

## Limitations of the Dodo Bird Verdict and the Role of Clinical Trials in Psychotherapy Research: Comment on Wampold et al. (1997)

Paul Crits-Christoph  
University of Pennsylvania

B. E. Wampold et al.'s (1997) meta-analysis provides a useful and methodologically sophisticated summary of the results of comparative psychotherapy outcome research. Despite its strengths, some limitations of the meta-analysis that may have biased the results against finding differences between treatments are pointed out in this article. In addition, the types of treatments and patient populations to which the results can be generalized are clarified through an analysis of the studies contained within the meta-analysis. The importance of exceptions to the Dodo bird verdict is emphasized. Disagreements with Wampold et al. on the implications of their meta-analysis for research and practice, in particular the role of clinical trials in psychotherapy research and the need for identifying treatments that are "empirically supported," are discussed.

The meta-analysis of Wampold et al. (1997) has many fine features. Among these assets include the focus on "bona fide" psychotherapies, the correction of some of the methodological problems of previous meta-analyses, the use of sophisticated meta-analytic statistics, and an attention to moderating variables. Thus, the Wampold et al. meta-analysis is an important contribution to the Dodo bird issue. Based on these results, it seems clear that the equivalent outcomes produced by different psychotherapies needs to be taken seriously.

In this commentary, I first address some limitations of the Wampold et al. (1997) review that may have affected the results. The set of treatments and patients populations to which the Dodo bird verdict appears to apply is clarified through a detailed examination of the studies contained within the meta-analysis. The importance of exceptions to the overall trend is presented. Finally, I discuss the implications of the meta-analysis in terms of the Division 12 task force effort criticized by Wampold et al., and I describe a role for clinical trials in psychotherapy research, despite the common finding of no difference between bona fide psychotherapies.

### Methodological Limitations to the Meta-Analysis

A number of decisions made by Wampold et al. (1997) may have biased their results against finding differences between treatment conditions. One issue is the inclusion of follow-up assessments. Although there are examples of when psychotherapy effects do not emerge until follow-up, in studies of clinical disorders there is typically some degree of relapse during the follow-up period. Because follow-up periods in psychotherapy outcome studies have almost always been naturalistic (i.e., pa-

tients are free to seek additional treatment), the differences between treatment conditions are often attenuated at the follow-up assessment because the patients who were most symptomatic (i.e., in a treatment that did not fare well) have had more treatment and some of the patients who were well at termination have now relapsed. Although it is certainly important to continue to investigate the long-term outcomes of psychotherapy treatments, the interpretative problems inherent in naturalistic follow-up need to be acknowledged.

At least one study (Lomont & Sherman, 1971) included by Wampold et al. (1997) did not find any pre-post changes for any of the treatments studied. It seems questionable to use a study that found all treatments to be completely ineffective in the context of a claim that different psychotherapies are equally effective.

A more important issue relates to the decision by Wampold et al. (1997) to average effect sizes across outcome measures within a study. A typical comparative psychotherapy outcome study includes a number of outcome measures tapping change on the presenting problem that is targeted by the treatments and then a variety of secondary outcome measures designed to explore the breadth of treatment effects. Not uncommon, one treatment may be superior to another on the target measures but shows little advantage on the secondary measures that were not a focus of treatment. I do not want to minimize the importance of examining the broader effects of psychotherapy treatments. Nevertheless, if one treatment is superior to another on the primary presenting problems but equal to the other treatment on other outcomes, the first treatment remains clinically more important—yet this difference is obscured through the averaging of outcome measures.

### Limitations on the Generalizability of the Dodo Bird Conclusion

The above issues of inclusion of follow-up assessments and averaging of outcome measures are likely to take on greater importance in the context of the treatment of relatively more

---

I would like to thank Robert DeRubeis and Dianne Chambless for their helpful comments on a draft of this comment.

Correspondence concerning this article should be addressed to Paul Crits-Christoph, Department of Psychiatry, University of Pennsylvania, Room 700, 3600 Market Street, Philadelphia, Pennsylvania, 19104. Electronic mail may be sent via Internet to crits@landru.cpr.upenn.edu.

severe clinical disorders. With studies of mildly impaired samples, there may be little relapsing over follow-up and therefore little, if any, seeking of additional treatment. In addition, with mild conditions, the nonspecific effects of treatments (therapeutic alliance, positive expectations about change, etc.) are likely to be powerful enough in themselves to affect both primary and secondary outcomes, leaving little room for the specific factors to play much of a role. Thus, the nature of the sample of studies as a whole contained within Wampold et al.'s (1997) article is important for determining the extent to which these biases may have occurred. A description of the sample of studies has even more importance in terms of setting the boundaries of the Dodo bird verdict. The Wampold et al. article did not provide a description of the studies other than to mention that there was a preponderance of cognitive and behavioral studies.

