
Paul Crits-Christoph  
University of Pennsylvania

G. Terence Wilson  
Rutgers University

Steven D. Hollon  
Vanderbilt University

D. Westen, C. M. Novotny, and H. Thompson-Brenner (2004) suggested that efforts to identify empirically supported treatments are misguided because they are based on assumptions that are not applicable for some types of treatment and patients. The authors of this comment argue that Westen and colleagues are simply incorrect when they assert that empirically supported treatments require that psychopathology must be highly malleable, that treatments must be brief, or that the samples studied are unrepresentative of the kinds of patients typically encountered in clinical practice—comorbidity is common in many clinical trials. Randomized controlled trials remain the most powerful way to test notions of causal agency.

Keywords: empirically supported psychotherapies, randomized clinical trials, comorbidity

Westen, Novotny, and Thompson-Brenner (2004) made a number of thoughtful and provocative observations about evidence-based psychotherapies and the randomized controlled trial (RCT) methodology on which they are based. Proponents of evidence-based treatment will endorse many of the points Westen et al. made. Among other suggestions, Westen et al. called for (a) improved reporting of the results of RCTs, especially reporting of the selection of patients and reasons for excluding patients at each stage of the recruitment and treatment process; (b) more detailed descriptions of the number and characteristics of participating clinicians; and (c) the empirical testing of approaches widely used clinically but not yet adequately examined in clinical trials. We agree with Westen et al. that such improvements would clarify the applicability of research to clinical practice. The recent adoption of the Consort RCT reporting guidelines by journals such as the Journal of Consulting and Clinical Psychology, as well as major initiatives by both the National Institute on Drug Abuse (for more information, visit the Clinical Trials Network at http://www.nida.nih.gov/CTN/Index.htm) and the National Institute of Mental Health in the direction of community-based research with high attention to generalizability, will help move published RCTs exactly in these directions (Norquist, Lebowitz, & Hyman, 1999).

Despite the important areas of agreement, we disagree with Westen et al. on numerous issues regarding the assumptions upon which empirically supported therapies (ESTs) are based and the future direction that the field of psychotherapy research should take. Also, what is often most telling in their “story” is what they selectively omitted from the narrative.

Assumptions of ESTs

Westen et al.’s (2004) arguments for moving away from validating treatments and toward “empirically informed psychotherapy” (p. 658) are primarily derived from an identification of certain assumptions of ESTs, and the RCT methodology used to test them, and then marshalling evidence to question these assumptions. If the assumptions are flawed, Westen et al. reasoned, then a new approach, not based on ESTs and RCTs, is needed. Because this is a crucial point and the foundation upon which Westen et al.’s arguments rest, we address in detail the key “assumptions” mentioned in their article.

Psychopathology Is Highly Malleable, and Most Patients Have Only One Problem

Two assumptions of ESTs are described by Westen et al.—that psychopathology is highly malleable and that most patients have primarily one problem or can be treated as if they do—are questioned by Westen et al., and are then offered as reasons why the dominant research model, which emphasizes short-term manual-based treatments of specific disorders, is problematic. But these are not inherent assumptions of RCTs or the EST movement; rather, they are issues that can (and should) be addressed by data. Westen et al. cited data on relapse rates in RCTs of depression, as well as naturalistic data on dose–response curves for patients with characterological problems, to illustrate that the assumption that psychopathology is highly malleable is not always true. The fact that not all patients respond to (or that some relapse following) short-term treatment has nothing to do with the definition, meth-
odology, or assumptions of ESTs or RCTs. What Westen et al. are confusing is the prototypic RCT in the psychotherapy literature with assumptions of ESTs.

