problem?
5. What treatments can be added (combined treatments) to optimize change?
6. What processes within or during treatment influence (mediate) outcome?
7. What child, parent, family, and contextual features influence (moderate) outcome?

used in parent training (e.g., reinforcement in various forms, time-out from reinforcement, ignoring) have rather extensive basic and applied research that provide empirical underpinnings of the procedures (i.e., experimental and applied analysis of behavior, respectively). Parent management training alone or combined with other procedures has been identified as an intervention with outcome evidence in its behalf in the present series of articles (Brestan & Eyberg, this issue; Pelham, Wheeler, & Chronis, this issue; Rogers, this issue). Research on parent management training has provided a model about how the problem may develop for many children, how many domains of functioning are affected, what central foci of treatment are needed in cases where inept practices are evident, and what to monitor during treatment (e.g., parent progress) if the outcomes are to be achieved. Parent management training for oppositional and conduct problems illustrates an area where several facets of our blueprint are addressed.

The blueprint we have outlined reflects the interplay of theory and research. However, more than that, the steps are intended to reflect a progression of research. The goal is to permit evaluation of the field in a quite different way from the usual literature reviews. Current reviews focus on what has been accomplished in research and react to what is provided by the available studies. Such reviews, although valuable, suffer from the fact that the accumulation of studies is haphazard. What we can say is based on tracing the path through which the research has wandered. If we, as investigators, have ignored many critical questions about therapy or our studies were routinely subject to design practices that are limiting (e.g., narrow range of outcomes), then the conclusions will be commensurately limited.

A more proactive stance for the field is one that begins with a model of what we need to know in moving from ignorance to knowledge about effective and disseminable interventions. Specifying some of the critical steps and movement from one step to another is likely to lead to much greater progress and, as important, to allow us to assess in subsequent reviews how much progress has been made. The blueprint we suggest offers a way to chart a course for treatment research. Subsequently, a systematic review of the range

of issues and questions that are addressed over time permits us to identify where we are on that voyage.

Expanding Research Methods and Foci

The conclusions reached about whether treatments are probably efficacious or well established can depend on a number of conditions related to how the psychological therapy research was conducted. As a way of identifying these conditions, we refer to them as potential methodological moderators—namely, conditions of the design (e.g., assessment, selection of participants) on which the conclusions might well depend. We identify key points here with the hope of fostering improved outcome research.

Assessment

Multiple outcome domains and measures. Children and adolescents are referred to treatment because of dysfunction in some domain and our attention as researchers is immediately drawn to symptoms (i.e., presenting problems and target complaints). Clearly, symptom reduction is a relevant and important outcome. Indeed, the vast majority of therapy studies recognize this and rely heavily on symptom reduction as the primary or sole outcome criterion (Kazdin et al., 1990). There are many other outcomes that are critically important because of their significance to the child; family; and, in many cases, society at large. Table 3 includes domains that are relevant for evaluating therapy outcomes.

The range of domains included in treatment should be broadened to address those areas covered in the table. Few studies go beyond symptom measures, although ratings by different informants in different settings (e.g., parents and teachers, home and school) may

be included. Perhaps we have embraced too eagerly that the goal of therapy is to reduce symptoms or focus on that almost exclusively. We know from longitudinal studies of child dysfunction that prognosis and long-term outcomes depend on many domains, in addition to symptom type, severity, and onset (e.g., Robins & Rutter, 1990). Consequently, in evaluating the effectiveness of treatment, evidence across a number of domains is likely to be critical.

