Psychological Bulletin
2005, Vol. 131, No. 3, 418–426

0033-2909/05/$12.00 DOI: 10.1037/0033-2909.131.3.418

Jousting With Straw Men: Comment on Westen, Novotny, and Thompson-Brenner (2004)

John R. Weisz
Judge Baker Children's Center, Harvard University

V. Robin Weersing
Yale University School of Medicine

Scott W. Henggeler
Medical University of South Carolina

Empirically supported treatments (ESTs) do not cure every patient, and the randomized trial is not a flawless methodology. Upon these often-noted and widely accepted points, D. Westen, C. M. Novotny, and H. Thompson-Brenner (2004a) built a critique of ESTs and EST research. However, important work developing effective, clinically relevant treatments for serious problems was omitted from the Westen et al. (2004a) review. Little documentation was offered for the purported "assumptions" of EST methodology that Westen et al. (2004a) criticized; and different review standards were applied to studies supporting versus those disagreeing with Westen et al.'s (2004a) views. Finally, the correlational research designs proposed as a remedy by Westen et al. (2004a) have far more serious weaknesses than randomized trials, thoughtfully applied to real-world clinical care.

Keywords: empirically supported treatments, psychotherapy, clinical practice, randomized controlled trials

Over the past 2 decades, a revolution has begun in psychotherapy research and practice. Researchers, practitioners, policymakers, and consumers have come together to try to understand and improve the quality of mental health care. Psychotherapy researchers have stepped out of their university laboratories and begun to conduct rigorous applied treatment research in community settings, with clear relevance to everyday clinical practice. Federal funding has been made available to clinical care providers to build practice-based research infrastructure and support the training of community therapists in empirically supported treatments (ESTs; e.g., U.S. Department of Health and Human Services, 2004a, 2004b). Policymakers at the national level have endorsed the importance of evidence-based quality mental health care (National Institute of Mental Health, 2001; Office of the Surgeon General, 1999; President's New Freedom Commission on Mental Health, 2003). Family advocacy groups and patient organizations have become increasingly vocal in advocating not just for access to mental health care but to interventions with demonstrated effectiveness and high patient satisfaction (Allness & Knoedler, 2003; Flynn, 2005; Hoagwood, 2003, in press), and states have developed initiatives to support the use of effective mental health services (National Association of State Mental Health Program Directors, 2004). This is a time of great change and, for many, great excitement. In this context, we were pleased to see a substantial portion of the August 2004 issue of *Psychological Bulletin* devoted to a discussion of EST in mental health. However, we were disappointed in the inaccurate content, selective reviewing, and combative tone of the lead article by Westen, Novotny, and Thompson-Brenner (2004a) critiquing EST research.

To be sure, there were a number of points in the Westen et al. (2004a) article with which we agree—indeed, which we and our colleagues have made repeatedly over the past 15 years—and anyone who has carried out meta-analyses would concur with Westen et al. that reporting in outcome studies needs to be much more complete and consistent. However, we do not find that the comments of Westen et al. (2004a, 2004b) add substantively to what has already been written or provide a constructive guide for how to actually improve psychotherapy research. Moreover, the credibility of the Westen et al. critique is weakened by their selective review of ESTs (failure to note ESTs that are quite inconsistent with their arguments), their selective review of the evidence base (omitting findings inconsistent with their points),

John R. Weisz, Judge Baker Children's Center, Harvard University; V. Robin Weersing, Yale University School of Medicine; Scott W. Henggeler, Family Services Research Center, Medical University of South Carolina.

Scott W. Henggeler (SWH) is a shareholder in MST Services, Inc., which is licensed by the Medical University of South Carolina for the transport of multisystemic therapy technology and intellectual property.

Preparation of this article and some of the work described herein was facilitated by support from the Annie E. Casey Foundation (SWH), the John D. and Catherine T. MacArthur Foundation (*Linking Science and Practice to Improve Youth Mental Health* [John R. Weisz; JRW] and *Research Network on Youth Mental Health* [JRW]), the Klingenstein Third Generation Foundation (V. Robin Weersing; VRW), the Robert Wood Johnson Foundation (VRW), the W. T. Grant Foundation (VRW), the Center for Substance Abuse Treatment (SWH), the National Institute on Alcoholism and Alcohol Abuse (AA122202 [SWH]), the National Institute on Drug Abuse (DA17487, DA015844, DA10079 [SWH]), and the National Institute of Mental Health (R01 MH 064503-01A1 [VRW], R01 MH 60663 [SWH], R01 MH57347 [JRW], R21-MH63302 [JRW], R01-MH068806 [JRW]).

Correspondence concerning this article should be addressed to John R. Weisz, Judge Baker Children's Center, Harvard University, 53 Parker Hill Avenue, Boston, MA 02120-3225.

and their use of shifting standards of evidence (being strict about findings and methods they critique but lax about findings and methods that fit their views). We also have concerns about a primary metamessage of the critique, which seems to be that because the randomized controlled trial (RCT) design is not a perfect methodology, the field should shift wholesale to alternative methodologies (such as correlational survey designs) that have even more significant limitations than RCTs.

Our Perspective on Clinical Practice, Clinical Science, and the Search for Best Practices

To establish a conceptual context for our comments, we note here some general principles that have guided our work and our thinking about science and practice over the years.

1. We believe that the search for best practices requires collaborative effort by the research and practice communities. The ideal outcome of this search is a set of interventions that meets both standards of evidence and requirements of utility and effectiveness in practice. We cannot imagine achieving this outcome without active practice–research collaboration. Our assumption is that most clinical practitioners and most clinical researchers are genuinely trying to improve mental health care for those who seek help. Both deserve respect for their good-faith efforts.

