

# Some Limitations on the External Validity of Psychotherapy Efficacy Studies and Suggestions for Future Research

GLENN D. SHEAN, Ph.D.

*Increased emphasis on identifying empirically supported treatments (ESTs) has enhanced the scientific basis for psychotherapy practice, but uncritical acceptance of ESTs as the basis for credentialing and policy decisions risks stifling innovation and creativity in the field. There are limitations inherent in efficacy studies of psychotherapy that can constrain external validity. This article discusses several limitations on the external validity of efficacy studies, as well as other issues related to evaluating psychotherapy outcome research. These limitations and concerns include: 1) the practice of maximizing homogeneity by selecting participants diagnosed with a single Axis I disorder; 2) the practice of requiring manualized therapies for efficacy research; 3) the assumption that lasting and meaningful changes occur and can be assessed within a relatively short time frame; 4) the assumption that valid assessments of outcome can be conducted in randomized control trials studies without concern for researcher allegiance; and 5) the view that evidence of effectiveness from non-RCT design studies can be ignored. Finally, alternative research approaches for studying psychotherapy that can potentially supplement knowledge gained from efficacy studies and foster continued innovation and creativity in the field are discussed.*

**KEYWORDS:** psychotherapy efficacy; empirically supported treatments; psychotherapy effectiveness; outcome

## INTRODUCTION

The purpose of psychotherapy research is twofold 1) to establish an empirical basis for psychological therapies and 2) to increase understanding of the processes that facilitate change. Emphasis on the importance of establishing an empirical basis for psychological therapies increased substantially in 1995 when the Task Force on Promotion and Dissemination

Professor of Psychology, College of William & Mary. **Mailing address:** College of William & Mary, PO Box 8795, Williamsburg, VA 23187-8795. e-mail: gdshea@gmail.com

AMERICAN JOURNAL OF PSYCHOTHERAPY, Vol. 66, No. 3, 2012

of Psychological Procedures, sponsored by Division 12 (Clinical Psychology) of the American Psychological Association (APA), published a list of criteria for identifying empirically validated treatments (later relabeled empirically supported treatments or ESTs). The criteria set for a “well established” EST were as follows:

a treatment should be manualized and demonstrated to be more effective than other treatments or placebo, or equivalent to an established EST in at least two randomized control group studies or in a number of single case design experiments conducted by different researchers (Chambless & Holton, 1998, p. 9).

These criteria were patterned after criteria established by the U.S. Food and Drug Administration for drug efficacy studies to include the additional requirement that ESTs are defined as “those clearly specified psychological treatments that have been shown to be efficacious in controlled research studies with a clearly delineated population” (Chambless & Holton, 1998, p. 9). Empirically supported treatments have been widely adopted as the standard of care required by many managed care organizations (MCOs) and state Medicaid programs (Carpinello, Rosenberg, Stone, Schwager, & Felton, 2002; Seligman & Levant, 1998). Randomized control trial studies have also become the “gold standard” for psychotherapy research (Sternberg, 2006), as indicated in the funding guidelines recently issued the National Institute of Mental Health (2008).

An impressive body of evidence has been accumulated since 1995 supporting the effectiveness of ESTs for treating a number of psychological dysfunctions (Barlow, Gorman, Shear, & Woods, 2000; Stuart, Treat, & Wade, 2000). This evidence is the basis for growing recognition of the importance of ESTs as indicators of competence among the public, policymakers, and training programs, and for reimbursement among MCOs (Laroche & Christopher, 2009). The enthusiasm of EST advocates is demonstrated by references to psychotherapies not included in evidence supported lists as, “less essential and outdated” (Calhoun, Moras, Pilkonis, & Rehm, 1998, p. 151) and articles in the popular press with titles such “Is Your Therapist a Little Behind the Times?” (Baker, McFall and Shorham, 2009) or “Ignoring the Evidence: Why Do Psychologists Reject Science” (Begley, 2009).

Nevertheless, there is growing concern about the potential legal ramifications and practice limitations associated with adoption of lists of ESTs as the basis for judging clinical competence and reimbursement (Rupert & Baird, 2004). There are also concerns about the applicability of ESTs to

the diverse range of patients seen in actual practice settings (LaRoche & Christopher, 2009; Sue, et al, 2006; Wampold, 2007; Westen, Novotny, & Thompson-Brenner, 2004).

The goal of grounding psychotherapy practice in research is wise from both scientific and public policy perspectives, but there is risk associated with overzealous generalizations about the external validity of ESTs. Restrictions on training, and practice based solely on psychotherapy efficacy studies, may result in adverse legal and practice effects and stifle innovation and creativity in the field. There are two methods that can be used to evaluate psychotherapy outcomes. *Efficacy* studies use random controlled trial (RCT) designs to compare the application of manualized therapies delivered over a fixed number of sessions to treat individuals diagnosed with specific uncomplicated disorders. *Effectiveness* studies are conducted in naturalistic settings without the use of manuals, or strict session limits, most often with individuals who often have multiple problems. As Seligman (1996) has indicated, there is a substantial inferential distance between efficacy studies of manualized therapies of fixed duration, with selected, patient samples and effectiveness studies of therapy as practiced in naturalistic settings.

