



Published in final edited form as:

Res Soc Work Pract. 2014 March ; 24(2): 213–223. doi:10.1177/1049731513503047.

Meta-Analysis and Systematic Review Assessing the Efficacy of Dialectical Behavior Therapy (DBT)

Patrick T. Panos¹, John W. Jackson², Omar Hasan², and Angelea Panos¹

¹University of Utah, Salt Lake City, UT, USA

²Harvard University, Boston, MA, USA

Abstract

Objective—The objective was to quantitatively and qualitatively examine the efficacy of DBT (e.g., decreasing life-threatening suicidal and parasuicidal acts, attrition, and depression) explicitly with borderline personality disorder (BPD) and using conservative assumptions and criteria, across treatment providers and settings.

Method—Five randomized controlled trials (RCTs) were identified in a systematic search that examined the efficacy of DBT in reducing suicide attempts, parasuicidal behavior, attrition during treatment, or symptoms of depression, in adult patients with BPD.

Results—Combining effect measures for suicide and parasuicidal behavior (five studies total) revealed a net benefit in favor of DBT (pooled Hedges' g -0.622). DBT was only marginally better than treatment as usual (TAU) in reducing attrition during treatment in five RCTs (pooled risk difference -0.168). DBT was not significantly different from TAU in reducing depression symptoms in three RCTs (pooled Hedges' g -0.896).

Discussion—DBT demonstrates efficacy in stabilizing and controlling self-destructive behavior and improving patient compliance.

Keywords

meta-analysis; outcome study; borderline personality disorder (BPD); randomized controlled trial (RCT); evidence-based practice; methodological article; mental health; field of practice; TAU; outcome study; adults; population; suicide prevention; attrition; dialectical behavioral therapy (DBT)

Introduction

Borderline personality disorder (BPD) is a substantial public mental health problem due to its high clinical prevalence, disproportionately high treatment utilization rate, and lack of responsiveness to traditional therapy. The high utilization rate for BPD occurs despite a

Reprints and permission: sagepub.com/journalsPermissions.nav

Corresponding Author: Patrick T. Panos, University of Utah, BSW Program, 395 South 1500 East, Salt Lake City, UT 84112, USA. patrickpanos@gmail.com.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Appropriate meta-analyses of DBT effectiveness is important with BPD because of the *DSM-IV-TR* system used to make the diagnosis for this psychiatric condition. Specifically, the polythetic nature (none of the criteria are either necessary or sufficient) of the diagnostic system (requiring five of the nine criteria to make a categorical BPD diagnosis) inescapably results in extensive heterogeneity both between- and within-BPD populations being treated and studied (Oldham, 2005). In other words, a meta-analysis and systematic review is critical to determine expected efficacy across treatment providers and settings. Furthermore, meta-analysis is needed to determine if DBT is efficacious across the full range of the diagnostic population that would be encountered across treatment settings. More importantly, meta-analysis can also be used to examine DBT theoretic components underpinning this therapeutic approach. Specifically, meta-analysis can be used to examine difference in specific target behaviors and skills (e.g., therapy-interfering behaviors) that DBT practitioners assert are underlying more the general mental health outcomes addressed across therapeutic approaches (e.g., depression).

In particular, it is our objective to provide a rigorous meta-analysis and systematic review that followed PRISM statement standards (Moher, Liberati, Tetzlaff, & Altman, 2009), using conservative assumptions and criteria, to examine the efficacy of DBT explicitly with BPD. More specifically, we examined the efficacy in DBT in regard to the three of the four distinct measurable behavioral targets upon which are the theoretical focus of the therapy: (1) life-threatening suicidal and parasuicidal acts; (2) decrease therapy-interfering behaviors, consistently measured as attrition rates within each study; and (3) depression, which was a common measure used across studies to measure the quality of life target. It can be easily argued that depression is a narrow measure for a broad target area. However, this measure is the only quality of life behavior target consistently assessed across studies. In general, the measures used in this meta-analysis are the same or similar to the measures used by Linehan in her original study (Linehan et al., 1991). The other behavioral target, “increasing behavioral skills,” although initially included as a variable in our study, was subsequently dropped from our meta-analysis because we found that it was not consistently operationalized or measured across studies. Multiple outcomes within individual studies were analyzed independently. The authors of this study conducted this review without compensation and have no conflicts of interests.

Method

Search Strategy

MEDLINE, EMBASE, PsychINFO, and the Cochrane Central Register of Controlled Trials databases were searched (1990 to August 2009) with the following keyword string:

(Borderline Personality Disorder or “Borderline Personality Disorder” or
Borderline State or “Borderline State” or BPD)

AND

(Dialectic Behavior Therapy or “Dialectical Behavior Therapy” or “Dialectical
Behavioral Therapy”)

AND

(Randomized Controlled Trial or Controlled Clinical Trial or Clinical Trial or Longitudinal Study or Follow Up Study).

The search strings were used consistently across all the databases that were searched, and no truncations were used. Review articles were searched for additional relevant citations. No authors were contacted.

Study Selection

We included all published RCTs that assessed the efficacy of DBT in reducing suicidal attempts, parasuicidal behavior, attrition, or depressive symptomatology in adult patients diagnosed with BPD. Studies were excluded if they lacked an adequate control group, such as “Treatment as Usual” (TAU) or “alternative psychotherapy being provided by the treatment facility,” or included patients without a diagnosis derived from either a published and validated semistructured interview or a clinical interview. We only included trials that assessed depressive symptomatology with the use of published and validated depression rating scales. Additionally, many studies published in languages other than English were not included in this meta-analysis and systematic review if the authors lacked sufficient knowledge of the language to properly assess if the study met the aforementioned criteria. The study selection process is summarized in Figure 1. After identifying the 27 studies that met the general criteria for a more detailed evaluation, each author next reviewed all the articles independently in detail. A consensus among the researchers was then obtained regarding which articles should be included in the data extraction.

There was some division among the authors whether the Turner study was comparable to the other trials. Specifically, there was disagreement whether or not it was appropriate to combine the results of the Turner group, which was designed to test comparative effectiveness against client-centered therapy (CCT), with the other trials that were designed to test efficacy. Because it met our a priori definition, “alternative psychotherapy being provided by the treatment facility,” we included it in the analysis and performed a sensitivity analysis to examine if any bias was introduced by its inclusion. Specifically, we performed repeated analyses where each study was excluded once, to determine the influence each study had on the combined effect size.

Data Abstraction

Data forms, coding, and data abstraction process followed the abstraction procedures as outlined by Littell, Corcoran, and Pillai (2008). Specifically, the coding structure that was needed to answer the primary research questions was designed; next pilot tested by entering data onto an electronic form, and then revised. Team members initially conducted data extraction and coding independently. Next, the authors compared and discussed their findings until a consensus about the data was obtained, and then the research team jointly entered the data into a Microsoft Excel table. Study characteristics of interest were age, gender, psychopharmacotherapy use, description of DBT and TAU components, and the therapist experience. For outcome data, we extracted event data for suicidal attempts (absolute count), mean rating scores, and standard errors of parasuicidal behavior (e.g., Suicide Attempt Self-Injury Interview, Parasuicide History Interview [PHI], BPD Severity Index, Overt Aggression Scale–Modified, and Daily Log of Events) and depression

symptomatology (e.g., Hamilton Rating Scale for Depression and Beck Depression Inventory) before treatment, and after 12 months of follow-up, for both the DBT and TAU treatment arms. Additionally, attrition rates were recorded for each arm. At least one outcome of interest was available for extraction in each study.