2. We also believe that the search for best practices requires methodological triangulation. Each method by which treatments are studied—including RCTs and correlational approaches—is imperfect and thus relatively easy to criticize for its specific limitations. The challenge is to improve methods where possible and otherwise to combine them in ways that ensure the limitations of one approach are counterbalanced by the strengths of another. This said, we do hold to the general principle that manipulating variables experimentally is the best method we have for establishing cause–effect relationships, and experimental manipulation and random assignment are compelling strengths of the RCT design.

3. We also believe that identification of beneficial practices is an evolving process, each product of which is destined to be imperfect in various ways. Thus, at this early stage in the evolving history of treatment development and testing, most individual treatments and their evidence base will be easy to criticize for a variety of limitations.

4. This being the case, it seems to us that the usefulness of any critique is best judged, not by whether it can identify limitations in methods or products, since such limitations are evident to many and often noted in the literature, but by whether the critique (a) provides a fair and accurate characterization, (b) correctly identifies causes of problems noted, and (c) proposes new approaches that will significantly improve the current state of affairs. In our view, the Westen et al. critique falls short in regard to a, b, and c. We address these three standards in the following sections.

Did Westen et al. Correctly Characterize the Evidence Base?

First, we do not believe that Westen and colleagues (2004a) accurately characterized the evidence base they were critiquing. Our particular expertise is investigating psychotherapy in real-world settings, and we focus our comments primarily on the Westen et al. summary of the effects of "treatment as usual" (TAU) and the relation between evidence-based treatments and real-world conditions of clinical practice.

We begin by examining the Westen et al. (2004a) critique of the Weersing and Weisz (2002a) benchmarking investigation. From the start, Westen et al. appear to have misunderstood the purpose of the study. The investigation was reviewed under the heading "*Studies Testing the Transportability of ESTs*" (Westen et al., (2004a, p. 648), when in fact, it was a naturalistic investigation of everyday clinical care. Weersing and Weisz (2002a) identified depressed children and adolescents seen for services in community mental health centers, prospectively followed these youths for 2 years, and measured both the services received and outcomes achieved using well-validated instruments and assessment methods. To anchor any observed changes in depression symptoms, the authors compared the symptom slope of youths treated in the community with the most stringent benchmark available—the outcomes of cognitive–behavioral therapy (CBT) for youths with depression, reported in published clinical trials. This nonexperimental comparison revealed a substantial outcome gap, with youths in clinical trials improving nearly twice as quickly as depressed youths treated in the community.

Westen et al. (2004a) criticized the study because the community clinic sample had a higher rate of comorbid diagnoses than the CBT sample; they suggested that this difference might account for the outcome difference reported. They did not report, however, that Weersing and Weisz (2002a) had considered this possibility, tested the effects of comorbidity (both general and specific) on symptom slope, and found that comorbidity did not impact depression outcome in this community care sample. Also omitted from the Westen et al. critique was the finding that there were no predictors within the community clinic sample that erased the gap in outcomes between the clinics and the CBT efficacy studies, but some variables did predict even worse outcome in the community. Finally, Westen et al. (2004a) suggested that the outcome gap uncovered by Weersing and Weisz (2002a) could have been due to differences in the treatments delivered. Indeed, this was precisely the point. Weersing and Weisz (2002a) were unable to test the effects of therapy type—CBT versus eclectic/insight-oriented care—within the community sample because none of the therapists were providing this evidence-based treatment. However, they did conclude that differences in therapy type might be a reasonable explanation for the discrepancy in outcomes between the community clinics and the published results of CBT. This conclusion came with multiple caveats regarding the nonexperimental nature of the design, other unmeasured variables that might affect outcome (e.g., maternal depression, socioeconomic status), and the dearth of data on how CBT would perform in the setting and sample of the community centers. The recommendation of the report was that precisely this type of real-world transportability research was needed to clarify the meaning of the findings.

We comment on Weersing and Weisz's (2002a) investigation not just to correct the misrepresentation of the study in Westen et al. (2004a) but because the results are in line with a growing body of research completely omitted from the Westen et al. critique. The finding that usual clinical care, not guided by empirical evidence, may not work well is emerging from a number of studies, using a variety of methods, in the child and adolescent treatment literature. For more than a decade, Weisz and colleagues have been searching that literature for studies of psychotherapy in service settings in which therapists were able to use their clinical judgment to deliver treatment as they saw fit, unconstrained by evidence-based interventions or manuals, and in which there was a comparison of this treatment to a control condition. Meta-analyses of these studies of usual clinical care have found effect sizes averaging about zero (see, e.g., Weisz, 2004; Weisz, Donenberg, Han, & Weiss, 1995). Consistent with this pattern, Weiss, Catron, Harris, and Phung (1999) found minimal effects of treatment delivered to disturbed school-aged youths by practitioner therapists hired by a local community clinic. Therapists were free to use whatever methods they preferred, and they averaged 60 individual sessions, 18 parent sessions, and 13 school consults. After all this treatment, guided by therapists' own choice of methods, youth outcomes were equivalent (actually, nonsignificantly inferior) to those of the randomly assigned control group who had received only academic tutoring. Additional evidence suggests that efforts to link individual TAUs together within what have been called "systems of care" may not be helpful either (Bickman, 1996; Bickman et al., 1995; Bickman, Noser, & Summerfelt, 1999). We are not aware of credible evidence on effects of TAU for adults, but for children and adolescents the evidence is not encouraging regarding the impact of usual clinical care that has not been derived from empirical treatment research.

Another aspect of the evidence base selectively reviewed by Westen et al. (2004a) is that dealing with the dose–response relationship in psychotherapy. Westen and colleagues argued (2004a) that studies on this topic "consistently find a dose–response relationship, such that longer treatments, particularly those of 1 to 2 years and beyond, are more effective than briefer treatments" (p. 633). This conclusion, based on research cited from the 1980s and 1990s, certainly fits Westen and colleagues' stated preference for lengthier treatment. However, this conclusion ignores more recent and more rigorous studies (Andrade, Lambert, & Bickman, 2000; Bickman, Andrade, & Lambert, 2002; Salzer, Bickman, & Lambert, 1999) with children and adolescents showing no dose–response relationship.

