

Weighing the Evidence for Psychotherapy Equivalence: Implications for Research
and Practice

Robin Westmacott John Hunsley

University of Ottawa

Abstract

In the past two decades, numerous meta-analyses have been published that examine the question of psychotherapy equivalence. Hunsley and Di Giulio (2002) critically reviewed this literature and concluded that there was abundant evidence that the Dodo bird verdict of equivalence across psychotherapies is false. In this article, we summarize and update Hunsley and Di Giulio's (2002) review of recent meta-analyses and comparative treatment studies relevant to this question. Taken together, the empirical evidence clearly indicates that psychotherapy *nonequivalence* is the rule, not the exception. We discuss these findings and their implications for psychological research and practice.

Keywords: psychotherapy equivalence, Dodo bird, comparative treatment studies, evidence-based treatments, effectiveness research

Weighing the Evidence for Psychotherapy Equivalence: Implications for Research and Practice

Since Rosenzweig (1936) asserted that all psychotherapy produced equivalent outcomes (and quoted the Dodo bird from *Alice in Wonderland* saying, "Everyone has won, and all must have prizes"), psychotherapy equivalence has been referred to as the Dodo bird verdict, and frequent claims have been made about the general equivalence of all forms of psychotherapy. Proponents of this perspective have argued that psychotherapy, in general, is effective and that there is no compelling evidence to suggest that some treatments are better than others for clinical problems (e.g., Bohart, O'Hara, & Leitner, 1998; Zinbarg, 2000). Accordingly, the various theoretical orientations are merely variations on a single theme and, although their distinctions may be important to clinicians and psychotherapy researchers, they are essentially meaningless with respect to actual treatment outcome.

Claiming all psychotherapies are equivalent is like suggesting that, for example, because applied behavioral analysis is useful for treating autistic disorder, any treatment provided for this disorder, be it thought field therapy or play therapy, is likely to be equally effective. Indeed, Luborsky et al. (2003) recently suggested that psychoanalysis, despite a lack of empirical comparisons with other treatments, may plausibly be assumed to be equivalent to other efficacious psychotherapies in light of the typical research finding of psychotherapy equivalence. Given the ubiquity of the claims for psychotherapy equivalence and the limited attention typically given to the actual research purporting to support the claim, there is the real possibility that

practitioners and students in mental health fields accept the Dodo bird verdict simply because it appears to be generally and uncritically accepted by others.

In the past two decades, numerous meta-analyses have been published that examine this question of psychotherapy equivalence. Hunsley and Di Giulio (2002) critically reviewed this literature and concluded that there was abundant evidence that the Dodo bird verdict is false. In this article, we begin by summarizing Hunsley and Di Giulio's (2002) review and then provide an updated review of relevant meta-analyses and comparative treatment studies published since Hunsley and Di Giulio's review. We consider evidence from (a) treatment outcome studies that compare the treated group with a control group to whom no services are provided (typically, a wait-list control group) and (b) comparative treatment studies that compare at least two active treatments (with a no-treatment control group sometimes, but not always, included). Clearly, comparative treatment studies are most relevant to the Dodo bird verdict as they provide a "head-to-head" comparison of treatments drawing on the same sample of clients randomly assigned to each condition. For the most part, the results we present use the d statistic for estimating the effect size of treatments (i.e., the difference between treatments or between treatment versus no treatment is expressed in standard deviation units); in some instances, when useful for interpretative purposes, we also provide information on other types of effect sizes. Finally, we briefly discuss the implications of these findings for psychotherapy research and current efforts to promote evidence-based psychotherapeutic practices.

Before considering the meta-analytic evidence, it is important to note that many authors have raised scientific cautions to consider when interpreting evidence that appears to indicate psychotherapy equivalence (Beutler, 1991; Cuijpers, 1998; Hsu,

2000; Norcross, 1995; Reid, 1997; Shadish & Sweeney, 1991; Stiles, Shapiro, & Elliott, 1986). Treatment fidelity, researcher theoretical allegiance, and measurement quality should all be considered before tentatively accepting that there may be no true difference between treatments in a given study. Sample size is also a critical element. Kazdin and Bass (1989) calculated that, based on the typical differences found between treatments, researchers wishing to compare two or more treatments should plan to include over 70 participants per treatment condition if they wish to have adequate power to detect treatment differences.