Assessment of Methodological Quality

The methodological quality of included studies was assessed using the Clinical Trials Assessment Measure (CTAM; Tarrier & Wykes, 2004). The CTAM was developed to function as a quality measure for clinical trials in psychiatry and mental health with “psychological treatments.” Specifically, the CTAM assesses 15 items of the experimental design over six areas: sample size and subject recruitment; assignment to treatment, outcome assessment, control group, description of treatments, and analysis. A maximum total score of 100 is possible, which would indicate that the quality of design met the highest possible standards according to Consolidated Standards of Reporting Trials guidelines (Moher, Schulz, & Altman, 2001). The CTAM also has good concurrent validity in that it correlates highly with other generic assessment of clinical trial quality, including Jadad et al. (1996; $r = .97$) and Chalmers et al. (1981; $r = .93$). The CTAM exhibits a good blind interrater agreement of 0.96 and appropriate internal consistency (Cronbach’s $\alpha = .697$). Each study selected for inclusion in the meta-analysis was rated by at least two member of the research team, and all members rated at least one study. All pairs of ratings of the same study were found to be within 5% of one another.

The CTAM scores allow for the possible identification of potential experimental bias due to poor research design. Although total CTAM scores permit an overall comparison of research design quality, a more practical method is to examine the particular design areas, if any, that receive low scores.

The seminal study first conducted by Linehan et al. (1991) had the greatest number of difficulties that affected research quality, resulting in a relatively low CMAT score of 65. Specifically, this research study used subjects who were referred into the study, permitting for possible selection bias and difficulties with generalizability of the obtained results. Additionally, there was a lack of an adequate description of the randomization process to ensure no bias entered into the allocation process. Next, there was no intention-to-treat analysis, leading to a possible increase in false-positive findings due to bias introduced by differences in treatment dropouts. Also, the sample size was too small to ensure adequate power; nevertheless, positive results were still found. Finally, verification and the description of rater blinding was not provided, introducing the possibility assessment bias due rater expectations.

The research study of next highest quality, in regard to research standards measured by the CTAM, was conducted by Turner (2000), achieving a relatively low score of 70. In particular, Turner recruited subjects who first were treated at an emergency room for suicidal ideation and then were referred to a community mental health center. Consequently, the method of recruitment was vulnerable to selection bias and limits generalizability. Additionally, Turner failed to adequately describe his randomization process, and there was no indication that the randomization process was conducted independently from the trial

research team, which could have introduced an expectation bias. Also, the sample size was small, with only 24 participants divided among the two treatment arms, greatly affecting the power of the study. Next, the blinding of the assessors was not adequately described or verified. Finally, an adequate analysis of the dropouts (10 of the 24) was not provided; however, an intent-to-treat analysis protocol was appropriately conducted.

Koons et al. (2001) used significantly stronger research protocols; nevertheless, research methodology difficulties were identified, resulting in a modest CTAM score of 83. Like the previous studies, Koons and colleagues specifically recruited by referral a low number of participants, resulting in only 10 participants being assigned to each treatment condition. Additionally, the researchers failed to describe blinding procedures and verification of the assessors, nor did they adequately describe dropouts or follow intention-to-treat analysis protocols.

In their 2006 study, Linehan and colleagues made significant changes to their methods, including improving recruitment procedures to improve generalizability, and increasing the number of participants in each treatment condition. The study only received a CTAM score of 85 due to the failure to verify rater blinding, and an inadequate investigation and handling of dropouts from the assessment. It was also unclear whether and intention-to-treat protocol was followed during the analysis.

Verheul et al. (2003) had the highest quality of research design, obtaining a CTAM score of 94. The only problem with their research article was the lack of an adequate description of the randomization process to ensure no bias entered into the allocation process. In particular, it was unclear if randomization occurred independently from the trial research team. Nevertheless, the standards that the researchers followed were praiseworthy.

Quantitative Data Analysis

Suicidality—The seven studies initially selected for inclusion in the meta-analysis presented a variety of different effect measures for suicidality (see Table 1). Two studies presented results in terms of a relative risk, where three presented preand postmean scores for each treatment arm, along with their respective standard errors. For studies reporting relative risks, odds ratios (ORs) were calculated from the event data and converted standardized mean difference d by multiplying the log OR by a constant of $(3/\pi)^{1/2}$ (Hasselblad & Hedge, 1995). In this calculation, we assume that the number of suicide attempts is a continuous trait that follows a logistic distribution. For studies reporting pre- and postmean scores for each treatment arm, the standardized mean difference was calculated by subtracting the postmean score from the premean score, and dividing it by the standard error of the mean difference. In computing this standard error, it was assumed that the correlation between parasuicidal behavior scores from the Parasuicide History Interview (PHI) was .39 for a 12-month period (Linehan et al., 1991). To correct for small sample bias, the standardized mean difference d was converted into Hedges g (Lipsey & Wilson, 2001). It was not possible to compute the standardized mean difference for one of the studies that reported β coefficients from hierarchical linear model (Clarkin, Levy, Lenzenweger, & Kernberg, 2007), and as a result this study was omitted from the meta-analysis. For each analysis, we performed a test for homogeneity and used both Mantel–Haenszel fixed effects

and DerSimonian and Laird random effects models to calculate the combined effect size. Heterogeneity was assessed using Cochran's Q and the I^2 statistics. In particular, Cochran's Q test is a nonparametric statistical test to verify if different treatments have identical effects (Conover, 1999). I^2 is a complementary approach to examining heterogeneity that estimates variability in the observed effect size that is not due to random error (Higgins & Thompson, 2002; Pereira, Patsopoulos, Salanti, & Ioannidis, 2010). Nevertheless, the heterogeneity between studies suggest that the random effects model would be the most appropriate approach in this study and that the true effect could vary from study to study.

To report our findings in clinically relevant language, we performed additional meta-analyses of the two studies reporting risk ratios, and another three studies reporting pre- and postmean scores with standard deviations.

Further Qualitative Review

During the final selection of articles, as mentioned previously, one study by Clarkin, Levy, Lenzenweger, and Kernberg (2007) was omitted from the analysis. It reported an effect size (a β and intercept) from a hierarchical linear model that estimated a growth trajectory for each individual. We believe that it would be inappropriate to pool this effect measure with nonparametric effect sizes that are not based upon the same modeling assumptions.