Among the many domains, perhaps special emphasis should be given to impairment (i.e., the extent to which the individual's functioning in everyday life is impeded). For a child or adolescent, meeting role demands at home and at school; interacting prosocially and adaptively with others; and being restricted in the settings, situations, and experiences in which he or she can function can vary considerably. Impairment is related to, but distinguishable from, symptoms and meeting criteria for a disorder (Sanford, Offord, Boyle, Peace, & Racine, 1992). The significance of impairment stems from its relation to seeking treatment. Impairment, more than symptoms, predicts the likelihood that a child is referred for mental health services (Bird et al., 1990). Also, among those treated, the level of impairment at the end of treatment predicts the likelihood of relapse (e.g., Lewinsohn, Seeley, Hibbard, Rhode, & Sack, 1996). Impairment is a key domain, and measures are readily available for use in treatment studies (e.g., Hodges, Bickman, Ring-Kurtz, & Rieter, 1991; John, Gammon, Prusoff, & Warner, 1987; Shaffer et al., 1983). By singling out impairment, we do not intend to imply that one criterion (impairment) should replace another (symptoms). Just the opposite, multiple outcomes should be considered in identifying empirically supported treatments.

Broadening the scope of assessment batteries is the primary message to improve outcome research (Kendall & Flannery-Schroeder, 1998). A broader scope permits the evaluation of changes outside the presenting symptoms as well as change on the targeted area. A broader scope, when measures tap different theoretical systems, permits analyses of what changed and what mediated change from both within and outside the guiding theoretical orientation of the treatment. For example, it would be quite informative to know that an effective treatment was mediated by changes consistent with one theoretical conceptualization of the problem but not consistent with another.

There is a more subtle point worth noting in passing. In evaluating whether treatments are probably efficacious or well established, it is quite conceivable that the domains and the level of change that are relevant or, most important, will vary by problem area (e.g., autism vs. depression). Also, among individuals referred for a given problem (e.g., anxiety disorder), it is possible that not all domains (symptoms, social functioning, and impairment) are equally relevant. At this point, these are

Table 3. *Criteria to Evaluate Treatment Outcome*

- | | |
|----|---|
| 1. | Child Functioning |
| | a. Symptoms |
| | b. Impairment |
| | c. Prosocial competence |
| | d. Academic functioning |
| | e. Peer relationships/social functioning |
| 2. | Parent and family functioning |
| | a. Dysfunction (e.g., symptoms) |
| | b. Contextual influences (e.g., stress, quality of life) |
| | c. Conditions that promote adaptation (e.g., family support, quantity and quality time) |
| 3. | Social impact measures |
| | a. Consequences on systems (e.g., school activities, attendance, truancy) |
| | b. Service use (e.g., reductions in special services or needed services) |
| | c. Monetary costs and gains (e.g., on or off social assistance, costs for services) |

nuances that focus on fine-grained distinctions. The initial challenge is to evaluate a range of domains as the basis of treatment evaluation. More than evaluating changes on a measure, attention to a profile of functioning would be an important line to pursue.

Multiple domains hint at a broader problem with multiple measures. There are many treatment outcomes and criteria that are relevant to the conclusions reached about treatment, and these can vary in terms of perspective or informant (Achenbach, McConaughy, & Howell, 1987; Strupp & Hadley, 1977), types of outcome measure (e.g., Kazdin & Wilson, 1978), and facets of the problem or disorder (e.g., relapse, recovery, remission; Frank et al., 1991). The striking feature of different outcomes is that valid measures of client change do not necessarily correlate highly with each other. For example, parent, teacher, and child ratings of symptoms do not correlate very highly with each other or very well with other indices such as archival records (e.g., arrest, truancy) or overt behavioral observations at home (e.g., Achenbach et al., 1987; Patterson, Dishion, & Chamberlain, 1993). This means that conclusions reached about treatment techniques, including the relative effectiveness of different techniques, may vary by the measures (e.g., Szapocznik et al., 1989).

The implications of these comments for identifying effective treatments are major. Probably efficacious and well-established treatments may not differ from no-treatment or nonspecific treatment control conditions on some measures but will differ on others. It is important to include multiple measures and assessment of multiple domains to permit evaluation of the impact across a range of domains and methods of assessment. At some point, there may be a need to prioritize measures or to decide the domains on which classification of efficacy and effectiveness will be based.