Westen et al. (2004a) also seemed quick to discount the findings of studies explicitly designed to test the main question of their review—that is, whether the findings and treatments emerging from RCTs are generalizable to the settings and samples of active clinical practice? In their review, they allocated less than a paragraph to two particularly seminal studies in this area—Wade, Treat, and Stuart's (1998; see Westen et al.'s, 2004a, discussion of Stuart, Treat, & Wade, 2000, which was the 1-year follow-up) investigation of the transportability of CBT for panic disorder and Franklin, Abramowitz, Levitt, Kozak, and Foa's (2000) investigation of exposure therapy for obsessive–compulsive disorder in a sample of patients excluded from an RCT. In the words of Westen and colleagues (2004a), the Stuart et al. (2000) study had "impressively low relapse rates," and the Franklin et al. sample was "superb" (p. 649), but Westen et al. (2004a) dismissed the results out of hand with the suggestion that the clinical assessors would have known the patients received treatment (and were subject to bias).

We certainly agree that blind assessment is desirable; however, we were quite surprised at the short shrift given to this work. Not only are these studies central to the thesis of Westen et al.'s (2004a) critique, but the methods of assessment used in Wade et al. (1998) and Franklin et al. (2000) were substantially more rigorous than other investigations of "naturalistic" therapy that were featured in the Westen et al. review, concur with the authors' arguments, and were presented at length as useful and thought provoking (pp. 647–648; a point we return to later). Furthermore, Westen et al. did not report a number of additional studies testing the generalizability/transportability of RCT treatments, including CBT for social phobia (Lincoln et al., 2003), CBT for panic disorder (Addis et al., 2004), interpersonal therapy for depression (Mufson et al., 2004), cognitive therapy for depression (Merrill, Tolbert, & Wade, 2003), and CBT for bulimia nervosa (Tuschen-Caffier, Pook, & Frank, 2001). These investigations, too, have generally demonstrated the robustness of evidence-based treatments across a variety of patient, provider, and setting parameters. Overall, Westen et al. appear to have provided a highly selective review of the evidence base on TAU and the transportability of ESTs, favoring, highlighting, and readily forgiving flaws in reports that supported their views and omitting, downplaying, or sharply criticizing reports that disagreed with their views.

Did Westen et al. Correctly Identify Causes of the Purported Problems (i.e., the "Assumptions")?

A second question to ask about the Westen et al. (2004a) critique is whether it correctly identified causes of the purported problems it presented. To address this question, we need to examine the four "assumptions" Westen and colleagues (2004a) claimed are made by those who use "EST methodology" (p. 633). Indeed, such an examination is crucial because Westen et al.'s critique is founded largely on their criticism of these purported assumptions that they believe underlie the development and validation of ESTs. Some of these assumptions may fit the picture of ESTs drawn by these authors, reflecting treatments that have not yet been designed to fit into clinical practice contexts, but the assumptions have little applicability to the ESTs we know that have been developed for use by practitioners in practice settings with clinically referred individuals. In our analysis of the assumptions purportedly underlying ESTs, we focus in particular on five EST models that have considerable empirical support and have addressed a broad range of serious clinical problems in adults and youths. We have chosen to focus on these interventions because they clearly violate the first three purported assumptions of ESTs that form the basis of the Westen et al. critique and because each one has been successfully tested using the RCT methods described as inadequate by Westen et al. (2004a; in their presentation of the fourth assumption).

> The *community reinforcement approach* (CRA; Budney & Higgins, 1998) is one of the most extensively validated treatments of adult drug abuse (National Institute on Drug Abuse, 1999; Roozen et al., 2004) and has been tested in more than 15 randomized trials.

> The *program of assertive community treatment* (PACT; Stein & Santos, 1998), for persons with severe and persistent mental illness, has been the subject of more than 25 randomized trials (Bond, Drake, Mueser, & Latimer, 2001; U.S. Department of Health and Human Services, 1999) and has received broad dissemination to practice settings (Gold et al., 2003).
>
> *Multisystemic therapy* (MST; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998), *functional family therapy* (Sexton & Alexander, 2004), and *multidimensional treatment foster care* (MTFC; Chamberlain, 2003) are the three effective treatments of juvenile offenders cited in Elliott's Blueprints series (Elliott, 1998) and the U.S. Surgeon General's report on youth violence (U.S. Public Health Service, 2001) and have been tested in more than a dozen RCTs.

Purported Assumption 1: "Psychological processes are highly malleable" (p. 633). Westen and colleagues (2004a) contended, "The assumption of malleability is implicit in the treatment lengths used in virtually all ESTs, which typically range from about 6 to 16 sessions" (p. 633). Obviously, some degree of malleability must be assumed by anyone embarking on treatment, but the use of RCTs to study psychological treatments of limited duration does not imply an assumption that psychological conditions are easy to change any more than use of RCTs to study cancer treatment of limited duration implies that cancer is easy to cure. Moreover, treatment duration in our example EST programs does not look like the picture Westen et al. (2004a) painted. These ESTs clearly consider the chronicity of certain types of problems, and their services are designed accordingly. PACT provides daily 24-hr availability of services, with no preset time limits. Similarly, MST is usually provided for at least 4 months, MTFC for 6 to 9 months, and CRA for 6 months with multiple treatment sessions per week and daily 24-hr availability of services in many cases. It is important to note that treatment completion within these ESTs is generally not fixed but criterion based—that is, dictated by the extent to which treatment goals have been attained or improved functioning demonstrated. This seems quite consistent with what Westen et al. (2004a) implied is needed. Good news: It is already being done within ESTs.