There is no inherent reason why ESTs have to be short-term or cannot be integrative or eclectic. In fact, the American Psychological Association Division 12 Task Force that created one of the first lists of ESTs identified an integrative treatment, dialectical behavior therapy (Linehan, 1993), as a probably efficacious treatment on the basis of the one efficacy study that was available at that time. If a treatment that is even longer than 1 year had acceptable efficacy data, it would have also been included as an EST. Short-term studies, of course, are easier to conduct and fit better within the framework of the typical duration of grants made by funding agencies. In addition, the problem of attrition may be significant in long-term studies. But there are potential solutions to these practical difficulties and the threats to internal validity that they pose. Jarrett et al. (2001) reported on an RCT that included 8 months of cognitive therapy following 12–14 weeks of acute treatment for recurrent major depressive disorder (MDD), and Frank et al. (1990) have conducted an RCT involving 3 years of interpersonal therapy following initial recovery among patients with recurrent depression. The fact that Westen and colleagues (2004) are aware of these longer term RCTs (the work of Frank is mentioned in their article) but continue to suggest that short-term treatment duration is an inherent aspect of RCTs and ESTs is perplexing.

The question here is not one of an assumption but where best to begin in attempting to understand the best way to help patients. The clinician–researchers who developed and tested short-term treatments did so because they had clinical experiences that led them to believe that such short-term treatment would help many patients. If short-term treatment can help many patients, this is the sensible place to begin, particularly when one considers that many patients do not stay in psychotherapy for a long period of time. Why add to the cost and time burden for patients and insurers if short-term treatment is sufficient?

In evaluating the results of short-term studies, Westen et al. provided in our view an overly pessimistic appraisal of the outcomes of ESTs. In regard to the treatment of MDD, they cited an earlier review of controlled trials in which the median effect size for ESTs relative to placebo or control conditions was 0.30 (Westen & Morrison, 2001). Comparable effect sizes for panic (0.80) and generalized anxiety disorder (GAD; 0.90) were seen as being considerably more impressive. What Westen and colleagues did not say in their review is that this rather unimpressive effect size for the treatment of depression was based on only three studies. Three studies do not provide a very sound basis for drawing conclusions about the magnitude of treatment effects, particularly when two of the three involve control conditions that include some potentially active ingredients. In the treatment of depression, comparisons to wait-list or minimal treatment controls typically produce effect sizes of a magnitude of 0.85 and above, whereas comparisons with nonspecific or placebo controls typically produce effect sizes about half that size (Robinson, Berman, & Neimeyer, 1990).

Given the hundreds of controlled trials pointing to the efficacy of antidepressant medications and the dozens of trials supporting the efficacy of psychosocial interventions like cognitive–behavioral therapy (CBT) or interpersonal therapy (Hollon, Thase, & Markowitz, 2002), it is difficult to understand why Westen and colleagues would be willing to present such an uncharacteristically low (and potentially unstable) estimate of effect size as being representative of treatment efficacy or why they failed to point out how inconsistent their estimate was from others reported in the literature. They did at least note that rates of improvement present a more varied picture, although the 37% rate they cited for intent-to-treat samples of depressed patients (based on only six studies) was considerably lower than the estimates (50% or greater) typically reported in other larger and more inclusive reviews (Depression Guideline Panel, 1993; Mulrow et al., 1999).

Westen et al. did make an important point when they noted that despite showing substantial improvement, some depressed patients remain symptomatic at the end of short-term treatment. The average scores for intent-to-treat samples of patients treated with either drugs or psychotherapy typically end treatment about a standard deviation above the mean for normative samples on either self-report or clinical ratings. Given that these groups usually start at least two to three standard deviations above the mean, the amount of change is impressive, although the absolute outcomes still leave much to be desired. If the point is that not everyone derives full benefit from brief treatment regimens, then we whole-heartedly agree. No one except for perhaps some managed care companies would argue that brief treatment is sufficient for all or even most depressed patients (Stricker et al., 1999). Nonetheless, it is sufficient for some. If the choice is between providing long-term treatment of unknown efficacy for all patients or short-term treatment of known efficacy that can be extended for those who do not respond, then the latter strategy would seem to be preferred.