Nevertheless, in comparing transference-focused psychotherapy, DBT, and supportive therapy in a RCT, the Clarkin research group (2007) obtained results consistent with the findings of this meta-analysis. First, they found that both transference-focused therapy and DBT were significantly associated with improvement in suicidality. Additionally, all three therapies were "predictive of the rate of change in a positive direction for depression, anxiety, global functioning, and social adjustment" (p. 927). The researcher concluded that the general equivalence of outcomes across the three therapies studied indicate that there may be different routes for change in people with BPD and that DBT is successful at teaching skills needed to regulate emotions and reduce symptoms. They further asserted the need for additional research to determine the mechanisms through which treatments lead to outcomes.

Results

Description of Included Studies

Table 1 summarizes the characteristics of the five included studies (top portion). We identified 247 individuals with BPD in these five RCTs, with 145 patients (59%) belonging to the two major RCTs conducted by Dr. Marsha M. Linehan, the creator of DBT. Four of these five studies (Koons et al., 2001; Linehan et al., 1991, 2006; Verheul et al., 2003) were confined to women, with the fifth, conducted by Dr. Ralph M. Turner, including only five men in a study totaling 24 participants; thus, men comprised only 2% of the overall sample. The majority of subjects were young women, with the Linehan studies restricting age range to 18–45; only one study, conducted by Dr. Roel Verheul, included older subjects, with an upper age limit of 70. These five studies constitute the range of people with BPD that have participated in the RCTs, and care must be taken in generalizing beyond this cohort.

Since all studies were RCTs, and targeted relatively heterogeneous populations, the use of a random model was chosen for this study. In particular, although sociodemographic and clinical characteristics were similarly distributed across the two comparison groups in each study, there were distinct differences between studies. For instance, studies differed in the type of randomization used, with the more recent studies (Linehan et al., 2006; Verheul et al., 2003) employing more sophisticated computerized adaptive minimization randomization techniques. Intake assessments included use of structured interviews, with or without the addition of standardized psychometric instruments, to accurately diagnose BPD using accepted clinical criteria across all five studies.

Of note, the patients included in all five RCTs were sampled from a population of patients referred for psychotherapy and do not reflect population-based probability samples of BPD patients. All patients had exhibited parasuicidal behaviors or attempted suicide prior to study enrolment. Studies varied in their descriptions of the use of psychotherapeutic medicines, with Turner (2000) and Verheul et al. (2003) reporting use of psychotherapeutic medicines in a comparatively larger proportion of the control groups. Nearly all studies followed subjects for 1 year of treatment and varied substantially in the magnitude of sample attrition rates. Additionally, differences in attrition rates of greater than 20% between the control and treatment groups in four of the five raise the possibility of threats to internal validity in the studies due to changes in the equivalences of the study groups. Nevertheless, each study asserted that there were no identifiable differences between the treatment and control groups.

TAU was used for the control groups in four studies (Koons et al., 2001; Linehan et al., 1991, 2006; Verheul et al., 2003), whereas the fifth (Turner, 2000) used a modification to the TAU approach in the comparison group. Outcomes were assessed using standardized psychological scales and structured interviews, commensurate with established professional standards, across all five studies. On a 0–100 point scale, the five studies earned CTAM quality scores between 70 and 100.

Reasons for excluding the Clarkin et al. (2007) study in Table 1 are described in the Method section.

Efficacy of DBT in Treating Parasuicidal Behavior

Pooled analyses were performed separately for parasuicidal behavior assessments measured with psychological scales and suicide attempt rates measured in absolute numbers to obtain interpretable effect measures, and subsequently combined using the Hedges' *g* to pool effects across these two types of measures.

Three studies (Koons et al., 2001; Linehan et al., 1991; Turner, 2000) measured participants' performance on psychological scales designed to assess parasuicidal behaviors in a pre- and postintervention design. Pooled estimates of the mean difference in pre- and postintervention scores on parasuicidal behavior psychological scales (mean difference score) were compared between the treatment and control arms, using the DerSimonian–Laird random effects statistical model. Since one of the three studies (Turner, 2000) used a different psychological scale for assessing parasuicidal behaviors and sample sizes were relatively small, effect measures were standardized using the standard error of the mean difference score and

multiplied by a constant to obtain Hedges' g values. This minimizes the bias from small sample sizes and permits an accurate comparison to be made across the three studies.

Despite the small number of studies, the p value for Cochran's Q was .07, with an I^2 of 62.5%, suggestive of significant heterogeneity among studies. This is consistent with an anticipated high degree of heterogeneity, given the temporal, geographic, and methodological differences among these three studies, and justifies the use of a random effects model. As seen in Figure 2, the 95% confidence interval (CI) for the pooled Hedges' g crosses the null value of 0, indicating a lack of substantial treatment advantage in reducing parasuicidal behavior, as assessed by psychological scales, from the use of DBT compared with TAU (pooled Hedges' g -0.636, 95% CI [-1.382, +0.111]).

However, the two studies (Linehan et al., 2006; Verheul et al., 2003) that compared the efficacy of DBT to TAU in reducing the relative risk for suicide attempts did show a significant reduction in suicide risk (Figure 3). Both studies were conducted rigorously, had large sample sizes, accounting for 64% of the individuals in our meta-analysis (159/247 subjects), and used computerized adaptive minimization randomization techniques to appropriately select subjects for the two comparison groups. Since only two studies were pooled in this analysis, both fixed and random effects models are expected to yield similar results. The pooled OR of 0.31 (95% CI [0.15, 0.67]) from these two studies indicates that DBT offers a significant advantage over TAU in reducing the hard end point of suicide attempts. To further evaluate the efficacy of DBT, we constructed a DerSimonian-Laird random effects model that pooled the results from all five suicide studies using the Hedges' g to quantify effect measures for comparability (Figure 4). The pooled Hedges' g from this model favored DBT with a lower confidence limit well above the null value, suggesting an overall benefit for DBT in reducing parasuicidal behaviors and mitigating suicide risk when compared with TAU (pooled Hedges' g -0.622, medium effect, 95% CI [-0.983, -0.260]).

Efficacy of DBT in Reducing Attrition

Five studies compared the risk difference for attrition during treatment between DBT and TAU (Figure 5). Obtained results indicate that DBT was only marginally better than TAU in reducing attrition during treatment in five RCTs (pooled risk difference -0.168, small effect, 95% CI [-0.323, -0.002]). It should be noted, however, the Cochran's Q was greater than the degrees of freedom (7.5 vs. 4) and the I^2 was fairly large (46.7%), indicating substantial heterogeneity among these studies; there did not seem to be any discernible pattern to this heterogeneity. Therefore, the efficacy of DBT in reducing treatment attrition may depend on the clinical population and therapy-related factors. Identifying factors that may predict patient response in regard to attrition will require further research.

Efficacy of DBT in Treating Depression

Three studies (Koons et al., 2001; Linehan et al., 2006; Turner, 2000) measured the impact of DBT on depression using the Hamilton Depression Rating scale (HAM-D), a commonly used psychological scale for measuring depression; all three used a pre- and postintervention design comparing DBT with TAU or a modification of TAU (Turner, 2000). Pooled estimates of the mean difference in pre- and postintervention HAM-D scores (mean

difference score) were compared between the treatment and control arms using the DerSimonian–Laird random effects statistical model. Due to the small sample sizes in all the three studies, we converted effect measures into Hedges' g values for accuracy of comparison across the three studies.