Effectiveness studies lack control over variables, such as number of sessions, therapeutic method, and sample characteristics, but have fewer problems associated with inferential distance because they more closely resemble and test what the research is designed to generalize to—actual practice. Additional concerns about the external validity of efficacy studies have been expressed by several prominent psychotherapy researchers (e.g., Beutler, 1998; Sue, et al., 2006; Wachtel, 2010; Wampold, 2007; Westen, et al., 2004). These concerns include:

- 1) the potential limitations of RCT studies that attempt to maximize homogeneity by selecting participants diagnosed with a single, uncomplicated Axis I disorder;
- 2) the practice of limiting studies to manualized therapies;
- 3) the practice of assessing changes solely in terms of symptom reduction, often assessed over a relatively short time period;
- 4) the failure to incorporate appropriate safeguards in RCT studies to minimize the effects of “therapist allegiance” and,
- 5) the dismissal of the relevance of evidence from effectiveness studies.

This paper will discuss these issues and suggest additional approaches to psychotherapy research that can contribute important knowledge about psychotherapy process and outcome.

### 1. SAMPLE RESTRICTIONS

The rationale for limiting participants to a single diagnostic category is that specification of the sample population allows for ease of replication. While replication of research evidence is important, the goal of easy replication must be balanced against evidence that there is substantial heterogeneity within diagnostic groups (Howard, et al., 1996). The practice of limiting efficacy studies to populations with specific problems does not mirror the realities of actual clinical practice. Evidence indicates that about one-third to one-half of people seeking mental health treatment do not meet criteria for any one diagnostic category (Howard, et al., 1996), and when specific symptoms are the focus of treatment about one-half of the patients add new target complaints or change their complaints during the course of treatment (Kazdin, 2008).

Restricting efficacy studies to individuals with a single DSM diagnosis is likely to result in evidence that cannot be generalized beyond specific study sample characteristics. Between 50% to 90% of Axis I diagnoses are comorbid with other Axis I or Axis II disorders (Kessler, Stang, Wittchen, Stein, & Walters, 1999; Zimmerman, McDermot, & Mattia, 2000; Thompson-Brenner & Westen, 2004a, 2004b; Westen & Hernden-Fischer, 2001). Axis I symptoms often are related to broad dispositional traits (e.g., negative affect, vulnerability, introversion) that may influence long-term outcomes and predispose individuals to developing symptoms under different circumstances (Blatt & Zuroff, 1992; Hammen, Ellicott, Gitlin, & Jamison, 1989; Krueger, 2002; Kwon & Whisman, 1998; Mineka, Watson, & Clark, 1992; Zinburg & Barlow, 1996). One cannot assume that individuals with a single diagnosis are matched for additional characteristics including: the presence and degree of family and marital difficulties, the extent of life and job frustrations, the presence of pretreatment trauma, chronic illnesses, substance abuse problems, ethnicity, culture, or issues related to transitions or life stage adjustment issues. If, for example, depressive symptoms remit in response to a targeted therapeutic intervention but dispositional and contextual problems are ignored, outcome may be quite different from what symptom-focused follow-up evaluation indicates. An additional issue regarding efficacy studies is the questionable validity of many DSM-IV diagnoses. Studies of psychotherapy outcome should include assessments of participants beyond DSM-based diagnostic symptoms to include broad-based assessments continued over several years. Assessment instruments are available that can provide

broad based assessments of function and outcome but, these instruments are not widely used (Westen & Shedler, 1999a, 1999b, 2007). Inclusion of these measures in addition to DSM symptoms would allow for more valid assessments of psychotherapy outcome.

In summary, the practice of restricting EST studies to a single diagnostic category limits the external validity of many efficacy studies because important comorbid problems and contributing dispositional and contextual factors may be ignored. People develop difficulties in different contexts and bring many attributes and assets to therapy including: their personal histories, ability to identify and describe their emotions and concerns, level of self-reflectiveness, affect tolerance, readiness and motivation for change, quality of social supports, coping abilities, and capacity for intimacy (Asay & Lambert, 1999; Prochaska & Norcross, 2002). These attributes and influences can have a significant impact on psychotherapy outcome. Therapy is not only about symptom reduction; often it is designed by mutual agreement to bring about broad changes, such as helping patients to face and to deal with emotional and interpersonal issues, to make positive life changes or to encourage the risk of being more honest with themselves and others, to develop more fulfilling relationships, to actualize talents and abilities, to tolerate and be more aware of affects, to understand self and others in more nuanced ways, and to be better prepared to deal with life's challenges with greater flexibility. The diversity of these goals can make it difficult to measure systematically outcome across individuals. Efficacy studies have ignored these complexities.