There are 114 articles in the reference list of the Wampold et al. (1997) article that were included in the meta-analysis. Not all of these were independent studies because often follow-up data were presented in a separate report from the original termination data. Of the 114 articles, 51 (about 45%) appear to target a *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., *DSM-IV*; American Psychiatric Association, 1994) disorder, whereas the rest of the studies used samples that were generally at milder levels of dysfunction (e.g., 11 studies on test anxiety, 4 studies on speech anxiety). Exactly half (57) of the 114 articles involved the treatment of some form of anxiety. Forty (about 35%) of the 114 articles had undergraduates (almost always solicited) as participants in their study sample.

An examination of the types of treatments contained in the pool of studies reveals that indeed the bulk of the studies compared different forms of individual cognitive-behavioral therapies (e.g., cognitive therapy, desensitization, exposure, relaxation, skills training, assertion training). By my count, 79 (about 69%) of the 114 articles involved these kinds of comparisons. Thus, Wampold et al.'s (1997) Dodo bird verdict mostly applies to comparisons of cognitive and behavioral treatments for anxiety problems. The lack of differences between cognitive and behavioral treatments in particular does not mean that other types of noncognitive or nonbehavioral treatments would fare equally well with patients having anxiety problems. To the extent that distorted cognitions about the consequences of taking some action (e.g., belief that one would lose control, die, be ridiculed) are involved in anxiety, both cognitive (i.e., using verbal methods to convince a patient of the irrationality of their beliefs, thereby encouraging them to try something new) and behavioral (exposure treatment that demonstrates that the feared outcome does not happen) techniques might be effective ways of changing patients' chronic expectations and beliefs and thereby reduce their anxiety. I suspect cognitive techniques would be rated by clinicians as quite dissimilar to some behavioral techniques (e.g., flooding), so the similarity ratings made by Wampold et al. (1997) are not likely to have sorted out the effect sizes in the way I propose.

### The Importance of Exceptions to the Trend

Although cognitive and behavioral studies of anxiety problems are overrepresented in the pool of available studies, they

do not account for all of the studies, suggesting that the Dodo bird verdict may have some generalizability beyond simply cognitive and behavioral treatments for anxiety problems. However, an examination of the specific studies that were not comparisons of individual cognitive and behavioral treatments suggests that the Dodo bird verdict does not always apply. Among the 114 studies, I counted 29 independent studies that did not involve undergraduate samples and were not comparisons of simple, individual cognitive-behavioral treatments with each other (i.e., at least one of the treatments was not cognitive-behavioral or if both were cognitive-behavioral a couples treatment component was included). At least 14 of these 29 studies reported some meaningful difference between the treatment conditions. Because these effects are an important counterpoint to the conclusions of Wampold et al. (1997), effect sizes from these studies are presented in Table 1. For the calculations of effect sizes, I compared posttreatment means of pairs of psychotherapies and divided by the larger standard deviation. A single outcome measure assessing the presenting problems was selected for each study.

The 14 studies displayed in Table 1 illustrate that the Dodo bird verdict does not always apply. The effects sizes in these studies are generally in the "large" range (or greater) for the behavioral sciences, using Cohen's (1969) standards. Wampold et al. (1997) expressed a concern that the few examples of differences between treatments might simply be chance deviations expected in the context of a distribution of studies, where the mean effect size is near zero. My concern is that the results of large numbers of studies of different forms of individual cognitive-behavioral treatments (frequently using mildly dysfunctional, undergraduate samples) might obscure the effects of other kinds of comparisons, particularly with patients who are more dysfunctional. It is not surprising that the literature is full of comparisons of individual cognitive-behavioral therapies with each other on undergraduate samples because such studies are far easier to conduct compared with research where two different forms of psychotherapy are compared on more clinically distressed samples. Even if the Dodo bird verdict is found with most studies of bona fide psychotherapies for clinical disorders, the exceptions should not be discounted as chance deviations from a mean effect of zero if several studies are consistent with a given effect. Although I would like to see some of the effects displayed in Table 1 replicated, I also would not want to dismiss these effects out-of-hand because of the lack of effects in other, unrelated studies.