Westen et al. (2004a) suggested that treatments of "1 to 2 years and beyond" (p. 633) may be essential, and for some chronic and refractory conditions, this may be the case, as in the PACT program for persons with persistent mental illness. However, for the majority of problems and conditions, long-term treatment may be more consistent with practice decades ago and, perhaps for a wealthy few today, than with the reality of current clinical care for most people. As an example, a recent review of a national database of over 6,000 adult patients in multiple forms of usual clinical care (e.g., counseling centers, local and national health maintenance organizations, community mental health centers) showed that the average number of sessions was less than 5 (see Hansen, Lambert, & Forman, 2002). Within this context, the Westen et al. estimate of 6–16 sessions for the average EST seems generous. An argument could be made that for many problems and conditions designing treatments that require 2 years or more to be effective may not be so helpful in a practice environment in which very few people actually receive that much service.

Finally, the notion that most psychological processes and disorders are "highly malleable" would be rejected out of hand by most of the treatment researchers we know. Most recognize that change in long-standing patterns is quite difficult, and many who treat the most severe problems aim not for completely normal functioning but for an attenuation of the debilitating effects of conditions that are often quite chronic. Treatment could go on virtually forever in many cases, but in a practice climate characterized by finite fiscal resources and a philosophy of empowerment for clients and their families, the potential gains of extended treatment must be balanced against the efficient and equitable use of available resources and the value of helping clients develop robust coping skills and indigenous support networks.

Purported Assumption 2: "Most patients have one primary problem or can be treated as if they do" (p. 634). Westen and colleagues (2004a) contended that EST researchers fail to consider comorbidities in the design and implementation of interventions. Considerable EST research has indeed used exclusionary criteria, and we too have been critical of this practice (e.g., Weisz, 2004). But the purported assumption could not be further from reality for the index ESTs noted above, which reflect the movement toward practice relevance. Juvenile offenders, drug abusing adults, and persons with severe and persistent mental illness present a broad range of emotional, interpersonal, and psychiatric comorbidities; and each of these ESTs has been designed with the flexibility to address possible co-occurring problems directly. The CRA manual (Budney & Higgins, 1998), for example, includes sections on improving time management, intimate partner relations, and vocational satisfaction. Likewise, the functional family therapy (Alexander & Parsons, 1982; Sexton & Alexander, 2004), MST (Henggeler et al., 1998; Henggeler, Schoenwald, Rowland, & Cunningham, 2002), and MTFC (Chamberlain, 1998, 2003) manuals present protocols for intervening at cognitive, family, peer, school, and community levels for the juvenile offenders and their families. PACT is similarly comprehensive, with interventions ranging from medication management to vocational training (Stein & Santos, 1998). Indeed, the comprehensive and broad-based nature of these ESTs may be critical to their established effectiveness with the multiproblem clients they serve.

Purported Assumption 3: "Psychological symptoms can be understood and treated in isolation from personality dispositions" (p. 636). We were mystified by this purported assumption. We find it hard to imagine that people who have actually carried out psychotherapy—be they researchers or practitioners—would endorse such an odd idea. A very significant part of learning to use an EST is learning to apply the procedures in ways that fit well with patient personality and with environmental factors. And some EST researchers (e.g., Persons & Tompkins, 1997) have strongly emphasized the need for ESTs to be formulation driven and thus individualized to fit each patient's personality, cognitive, affective, and social characteristics and context. In the five index treatments noted above, individualizing treatment procedures to fit each individual's personality and situation is a core element.

Purported Assumption 4: "Controlled clinical trials provide the gold standard for assessing therapeutic efficacy" (p. 637). We would argue that random assignment and experimental control are indeed optimal methods for establishing causality. However, it appears to us that Westen et al. (2004a) did not so much critique experimentation per se as criticize some of the specific ways experimental methods have been used in some EST research. Thus, in this section, we assess whether the Westen et al. description of how RCT methods have been applied (a) accurately characterizes the field and (b) accurately identifies procedures that are necessary for the conduct of an RCT (e.g., Is it true that to carry out a controlled trial, one must use a didactic treatment manual?).

Much of the Westen et al. (2004a) critique of RCTs is actually a discussion of EST treatment manuals. The picture painted by Westen et al. (2004a) depicts EST manuals as rigidly structured documents that minimize the patient's active involvement in the treatment process, prevent therapists from using clinical judgment, reduce the therapist to a "research assistant" whose job is to "run subjects" (p. 639), and are incompatible with an emphasis on broad principles of change. One can certainly find manuals for which some of these features are evident. However, as we noted previously, one can also find examples—in some of the most thoroughly evolved and tested ESTs for some of the most serious real-world problems—of principle-based treatment protocols in which therapist, client, and the client's significant others play active roles in fashioning the specifics of the intervention.

As a careful reading of the manuals for our example ESTs would reveal, the manuals set the playing field and provide guidelines and principles for the design and implementation of interventions. The manuals do not provide a lock-step description of therapist behavior in the ways described by Westen et al. By all means, logical sequencing of interventions is used (e.g., assessment and engagement precede interventions), principles of practice and suggested techniques are described, and specific procedures that appear to be ineffective or counterproductive are proscribed (e.g., MTFC and MST do not allow group therapy with juvenile offenders, given evidence that such approaches may generate "delinquency training"; see Arnold & Hughes, 1999; Dishion, McCord, & Poulin, 1999); but within such logically and empirically guided frameworks clinicians have considerable flexibility—indeed, a mandate—to use their own creativity and resources to achieve desired clinical outcomes.