## 2. BRIEF SYMPTOM FOCUSED INTERVENTIONS AND OUTCOME

Psychotherapy efficacy studies target outcomes of treatment achieved between sessions 6 and 16 (Morrison, Bradley, & Westen, 2003; Westen, Novotny, & Thompson-Brenner, 2004). The idea that many standing problems can be permanently resolved in 6 to 16 sessions is not consistent with evidence of a significant psychotherapy dose-response relationship (Howard, Kopta, Krause, & Orlinsky, 1986; Kopta, Howard, Lowry, & Deutler, 1994; Seligman, 1995). The median treatment length for depression in outpatient settings is 75 sessions, a marked contrast to the average 6-16 sessions reported by efficacy studies. Field studies of an EST, such as CBT therapy for depressed patients, report an average of 69 sessions, a marked contrast to the average of efficacy studies (Morrison, Bradley, & Westen, 2003). Evidence indicates there is substantial variability in response to treatment and that we do not understand the contributors to this variability (Baldwin, Berkelson, Atkins, Olsen and Nielsen, 2009). For

example, change in therapy has been demonstrated to be a function of factors such as therapist competence and patient dispositional traits (Howard, Lueger, Maling, & Martinovitch, 1993; Kopta et al. (1994). Mental disorders are also characterized by intervals of remission and relapse, rendering the validity of relatively short-term follow-up evaluations potentially problematic. The well-designed NIMH Collaborative Research Program, for example, included a large sample size, an active medical management placebo control group, and employed investigators with allegiance to each approach to therapy under investigation (Elkin et al., 1989). Short-term assessments of depressed patients in treatment indicated promising responses to 16 weeks of both CBT and Interpersonal Therapy (IPT). Follow-up at 18 months indicated that the majority those treated either relapsed or sought further treatment (Shea, Elkin, et al., 1992).

### **3. THERAPY FROM A MANUAL**

The purpose of using therapy manuals in efficacy research is to reduce variability by standardizing interventions across participants and sites. Task force guidelines call for the use of “treatment manuals or their equivalent in the form of a clear description of the treatment” (Chambless & Ollendick, 2011) and assert that studies that do not include therapy manuals are “acceptable as evidence” only in “specific and rare exceptions” (Chambless et al. 1996, p. 6). The requirement of therapy manuals implies that psychotherapy can be standardized and administered in a manner parallel to drugs and other medical treatments but, the similarities between mental disorders and medical disorders are limited. Psychotherapy is rarely a “one-size-treats-all” process. Therapies for patients who are not preselected do not lend themselves to manualization, nor can providers be matched in study designs. Meta-analyses of effectiveness studies indicate that differences between approaches to psychotherapy account for less than 10% of the total variance among outcomes, suggesting that the “active ingredients” of therapy are not necessarily those associated with a particular treatment model (Lambert & Ogles, 2004; Luborsky, et al., 2002; Wampold, et al., 1997; 2002). The overlap between approaches in practice is such that raters of session transcripts often have difficulty determining which manualized therapy was being provided (Ablon & Jones, 2002). The lack of clarity about the processes that contribute to change with manualized therapy was described in a review of cognitive therapy outcome research as follows, “Perhaps we can state more confidently now than before that whatever mechanisms of changes with cogni-



tive therapy, it does not seem to be the cognitions as originally proposed” (Kazdin, 2007, p. 8).

A study of archival treatment records of therapist adherence to either a psychodynamic or cognitive prototype, regardless of therapist’s beliefs about what model they were following, indicated that several prototypic dynamic practices best predicted successful outcome (e.g., encouraging open-ended dialogue; identifying recurring themes in the patient’s life; linking feelings and perceptions to past experiences; drawing attention to feeling states experienced as unacceptable; identifying defensive responses; interpretations; and discussing potential connections between the therapy relationship and other relationships; (Ablon & Jones, 1998). Therapists’ adherence to CBT practices (e.g., focus on cognitive themes such as thoughts and belief systems; discussion of specific homework tasks and explanations of the rationale behind treatment and techniques; therapist introducing topics; therapist functioning in a didactic manner; discussion of specific treatment goals; and focus on the patient’s current life situation) did not relate to successful outcome regardless of therapist stated orientation (Ablon & Jones, 1998). A study of specific process predictors of positive outcome using the Psychotherapy Process Q-Set, indicated that adherence to cognitive-behavioral process was most characteristic, adherence to interpersonal and psychodynamic process; however, was most predictive of positive outcome (Ablon, Levy, & Katzenstein, 2006).

Surveys of practitioners indicate that most describe themselves as theoretically eclectic. Cognitive-behavioral clinicians report using psychodynamic strategies, exploring relationships and unconscious processes with patients evidencing higher levels of emotional dysregulation. Psychodynamic clinicians report that they use CBT treatment strategies with emotionally constricted patients (Thompson-Brenner & Westen, 2004a, 2004b). Psychotherapy outcome studies in naturalistic settings attempt to assess what is more often than not an individualized and fluid process that is not compatible with the use of specific therapy manuals as included in efficacy studies.