Despite the small number of studies, the p value for Cochran's Q was .006, with an I^2 of 80.7%, suggestive of significant heterogeneity among studies. This is consistent with an anticipated high degree of heterogeneity, given the geographic and methodological differences among these three studies, and justifies the use of a random effects model. As seen in Figure 6, the 95% CI for the pooled Hedges' g crosses the null value of 0, indicating a lack of substantial treatment advantage of DBT in reducing depression (pooled Hedges' g -0.896 , 95%, large effect, CI $[-1.903, +0.116]$). It should be noted that the pooled Hedges' g indicates a large effect size. Nevertheless, in examining the data, we found this large effect was due to the strong results found in a single study (Turner, 2000), and the variability between studies resulted in the lack of significance between DPT and TAU in treating depression.

Sensitivity Analyses

For the five suicide studies, we performed two types of sensitivity analyses. We performed an analysis with “one study removed,” using Hedges' g as the effect measure, to determine if the pooled results were heavily dependent on any one particular study. As seen in Figure 7, exclusion of any one study did not result in pooled estimates that were qualitatively different from the results of the overall meta-analysis.

Since the two major RCTs that established the benefits of DBT in reducing suicide attempts and parasuicidal behaviors were also conducted by Dr. Marsha M. Linehan, the creator of this new form of psychotherapy, we compared pooled results from her two studies with pooled results from the remaining three studies. Separate pooled estimates from these two groups of studies were qualitatively similar and favored DBT (results not shown). More significantly, non-Linehan studies showed slightly greater efficacy, suggesting that other therapists and treatment facilities can successfully use this form of psychotherapy.

Publication Bias

Publication bias can occur during a meta-analysis because a study with positive results is more likely to be published than ones with negative results. Bias in the systematic review can therefore occur because of the overrepresentation of studies with positive results. Typical tests of potential publication bias were meaningless in this meta-analysis, due to the extremely low number of studies examining the efficacy of DBT. For example, an attempt at producing funnel plots with only 5 points did not present a clear pattern that would allow interpretation. Therefore, formal analyses of publication bias are not presented.

Nevertheless, a qualitative review of efficacy studies on DBT suggests that the potential for publication bias is low. DBT is a relatively new form of psychotherapy, requiring considerable time and resources to train psychotherapists appropriately. Accordingly, the number of centers with the resources available for conducting well designed and appropriately executed RCTs of this treatment modality is very limited. Therefore, it is not

unreasonable to conclude that the number of unpublished studies in this field is likely to be negligibly small.

Since the number of RCTs comparing DBT versus TAU for reducing parasuicidal behaviors, suicide attempts, and depression are very few, with each study reporting on a battery of psychological evaluations, and the resulting study reports include a reasonable proportion of nonsuperiority conclusions, we believe that our meta-analysis is not substantially affected by publication bias.

Discussion

DBT is partially based upon the theoretical belief that people with BPD significantly lack the ability to regulate emotions or to tolerate stress. Additionally, personal and environmental factors often block and inhibit the use of behavioral skills those people may have, thereby reinforcing dysfunctional behaviors. These deficits and difficulties combine and result in pervasive instability in moods and extensive pattern of self-harming acts. Therefore, a main overarching goal of DBT is to stabilize the clients and help them achieve behavioral control. Consequently, as noted previously, DBT has four main behavioral targets: (1) decrease life-threatening suicidal and parasuicidal acts, (2) decrease therapy-interfering behaviors, (3) decrease quality of life interfering behaviors, and (4) increase behavioral skills. Individually, the five RCTs examined in this meta-analysis have uniformly shown the overall efficacy of DBT in treating BPD. As mentioned previously, it was pointed out by the *APA Practice Guidelines* what these studies have failed to show was whether the reported efficacy was “derived from specific ingredients of dialectical behavior therapy” (APA, 2001). It was our objective, therefore, to conduct a rigorous meta-analysis and systematic review, using conservative assumptions and criteria, to examine the efficacy of DBT explicitly in achieving its different behavioral targets across a range of clients, therapists, and treatment settings.

Obtained results indicate that DBT is efficacious in helping clients with BPD to significantly decrease suicidal and self-harming acts. Based upon the pooled OR, suicide attempts appear to be reduced by approximately two thirds. The goal of stabilizing clients and help them achieve behavioral control is clearly met by the accomplishment of this behavioral target. As shown by the meta-analysis in pooling the outcomes for parasuicidal and suicidal acts, the behavioral target of decreasing self-harm and suicidal attempts is efficaciously obtained across studies. Considering the high utilization rates that people with BPD of psychiatric services because of self-harm, this finding suggests that mental health resources could be conserved through the use of DBT.

In regard to decreasing therapy-interfering behaviors, we obtained results showing that DBT was effective in reducing treatment attrition in some studies. This could be due to particular aspects of the study execution, or result from valid differences in efficacy between study populations. For example, the amount of DBT expertise may be lower in some treatment plans than others.

When examining if DBT decreases quality of life-interfering behaviors (such as depression), despite positive improvement, the findings across studies were not significant. There are several possibilities to explain this finding. First, this behavioral target is more broadly defined and more difficult to operationalize and measure. Therefore, change is more difficult to detect and requires more statistical power. Second, because of the broad nature of this behavioral target, DBT as an intervention is less focused. Behavioral change is therefore more difficult to achieve and is likely diffused across the multiple dimensions of the behavioral target. Finally, unlike the previous behavioral targets that focused on stabilization and achieving control, this target focuses on helping people with BPD develop skills and abilities, which it is assumed they are lacking. Additionally, the clients must generalize these new skills and abilities across a wide range of life stressors. In brief, despite positive trends, results obtained during the meta-analysis fail to demonstrate that DBT is efficacious in decreasing quality of life-interfering behaviors across studies. The fourth behavioral target was not examined in this study.

There are three other findings obtained during our analysis that are worth noting. First, despite the fact that Dr. Linehan is the founder of DBT, other therapists are fully capable of equaling or surpassing her success in therapy. This is likely due to published training materials available to therapist wishing to learn this type of therapy. Second, quality of the study was mildly correlated with outcome in a positive direction. This may suggest that the level of care that people use in setting up and following the therapy is indicative of the fidelity with which they follow the model. Finally, the Turner group was consistently an outlier and although its removal did not change the parasuicide Hedges g , it did decrease the depression hedges g by 44%. This study, besides the only trial to test comparative effectiveness, employed therapists with substantially more experience than the others (22 years on average). Even though they were experienced at administering CCT, three of the four therapists in this study achieved greater effectiveness with DBT than CCT. The authors suggest that, among other possibilities, “differences in the patients’ beliefs and expectancies might have contributed to differential patient outcomes” (Turner, 2000, p. 419).

There are significant limitations to this study. Clearly, the small number of studies, as well as the small number of measure behavioral targets, makes cross-study comparisons difficult. There is a significant need for additional RCTs with larger patient samples.