Equally far from reality is Westen and colleagues' (2004a) notion that experimental control requires "that therapy is something done to a patient—a process in which the therapist applies interventions—rather than a transactional process in which patient and therapist collaborate" (p. 639). In each of the aforementioned ESTs, therapist–client collaboration is viewed as absolutely essential to the specification of treatment goals and the development and implementation of plans to achieve those goals. As further evidence of the emphasis that many ESTs place on client collaboration, major consumer advocacy groups such as the National Alliance for the Mentally Ill have supported ESTs such as MST and PACT (Allness & Knoedler, 2003; National Alliance for the Mentally Ill, 2003).

Moreover, for these treatments, in contrast to the arguments of Westen et al. (2004a), (a) randomized trials have usually included clients with multiple co-occurring problems and/or disorders, (b) the trials have often been based in community intervention settings, (c) it is not the case that secondary analyses have been required to identify effects, and (d) treatment fidelity has predicted outcome (see, e.g., Bond, McGrew, & Fekete, 1995; Gold et al., 2003; Henggeler, Melton, Brondino, Scherer, & Hanley, 1997; Henggeler, Pickrel, & Brondino, 1999; Huey, Henggeler, Brondino, & Pickrel, 2000; Latimer, 1999; McGrew & Bond, 1997; Schoenwald, Sheidow, & Letourneau, 2004; Sexton & Alexander, in press).

One additional concern raised by Westen et al. (2004a) in this section warrants attention. Westen et al. (2004a) stated (but did not document) that most comparisons of ESTs to TAU are not true TAUs. We certainly agree that TAU conditions should provide the most representative picture possible of actual treatment in the community. Indeed, a critical public health question to ask of any new treatment program is whether it can produce better results than the interventions people would have received in its absence. However, as discussed earlier, we find Westen et al.'s view of TAU—long-term treatment, delivered by expert psychologists, unconstrained by patient or payor time limits—quite unlike the data on usual care that we know. For both children and adults, real-world treatment for psychological problems tends to be short term, and patients are more likely to receive care from nonpsychologists (especially social workers) working in schools, public sector agencies, or primary care offices than from psychologists in independent practice offices (e.g., Burns et al., 1995).

In concluding this section of our commentary, we return to the question of whether Westen et al. (2004a) correctly identified causes of the purported problems on which they focused. Some of their proposed causes make sense to us, but the four assumptions they articulated do not work as a sweeping characterization of ESTs or clinical researchers. Most clinical researchers we know do not assume that psychological processes are highly malleable, and ESTs vary more widely in the design of their protocols and in their duration than Westen et al. (2004a) acknowledged. Most EST researchers we know do not believe that most patients have one primary problem or can be treated as if they do, and some of the best ESTs are focused—increasingly so in recent years—on the co-occurring problems, disorders, and life circumstances that can make conditions so complex for treated individuals. In a previous article on Freud's legacy, Westen (1998) commented that critics "attack an archaic version of psychodynamic theory that most clinicians . . . consider obsolete" (p. 333). In their recent critique of ESTs, Westen et al. appear to have done something similar, attacking a version of EST science that is obsolete from the perspective of effectiveness researchers who are working to keep science and practice closely linked.

To continue our summary regarding Westen et al.'s (2004a) assumptions, we stress that most EST researchers we know would never endorse the assumption that "psychological symptoms can be understood and treated in isolation from personality dispositions" (p. 636). Indeed, this assumption struck us as missing the mark so widely that we wondered if Westen and colleagues had experience using very many ESTs with real patients. Those who have such experience certainly know that finding ways to fit intervention principles and procedures into the context of the patient's personal style and environment is fundamental to effective treatment with ESTs. As for the purported assumption that "controlled clinical trials provide the gold standard for assessing therapeutic efficacy" (Westen et al., 2004a, p. 637), we find the content of Westen et al.'s critique much less focused on inherent characteristics of controlled experimental methods than on some of the ways those methods have been used by some investigators. Westen et al. primarily criticized the ways some investigators have designed manuals, structured some experimental tests, and selected some treatments to be the target of those tests. In their critique, they failed to note some of the most prominent ESTs to which their criticisms do not apply, and they failed to acknowledge that the solution to most of the problems they highlighted does not require either suspending or abandoning controlled clinical trials.

Did Westen et al. Propose New Approaches That Will Constitute Genuine Improvements?

As a final point, we examine the models of research put forth by Westen and colleagues (2004a) as alternatives to identifying ESTs through RCT methods. In their original review and in response to commentaries, Westen et al. (2004a, 2004b) proposed that psychotherapy outcome research substantially abandon its experimental tradition and adopt descriptive, correlational approaches. For example, Westen et al. (2004a) reviewed at length their own naturalistic therapy research, focusing on surveys that replaced standardized assessment with therapists' self-reports of their practice and judgments about their patients' diagnoses. Therapists also rated patients' "clinically significant change" (Morrison, Bradley, & Westen, 2003, p. 114) and whether outcomes "would likely be lasting" (Morrison et al., 2003, p. 115; see also Thompson-Brenner & Westen, in press). These sorts of practice surveys may have merit as hypothesis-generating tools; however, they suffer from fatal methodological flaws if the goal of the research is to draw inferences about what works for whom in the real world. Generalizing across large numbers of respondents does not erase these faults, particularly given low response rates and uncertainty about the representativeness and validity of responses. Readers are referred to evidence on the poor validity of diagnostic judgments by clinicians under usual clinical practice conditions (see, e.g., Aronen, Noam, & Weinstein, 1993; Jensen & Weisz, 2002); the diagnostic system may simply contain more disorders than can be carefully reviewed under usual clinical practice time constraints. Readers are also referred to the extended critiques of the *Consumer Reports* survey (Seligman, 1995) of therapy "effectiveness," published nearly 10 years ago, for a discussion of threats to inference in the kind of design proposed by Westen et al. (2004a; see, e.g., Jacobson & Christensen, 1996; Mintz, Drake, & Crits-Christoph, 1996). Although the *Consumer Reports* survey used patients as respondents, the same issues of self-report biases, retrospective reporting of treatment characteristics and outcomes, lack of randomization, and self-selection of respondents (and of the episode of care they choose to report on) would seem to apply as or more strongly to the therapist respondents favored by Westen et al. It also was surprising to us that this approach was presented so favorably, given their damning assessment of EST "transportability studies" using nonblind outcome assessors, individuals presumably less motivated to shape their responses in a particular direction than the therapists who provided the treatment that was being evaluated.