Timing of assessment. Conclusions about the efficacy and effectiveness of treatment are based on the status of participants at posttreatment. As a number of authors in this issue have noted, with few exceptions (e.g., 3.3-year follow-up; see Kendall & Southam-Gerow, 1996), there is a paucity of follow-up data. It is still the case that most outcome studies with children and adolescents do not assess follow-up, and when follow-up is gathered the duration is a matter of months after posttreatment (Kazdin et al., 1990; Weisz, Weiss, Han, Granger, & Morton, 1995). Yet, in the context of identifying ESTs, follow-up data are essential. The conclusions about the efficacy and effectiveness of treatment relative to no-treatment or about the relative efficacy of two or more treatments can vary at posttreatment and at follow-up and over the course of two different follow-up periods (e.g., Kolvin et al., 1981; Meyers, Graves, Whelan, & Barclay, 1996; Newman, Kenardy, Herman, & Taylor, 1997). This fact means

that, in principle, a treatment identified as probably efficacious or well established based on posttreatment data might not be classified in the same way if follow-up data were obtained. The data are much too sparse to make any sweeping claims except perhaps to note that conclusions about treatment efficacy drawn in the previous articles usually are based on treatment effects assessed at posttreatment.

These comments might imply that some treatments may be efficacious in terms of short-term outcome but may not be so classified with long-term follow-up. Even if this were the case, short-term effects can be very important, and treatments that are efficacious (e.g., different from no-treatment) at posttreatment but not at follow-up are valuable. For example, most children who are oppositional in early childhood or who wet their beds (e.g., to age 4 or 5) do not retain these behaviors into adulthood. For these behaviors, an efficacious intervention, based on posttreatment data, might be worth identifying even if 5-year follow-up showed no difference between treatment and no-treatment groups on certain measures. Nevertheless, in identifying efficacious and effective treatments, we underscore the importance of providing follow-up data.

Clinical Significance

When striving to identify effective treatments, it is important to consider in greater depth the extent to which changes on various outcome measures translate to palpable benefits or meaningful changes for children and families. Measures commonly used (e.g., symptom checklist and rating scales, completed by parents, teachers, and children) reflect constructs about which we have keen interest. Yet, we sometimes have little evidence that these measures translate to meaningful change or clear benefits in everyday life. Evaluation of the clinical significance of change is intended to redress this concern by reflecting the extent to which the changes over the course of treatment are clinically meaningful or important.

Several methods have been used to evaluate clinical significance, including (a) the extent to which or whether initially deviant children return to normative (nondeviant) levels of functioning on standardized measures (e.g., Kazdin, 1977; Kendall & Grove, 1988), (b) the amount changes individuals make from pre- to posttreatment, in standard deviation units on the outcome measures (e.g., Jacobson & Revenstorf, 1988), (c) whether children no longer meet criteria for a diagnosis at the end of treatment, (d) evaluation by individuals in contact with the client who judge whether the changes are important or make a difference (Wolf, 1978), and (e) changes on social impact measures (e.g., arrest, truancy, rehospitalization; Kazdin, 1998). Use of these methods is important and is to be encouraged

(Kendall, 1997). Nevertheless, our comments, although not intended to detract from their adoption and expansion, do suggest that the meaning of clinically significant change is by no means obvious.

Clinical significance has been defined by researchers as a way to move beyond merely showing statistically significant changes. Yet, whether individuals who have made a "clinically significant change," as commonly defined, are actually better off has not been firmly established. Stated another way, clinically significant change has been operationalized by researchers in ways that have not fully appreciated what might be called *clinical validity*, (i.e., establishing that the measures make a difference to persons who receive treatment or live with persons that do or that are noticeable). Conclusions about treatments that are empirically supported, whether from individual studies or meta-analyses, are based on effecting changes that are significant, usually statistically, but sometimes clinically. In the prioritization of treatments, we certainly want to delineate those interventions that have produced replicable and statistically reliable changes. Yet, we also want to have some idea about whether the changes are meaningfully reflected in everyday life. Interpretation of the measures and clinically significant change on the measures warrants further validation.