Westen and colleagues’ (2004) suggestion that investigators who have examined short-term treatments for depression have engaged in a “circular” argument and that the short-term studies “provide a definitive disconfirmation of the central hypothesis that motivated this line of research” (p. 641) fails to take into account that the short-term studies have provided evidence that these treatments help some patients considerably. The development of this area of research has in fact been linear, not circular: Early studies identified that short-term treatment helped some, but not all patients; subsequent studies have attempted to improve response rates (e.g., Teasdale et al., 2000) or provide longer term treatment for the subset of patients prone to relapses/recurrences (i.e., those with highly recurrent depression; e.g., Frank et al., 1990; Jarrett et al., 2001). Thus, the limited response rate data for short-term treatments does not suggest we should steer the field away from ESTs or the RCT methodology. Rather, the RCT–EST literature has provided a clear way to help some people and stimulated new directions to enhance outcomes.

In regard to bulimia nervosa (BN), Westen et al. once again relied on their own selective review and ignored previous reviews and meta-analyses (e.g., Fairburn & Harrison, 2003; Whittal, Agras, & Gould, 1999). The most comprehensive analysis of the evidence for treatment of eating disorders was completed by the National Institute for Clinical Excellence (NICE) in the United Kingdom (NICE, 2004). The NICE guidelines comprise a series of recommendations that are the product of an interdisciplinary and rigorous process that includes professional mental health organizations, academic institutions, and NICE itself. The recommendations are given a grade ranging from “A” (reflecting strong empirical support provided by well-conducted RCTs) to “C”
(reflecting expert opinion in absence of strong empirical data). The scope and rigor of this analysis of the data protect against the possible agendas or biases inherent in guidelines developed by particular professional organizations or individual reviewers. The NICE guidelines assigned manual-based CBT for BN (Fairburn, Marcus, & Wilson, 1993)—a theory-driven approach derived from basic and applied research as well as clinical experience—the rarely given methodological grade of “A.”

Despite relying on different reviews that cast a somewhat different view of the empirical literature, we fully agree with Westen et al.’s (2004) conclusion that short-term CBT treatment does not help everyone with BN. However, the limited success of existing treatment of BN is old news (e.g., Wilson, 1996). As with depression, researchers need to continue to test treatments for those patients that do not respond well to existing ESTs for BN (e.g., Wilson, 2004). Westen et al. (2004) ignored prior conceptual analyses and empirical research focused on improving existing ESTs for BN. The most serious omission is any mention of the development and evaluation of an enhanced CBT manual for the treatment of all eating disorders (Fairburn, Cooper, & Shafran, 2003). Fairburn et al. (2003) have expanded the cognitive-behavioral model of the mechanisms that maintain BN and other eating disorders. Each new mechanism has a module associated with it. The goal of this theory-driven treatment, then, is to flexibly utilize the relevant modules to tailor the therapy to each particular patient’s needs. An initial report on the outcome of this enhanced CBT approach indicates that it may be significantly more effective than the 1993 manual-based treatment (Fairburn, 2004). Through additional RCTs we can expect that existing ESTs will continue to be modified, or new ESTs established, with incremental advances in the amount of clinical response and sustained remission obtained. If Westen et al. believe that a longer term treatment, rather than a modified short-term treatment, will yield higher success rates for those who do not benefit considerably from short-term treatment for BN, we suggest that they develop a treatment module that can be taught to therapists and tested empirically and see if the data support their hypothesis.

The same issues apply to GAD. It has been known for some time that short-term CBT treatments for GAD, while clearly showing efficacy relative to comparison conditions, leave about half of patients with clinically significant symptoms after treatment (Chambless & Gillis, 1993). These results have led to several new ideas by clinically sophisticated investigators who are interested in improving response rates in the treatment of GAD. Ladouceur et al. (2000) are exploring the value of adding a component to CBT that focuses on coping with uncertainty in life. Thus far, this treatment has been shown to be superior to a wait-list control (Dugas et al., 2003; Ladouceur et al., 2000). Wells (1999) has developed a variation of CBT that focuses on patients’ positive and negative beliefs about worry, especially beliefs that may act to maintain worry. A trial of this approach is ongoing.