Without a doubt, BPD is one of the most difficult psychiatric conditions to treat, and there have been few treatment options available to therapist. DBT demonstrates great efficacy in stabilizing and controlling self-destructive behavior and improving patient compliance. There is insufficient data to conclude that DBT is successful at teaching the behavioral skills necessary to make long-term improvement in the quality of life of patients. More extensive research is needed to make this determination.

Acknowledgments

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: Author; 2000. text rev.
- American Psychiatric Association. 2001; Practice guideline for the treatment of patients with borderline personality disorder: American Psychiatric Association Practice Guidelines. *American Journal of Psychiatry*. 158:1–52.
- Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. 1981; A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials*. 2:31–49. [PubMed: 7261638]
- Choi-Kain LW, Gunderson JG. 2009; Borderline Personality Disorder and resistance to treatment: The primary sources of resistance. *Psychiatric Times*. 26:35–36.
- Clarkin JF, Levy KN, Lenzenweger MF, Kernberg OF. 2007; Evaluating three treatments for borderline personality disorder: A multiwave study. *American Journal of Psychiatry*. 164:922–928. [PubMed: 17541052]
- Conover, WJ. Practical nonparametric statistics. 3. New York, NY: John Wiley; 1999.
- Dimeff L, Linehan MM. 2001; Dialectical behavior therapy in a nutshell. *The California Psychologist*. 34:10–13.
- Gross R, Olfson M, Gameroff M, Shea S, Feder A, Fuentes M, Weissman MM. 2002; Borderline personality disorder in primary care. *Archives of Internal Medicine*. 162:53–60. [PubMed: 11784220]
- Hasselblad V, Hedge LV. 1995; Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*. 117:167–178. [PubMed: 7870860]
- Higgins JP, Thompson SG. 2002; Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*. 21:1539–1558. [PubMed: 12111919]
- Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. 1996; Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*. 17:1–12. [PubMed: 8721797]
- Koerner K, Dimeff LA. 2000; Further data on dialectical behavior therapy. *Clinical Psychology: Science & Practice*. 7:104–112.
- *. Koons CR, Robins CJ, Tweed JL, Lynch TR, Gonzalez AM, Morse JQ, Bastian LA. 2001; Efficacy of dialectical behavior therapy in women veterans with borderline personality disorder. *Behavior Therapy*. 32:371–390.
- Kosfelder J. 2007; A meta-analysis of the efficacy of dialectical behavior therapy in borderline personality disorder. *Zeitschrift Für Klinische Psychologie, Psychopathologie Und Psychotherapie: Roschung und Praxis*. 36:11–17.
- Lenzenweger M, Lane M, Loranger A, Kessler R. 2007; DSM-IV personality disorders in the national comorbidity survey replication. *Biological Psychiatry*. 62:553–564. [PubMed: 17217923]
- Linehan, MM. Cognitive behavioral therapy of borderline personality disorder. New York, NY: Guilford Press; 1993a.
- Linehan, MM. Skills training manual for treating borderline personality disorder. New York, NY: Guilford Press; 1993b.
- *. Linehan MM, Armstrong HE, Suarez A, Allmon D, Heard HL. 1991; Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Archive of General Psychiatry*. 48:1060–1064.
- *. Linehan MM, Comtois KA, Murray AM, Brown MZ, Gallop RJ, Heard HL, Lindeboim N. 2006; Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*. 63:757–766. [PubMed: 16818865]
- Linehan MM, Tutek DA, Heard HL, Armstrong HE. 1994; Interpersonal outcome of cognitive behavioral treatment for chronically suicidal borderline patients. *American Journal of Psychiatry*. 151:1771–1776. [PubMed: 7977884]

- Lipsey, MW, Wilson, DB. Practical meta-analysis. Thousand Oaks, CA: Sage; 2001.
- Littell, JH, Corcoran, J, Pillai, V. Systematic reviews and meta-analysis. New York, NY: Oxford University Press; 2008.
- Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group. 2009; Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Open Medicine*. 3:123–130. Retrieved July, 19, 2011
- Moher D, Schulz KF, Altman DG. for the CONSORT Group. 2001; The CONSORT statement: Revised recommendations for improving the quality of reports of parallel group randomized trials. *Journal of American Medical Association*. 285:987–91.
- Oldham, JM. Guideline watch: Practice guideline for the treatment of patients with borderline personality disorder. Arlington, VA: American Psychiatric Association; 2005.
- Ost LG. 2008; Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behaviour Research & Therapy*. 46:296–321. [PubMed: 18258216]
- Pereira TV, Patsopoulos NA, Salanti G, Ioannidis JPA. 2010; Critical interpretation of Cochran's *Q* test depends on power and prior assumptions about heterogeneity. *Research Synthesis Methods*. 1:149–161. [PubMed: 26061380]
- Sansone, RA, Sansone, LA. Borderline personality: A psychiatric overview. In: Sansone, RA, Sansone, LA, editors. *Borderline personality in the medical setting: Unmasking and managing the difficult patient*. New York, NY: Nova Science; 2007. 3–36.
- Scheel KR. 2000; The empirical basis of dialectical behavioral therapy: Summary, critique, and implications. *Clinical Psychology: Science & Practice*. 7:68–86.
- Tarrier N, Wykes T. 2004; Is there evidence that cognitive behaviour therapy is an effective treatment for schizophrenia? A cautious or cautionary tale? *Behaviour Research & Therapy*. 42:1377–1401. [PubMed: 15500811]
- Tucker L, Bauer SF, Wagner S, Harlam D, Sher I. 1987; Long-term hospital treatment of borderline patients: A descriptive outcome study. *American Journal of Psychiatry*. 144:1443–1448. [PubMed: 3674226]
- *. Turner RM. 2000; Naturalistic evaluation of dialectical behavior therapy-oriented treatment for borderline personality disorder. *Cognitive and Behavioral Practice*. 7:413–419.
- *. Verheul R, van den Bosch LMC, Koeter MWJ, de Ridder MAJ, Stinen T, van den Brink W. 2003; Dialectical behavior therapy for women with borderline personality disorder: 12-Month, randomized clinical trial in the Netherlands. *British Journal of Psychiatry*. 182:135–140. [PubMed: 12562741]