#### 4. THE “WILD CARD” EFFECT IN PSYCHOTHERAPY RESEARCH

A limitation of nearly all efficacy studies that form the basis of EST lists is the omission of double-blind outcome ratings, an important safeguard that is included in RCT designs. Double-blind ratings in psychotherapy efficacy studies are admittedly difficult, expensive, and often impractical to include, however, the lack of double-blind ratings leaves results susceptible to the “wild card” effect of therapist alle-

giance. This is not a minor issue since therapist allegiance has been shown to account for about 69% of the variance in psychotherapy outcome studies (Luborsky et al, 1999). Double-blind ratings may be difficult to include in psychotherapy studies but several practical suggestions for controlling the effects of therapist allegiance were provided by Luborsky and colleagues; these suggestions have been ignored. Luborsky et al. (1999) suggest that at minimum comparative treatment studies be conducted using raters of varied theoretical persuasions and researchers with minimal allegiance to the approaches studied. In addition, therapists for each treatment mode should be selected and supervised by those who represent the same treatment mode. Therapists should be assigned to each mode of treatment on the basis of ratings of their effectiveness. Outcome criteria should be developed based on the input of the therapists of all persuasions under study, and long-term functional follow-up evaluations should be conducted using consistent outcome criteria. Luborsky et al., (1999) recommend that if all else fails efficacy studies include researcher/therapist allegiance as a variable in all analyses.

#### **5. DISMISSAL OF EVIDENCE FROM TRIALS THAT ARE NOT RANDOMIZED AND CONTROLLED**

A study of patients with panic disorder by self-identified psychodynamic clinicians indicated rates of remission and change scores commensurate with those of empirically supported therapies for panic disorder. Treatment gains were maintained at 6-month follow-up (Ablon, Levy, & Katzenstein, 2006). Meta-analytic analysis of comparative outcome studies indicated that there no large differences exist in efficacy among the major psychotherapies (e.g., cognitive-behavioral, interpersonal, behavioral activation, psychodynamic, problem solving, or social skills training) for mild to moderate depression (Benish, Imel, & Wampold, 2008; Cuijpers, vanStraten, Andersson, and van Oppen, 2008; Luborsky et al., 2002; Westen and Morrison, 2001). There is substantial evidence available to support of the effectiveness of dynamic therapies (e.g., Clarkin et al., 2007; Leichenring, 2001; 2005; Leichenring & Leibing, 2003; Leichenring & Rabung, 2008; Leichenring, Rabung, & Leibing, 2004; Levy & Ablon, 2008; Lewis, Dennerstein, & Gibbs, 2008; Seligman, 1996; Shedler, 2010; Wampold, 2008). Evidence from effectiveness studies of the effects of psychodynamic therapies is ignored.



### THE PRESENT

Efficacy studies provide reasonable assurance of internal validity, (i.e., the degree to which results reported can be attributed to the therapeutic approach studied) but do not necessarily allow for conclusions about effectiveness. Psychotherapy outcome research has been based on the assumptions of a linear causal model where patient symptoms → manualized therapeutic approach → outcome (symptom reduction). But psychotherapy as practiced in the field often involves more than the application of specific strategies to fix specific problems; it is a complex series of interactions that take place over time between a therapist who has unique characteristics and a patient who has unique characteristics; it is a process that takes place in a particular context and that results in a reciprocal, unfolding causal process with unpredictable emergent properties (Wampold, Hollon, & Hill, 2011). Questions about outcome in terms of symptom reduction are important, but it is also relevant to evaluate the consequences of therapy for the patient's life situation beyond symptom reduction, for example, whether long-term patterns have changed and the quality and character of the patient's life improved so that life has become more satisfying, more meaningful and fulfilling, and more socially productive (Orlinsky, 2009).

Psychotherapy research is also not only about questions of outcome, but there also are many important questions that must be studied about how therapy works and how treatments can be improved. Process research examines what occurs during therapy. Process variables, such as patient characteristics, context, therapist responses and the therapeutic alliance, cannot be experimentally manipulated; nevertheless, they can influence outcome. Research indicates, for example, that the quality of the therapeutic relationship accounts for about 30% of the variance in outcome, therapist techniques about 15%; expectancy (hope), therapist credibility about 15%, and environmental and patient characteristics (e.g., readiness for change, openness, engagement, active participation, ability to verbalize feelings) about 40% (Asay and Lambert, 1999). Characteristics of the therapist have a significant effect on outcome (Beutler, et al., 2004; Horvath & Bedi, 2002; Wampold, 2001), and within treatments, therapists vary considerably in their outcomes (Huppert, et al, 2001; Kim, Wampold, & Bolt, 2006; Wampold & Brown, 2005). We do not know enough about what highly effective therapists do that has a significant positive impact on outcome but, case studies suggest that process variables such as therapist immediacy (e.g., inquiring about reactions to the therapy relationship,

pointing out parallels between the therapy relationship and other relationships, processing boundary crossings, disclosure of feelings) have a significant effect on outcome (Hill, et al., 2008; Kasper, Hill, & Kivlighan, 2008).