Another approach is evaluating the utility of incorporating acceptance and mindfulness techniques into traditional CBT for GAD (Roemer & Orsillo, 2002), but no data are yet available. Acceptance-based techniques include teaching patients to accept unwanted internal feelings and to practice present-moment awareness in lieu of future-focused worry. An additional new approach to the treatment of GAD has focused on the addition of interpersonal/emotional processing techniques to standard CBT for GAD. The added component helps patients to identify their interpersonal needs and fears, attend to their emotions related to interpersonal relationships, and develop more successful interpersonal behaviors (Newman, Castonguay, Borkovec, & Molnar, 2004). Preliminary results from an RCT have suggested that this new package leads to significantly better long-term outcomes than standard CBT (Newman, Castonguay, & Borkovec, 2002). Whether continued improvements in response and remission rates will come from these approaches or from longer term treatment of a psychodynamic, cognitive, experiential, or other flavor remains to be explored through further clinical experiments.

As can be seen from the above new innovations currently being empirically tested in RCTs, Westen et al.’s view that the EST movement assumes that most patients have primarily one problem or can be treated as if they do is a misrepresentation of most clinical models. This misrepresentation applies not only to these newer approaches that address underlying issues that might cut across diagnostic syndromes or include modules that can be applied depending on patient characteristics, but it also applies to many versions of standard CBT as well. For example, A. T. Beck’s cognitive therapy (A. T. Beck & Emery, 1985; A. T. Beck, Rush, Shaw, & Emery, 1979), which is the most widely empirically tested psychotherapy for both MDD and GAD, allows for individual case formulation and addresses the types of issues that might be found with difficult patients who have personality pathology. The flexibility and range of cognitive therapy is especially evident in newer writings on cognitive therapy (e.g., J. S. Beck, 1995).

Comorbidity and the Misperception of Sample Purity

Westen et al. asserted that because of the problem of comorbidity, investigators use only “pure” samples for treatment studies and that such samples have little generalizability to clinical practice. Moreover, Westen and colleagues described the use of noncomorbid samples as a requisite for experimental control in the “EST methodology.” Neither of these assertions is true. Although it is true that some studies in the literature have used a relatively noncomorbid sample, many studies, especially more recent studies, have used highly comorbid samples. For example, in the case of cognitive therapy for panic disorder, early studies used relatively pure samples because of the desire to specifically test the theoretical relevance of the model to panic per se (i.e., Clark et al., 1994). Later studies, however, addressed the clinical question of how this treatment fares with more typically comorbid samples (Barlow, Gorman, Shear, & Woods, 2000; Williams & Falbo, 1996). Similarly, studies supporting the efficacy of ESTs for MDD (DeRubeis et al., in press), GAD (Borkovec, Newman, Pincus, & LYTLE, 2002), and BN and binge eating disorder (AGRA, WALSH, FAIRBURN, WILSON, & KRAEMER, 2000; WILFLEY et al., 2000) have not excluded most Axis I and Axis II comorbidity. For example, in a recent study of cognitive therapy and pharmacotherapy for moderate to severe MDD, nearly three quarters of the 240 patients randomized to treatment met criteria for at least one other Axis I disorder and nearly half met criteria for at least one Axis II disorder and over a third met criteria for current or past substance abuse or dependence (DeRubeis et al., in press). In fact, only 16.5% of the sample met criteria for MDD alone.

Westen and colleagues (2004) tied the use of noncomorbid samples to high rates of excluding patients from RCTs to make the
point that difficult to treat patients are excluded from efficacy studies, and therefore the generalizability and clinical usefulness of these studies is limited. However, there is clear documentation, not cited by Westen et al., that the majority of patients are excluded from RCTs because their disorder is not severe enough (Jacobson & Christensen, 1996; Stirman, DeRubeis, Crits-Christoph, & Brody, 2003). For example, Stirman et al. (2003) examined 126 patient charts in a consortium of nonresearch clinics and found that only 2 patients with Axis I comorbidity would have failed to map onto some treatment literature and that the most common reasons for noninclusion was having an underresearched diagnosis (e.g., adjustment disorder) or insufficient symptom severity. There was simply little evidence that patients seen in clinical practice had symptoms that were more severe or more complicated than those treated in the empirical trials. Moreover, as Westen and colleagues acknowledged, Axis I and Axis II comorbidity have not always been found to be associated with poor outcome of short-term treatments (i.e., these patients are not always more difficult to treat). The fact that comorbidity has often been examined as a predictor of outcome of short-term trials explicitly refutes the assertion that investigators were “assuming” comorbidity is not relevant. In a previous publication, Westen and Morrison (2001) made the same faulty arguments about comorbidity and exclusion criteria in RCTs, and these arguments were addressed by DeRubeis and Stirman (2001). It is striking that Westen et al. did not even attempt to discuss this previous challenge to their mistaken understanding of the types of patients who are predominantly excluded from RCTs.