By the end of their review, Westen et al. suggested several refinements to this descriptive method, including taping random therapy sessions and use of independent interviewers, as a way to gain some verification of therapist behavior and patient outcomes. However, actual manipulation of therapy processes and random assignment of patients would apparently be delayed for many years in the Westen et al. plan, after many iterations of correlating descriptive process variables with outcomes. To us, this research agenda ignores the substantial, iterative contributions to the psychotherapy outcome literature that have accrued through experimentation. Indeed, other proponents of the use of practice as a "natural lab" have stressed the importance of building RCTs into the very structure of field work (e.g., use of randomized dismantling designs; Borkovec, Echemendia, Ragusea, & Ruiz, 2001). Furthermore, even among therapy process researchers there is continuing debate about the value of correlational designs in revealing causal relationships. For example, if practicing therapists are perfectly responsive to their patients' needs, then significant relationships between therapist behavior and patient outcomes may disappear from view—a hypothesis that has been put forth to explain null findings in previous, correlational process–outcome studies (Stiles & Shapiro, 1994; cf. Hayes, Castonguay, & Goldfried, 1996). In addition, there is general agreement among methodologists that mediation analyses, and other correlational tests examining component processes or active ingredients of treatment, are greatly enhanced when an experimentally manipulated (e.g., randomly assigned treatment) variable can be used to predict change in mediator and change in outcome (e.g., Kraemer, Wilson, Fairburn, & Agras, 2002; Rose, Holmbeck, Coakley, & Franks, 2004; Weersing & Weisz, 2002b).

As we stated at the beginning of this commentary, we fully agree that the RCT design has limitations and is not the only pathway to truth and that reporting of outcome research findings should be more complete and more uniform across studies (see Weisz, Huey, & Weersing, 1998). In our own work, we have supplemented the use of RCTs with correlational process analyses (Huey et al., 2000), descriptive surveys (e.g., Weersing, Weisz, & Donenberg, 2002), meta-analyses (e.g., Weisz, Weiss, Han, Granger, & Morton, 1995), narrative literature reviews (e.g., Weersing & Weisz, 2002b), and benchmarking (e.g., Weersing & Weisz, 2002a). However, using these techniques as an important supplement to experimentation seems to us a far cry from attempts to rewrite the psychotherapy research literature from the ground-up, beginning with the weakest weapons in our methodological armory.

Instead of a retreat from randomized trials, we believe the field is better served by building on their strengths while focusing on the goal of developing treatment that works well with the clients, therapists, and conditions of everyday clinical practice. One approach to this goal is to extend treatments with a strong evidence base into an ever-broader range of practice settings and cultural contexts, increasing their range of clinical utility while retaining the use of RCTs as a continuing check on how well they work; see for example, the work extending MST to such populations as individuals referred to hospital inpatient units (Henggeler, Schoenwald, Rowland, & Cunningham, 2002) and to such cultural contexts as Norway (Ogden & Halliday-Boykins, 2004). Another approach is the development and application of an EST treatment development model designed to orient RCTs toward the creation of the most practice-ready treatments possible across a broad range of intervention methods and problem foci. This is the objective of the deployment-focused model of intervention development and testing (Weisz, 2004; Weisz, Southam-Gerow, Gordis, & Connor-Smith, 2003). In this model, initial intervention design relies on the interplay of clinical expertise and prior empirical evidence; the resulting protocol is tested via an initial efficacy trial to establish its potential for benefit, then used with and adapted to fit the kinds of treated individuals, treating clinicians, and treatment contexts for which the intervention is ultimately intended. These adaptations are followed by partial and ultimately full effectiveness trials (all RCTs) to assess the extent to which the adapted intervention is working well in its intended practice context. As this model and our other work suggests, we believe that the experimental method in general and the randomized trial in particular offer causal inference potential that should be embraced, not rejected, in the search for interventions that work well in the real world of clinical practice.

References

Addis, M. E., Hatgis, C., Krasnow, A. D., Jacob, K., Bourne, L., & Mansfield, A. (2004). Effectiveness of cognitive–behavioral treatment for panic disorder versus treatment as usual in a managed care setting. *Journal of Consulting and Clinical Psychology, 72,* 625–635.

Alexander, J. F., & Parsons, B. V. (1982). *Functional family therapy: Principles and procedures.* Carmel, CA: Brooks & Cole.

Allness, D. J., & Knoedler, W. H. (2003). *A manual for ACT start-up.* Arlington, VA: National Alliance for the Mentally Ill.

Andrade, A. R., Lambert, W., & Bickman, L. (2000). Dose effect in child psychotherapy: Outcomes associated with negligible treatment. *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 161–168.

Arnold, M. E., & Hughes, J. N. (1999). First do no harm: Adverse effects of grouping deviant youth for skills training. *Journal of School Psychology, 37,* 99–115.

Aronen, E. T., Noam, G. G., & Weinstein, S. R. (1993). Structured diagnostic interviews and clinicians' discharge diagnoses in hospitalized adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 32,* 674–681.

Bickman, L. (1996). A continuum of care: More is not always better. *American Psychologist, 51,* 689–701.

Bickman, L., Andrade, A. R., & Lambert, E. W. (2002). Dose response in child and adolescent mental health services. *Mental Health Services Research, 4,* 57–70.