Therapist effects have been ignored in most EST research, an omission that limits generalizability (Serlin, Wampold, & Levin, 2003), inflates treatment effects (Wampold & Serlin, 2000), and obscures understanding how process variables are related to outcome (Baldwin, Wampold & Imel, 2007). The APA recently created a new Joint Task Force to develop guidelines for research on *evidence-based psychotherapy relationships*, the goals of which are to foster research efforts to identify elements of effective therapy relationships, and to identify effective methods of adapting therapy to the characteristics and needs of the patient (other than diagnosis; Norcross, 2002). These guidelines may help broaden the focus of psychotherapy research to include questions about variables such as the influence of the therapist-patient relationship, therapy process variables, and the impact of patient characteristics beyond diagnostic symptoms on outcome. We know a fair amount about the processes that contribute to therapeutic effectiveness including: the importance of developing a positive therapeutic alliance, fostering a sense of hope and self-efficacy, and encouraging relevant emotional expression but, there is much more to be learned about the particular skills, processes and practices associated with different conceptual models, how they may contribute to positive change, and how they interact with individual differences (Beutler et al., 2011). We need to learn more about the effects of other moderator variables on psychotherapy effectiveness, including variables such as: levels of distress and impairment, co-occurring problems, dispositional traits, self-reflectiveness, and openness to experience, as well as access to social supports, coping skills, and how these moderators interact with therapeutic approaches (Goldfried and Eubanks-Carter, 2004). We also do not adequately understand the proper conditions for effective application of therapeutic interventions such as the two-chair technique, transference interpretation, enactment, and the use of metaphor and paradox. Shedler (2010) described the key features that distinguish dynamic psychotherapy: as: a focus on affect and expression of emotion; exploration of attempts to avoid distressing thoughts and feelings; identification of recurring themes and patterns; discussion of past experiences with a developmental focus; focus on interpersonal relations; focus on the therapy relationship; and explo-

ration of fantasy life. Research is needed to identify when, how, in what context and with whom these strategies are most likely to be effective, and whether or not these strategies can be effective across theoretical approaches. Clinical studies of comparable cases can be of value as illustrations of therapeutic principles (Goldfried and Wolfe, 1998). We can also learn about processes that bring about change by systematically studying the recorded therapy sessions of acknowledged expert clinicians. There is a need for in-depth within-model comparisons of therapies and comparisons of components of therapy models may help us better understand change mechanisms that operate within a particular model, as well as how to best maximize efficacy in the application of these mechanisms (Jacobson, 1999; Jacobson & Addis, 1993).

*The APA Presidential Task Force on Evidence Based Practice* (2005) defined evidence based practice as “the integration of the best available research with clinical expertise in the context of patient characteristics, culture and preferences” (APA, 2005, p. 273). This broader definition was apparently developed as a response to growing concerns that ESTs were being narrowly interpreted and misused as justification for questionable restrictions on access to care and treatments of choice (Norcross, Koocher, & Garofalo, 2006). The redefinition of ESTs was intended to incorporate the strengths of EST studies while also recognizing the value of diverse methodological approaches that “require an appreciation of the value of multiple sources of scientific evidence” (APA, 2006, p. 280). Much more research is needed to help us better understand the interaction between process and outcome in terms of the impact of the therapist, how and when certain interventions are most likely to be effective, and how the process of therapy changes over time.

Wampold et al. (2011) suggest that multilevel statistical models should be applied in psychotherapy research to estimate the proportion of variance in outcomes because of different levels of variables (e.g., patient characteristics, therapist characteristics, and treatments) and to assess the impact of these variables on outcome (Hox, 2002). Multilevel statistical models can be used to analyze longitudinal process and outcome data that are nested within therapists so that the effects of temporal and nested components of therapy can be identified and evaluated (Klein, et al., 2003). Once process variables, such as the characteristics and actions of effective therapists, are better understood within and across therapeutic approaches and evidence are accumulated about how these characteristics interact with patient variables across time, we will be better able to train effective therapists, and to assess and make policy decisions about therapeutic

effectiveness. Until that time it seems most appropriate to remain somewhat circumspect about the practice and policy implications of psychotherapy efficacy studies.

