The Efficacy of Dialectical Behavior Therapy for Borderline Personality Disorder
Drew Westen, Harvard Medical School and The Cambridge Hospital

The most scientifically appropriate conclusion to draw from the body of research on dialectical behavior therapy (DBT) is that it is probably useful in producing initial improvement on multiple variables relevant to outcome in borderline personality disorder, particularly parasuicidal behavior. These are promising and clinically relevant findings. However, its genuine efficacy in producing changes that last after the treatment has been discontinued is unknown. A major gap in the literature is the absence of any follow-up data at intervals appropriate for assessing the efficacy of a treatment for a disorder that is chronic by definition. The gap between the data and widespread beliefs about the efficacy of DBT is not, however, specific to DBT. A recent meta-analysis suggests that similar disparities are the norm in most research on empirically supported therapies, such as brief treatments for depression.

Key words: dialectical behavior therapy, borderline personality disorder, empirically validated therapies. [Clin Psychol Sci Prac 7:92-94, 2000]

Scheel (this issue) has done the field a tremendous service with this critical review of dialectical behavior therapy (DBT) for borderline personality disorder (BPD), in both the content of her review and the tone. The story she tells could easily have read as a one-sided polemic, yet in the end we are left with the kind of dialectical call for acceptance and change that Linehan has championed in the treatment of BPD. As a researcher and clinician who arrived, from a very different theoretical standpoint, at a view similar to Linehan’s regarding the importance of targeting affect regulation strategies in BPD (Westen, 1991; Westen, Muderrisoglu, Fowler, Shedler, & Koren, 1997), I have wanted to believe from the start that Linehan had developed a breakthrough in treating one of the most important aspects of the pathology of these patients.

Yet, with Scheel, I have become increasingly concerned about what at last count appeared to be roughly a 20:1 ratio of theoretical/clinical papers to empirical studies on DBT,¹ and the kind of bandwagon-jumping that many with an empirical bent of mind have decried with other treatments such as eye movement desensitization. Indeed, the evidence Scheel lays out for DBT for borderline patients is considerably weaker than the evidence for psychodynamic treatment for distressed couples (Snyder, Wills, & Grady-Fletcher, 1991)—and even for psychoanalysis for the medical control of juvenile diabetes (Fonagy & Moran, 1990). (I do not mean this as an apology for psychoanalytic clinicians who see no need to test the utility of their preferred mode of therapy, only to suggest a comparison to treatments that most readers of this journal probably assume to be empirically unvalidated if not invalidated.)

One piece of data—or perhaps its absence—is of major concern, namely, the lack of 2-, 3-, and 5-year follow-up data on patients with a disorder that by definition (as a personality disorder) is enduring and persistent. The first major empirical paper in the Archives of General Psychiatry appeared in 1991, and it is now 1999. As yet, we have seen a one-year follow-up on that sample with somewhat mixed results. The tendency of patients in the DBT group to show reductions in hospital stays and parasuicides in the year following treatment is an impressive and powerful finding, particularly for a patient group traditionally viewed by many as intractable, even with a comparison sample of patients who received what appears to be a messy melange of the good, the bad, the therapeutically ugly. But as Scheel points out, the DBT patients in this study were very unusual in the sheer care and nurturance they received by a team of people committed to them and to a new and exciting form of treatment. Whether this factor—or whether the skills training, the contingency management, the nonpathologizing attitude toward these patients’ desperate efforts to escape their pain, the supervision of therapists, the support for therapists working with very difficult patients, or simply Linehan’s charisma or brilliance—accounts for the positive findings that did emerge at one year is unknown. And data at 2 years and beyond are crucial for evaluating a treatment for a chronic disorder that produced suggestive but somewhat equivocal results at one-year follow-up.

At this point, it seems that the most scientifically appropriate conclusion to draw from the body of research Scheel reviews is the following. DBT is useful in produc-
ing initial improvement on multiple variables relevant to outcome in BPD, particularly parasuicidal behavior. Its genuine efficacy in treating whatever leads these patients to cut themselves, and in treating them in such a way as to produce changes that last after the treatment has been discontinued, is unknown. It seems likely, based on the available data, that DBT has particular utility in decreasing parasuicidal behavior. Once again, however, adequate comparison conditions need to be included to draw any strong conclusions.

Although I believe Linehan has been quite circumspect in her claims for DBT, this conclusion is very different from the sentiments I have heard from many in both the clinical and research communities, who have, for example, begun to assert that only DBT should be taught to graduate students in clinical psychology learning how to work with patients with BPD (or personality disorders in general), because it is the only empirically supported therapy for the disorder. I do not believe the data Scheel reviews can be taken by any scientist as support for the proposition that we should use DBT instead of other treatment modalities, other than perhaps unsystematic, poorly supervised treatment by community mental health counselors with unknown training and huge caseloads—the comparison group in the only major study to date.