Bickman, L., Guthrie, P. R., Foster, E. M., Lambert, E. W., Summerfelt, W. T., Breda, C. S., & Heflinger, C. (1995). *Evaluating managed mental health services: The Fort Bragg experiment.* New York: Plenum Press.

Bickman, L., Noser, K., & Summerfelt, W. T. (1999). Long-term effects of a system of care on children and adolescents. *Journal of Behavioral Health Services and Research, 26,* 185–202.

Bond, G. R., Drake, R. E., Mueser, K. T., & Latimer, E. (2001). Assertive community treatment for people with severe mental illness: Critical ingredients and impact on patients. *Disease Management and Health Outcomes, 9,* 141–159.

Bond, G. R., McGrew, H. H., & Fekete, D. M. (1995). Assertive outreach for frequent users of psychiatric hospitals: A meta-analysis. *Journal of Mental Health Administration, 22,* 4–16.

Borkovec, T. D., Echemendia, R. J., Ragusea, S. A., & Ruiz, M. (2001). The Pennsylvania Practice Research Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clinical Psychology: Science and Practice, 8,* 155–167.

Budney, A. J., & Higgins, S. T. (1998). *A community reinforcement plus vouchers approach: Treating cocaine addiction.* Rockville, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Institute on Drug Abuse.

Burns, B., Costello, E., Angold, A., Tweed, D., Stangl, D., Farmer, E., & Erklani, A. (1995). Children's mental health service use across service sectors. *Health Affairs, 14,* 147–159.

Chamberlain, P. (1998). *Family connections: A treatment foster care model for adolescents with delinquency.* Eugene, OR: Northwest Media.

Chamberlain, P. (2003). *Treating chronic juvenile offenders: Advances made through the Oregon multidimensional treatment foster care model.* Washington, DC: American Psychological Association.

Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist, 54,* 755–764.

Elliott, D. S. (Series Ed.). (1998). *Blueprints for violence prevention.* Boulder: University of Colorado, Center for the Study and Prevention of Violence.

Flynn, L. (2005). Family perspectives on evidence-based practices. *Child and Adolescent Psychiatric Clinics of North America, 14,* 217–224.

Franklin, M. E., Abramowitz, J. S., Levitt, J. T., Kozak, M. J., & Foa, E. B. (2000). Effectiveness of exposure and ritual prevention for obsessive–compulsive disorder: Randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology, 68,* 594–602.

Gold, P. B., Meisler, N., Santos, A. B., Keleher, J., Becker, D. R., Knoedler, W. H., et al. (2003). The program of assertive community treatment: Implementation and dissemination of an evidence-based model of community-based care for persons with severe and persistent mental illness. *Cognitive and Behavioral Practice, 10,* 290–303.

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose–response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9,* 329–343.

Hayes, A. M., Castonguay, L. G., & Goldfried, M. R. (1996). The study of change in psychotherapy: A reexamination of the process-outcome correlation paradigm. Comment on Stiles and Shapiro (1994). *Journal of Consulting and Clinical Psychology, 64,* 909–914.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65,* 821–833.

Henggeler, S. W., Pickrel, S. G., & Brondino, M. J. (1999). Multisystemic treatment of substance abusing and dependent delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research, 1,* 171–184.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents.* New York: Guilford Press.

Henggeler, S. W., Schoenwald, S. K., Rowland, M. D., & Cunningham, P. B. (2002). *Serious emotional disturbance in children and adolescents: Multisystemic therapy.* New York: Guilford Press.

Hoagwood, K. E. (2003, Fall). Evidence-based practices in child and adolescent mental health: Its meaning, application, and limitations. *NAMI Beginnings, 1*(3), 3–7.

Hoagwood, K. E. (in press). Family-based services in children's mental health: A research review and synthesis. *Journal of Child Psychology and Psychiatry.*

Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68,* 451–467.

Jacobson, N. S., & Christensen, A. (1996). Studying the effectiveness of psychotherapy: How well can clinical trials do the job? *American Psychologist, 51,* 1031–1039.

Jensen, A. L., & Weisz, J. R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology, 70,* 158–168.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59,* 877–883.

Latimer, E. (1999). Economic impacts of assertive community treatment: A review of the literature. *Canadian Journal of Psychiatry, 44,* 443–454.

Lincoln, T., Rief, W., Hahlweg, K., Frank, M., von Witzleben, I., Schroeder, B., & Fiegenbaum, W. (2003). Effectiveness of an empirically supported treatment for social phobia in the field. *Behaviour Research and Therapy, 41,* 1251–1269.

McGrew, J. H., & Bond, G. R. (1997). The association between program characteristics and service delivery in assertive community treatment. *Administration and Policy in Mental Health, 25,* 175–189.

Merrill, K., Tolbert, V., & Wade, W. (2003). Effectiveness of cognitive therapy for depression in a community mental health center: A benchmarking study. *Journal of Consulting and Clinical Psychology, 71,* 404–409.

Mintz, J., Drake, R. E., & Crits-Christoph, P. (1996). Efficacy and effectiveness of psychotherapy: Two paradigms, one science. *American Psychologist, 51,* 1084–1085.

Morrison, K. H., Bradley, R., & Westen, D. (2003). The external validity of controlled clinical trials of psychotherapy for depression and anxiety: A naturalistic study. *Psychology and Psychotherapy: Theory, Research, and Practice, 76,* 109–132.

Mufson, L., Dorta, K. P., Wickramaratne, P., Nomura, Y., Olfson, M., & Weissman, M. M. (2004). A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry, 61,* 577–584.

National Association of State Mental Health Program Directors. (2004). *New Freedom Commission on Mental Health: State implementation activities*. Available from: http://www.nasmhpd.org/general_files/State%20table5.pdf

National Institute of Mental Health. (2001). *Blueprint for change: Research on child and adolescent mental health. Report of the National Advisory Mental Health Council's Workgroup on Child and Adolescent Mental Health Intervention Development and Deployment.* Rockville, MD: U.S. Department of Health and Human Services.