### Implications for Research and Practice

Wampold et al. (1997) stated a number of limitations of their meta-analysis, including the overrepresentation of behavioral and cognitive-behavioral studies and that it would be "unwarranted to conclude from this study that all therapies are equally effective with all disorders" (p. 210), that many therapies actually practiced were not represented very much—if at all—in the meta-analysis, and that inclusion of dependent variables that were unimportant or less sensitive to change may have created a bias. Despite these limitations, Wampold et al. (1997) believe that their results have "profound implications for research and practice" (p. 211) and that the strategy of attempting to empiri-

Table 1  
*Treatment Differences Involving at Least One Noncognitive–Behavioral Comparison*

Study	Treatment	Finding	Outcome measure	Effect size
Andrews (1971)	Desensitization and reinforcement (DR), client centered (CC)	DR > CC	Anxiety	0.49
Arean et al. (1993)	Problem solving (PS), reminiscence therapy (RT)	PS > RT	HAMD	0.89
Arnouk et al. (1985)	Partner-assisted exposure therapy (PE) + exposure relaxation (ER), PE + exposure communications training (EC)	PE + EC > PE + ER	Unaccompanied excursions out of the home	0.84
Beach & O'Leary (1992)	Marital therapy (MT), cognitive therapy (CT)	MT > CT	Wives' rating of marital adjustment	1.02
Borkovec & Costello (1993)	Nondirective (ND), cognitive-behavioral (CB)	CB > ND	HAMA	0.90
Borkovec & Mathews (1988)	Coping desensitization (CD), CB	CB > CD	HAMA	0.66
Clarke & Greenberg (1986)	Affective gestalt (AG), PS	AG > PS	Indecision	0.63
Fairburn et al. (1986)	CB, short-term focal (SF)	CB > SF	Global clinical	0.90
Foa et al. (1991)	Stress inoculation (SI), supportive counseling (SC)	SI > SC	Posttraumatic stress disorder severity	0.98
Gallagher-Thompson & Steffen (1994)	CB, brief psychodynamic (BP)	BP > CB: for short durations of caretaking CB > BP: for long durations of caretaking	Beck Depression Inventory	0.39 1.60
Goldman & Greenberg (1992)	Integrated systemic (IS), emotionally focused (EF)	EF > IS	Target complaints	0.43
Jacobson et al. (1991)	Cognitive therapy (CT), behavior marital therapy (BMT)	CT > BMT: for maritally nondistressed couples	HAMD	0.64
Kazdin et al. (1987)	CB, nondirective relationship therapy (NR)	CB > NR	Total behavior problems	0.81
Mavissakalian et al. (1983)	Paradoxical intention (PI), self-statement training (SS)	PI > SS	Anxiety during behavioral avoidance test	0.74

Note. HAMD = Hamilton Rating Scale for Depression; HAMA = Hamilton Anxiety Rating Scale.

cally validate specific psychotherapies “weakens support for psychotherapy as a mental health treatment rather than strengthens it” (p. 211). In my article, I disagree with Wampold et al. most strongly because I believe the acknowledged limitations of the meta-analysis in conjunction with further limitations mentioned above restrict the implications of Wampold et al.’s meta-analysis for research and practice.

First, it is unclear to me why Wampold et al. (1997) believe that their work has profound implications for the practice of psychotherapy, when some of the most common forms of psychotherapy in practice are hardly included within their review. There are only four studies on psychodynamic therapy, three studies of child or adolescent treatment, and zero studies of family therapy in the meta-analysis. Thus, the authors were generalizing beyond the available data.

In regard to implications for research, Wampold et al. (1997) believe that their results suggest that attempts to identify which psychotherapy treatments have been empirically validated along the lines of the work of the American Psychological Association Division 12 task force (Chambless et al., 1996) are misguided. Wampold et al. (1997) stated that “if one . . . places faith in the scientific evidence that psychotherapy in general is extremely

efficacious . . . , research in psychotherapy would differ considerably from the present focus on clinical trials” (p. 211). I believe Wampold et al. mischaracterized the Division 12 task force’s effort on empirically validated treatments and were too quick to dismiss the role of clinical trials in psychotherapy research, given the limitations of their meta-analysis.

In regard to the Division 12 task force’s effort, Wampold et al. (1997) stated that

the goal of the empirical validation movement is to identify a small set of treatments that satisfy criteria, which are based on the assumption that the unique ingredients of the treatment are responsible for the efficacy of the treatment (Wampold, 1997). (p. 211)

The intention of the task force was to identify all treatments for which there is empirical support, not just a “small set” (Task Force on Promotion and Dissemination of Psychological Procedures, 1995). In the draft of the Division 12 task force report currently being prepared, there are 21 well-established treatments and 36 probably efficacious treatments. Thus, the empirical validation strategy, rather than weakening support for psychotherapy, has provided some support for at least 57 specific psychotherapies for specific problems! Moreover, studies that

only compared a psychotherapy with a no-treatment control, and therefore did not document that the unique ingredients were responsible for efficacy, were considered by the task force and included in the "probably efficacious" category.