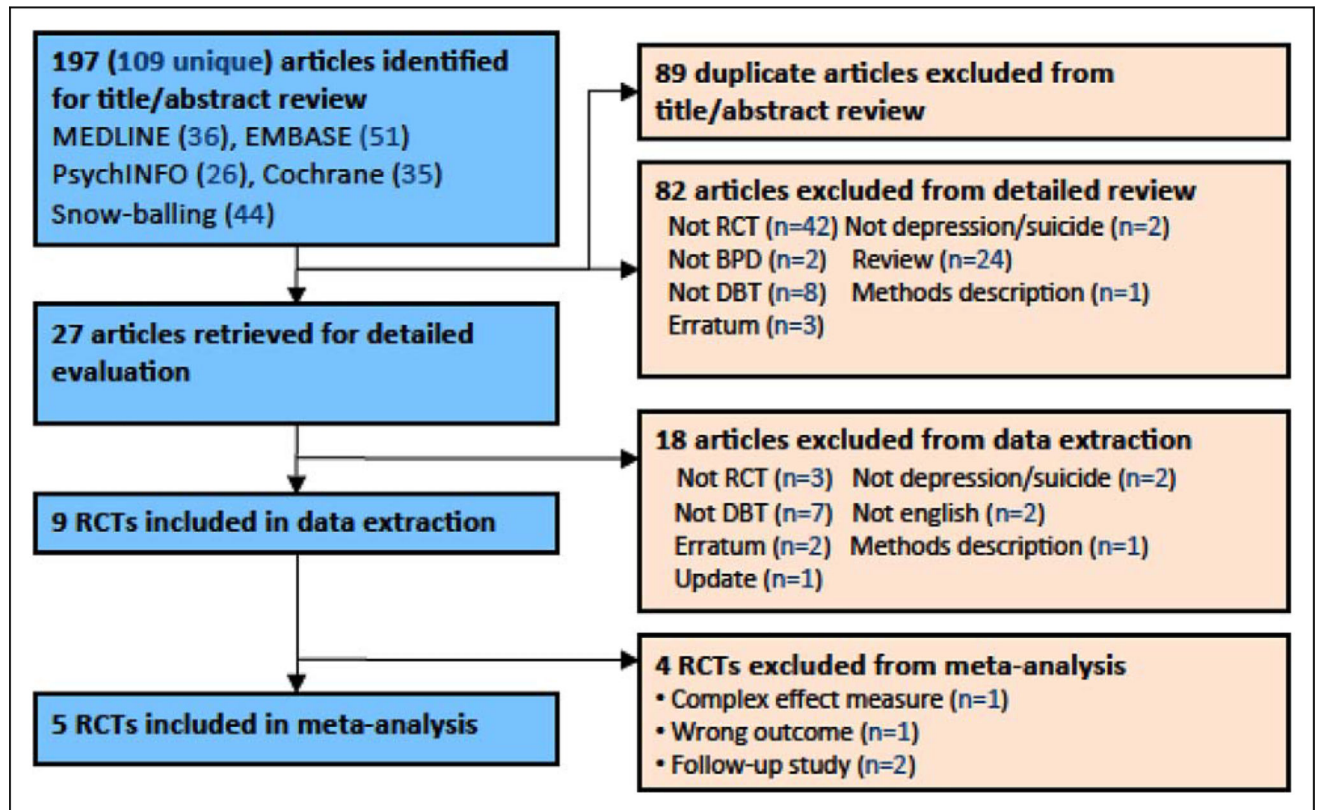


Figure 1.
Study flow diagram.

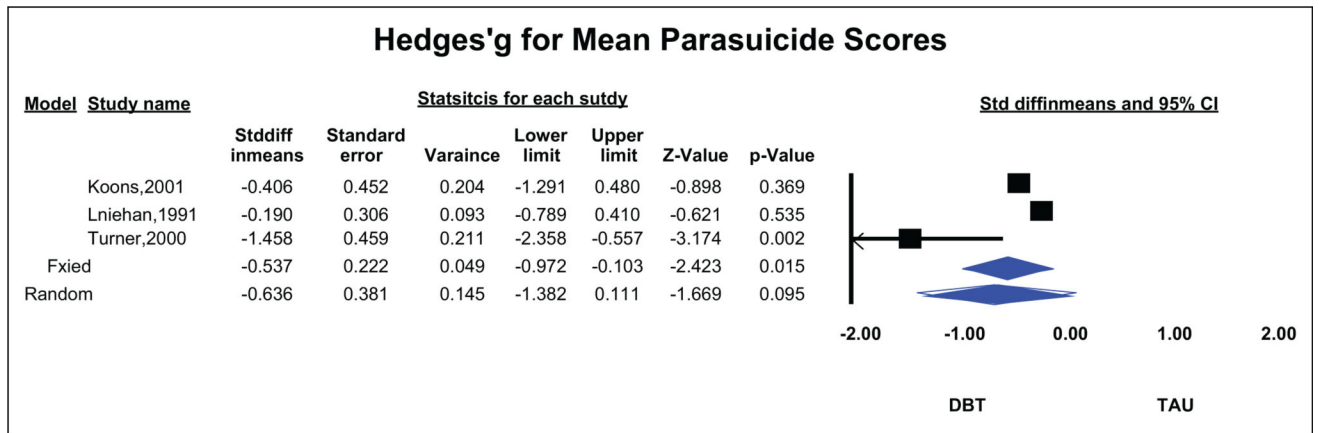


Figure 2.
Hedges' *g* for mean parasuicide scores.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

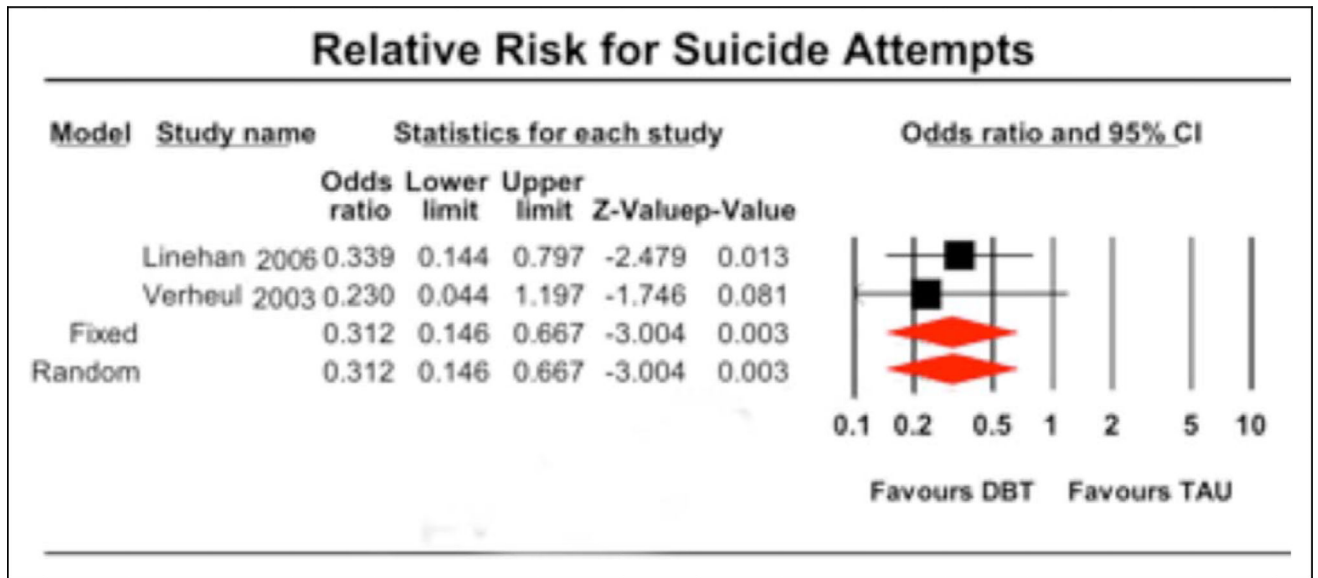


Figure 3.
Relative risk for suicide attempts.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