National Institute on Drug Abuse. (1999). *Principles of drug addiction treatment: A research-based guide* (NIH Publication No. 99-4180). Rockville, MD: National Institutes of Health.

Office of the Surgeon General. (1999). *Mental health: A report of the Surgeon General.* Rockville, MD: U.S. Department of Health and Human Services.

Ogden, T., & Halliday-Boykins, C. A. (2004). Multisystemic treatment of antisocial adolescents in Norway: Replication of clinical outcomes outside of the US. *Child and Adolescent Mental Health, 9,* 77–83.

Persons, J., & Tompkins, M. (1997). Cognitive–behavioral case formulation. In T. D. Ells (Ed.), *Handbook of psychotherapy case formulation* (pp. 314–339). Oakland, CA: Center for Cognitive Therapy.

President's New Freedom Commission on Mental Health. (2003). *Achieving the promise: Transforming mental health care in America. Final Report.* Rockville, MD: U.S. Department of Health and Human Services.

Roozen, H. G., Boulogne, J. J., Van Tulder, M. W., Van Den Brink, W., De Jong, C. A., & Kerkhof, A. J. (2004). A systematic review of the effectiveness of the community reinforcement approach in alcohol, cocaine, and opioid addiction. *Drug and Alcohol Dependence, 74,* 1–13.

Rose, B. M., Holmbeck, G. N., Coakley, R. M., & Franks, E. A. (2004). Mediator and moderator effects in developmental and behavioral pediatric research. *Developmental and Behavioral Pediatrics, 25,* 58–67.

Salzer, M. S., Bickman, L., & Lambert, E. W. (1999). Dose–effect relationship in children's psychotherapy services. *Journal of Consulting and Clinical Psychology, 67,* 228–238.

Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology, 33,* 94–104.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The *Consumer Reports* study. *American Psychologist, 50,* 965–974.

Sexton, T. L., & Alexander, J. F. (2004). *Functional Family Therapy clinical training manual.* Seattle, WA: FFT.

Sexton, T. L., & Alexander, J. F. (in press). Functional family therapy for externalizing disordered youth. In J. Lebow (Ed.), *Handbook of clinical family therapy.* New York: Wiley.

Stein, L. I., & Santos, A. B. (1998). *Assertive community treatment of persons with severe mental illness.* New York: Norton.

Stiles, W. B., & Shapiro, D. A. (1994). Disabuse of the drug metaphor: Psychotherapy process–outcome correlations. *Journal of Consulting and Clinical Psychology, 62,* 942–948.

Stuart, G. L., Treat, T. A., & Wade, W. A. (2000). Effectiveness of an empirically based treatment for panic disorder delivered in a service clinic setting: 1-year follow-up. *Journal of Consulting and Clinical Psychology, 68,* 506–512.

Thompson-Brenner, H., & Westen, D. (in press). A naturalistic study of bulimia nervosa, Pt. 1: Comorbidity and therapeutic outcome. *Journal of Nervous and Mental Disease.*

Tuschen-Caffier, B., Pook, M., & Frank, M. (2001). Evaluation of manual-based cognitive-behavioral therapy for bulimia nervosa in a service setting. *Behaviour Research and Therapy, 39,* 299–308.

U.S. Department of Health and Human Services. (1999). *Mental health: A report of the Surgeon General.* Rockville, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health.

U.S. Department of Health and Human Services. (2004a, March 17). *National training and technical assistance center for child and adolescent mental health cooperative agreement* (SM 04–002). Available at http://alt.samhsa.gov/grants/2004/nofa/sm04–002_inf_NTTAC.asp

U.S. Department of Health and Human Services. (2004b, June 7). *State implementation of evidence-based practices II—Bridging science and service* (RFA-MH-05–004). Available from: http://grants1.nih.gov/grants/guide/rfa-files/RFA-MH-05–004.html

U.S. Public Health Service. (2001). *Youth violence: A report of the Surgeon General.* Washington, DC: Author.

Wade, W. A., Treat, T. A., & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology, 66,* 231–239.

Weersing, V. R., & Weisz, J. R. (2002a). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology, 70,* 299–310.

Weersing, V. R., & Weisz, J. R. (2002b). Mechanisms of action in youth psychotherapy. *Journal of Child Psychology and Psychiatry, 43,* 3–29.

Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child and Adolescent Psychology, 31,* 168–180.

Weiss, B. H., Catron, T., Harris, V., & Phung, T. M. (1999). The effectiveness of traditional child psychotherapy. *Journal of Consulting and Clinical Psychology, 67,* 82–94.

Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples.* Cambridge, England: Cambridge University Press.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between lab and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 688–701.

Weisz, J. R., Huey, S. M., & Weersing, V. R. (1998). Psychotherapy outcome research with children and adolescents: The state of the art. In T. H. Ollendick & R. J. Prinz (Eds.), *Advances in clinical child psychology* (Vol. 20, pp. 49–92). New York: Plenum Press.

Weisz, J. R., Southam-Gerow, M. A., Gordis, E. B., & Connor-Smith, J. (2003). Primary and secondary control enhancement training for youth depression: Applying the deployment-focused model of treatment development and testing. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 165–183). New York: Guilford Press.

Weisz, J. R., Weiss, B., Han, 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.

Westen, D. (1998). The scientific legacy of Sigmund Freud: Toward a psychodynamically informed psychological science. *Psychological Bulletin, 124,* 333–371.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004a). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin, 130,* 631–663.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004b). The next generation of psychotherapy research: Reply to Ablon and Marci (2004), Goldfried and Eubanks-Carter (2004), and Haaga (2004). *Psychological Bulletin, 130,* 677–683.

Received October 8, 2004
Accepted December 30, 2004 ■