Wampold et al. (1997) cited Klein (1996) as saying that no current psychotherapy would be approvable by the Food and Drug Administration standards and, therefore, that one should reject "the necessity of validating psychotherapy based on the active ingredients" (Wampold et al., 1997, p. 211). Although there are few studies demonstrating superiority of a psychotherapy to a pill placebo (e.g., Power et al., 1990)—which would be Klein's preference—mostly because these studies have not been performed, an increasing number of studies have documented the superiority of a psychotherapy to a control condition for nonspecific aspects of psychotherapy. These studies were excluded from the Wampold et al. meta-analysis.

Examples of where there is replicated evidence for the specific effects (i.e., superiority to comparison conditions that control for nonspecific elements) of psychotherapeutic techniques include cognitive therapy for panic disorder, cognitive therapy for depression, exposure therapy for agoraphobia, exposure and response prevention for obsessive-compulsive disorder, and cognitive-behavior therapy for generalized anxiety disorder (see review by DeRubeis & Crits-Christoph, in press). These treatments would meet the Food and Drug Administration's approval standards if the requirement was superiority to a control condition that accounts for nonspecific effects. There is, therefore, no need to throw out the clinical trial methodology and the strategy of empirically validating specific psychotherapies because of a concern that this approach has not yielded any support for specific psychotherapies. Considerable support from clinical trials for specific psychotherapies does exist. The problem I identify here is that Wampold et al. (1997) conducted a review of comparative studies in which they attempted to draw implications for issues that can only be resolved through comparisons of active treatments with control groups.

What should the role of clinical trials in psychotherapy research be? Simple comparisons of bona fide psychotherapies should probably be the least common form of investigation rather than the most common. This is because without other control groups, such comparisons do not provide information on whether the lack of difference was due to both treatments being ineffective, being effective due to specific ingredients of each, or being effective due to nonspecific elements. Much more useful are designs that hold everything constant except one factor, so the potential causal impact of that one factor can be interpreted. Such additive or dismantling designs were excluded from the Wampold et al. (1997) review, although such studies would provide clearer evidence on the role of specific techniques compared with comparative designs. If more clinical trials using adequate control conditions were performed, the field of psychotherapy research would not have to dismiss the role of clinical trials and hope that naturalistic studies or other forms of research would satisfy critics such as Klein (1996) as well as managed care and government agencies interested in which psychotherapy works for what kind of problem.

Despite their limitations, comparative studies should continue to have a role. One role would be when a new treatment has been developed and needs to be compared with an existing,

already validated treatment. Comparisons of specific psychotherapies with medications are also very much needed. Such studies, however, would be enhanced by the inclusion of psychological or pill placebo control conditions as well. Another role of comparative studies is when there is a specific hypothesis about an Aptitude  $\times$  Treatment interaction. Although Wampold et al. (1997) included a few such studies in their meta-analysis, the study of Aptitude  $\times$  Treatment interactions using clinical samples is a new development, with very few studies yet performed. Indeed, given that within the Wampold et al. review I identified only 29 studies across a wide range of disorders that involved a noncognitive-behavioral treatment with a noncollege student sample and only 51 studies of any treatment type that targeted one of the many *DSM-IV* disorders, substantial further work appears to be needed in evaluating specific psychotherapies for specific disorders using a clinical trial methodology with adequate control conditions. When one realizes, for example, that not a single controlled study with adequate statistical power exists on standardized family or psychodynamic therapies for such common problems as social phobia, obsessive-compulsive disorder, major depression (nongeriatric sample), generalized anxiety disorder, agoraphobia, panic disorder, or borderline personality disorder and that even cognitive-behavioral treatments of these disorders have modest limited success, the need for standardizing and testing both existing and new treatments takes on even greater importance and the limits on generalization of the Wampold et al. meta-analysis come into clear focus.

The Dodo bird verdict may indeed be correct regarding some patient problems, but much more research is needed before it is known whether this verdict applies to the major clinical disorders. Moreover, the need to develop and experimentally test new treatments that might enhance the effectiveness of existing treatments leaves an important ongoing role for the clinical trial methodology in psychotherapy research. Of course, this is not to say that clinical trials are the only worthwhile methodology within psychotherapy research. Important research questions about psychotherapy can also be addressed through process studies and naturalistic studies as well as a variety of single-case and quasi-experimental designs.