## REFERENCES

- Ablon, J.S., & Jones, E.E. (1998). How expert clinicians' prototypes of an ideal treatment correlate with outcome in psychodynamic and cognitive-behavioral therapy. *Psychotherapy Research*, 8, 71–83.
- Ablon, J.S., & Jones, E.E. (2002). Validity of controlled clinical trials of psychotherapy: Findings from the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry*, 159, 775–783.
- Ablon, J.S., Levy, R.A., & Katzenstein, T. (2006). Beyond brand names of psychotherapy: Identifying empirically supported change processes. *Psychotherapy: Theory, Research, Practice, Training*, 43, 216–231.
- American Psychological Association Task Force on Psychological Intervention Guidelines. (1995). *Template for developing guidelines: Interventions for mental disorders and psychosocial aspects of physical disorders*. Washington, DC: American Psychological Association.
- American Psychological Association (2005). Evidence-based practice in psychology: APA presidential task force on evidence-based practice. *American Psychologist*, 61, 271–285.
- Asay, T.P., & Lambert, M.J. (1999). The empirical case for the common factors in therapy: Quantitative findings. In M.A. Hubble, B.L. Duncan, & S.D. Miller (Eds.). *The heart and soul of change: What works in therapy* (pp. 23–55). Washington, DC: American Psychological Association.
- Baker, T., McFall, R., & Shoram, V. (2009, November 15). *Is your therapist a little behind the times?* Washington D.C.: Washington Post.
- Baldwin, S.A., Berkeljon, P., Atkins, J., Olsen, A., & Nielsen, T. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology*, 77, 203–211.
- Baldwin, S.A., Wampold, B.E., & Imel, Z.E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*, 75, 842–852.
- Barlow, D.H., Gorman, J.M., Shear, M.K., & Woods, S.W. (2000). Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: A randomized controlled trial. *Journal of the American Medical Association*, 283, 2529–2536.
- Begley, S. (2009). Ignoring the evidence: Why do psychologists reject science? *Newsweek*, October 12, 2009.
- Benish, S.G., Imel, Z.E., & Wampold, B.E. (2008). The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons. *Clinical Psychology Review*, 28, 746–758.
- Beutler, L.E. (1998). Identifying empirically supported treatments: what if we didn't? *Journal of Consulting and Clinical Psychology*, 66, 113–120.
- Beutler, L.E., Malik, M., Alimohamed, S., Harwood, T.M., Talbi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M.J. Lambert (Ed.) *Handbook of psychotherapy and behavior change* (5<sup>th</sup> ed. Pp. 194–226). New York, NY: Wiley.
- Beutler, L.E., Harwood, T.M., Michelson, A. & Song, X. (2011). Reistance/Reactance level. *Journal of Clinical Psychology*, 67, 133–142.
- Blatt, S., & Zuroff, D. (1992). Interpersonal relatedness and self-definition: Two prototypes for depression. *Clinical Psychology Review*, 12, 527–562.
- Calhoun, K.S., Moras, K., Pilkonis, P.A., & Rehm, L. (1998). Empirically supported treatments: Implications for training. *Journal of Consulting and Clinical Psychology*, 66, 151–162.
- Carpinello, S.E., Rosenberg, L., Stone, J., Schwager, M., & Felton, C.J. (2002). New York State's campaign to implement evidence based practice practices for people with serious mental disorders. *Psychiatric Services*, 53, 153–155.

## Psychotherapy Efficacy Studies Limitations and Suggestions

---

- Chambless, D.L., Sanderson, W.C., Shohan, V., Bennet, B.E., Johnson, S., & Pope, K.S. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, 49, 5–18.
- Chambless, D.L., & Hollon, S.D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7–18.
- Chambless, D.L., & Ollendick, T.H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, 52, 685–716.
- Clarkin, J.F., Levy, K.N., Lenzenweger, M.F., & Kernberg, O.F. (2007). Evaluating three treatments for borderline personality disorder: A multiwave study. *American Journal of Psychiatry*, 164, 922–928.
- Cuijpers, J., van Straten, O., Andersson, H., & van Oppen, C. (2008). Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology*, 76, 909–922.
- Davison, G.C. (2000). Stepped care: Doing more with less? *Journal of Consulting and Clinical Psychology*, 68, 580–923.
- Elkin, I., Shea, T., Watkins, J., Imber, S., Sotsky, S., & Collins, J. (1989). National Institute of Mental Health treatment of Depression Collaborative Research Program. *Archives of General Psychiatry*, 46, 971–982.
- Hammen, C., Ellicott, A., Gitlin, M., & Jamison, K.R. (1989). Sociotropy/autonomy and vulnerability to specific life events in patients with unipolar depression and bipolar disorders. *Journal of Abnormal Psychology*, 98, 154–160.
- Hill, C.E., Sim, W., Spangler, P., Stahl, J., Sullivan, C., & Teyber, E. (2008). Therapist immediacy in brief psychotherapy therapy: Case Study II. *Psychotherapy: Theory, Research, Practice, Training*, 45, 298–316.
- Horvath, A.O., & Bedi, R.P. (2002). The alliance. In J.C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37–70). New York, NY: Oxford University Press.
- Howard, K.I., Kopta, S.M., Krause, M.S., & Orlinsky, D.E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist*, 41, 159–164.
- Howard, K.I., Lueger, R., Maling, M., Martinovich, Z. (1993). A phase model of psychotherapy: Causal mediation of outcome. *Journal of Consulting and Clinical Psychology*, 54, 106–110.
- Howard, K.L., Comille, T.A., Lyons, J.S., Vessey, J.T., Lueger, R.J., & Saunders, S.M. (1996). Patterns of mental health service utilization. *Archives of General Psychiatry*, 53, 696–703.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahway, NJ: Erlbaum.
- Huppert, J.D., Bufka, L.F., Barlow, D.H., Gorman, J.M., Shear, M.K., & Woods, S.W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcomes in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology*, 69, 747–755.
- Jacobson, N.S. (1999). The role of the Allegiance Effect in psychotherapy research: Controlling and accounting for it. *Clinical Psychology: Science and Practice*, 6, 116–119.
- Jacobson, N.S. & Addis, M.E. (1993). Research on couples and couple therapy: What do we know? Where are we going? *Journal of Consulting and Clinical Psychology*, 61, 85–93.
- Kasper, L., Hill, C.E., & Kivlighan, D.M. Jr. (2008). Therapist immediacy in brief psychotherapy therapy: Case Study I. *Psychotherapy: Theory, Research, Practice, Training*, 45, 281–297.
- Kazdin, A.E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, 3, 1–27.
- Kazdin, A.E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63, 146–159.
- Kessler, R.C., Stang, P., Wittchen, H.U., Stein, M., & Walters, E.E. (1999). Lifetime comorbidities between social phobia and mood disorders in the U.S. Comorbidity Survey. *Psychological Medicine*, 29, 555–567.
- Kim, D.M., Wampold, B.E., & Bolt, D.M. (2006). Therapist effects in psychotherapy: A random effects modeling of the NIMH TDCRP data. *Psychotherapy Research*, 16, 161–172.
- Klein, D.N., Schwartz, J.E., Santiago, N.J., Vivian, D., Vocisano, C., Castonguay, L.G., & Keller, M.B. (2003). Therapeutic alliance in depression treatment: Controlling for prior change and patient characteristics. *Journal of Consulting and Clinical Psychology*, 71, 997–1006.
- Kopta, S., Howard, K., Lowry, J., & Beutler, L. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of Consulting and Clinical Psychology*, 62, 1009–1016.