Next Steps for Research

The next generation of treatment outcome research will provide valuable information for the science and practice of child and adolescent therapy. True progress will require building on past successes and prior knowledge; avoiding past mistakes and nonproductive paths; and, no doubt, incorporation of a new generation of ideas and issues. To pace future research, we see merit in at least two steps: replication and transportability.

Taking a closer look at the conclusions we seek to reach requires replication. The two types of replication for consideration are *identical* and *conceptual*. In an identical replication the investigator undertakes to conduct an exact copy of the prior study (with new participants), using the same methods and measures. The results inform us about the ability of the treatment to again have certain effects and can help to establish a result that may have been in question. A conceptual replication mirrors the hypothesis of the initial study but uses different measures (or methods) to examine the outcomes. The concept (treatment) that is tested is the same, but it is gone about in a different manner. Although an identical replication helps establish a result, a conceptual replication, where the results cannot be considered an artifact of the methodology, adds to the weight of the confidence in the underlying theory or mechanism of change. Replications are required not only for the ultimate determination of the treatments that qualify as probably efficacious and well estab-

lished, but also for advancement of the science of treatment development.

Given that a treatment applied in a research setting can be considered efficacious, we must next move to determine if that intervention will be comparably efficacious/effective in a service-oriented clinical setting. Although difficulties inherent in the process of transportability (Kendall & Southam-Gerow, 1995) may explain the dearth of research, one study with adult panic sufferers (Wade, Treat, & Stuart, 1998) supports the transportability of a manual-based treatment to a clinical setting. Potential barriers and real speed bumps exist along the path from research to practice, but the future is nevertheless well served by the study and actual transport of efficacious interventions and well-established treatments to the world of clinical practice.

Conclusions

This special issue on which we have commented has provided important information about the current status of treatments in the respective areas that were reviewed. Our comments are intended to be constructive: to build on the articles by clarifying the overall mission or goals of treatment research and to convey key issues to improve the yield from studies. We proposed a blueprint or plan for developing and evaluating efficacious and effective treatments. Our plan lobbies for the accumulation of studies in several areas, beyond outcome questions and criteria used to define probably efficacious or well-established treatments. Impetus for proposing an overall plan stems from the directions and status of current research, which, by and large, has been limited in what is studied, how treatment is delivered, and who is served. Further analyses or reviews of the literature, however quantitative or comprehensive, will not provide the requisite information, if the constituent studies have not been completed. Conceptual and empirical work is needed to address a range of questions about how treatment works (mediators) and for whom (moderators). The blueprint is designed to convey the range of questions and to serve as a basis for charting or evaluating progress of research over time.

Apart from a proposed blueprint, several issues were identified to improve individual studies and the likelihood that the findings are clinically useful, important, and disseminable. Ideally, those treatments identified as probably efficacious or well-established would be suitable or the treatments of choice for clinical work. Much of the research is conducted under conditions that depart markedly from those to which we wish to generalize. Additional research with clinical samples is sorely needed. In addition, we identify several areas to integrate into outcome research to improve the strength and clinical relevance of the conclusions that can be drawn.