## References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Andrews, W. R. (1971). Behavioral and client-centered counseling of high school underachievers. *Journal of Counseling Psychology, 18*, 93-96.
- Arean, P. A., Perri, M. G., Nezu, A. M., Schein, R. L., Christopher, F., & Joseph, T. X. (1993). Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. *Journal of Consulting and Clinical Psychology, 61*, 1003-1010.
- Arnou, B. A., Taylor, C. B., Agras, W. S., & Telch, M. J. (1985). Enhancing agoraphobia treatment outcomes by changing couple communication patterns. *Behavior Therapy, 16*, 452-467.
- Beach, S. R. H., & O'Leary, K. D. (1992). Treating depression in the context of marital discord: Outcome and predictors of response of marital therapy versus cognitive therapy. *Behavior Therapy, 23*, 507-528.
- Borkovec, T. D., & Costello, E. (1993). Efficacy of applied relaxation and cognitive-behavioral therapy in the treatment of generalized anxiety disorder. *Behavior Therapy, 26*, 257-270.

- ety disorder. *Journal of Consulting and Clinical Psychology*, 61, 611–619.
- Borkovec, T. D., & Mathews, A. M. (1988). Treatment of nonphobic anxiety disorders: A comparison of nondirective, cognitive, and coping desensitization therapy. *Journal of Consulting and Clinical Psychology*, 56, 877–884.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L., Williams, D. A., & McCurry, S. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, 49, 5–18.
- Clarke, K. M., & Greenberg, L. S. (1986). Differential effects of the gestalt two-chair intervention and problem solving in resolving decisional conflict. *Journal of Counseling Psychology*, 33, 11–15.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- DeRubeis, R. J., & Crits-Christoph, P. (in press). Psychological treatments for adult disorders. *Journal of Consulting and Clinical Psychology*.
- Fairburn, C. G., Kirk, J., O'Connor, M., & Cooper, P. J. (1986). A comparison of two psychological treatments for bulimia nervosa. *Behaviour Research and Therapy*, 24, 629–643.
- Foa, E. B., Rothbaum, B. O., Riggs, D. S., & Murdock T. B. (1991). Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive-behavioral procedures and counseling. *Journal of Consulting and Clinical Psychology*, 59, 715–723.
- Gallagher-Thompson, D., & Steffen, A. M. (1994). Comparing effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *Journal of Consulting and Clinical Psychology*, 62, 543–549.
- Goldman, A., & Greenberg, L. (1992). Comparison of integrated systemic and emotionally focused approaches to couples therapy. *Journal of Consulting and Clinical Psychology*, 60, 962–969.
- Jacobson, N. S., Dobson, K., Fruzzetti, A. E., Schmalings, K. B., & Salusky, S. (1991). Marital therapy as a treatment for depression. *Journal of Consulting and Clinical Psychology*, 59, 547–557.
- Kazdin, A. E., Esveldt-Dawson, K., French, N. H., & Unis, A. S. (1987). Problem-solving skills and relationship therapy in the treatment of antisocial child behavior. *Journal of Consulting and Clinical Psychology*, 55, 76–85.
- Klein, D. F. (1996). Preventing hung juries about therapy studies. *Journal of Consulting and Clinical Psychology*, 64, 81–87.
- Lomont, J. F., & Sherman, L. J. (1971). Group systematic desensitization and group insight therapies for test anxiety. *Behavior Therapy*, 2, 511–518.
- Mavissakalian, M., Michelson, L., Greenwald, D., Kornblith, S., & Greenwald, M. (1983). Cognitive-behavioral treatment of agoraphobia: Paradoxical intention vs. self-statement training. *Behavior Research and Therapy*, 21, 75–86.
- Power, K. G., Simpson, R. J., Swanson, V., Wallace, L. A., Feistner, A. T. C., & Sharp, D. (1990). A controlled comparison of cognitive-behaviour therapy, diazepam, and placebo, alone and in combination, for the treatment of generalised anxiety disorder. *Journal of Anxiety Disorders*, 4, 267–292.
- Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatment: Report and recommendations. *The Clinical Psychologist*, 48, 2–23.
- Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Hyun-nie, A. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin*, 122, 203–215.

Received March 25, 1997

Revision received April 15, 1997

Accepted April 15, 1997 ■

### Dannemiller Appointed Editor of *Developmental Psychology*, 1999–2004

The Publications and Communications Board of the American Psychological Association announces the appointment of James L. Dannemiller, PhD, University of Wisconsin, as editor of *Developmental Psychology* for a 6-year term beginning in 1999.

Effective January 1, 1998, manuscripts should be directed to

James L. Dannemiller, PhD  
Department of Psychology  
University of Wisconsin—Madison  
1202 W. Johnson Street  
Madison, WI 53706-1611  
email: jldannem@facstaff.wisc.edu