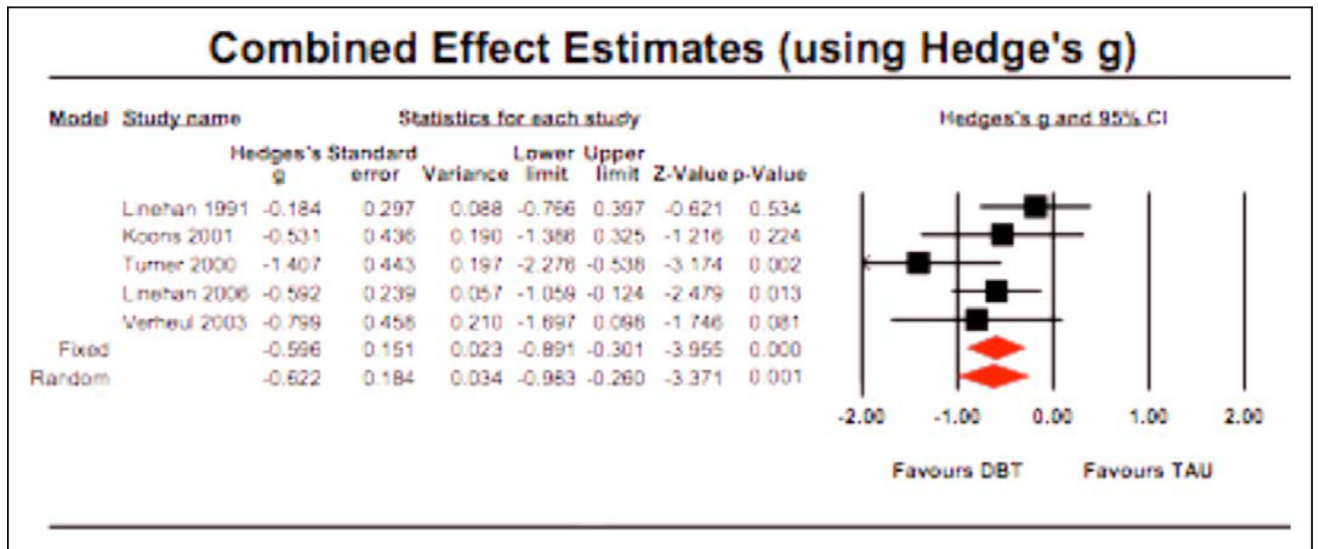


Figure 4.
 Combined effect estimated (using Hedge's *g*).

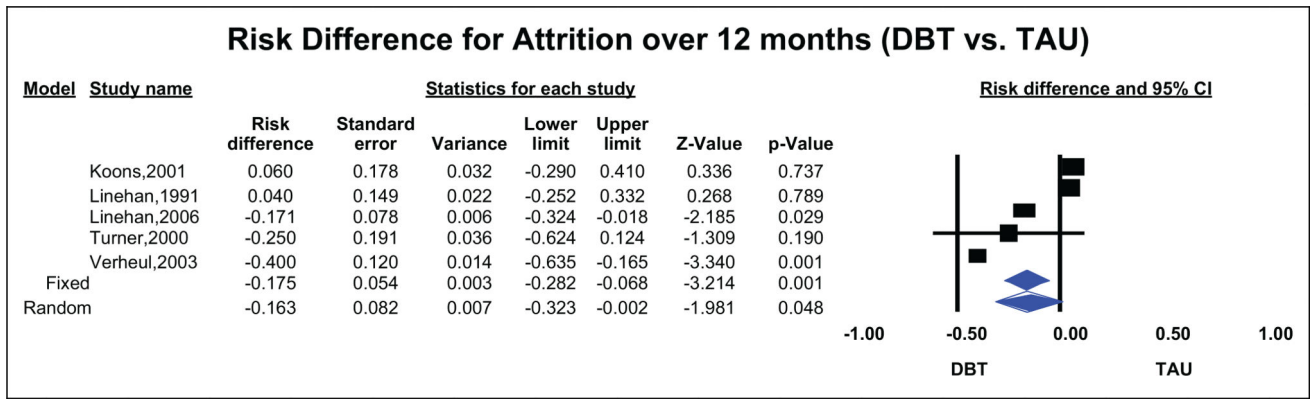


Figure 5. Risk difference for attrition over 12 months (DBT vs. TAU). DBT = Dialectical behavior therapy; TAU = treatment as usual.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Hedges'g for Mean Change in HAM-D Depression Score

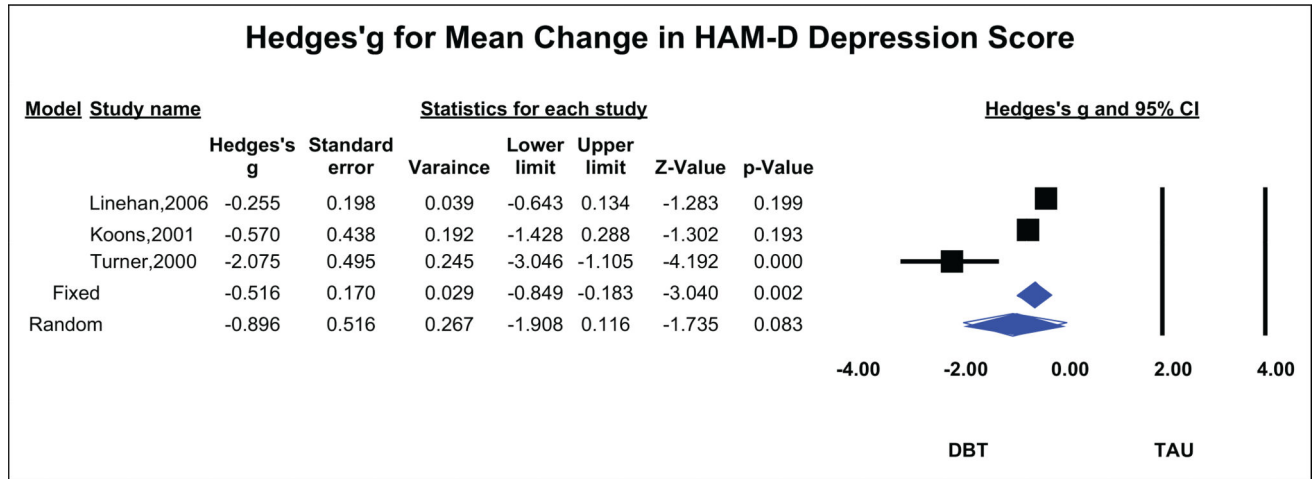


Figure 6. Hedges' *g* for mean change in HAM-D depression score. HAM = Hamilton Depression Rating scale.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