Consider what the attitude of the scientific community might be if the data reviewed by Scheel were the primary data supporting the efficacy of a new treatment for HIV. Advocates of the treatment could point to temporary reductions in certain classes of disease to which patients with HIV often fall ill because of their weakened immune system (but that are not typically fatal). They could also claim some limited evidence for its advantages over whatever treatments patients could afford if they had no health insurance, with some of these advantages present at 6-month follow-up and some at one-year follow-up, and no known benefits thereafter. I leave to readers to decide how enthusiastic a response such a treatment would receive from physicians, patients with HIV, and their families. At the very least, one would suspect that the NIH would commission several studies comparing the treatment to the major alternative treatments in widespread use.

But there is a much broader story to be told here than the response of the academic, clinical, and public policy-making communities to the promising but limited evidence for DBT for treating some of the symptoms of BPD. The gap between the data and the conclusions many have drawn from them is not unique to DBT. The same is true for the vast majority of what are typically described as empirically validated or empirically supported treatments for many disorders.

In a recently completed meta-analysis, in which we attended to variables not typically attended to by proponents of these treatments, Kate Morrison and I found that the data are just as inconclusive for the short-term manualized treatments for depression widely described as empirically validated (Westen & Morrison, 1999). The claim that these treatments have superior efficacy to other psychosocial treatments, such as the long-term psychodynamic psychotherapies widely practiced in the community (often referred to as “treatment as usual”), is undercut by three findings. First, data on follow-up at 2 years and beyond are sparse, and the follow-up data that do exist show that, on average, of patients who enter treatment, only 37% show initial clinical improvement, and only 27% maintain this improvement at 24 months. The major, and methodologically strongest, study that has compared patients in the treatment groups to placebo controls at 18-month or longer follow-up (the NIMH Treatment of Depression Collaborative Research Program) found no differences at 18 months between any of the conditions (Shea et al., 1992). For a disorder known to remit spontaneously after 20 weeks (Judd, 1997), and to recur in roughly half of patients within 3–5 years (Maj, Veltro, Pirozzi, Lobrace, & Magliano, 1992), the sparseness of data at follow-up intervals beyond 12 months in the period we examined (1990–1998) is striking in light of the claims made for these treatments.

Second, a virtue of Linehan's treatment is that she did not exclude patients because they were polysymptomatic. In contrast, the average study of empirically supported treatments for depression excludes over two thirds of patients who present with the disorder because of comorbid conditions deemed to make them unsuitable for inclusion in the study. Unfortunately, exclusion rates and improvement rates are positively correlated across studies of short-term manualized treatments, which means that these treatments may have less efficacy for the majority of patients who present with major depression even at initial completion of the treatment—and no known efficacy for whatever percentage of patients present for chronic or intermittent depression but do not meet criteria for major depressive disorder. Given the high rates of comorbidity found in virtually all epidemiological studies of depression (e.g., Mineka, Watson, & Clark, 1998), the exclusion of comorbid patients in these studies presents an enormous threat to external validity.
Third, to conclude that one treatment is more efficacious than another requires that researchers study both treatments and compare them. The indifference of many psychodynamic clinicians to testing their therapies, and the imposition of methodological strictures preventing the testing of longer term, typically psychodynamically oriented treatments of depression by review committees at NIMH (such as manualization, which may not be appropriate for treatments that are more open ended and exploratory and are aimed at addressing polysymptomatic patients whose problems may not correspond primarily to a single Axis I disorder), has prevented the possibility of answering—or even asking—the question of whether longer term treatments would do worse, as well, or better at clinically meaningful follow-up intervals than the treatments tested in the laboratory. Until attitudes change on both sides, the issue of relative efficacy will remain in the province of ideology rather than science.

Above all else, science is about holding a critical attitude, not just toward other people’s hypotheses, but especially toward one’s own. Recent research suggests that the well-known allegiance effect in psychotherapy research—in which the investigator’s preferred therapy “wins” most of the time against rival treatments—is even more powerful than once believed, accounting for 69% of the variance in treatment outcome in high-quality studies (Luborsky et al., 1999). This leaves very little variance to be accounted for by any other variable. What Scheel’s article makes clear is that the same kind of allegiance effects appear to be operating at a broader level—for example, in the way research is summarized in review articles and findings are described to graduate students. As Scheel’s review demonstrates, even when an individual investigator is honest in the way she or he reports her data, as Linehan has been, the fervor that characterizes much of the literature on empirically supported treatments—for example, in the way she or her report her data, as Linehan has been, demonstrates, even when an individual investigator is honest in the way she or he reports her data, as Linehan has been, the fervor that characterizes much of the literature on empirically supported treatments—well intentioned as it is, in its effort to tie practice to research findings, and to challenge the Goliath of the pharmaceutical industry, which can fund dozens more studies of any new medication than NIMH can fund any new psychotherapy—can readily convert data into dogma when researchers become too interested in demonstrating their hypotheses rather than in testing them.

NOTES

1. This is still better, of course, than the ratio for other treatments for BPD, for which the numerator is typically zero.
2. We did not, in this meta-analysis, compare psychosocial treatments to medications. However, the effect sizes tend to be similar, whereas the attrition rates tend to be much lower for psychosocial treatments, suggesting that the same problems described here are applicable (if not more so, given the low effective success rates when attrition, screening of patients, and follow-up interval are taken into consideration).

3. This includes the imipramine condition.

REFERENCES


Received June 30, 1999; revised September 7, 1999; accepted September 7, 1999.