- Kreuger, R.F., (2002). The structure of common mental disorders. *Archives of General Psychiatry*, 59, 570–571.
- Kwon, P., & Whisman, M.A. (1998). Sociotropy and autonomy as vulnerabilities to specific life events: Issues in life event categorization. *Cognitive Therapy and Research*, 22, 353–362.
- Lambert, M.J., & Ogles, B.M. (2004). The efficacy and effectiveness of psychotherapy. In M. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (5<sup>th</sup> ed., pp. 139–193) New York, NY: Wiley.
- LaRoche, M.J. & Christopher, M.S. (2009). Changing paradigms from empirically supported treatment to evidence-based practice: A clinical perspective. *Professional Psychology: Research and Practice*, 40, 396–402.
- Leichsenring, F. (2001). Comparative effects of short-term psychodynamic psychotherapy and Cognitive-behavioral therapy in depression: A meta-analytic approach. *Clinical Psychology Review*, 21, 401–419.
- Leichsenring, F. (2005). Are psychodynamic and psychoanalytic therapies effective? A review of empirical data. *International Journal of Psychoanalysis*, 86, 841–868.
- Leichsenring, F. & Leibing, E. (2003). The effectiveness of psychodynamic psychotherapy and cognitive behavior therapy in the treatment of personality disorders: A meta-analysis. *American Journal of Psychiatry*, 160, 1223–1232.
- Leichsenring, F., Rabung, S., & Leibing, E. (2004). The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders: A meta-analysis. *Archives of General Psychiatry*, 61, 1208–1216.
- Leichsenring, F., & Rabung, S. (2008). Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis. *Journal of the American Medical Association*, 300, 1551–1565.
- Levy, R., & Ablon, S.J. (Eds.). (2008). *Handbook of evidence-based psychodynamic psychotherapy: Bridging the gap between science and practice*. New York: Humana Press.
- Lewis, A.J., Dennerstein, M., & Gibbs, P.M. (2008). Short-term psychodynamic psychotherapy: review of recent process and outcome studies. *Australian and New Zealand Journal of Psychiatry*, 42, 445–455.
- Luborsky, L., Diguer, L., Seligman, D., Rosenthal, R., Krause, E.D., Johnson, S., et al. (1999). The researcher's own therapy allegiances: A “wild card” in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice*, 6, 95–106.
- Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T.P., & Berman, J.S. (2002). The dodo bird verdict is alive and well-mostly. *Clinical Psychology Science and Practice*, 9, 2–12.
- Mineka, S., Watson, D., & Clark, L.A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*, 49, 377–412.
- Morrison, C., Bradley, R., & Westen, D. (2003). The external validity of efficacy trials for depression and anxiety: A naturalistic study. *Psychology and Psychotherapy: Theory, Research and Practice*, 76, 109–132.
- National Institutes of Mental Health (2008). *NIMH strategic plan*. Rockville, MD: U.S. Department of Health and Human Services, NIH Publication No. 08-6368.
- Norcross, J.C. (2001). Purposes, processes and products of the task force on empirically supported therapy relationships. *Psychotherapy: Theory, Research, Practice, Training*, 38, 345–356.
- Norcross, J. (Ed.). (2002). *Psychotherapy relationships that work: Therapists contributions and responsiveness to patients*. New York: Oxford University Press.
- Norcross, J.C., Koocher, G.P., & Garofalo, A. (2006). Discredited psychological treatments and tests: A Delphi poll. *Professional Psychology: Research and Practice*, 37, 515–522.
- Norcross, J.C., & Wampold, B.E. (2011). What works for whom: Tailoring psychotherapy to the person. *Journal of Clinical Psychology*, 67, 127–132.
- Orlinsky, D.E. (2009). The “generic model of psychotherapy” after 25 years: Evolution of a research-based metatheory. *Journal of Psychotherapy Integration*, 19, 319–339.
- Prochaska, J.O., & Norcross, J.C. (2002). Stages of change. In J.C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37–70). New York, NY: Oxford University Press.
- Rupert, P., & Baird, R. (2004). Managed care and the independent practice of psychology. *Professional Psychology: Research and Practice*, 35, 185–193.
- Seligman, M.E. P. (1995). The effectiveness of psychotherapy. *American Psychologist*, 50, 965–974.
- Seligman, M.E.P. (1996) Science as an ally of practice. *American Psychologist*, 51, 1072–1079.