Meta-Analytic Evidence Presented by Hunsley and Di Giulio (2002)

Smith, Glass, and Miller (1980) conducted the first meta-analysis of psychotherapy. Based on several hundred treatment outcome and comparative treatment studies, they found strong evidence for significant differences among effects of different types of therapy (Table 5-4, p.94): Treatment outcome studies indicated that cognitive and cognitive-behavioral treatments had the largest effect sizes (mean d values of 1.31 and 1.21, respectively), followed by behavioral and psychodynamic treatments (0.91 and 0.78), humanistic treatments (0.63), and developmental treatments (including vocational-personal development counseling and "undifferentiated counseling"; 0.42). The authors then analyzed their data based on client diagnoses and again found substantial differences among treatment types (Table 5-5, p.96).

These, however, are not the results presented by advocates of the dodo bird verdict; instead, they focus on Smith et al.'s (1980) analyses conducted on therapy "classes," in which, based on treatment outcome studies, behavioral (mean $d = 0.98$) and verbal (mean $d = 0.85$) treatments produced comparable effects. These

“classes” were constructed by grouping cognitive-behavioral, behavioral, behavior modification, systematic desensitization, and other behavioral treatments in the behavioral class and grouping psychodynamic, humanistic and cognitive treatments in the verbal class. The logic of classifying cognitive treatments with psychodynamic and humanistic treatments is highly questionable, yet it is only in these analyses, among the dozens reported by Smith et al., that psychotherapy equivalence was found. In other words, the strongest evidence for the Dodo bird verdict from Smith et al. is based on a very questionable classification strategy! They also conducted analyses on 56 comparative treatment outcome studies involving behavioral and verbal classes of treatment. Even with the questionable classification strategy, behavioral treatments (mean $d = 0.96$) were significantly superior to the verbal treatments (mean $d = 0.77$; Table 5-14, p.108).

Weisz and colleagues conducted a series of meta-analyses focusing specifically on the child and adolescent treatment literature. Weisz, Weiss, Alicke, and Klotz (1987) meta-analyzed treatment outcome studies published between 1958 and 1984 and concluded that there was strong evidence for the superiority of behavioral treatments (including cognitive treatments) over nonbehavioral treatments. Subsequently, Weisz, Weiss, Han, Granger, and Morton (1995) meta-analyzed 150 child and adolescent treatment outcome studies published between 1983 and 1993. Behavioral treatments (cognitive, cognitive-behavioral, parent training, operant methods, respondent methods, and social skills training) yielded a mean d of 0.54, significantly greater than the mean d of 0.30 for the nonbehavioral treatments (client-centered and insight-oriented therapies). Taking into account important methodological features (such as random assignment, attrition, and

therapist experience), Weiss and Weisz (1995) evaluated the relative effectiveness of behavioral (including cognitive) versus nonbehavioral (psychodynamic and humanistic) treatments in a subset of the studies used by Weisz et al. (1987). This meta-analysis included 105 studies of treatments for anxiety disorders, depression and social skills deficits. Controlling for methodological quality, the mean d values of behavioral and nonbehavioral treatments were 0.86 and 0.38, respectively, with the relative difference even greater in the 10 comparative treatment studies in the sample that directly compared behavioral to nonbehavioral treatments (mean d values of 0.76 and 0.17, respectively).

Reid (1997) reviewed findings from 42 focused meta-analyses that examined treatments for specific conditions such as depression, insomnia, smoking cessation, and bulimia. He concluded that 74% showed evidence of differential treatment effects. He noted that behavioral (including cognitive and cognitive-behavioral) treatments showed clear superiority to other treatments for child maladaptation, child abuse, juvenile delinquency, and panic-agoraphobia. On the basis of his review, Reid concluded that there was little evidence to support the Dodo bird verdict.

In the most direct test of the Dodo bird verdict to date, Wampold, Mondin, Moody, Stich, Benson, et al. (1997) conducted a meta-analysis of comparative treatment studies published between 1970 and 1995. The authors calculated all effect size values between pairs of treatments and then calculated their d values in two ways. First, they aggregated all the absolute values of the obtained effect sizes, and divided by the number of effect sizes. Second, they calculated a mean d value by randomly assigning a positive or negative sign to each obtained effect size and dividing the aggregate of these values by the number of obtained effect sizes. They

reported a mean d of 0.19 for their first estimate (significantly different from zero) and a mean d of 0.0021 for their second (a nonsignificant effect).

Although emphasizing that their results strongly supported the Dodo bird verdict, Wampold et al. explicitly cautioned that their results were not evidence that all psychotherapies found in professional practice are equally efficacious or as efficacious as those included in their sample. In fact, closer examination of their results actually reveals that their data provided strong evidence for a lack of treatment equivalence. As Crits-Cristoph (1997) and Hunsley and Di Giulio (2002) pointed out, the majority of studies included in their analysis compared one type of cognitive-behavioral treatment to another cognitive-behavioral treatment; thus, even if warranted, the conclusion of psychotherapy equivalence could only be confidently applied within the family of cognitive-behavioral treatments (CBT), not to psychotherapy treatments in general. More importantly, Wampold et al. erred greatly in their calculations, as their second method for calculating the mean d value could, by definition, *only* yield a mean value of zero regardless of the true value (cf. Howard, Krause, Saunders, & Kopta, 1997).