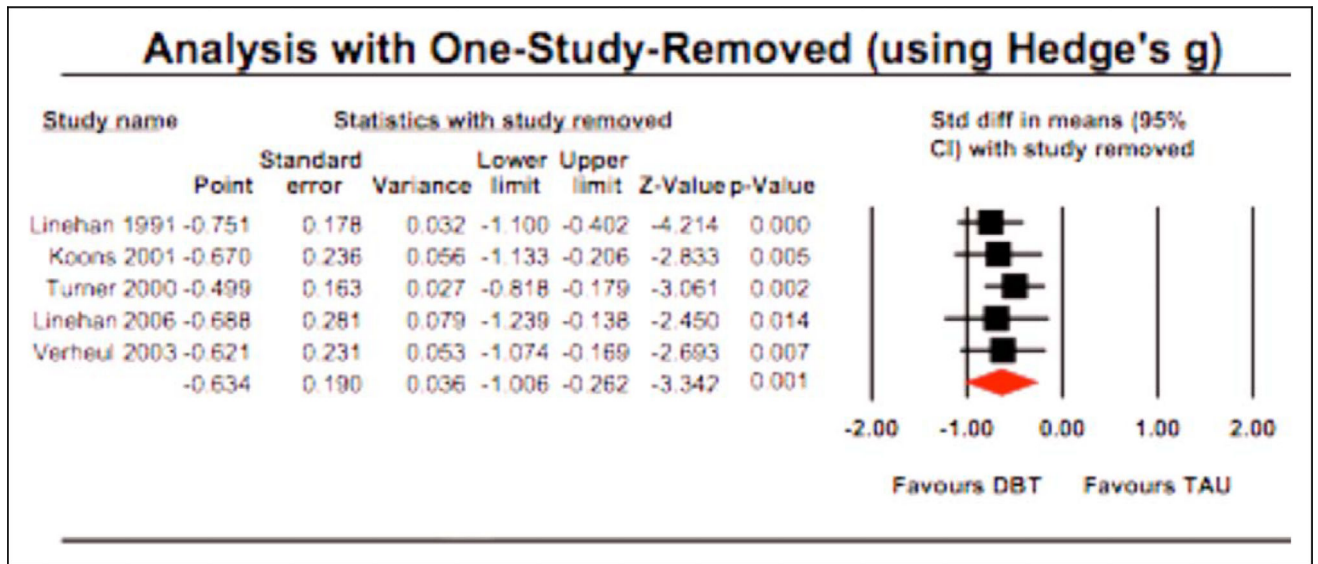


Figure 7.
Analysis with one study removed (using Hedge's *g*).

Table 1

Study Characteristics.

Study	CTAM Score	Sample Size	Gender	Mean Age (Minimum, Maximum)	Sample Description	Rx Use	BPD Assessment	Parasucide/Suicide Assessment	Depression Assessment	Therapist Experience	Treatment Comparison	Follow-Up (Month)	12-Month Attrition Rate	12-Month Suicide Attempt/Parasucide Behavior	12-Month Depression
Linehan, Armstrong, Suarez, Allmon, and Heard (1991)	65	44	Women	— (18, 45)	Clinically referred, 2 parasucide incidents in last 5 years	Not reported	DIB TAU	PHI—average events/person	n/a	Not reported	DBT	4	DBT: 45% TAU: 41%	Mean difference DBT: -2.95 TAU: -6.58 σ TAU Pre = 26.95 σ TAU Post = 25.02	Not assessed
Linehan et al. (2006)	85	101	Women	29.3 (18, 45)	2 parasucide incidents in last 5 years, one in past 8 weeks	Not reported	SCID	SASI Number of suicide attempts	HAM-D	25% DBT > 10 years, 56% CTBE > 10 years	DBT CTBE	12	DBT: 11.5% CTBE: 28.6%	% Attempted DBT: 23.1 CTBE: 46.0 Hazard ratio (95% CI) = 2.66 [2.4, 18.07] σ TAU Pre = 7.3 σ CBTE Post = 8.2	Mean difference DBT: -6.2 CTBE: -4.7 σ TAU Pre = 7.3 σ CBTE Post = 8.2
Koons et al. (2001)	83	28	Women	35.5 (21, 46)	Clinically referred, VA Ctr, 40% parasucide in previous 6 months	Not reported	SCID-I SCID-II	PHI	BDI HAM-D	Average 8.2 years' experience	DBT TAU	12	DBT: 23% TAU: 17%	Mean difference DBT: -4.7 TAU: +.3 σ TAU Pre = 1.3 σ TAU Post = 2.2	Mean difference DBT: -12.6 TAU: -8.3 σ TAU Pre = 9.7 σ TAU Post = 7.8
Verheul et al. (2003)	94	58	Women	34.9 (18, 70)	Clinically referred, no source restriction	DBT: 52% TAU: 61%	BPDSI	BPDSI, self-report	n/a	Not reported	DBT TAU	12	DBT: 37% TAU: 77%	% Attempted DBT: 7% TAU: 26% $\chi^2 = 3.24$	Not assessed σ TAU Post = 2.2
Turner (2000)	70	24	Both	22 (18, 27)	Clinically referred, outpatient suicide attempt treatment	DBT: 75% CCT: 92%	DIB SCID	Daily log of events	BDI HAM-D	Avg. 22 yrs Family systems, clientcentered, psychodynamic	DBT CCT	6 12	DBT: 25% CCT: 50%	Mean difference DBT: -5.67 CCT: -3.00 σ TAU Pre = .75 σ TAU Post = 2.18	Mean difference DBT: -13.25 CCT: -4.84 σ TAU Pre = 4.46 σ TAU Post = 2.18
Clarkin, Levy, Lenzeweger, and Kernberg (2007)	74	90	Both	30.9 (18, 50)	Recruited from NY tristate area, history of parasucide	TFP: 52%	IPE	OAS-m Suicidality	BDI	2 years experience w/BPD patients and treatment type	TFP	4	8.1% overall	HLM β_1 , effective size σ TAU Post = 3.90	HLM β_1 , effective size

Study	CTAM Score	Sample Size	Gender	Mean Age (Minimum, Maximum)	Sample Description	Rx Use	BPD Assessment	Parasuicide/Suicide Assessment	Depression Assessment	Therapist Experience	Treatment Comparison	Follow-Up (Month)	12-Month Attrition Rate	12-Month Suicide Attempt/Parasuicidal Behavior	12-Month Depression
						DBT: 70%					DBT	8			
						DST: 65%					DST	12		TFP: -.05, .33 DBT: -.09, .34 DST: -.03, .43	TFP: -.55, .50 DBT: -.59, .38 DST: -.69, .49

Note. DDI = Beck Depression Inventory; BPD = borderline personality disorder; BPDSI = BPD Severity Index; CCT = client-centered therapy; CTAM = Clinical Trials Assessment Measure; CTBE = Community Treatment By Experts; DBT = dialectical behavior therapy; DST = Dynamic Supportive Treatment; HAM-D = Hamilton Depression Rating scale; HLM = Hierarchical Linear Modeling; IPE = Initial Psychiatric Examination; OAS-m = Overt Aggression scale (modified); PHI = Parasuicide History Interview; Rx = Treatment; SASI = Suicide Attempt Self-Injury Interview; SCID = Structured Clinical Interview for DSM-IV Disorders TAU = treatment as usual; TFP = transference-focused psychotherapy.