References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101, 213-232.
- Bird, H. R., Yager, T. J., Staghezza, B., Gould, M. S., Canino, G., & Rubio-Stipec, M. (1990). Impairment in the epidemiological measurement of psychopathology in the community. *Journal of the American Academy of Child and Adolescent Psychiatry*, 29, 796-803.
- Casey, R. J., & Berman, J. S. (1985). The outcome of psychotherapy with children. *Psychological Bulletin*, 98, 388-400.
- Chambless, D. L., & Hollon, S. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7-18.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L., Williams, D. A., & McCurry, S. (1996). An update on empirically validated treatments. *The Clinical Psychologist*, 49, 5-18.
- Dishion, T. J., Patterson, G. R., & Kavanagh, K. A. (1992). An experimental test of the coercion model: Linking theory, measurement, and intervention. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior* (pp. 253-282). New York: Guilford.
- Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development*, 65, 649-655.
- Durlak, J. A., Wells, A. M., Cotten, J. K., & Johnson, S. (1995). Analysis of selected methodological issues in child psychotherapy research. *Journal of Clinical Child Psychology*, 24, 141-148.
- Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 16, 319-324.
- Forgatch, M. S. (1991). The clinical science vortex: A developing theory of antisocial behavior. In D. J. Pepler & K. H. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 291-315). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Frank, E., Prien, R. F., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P. W., Rush, A. J., & Weissman, M. M. (1991). Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Archives of General Psychiatry*, 48, 851-855.
- Gaston, L., & Gagnon, R. (1996). The role of process research in manual development. *Clinical Psychology: Science and Practice*, 3, 13-24.
- Hodges, K., Bickman, L., Ring-Kurtz, S., & Rieter, M. (1991). Multidimensional measure of level of functioning for children and adolescents. In A. Algarian & R. Freidman (Eds.), *A system of care for children's mental health: Expanding the research base* (pp. 149-154). Tampa, FL: Research and Training Center for Children's Mental Health.
- Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology*, 65, 599-610.
- Jacobson, N. S., & Revenstorf, D. (1988). Statistics for assessing the clinical significance of psychotherapy techniques: Issues, problems, and new developments. *Behavioral Assessment*, 10, 133-145.
- John, K., Gammon, G. D., Prusoff, B. A., & Warner, V. (1987). The Social Adjustment Inventory for Children and Adolescents (SAICA): Testing of a new semistructured interview. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 898-911.
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1, 427-452.
- Kazdin, A. E. (1995). Child, parent, and family dysfunction as predictors of outcome in cognitive-behavioral treatment of antisocial children. *Behaviour Research and Therapy*, 33, 271-281.
- Kazdin, A. E. (1997). Parent management training: Evidence, outcomes, and issues. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 1349-1356.
- Kazdin, A. E. (1998). *Research design in clinical psychology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
- Kazdin, A. E., Bass, D., Ayers, W. A., & Rodgers, A. (1990). The empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology*, 58, 729-740.
- Kazdin, A. E., & Crowley, M. (1997). Moderators of treatment outcome in cognitively based treatment of antisocial behavior. *Cognitive Therapy and Research*, 21, 185-207.
- Kazdin, A. E., & Wilson, G. T. (1978). Criteria for evaluating psychotherapy. *Archives of General Psychiatry*, 35, 407-416.
- Kendall, P. C. (in press). Directing misperceptions: Researching the issues facing manual-based treatments. *Clinical psychology: Science and Practice*.
- Kendall, P. C. (1997). Editorial. *Journal of Consulting and Clinical Psychology*, 65, 3-5.
- Kendall, P. C., Chu, B., Gifford, A., Hayes, C., & Nauta, M. (in press). Breathing life into a manual. *Cognitive and Behavioral Practice*.
- Kendall, P. C., & Flannery-Schroeder, E. (1998). Methodological issues in treatment research for anxiety disorders in youth. *Journal of Abnormal Child Psychology*, 26, 27-38.
- Kendall, P. C., & Grove, W. M. (1988). Normative comparisons in therapy outcome. *Behavioral Assessment*, 10, 147-158.
- Kendall, P. C., & Southam-Gerow, M. (1995). Issues in the transportability of treatment: The case of anxiety disorders in youth. *Journal of Consulting and Clinical Psychology*, 63, 702-708.
- Kendall, P. C., & Southam-Gerow, M. (1996). Long-term follow-up of treatment for youth with anxiety disorders. *Journal of Consulting and Clinical Psychology*, 64, 724-730.
- Kiesler, D. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin*, 65, 110-136.
- Kolvin, I., Garside, R. F., Nicol, A. E., MacMillan, A., Wolstenholme, F., & Leitch, I. M. (1981). *Help starts here: The maladjusted child in the ordinary school*. London: Tavistock.
- Levitt, E. E. (1957). The results of psychotherapy with children: An evaluation. *Journal of Consulting Psychology*, 21, 189-196.
- Levitt, E. E. (1963). Psychotherapy with children: A further evaluation. *Behaviour Research and Therapy*, 60, 326-329.
- Lewisohn, P. M., Seeley, J. R., Hibbard, J., Rhode, P., & Sack, W. H. (1996). Cross-sectional and prospective relationships between physical morbidity and depression in older adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 1120-1129.
- Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that "everyone has won and all must have prizes"? *Archives of General Psychiatry*, 32, 995-1008.
- Meyers, A. W., Graves, T. J., Whelan, J. P., & Barclay, D. (1996). An evaluation of television-delivered behavioral weight loss program. Are the ratings acceptable? *Journal of Consulting and Clinical Psychology*, 64, 172-178.
- Nathan, P. W., & Gorman, J. M. (Eds.). (1998). *A guide to treatments that work*. New York: Oxford University Press.
- Newman, M. G., Kenardy, J., Herman, S., & Taylor, C. B. (1997). Comparison of palmtop-computer-assisted brief cognitive-behavioral treatment to cognitive-behavioral treatment for panic disorder. *Journal of Consulting and Clinical Psychology*, 65, 178-183.

- Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.
- Patterson, G. R., Dishion, T. J., & Chamberlain, P. (1993). Outcomes and methodological issues relating to treatment of antisocial children. In T. R. Giles (Ed.), *Handbook of effective psychotherapy* (pp. 43-87). New York: Plenum.
- Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys*. Eugene, OR: Castalia.
- Robins, L., & Rutter, M. (Eds.). (1990). *Straight and devious pathways from childhood to adulthood*. Cambridge, England: Cambridge University Press.
- Roth, A., & Fonagy, P. (1996). *What works for whom: A critical review of psychotherapy research*. New York: Guilford.
- Sanford, M. N., Offord, D. R., Boyle, M. H., Peace, A., & Racine, Y. A. (1992). Ontario Child Health Study: Social and school impairments in children aged 6-16 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 60-67.
- Shaffer, D., Gould, M. S., Braisic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry*, 40, 1228-1231.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- Smith, M. L., Glass, G. V., & Miller, T. L. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
- Strupp, H. H., & Anderson, T. (1997). On the limitations of therapy manuals. *Clinical Psychology: Science and Practice*, 14, 76-82.
- Strupp, H. H., & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes. *American Psychologist*, 32, 187-196.
- Szapocznik, J., Rio, A., Murray, E., Cohen, R., Scopetta, M., Rivas-Vasquez, A., Hervis, O., Posada, V., & Kurtines, W. (1989). Structural family versus psychodynamic child therapy for problematic Hispanic boys. *Journal of Consulting and Clinical Psychology*, 57, 571-578.
- Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically validated psychological treatments: Report and recommendations. *The Clinical Psychologist*, 48, 3-23.
- Treadwell, K. R. H., & Kendall, P. C. (1996). Self-talk in anxiety disordered youth: States of mind, content specificity, and treatment outcome. *Journal of Consulting and Clinical Psychology*, 64, 941-950.
- Wade, W., Treat, T., & Stuart, G. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology*, 66, 231-289.
- Webster-Stratton, C. (1996). Early intervention with videotape modeling: Programs for families of children with oppositional defiant disorder or conduct disorder. In E. D. Hibbs & P. Jensen (Eds.), *Psychosocial treatment research of child and adolescent disorders: Empirically based strategies for clinical practice* (pp. 435-474). Washington, DC: American Psychological Association.
- Webster-Stratton, C., & Hammond, M. (1990). Predictors of treatment outcome in parent training for families with conduct problem children. *Behavior Therapy*, 21, 319-337.
- Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*, 55, 542-549.
- Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, 117, 450-468.
- Wilson, G. T. (in press). Manual-based treatment and clinical practice. *Clinical Psychology: Science and Practice*.
- Wolf, M. M. (1978). Social validity: The case of subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203-214.

Received December 23, 1997

Accepted January 13, 1998