Empirically Informed Treatment

In discussing future directions, Westen et al. (2004) proposed a model of empirically informed treatments as an alternative to manual-based treatments tested in RCTs. Although such a model might seem reasonable, it begs the question, What evidence will inform whom and how will it be evaluated? Actually, such a model already exists in the clinical psychology accreditation criteria of the American Psychological Association. To be accredited, doctoral programs in clinical psychology are required to expose students to the scientific underpinnings of psychology, but it is left to individual programs to adopt any philosophy of clinical training they wish, provided they articulate it in a coherent manner. We can do better. Groups such as the Academy of Psychological Clinical Science have proposed a different view of a connection between science and practice from Westen et al.’s that emphasizes the importance of training in and dissemination of evidence-based treatment.

At the end of the day, Westen et al.’s (2004) proposal for “empirically informed therapies” (p. 631) reduces to an appeal to let the well-trained, clinically sophisticated therapist ultimately decide what to do. This is not the place to revisit the serious problems with intuitive clinical judgment (Dawes, 1994; Garb, 1998) or the persisting lack of evidence that therapist experience per se improves treatment outcome (Bickman, 1999). But the issues cannot be avoided. In addition, for better or worse, the fact that increasingly psychotherapy is being provided by master’s level counselors from a wide range of disciplines, the ideal of a doctoral level clinical psychologist who combines clinical skill and sensitivity, and who is informed by the relevant scientific research, is not a practical solution to the provision of mental health services. The role of less highly trained or credentialed therapists, not to mention self-help and less costly interventions with a stepped-care framework, will likely be important parts of improved overall care. Evidence-based protocols are likely to feature prominently within this larger scheme.

Conclusion

The RCT literature on psychotherapy has provided a wealth of scientific information and created clinical options, not previously available, that dramatically help many patients. Westen et al. (2004) have not provided any compelling evidence why RCTs should not be the gold standard for understanding the efficacy and effectiveness of psychological interventions. Nor have they provided any evidence to refute the fact that existing ESTs do help some patients. We agree with Westen et al. that there is more work to be done, particularly for some disorders. The RCT methodology remains the best way to test new short-term, and long-term, treatments as ways to improve upon existing ESTs. This is not to say that other research designs cannot contribute meaningfully to the advancement of science. The exploratory correlational studies suggested by Westen et al. that examine techniques used in actual clinical practice settings by outstanding therapists have been recommended previously (Gendlin, 1986) and might usefully generate hypotheses that can be subsequently tested using more rigorous designs (e.g., RCTs) that have greater likelihood of controlling for potentially confounding variables.

Westen et al. (2004) asserted that there have been “massive shifts in training, practice, and third-party payment” (p. 650) as a result of ESTs. Actually, this is far from the truth. As mentioned, the American Psychological Association allows individual clinical psychology training programs to adopt any philosophy of clinical training they wish as long as it is articulated in a coherent manner. Other than a relatively limited number of state-sponsored programs for the seriously mentally ill and specialized clinics that might implement a particular EST, normal community-based treatment of anxiety, depressive, and eating disorders has not shifted to ESTs, and third-party payers are not requiring practitioners to deliver ESTs. In the case of anxiety disorders, research indicates that ESTs are not common in the community and, if anything, decreased in use during the 1990s (Goisman, Warshaw, & Keller, 1999). Given this current minimal penetration of ESTs into clinical practice and training and the enormous difficulties in disseminating ESTs, Westen et al.’s (2004) suggestion regarding “empirically informed psychotherapy” (p. 658) is likely to prevent, not facilitate, the adoption of evidence-based practice.

References


Received October 15, 2004
Accepted December 10, 2004