## Psychotherapy Efficacy Studies Limitations and Suggestions

---

- Seligman, M.E.P., & Levant, R. (1998). Managed care policies rely on inadequate science. *Professional Psychology: Research and Practice*, 29, 211–212.
- Serlin, R.C., Wampold, B.E., & Levin, J.R. (2003). Should providers of treatment be regarded as a random factor? If it ain't broke, don't "fix" it: A comment on Siemer and Joorman (2003). *Psychological Methods*, 8, 524–534.
- Shea, M., Elkin, I., Imber, S., Sotsky, S., Watkins, J., Collins, J., et al. (1992). Course of depressive symptoms over follow-up: Findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry*, 49, 782–787.
- Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist*, 65, 98–109.
- Sternberg, R.J. (2006). Evidence-based practice: Gold standard, gold plated, or fool's gold? In C.D. Goodheart, A.E. Kazdin, & R.J. Sternberg (Eds.), *Evidence-based psychotherapy: Where practice and research meet* (pp. 261–271). Washington, DC: American Psychological Association.
- 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.
- Sue, S., Zane, N., Levant, R.R., Silverstein, L.B., Brown, L.S., & Olkin, R., et al. (2006). How well do both evidence-based practices and treatment as usual satisfactorily address the various dimensions of diversity? In J.C., Norcross, L.E. Beutler, & R.F. Levant (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 329–337). Washington, D.C.: American Psychological Association.
- Thompson-Brenner, H., & Westen, D. (2004a). *Accumulating evidence for personality subtypes in eating disorders: Differences in comorbidity, adaptive functioning, treatment response, and treatment interventions in a naturalistic sample*. Unpublished manuscript, Boston University, Boston.
- Thompson-Brenner, H., & Westen, D. (2004b). *A naturalistic study of psychotherapy for bulimia nervosa: Comorbidity, outcome, and therapeutic interventions in the community*. Unpublished manuscript. Boston University, Boston.
- Wachtel, P.L. (2010). Beyond "ESTs": Problematic assumptions in the pursuit of evidence-based practice. *Psychoanalytic Psychology*, 27, 251–272.
- Wampold, B.E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Wampold, B.E. (2007). Psychotherapy: The humanist (and effective) treatment. *American Psychologist*, 62, 857–873.
- Wampold, B.E., & Serlin, R.C. (2000). The consequences of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods*, 5, 425–433.
- Wampold, B.E., Minami, T., Baskin, T.W., & Callen, T., & Tierney, S. (2002). A meta- (re)analysis of the effects of cognitive therapy versus "other therapies" for depression. *Journal of Affective Disorders*, 68, 159–165.
- Wampold, B.E., & Brown, G.S. (2005). Estimating therapist variability: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, 73, 914–923.
- Wampold, B.E., Hollon, S.D., & Hill, C.E. (2011). Unresolved questions and future directions in psychotherapy research. Chapter 11 in C.J. Gelso, (Ed.). *The real relationship in psychotherapy: The hidden foundation of change* (pp. 333–355). Washington, DC.: American Psychological Association.
- Watson, D., & Clark, L.A. (1992). Affects separable and inseparable: On the hierarchical arrangement of the negative affects. *Journal of Personality and Social Psychology*, 62, 489–505.
- Westen, D., & Shedler, J. (1999a). Revising and assessing Axis II, Part 1: Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry*, 156, 258–272.
- Westen, D., & Shedler, J. (1999b). Revising and assessing Axis II, Part2: Toward an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry*, 156, 273–285.
- Westen, D., & Harnden-Fischer, J. (2001). Personality profiles in eating disorders: Rethinking the distinction between Axis I and Axis II. *American Journal of Psychiatry*, 165, 547–562.
- Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 69, 875–899.

- Westen, D., Novotny, C.M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, 130, 631–663.
- Westen, D., & Shedler, J. (2007). Personality diagnosis with the Shedler-Westen Assessment Procedure (SWAP): Integrating clinical and statistical measurement and prediction. *Journal of Abnormal Psychology*, 116, 810–822.
- Zimmerman, M, McDermt, W., & Mattia, J. (2000). Frequency of anxiety disorders in psychiatric outpatients with major depressive disorder. *American Journal of Psychiatry*, 157, 1337–1340.
- Zinbarg, R.E., & Barlow, D.H. (1996). Structure of anxiety and the anxiety disorders: A hierarchical model. *Journal of Abnormal Psychology*, 105, 181–193.