The final meta-analysis reviewed by Hunsley and Di Giulio (2002) was that of Shadish, Matt, Navarro, and Phillips (2000). These researchers meta-analyzed 90 treatment outcome studies of *clinically representative psychotherapy*, only selecting studies in which clients, treatments, and therapists were representative of typical clinical settings. Shadish et al. found overall evidence of significant treatment effects in the sampled studies (mean $d = 0.41$). Using a random-effects model to predict treatment effect sizes, treatment orientation (i.e., behavioral vs. nonbehavioural) was found to be a significant predictor. In other words, treatment effect sizes were

larger for behavioral than for nonbehavioral treatments as practiced in typical treatment settings with typical clients and therapists.

Updating the Review: Article Search and Review Criteria

Our search for meta-analyses comparing psychotherapy treatments was conducted via a computer-based literature search of PsycInfo and Medline databases. Our search labels, specifying years 2002 – 2007, included “psychotherapy and equivalence,” “dodo bird,” “psychotherapy and meta-analysis,” “empirically supported treatments,” “allegiance,” “psychotherapy efficacy,” “differential treatment,” “common factors,” and “comparative treatment.” We then searched among the meta-analyses generated by this search strategy and selected only those studies in which the effects of different psychotherapies were compared via statistical analysis, not simply visual inspection.

We found only one comprehensive meta-analysis published since 2002 that included a range of treatments for a range of client conditions (Luborsky, et al., 2002) and six other more focused meta-analyses that examined treatment effects for short-term psychodynamic psychotherapy versus other treatments for various patient conditions (Leichsenring, Rabung, & Leibing, 2004), sex-offenders (Lösel & Schmucker, 2005), CBT for panic disorder with and without agoraphobia (Mitte, 2005), CBT and self-regulatory treatments for chronic low back pain (Hoffman, Papas, Chatkoff, & Kerns, 2007), and child and adolescent disorders (mostly externalizing problems and depression; Weisz, Jensen-Doss, & Hawley, 2006; Weisz, Valeri, & McCarty, 2006).

Luborsky and colleagues' comprehensive meta-analysis. Luborsky et al. (2002) examined 17 meta-analyses of comparative treatment studies and reported a mean

d value of 0.20; they described this value as nonsignificant, although the precise nature of the statistical analysis used to reach this conclusion is not clear. Although not discussed in their review, there were a number of primary research studies that were used in more than one of the meta-analyses they examined. Accordingly, the effect sizes reported in the 17 meta-analyses are not independent of each other. The precise impact these dependencies had on the accuracy and generalizability of their findings is hard to estimate, especially as only three of the meta-analyses contained over 10 studies, but it does raise questions about the accuracy of the .20 value they reported.

When commenting on their findings in a subsequent article, they stated “Our impression is that the occasional differences are likely to be attributable to chance factors, after all results are taken together” (Luborsky et al., 2003, p.458). However, in our view, these differences are important given that the overall effect size estimate of 0.20 was derived from comparing one efficacious treatment to another efficacious treatment. To put this result in context, it is informative to convert d to the metric of number needed to treat (NNT) commonly used in medicine. NNT provides information on the number of patients one would need to treat with the target treatment to have one more successful patient outcome than would be possible with the comparison treatment. Converting an effect size d value of 0.20 to NNT yields a value of approximately 9 (8.892; see Kraemer & Kupfer, 2006). Thus, the relative benefits of the more efficacious treatment become evident before even 10 patients are treated. In this light, a d value of 0.20 may well be important in a clinical context.

Were obtained differences between treatments due primarily to chance factors? Upon further examination of Luborsky et al.'s (2002) findings, it seems unlikely, as it is possible to discern some distinctive patterns in their results. Sixteen meta-analytic estimates involved comparisons among psychotherapies, with one involving a comparison between psychotherapy and pharmacotherapy. Of the 16 effect sizes, 5 involved comparisons within the family of CBT approaches (e.g., cognitive vs. behavioral), with only 1 being statistically significant. There were three significant comparisons between variants of CBT and a group of treatments labeled as "general verbal" treatments, with all three favoring CBT. Four meta-analytic results compared the CBT family of treatments to the psychodynamic family of treatments, with only one significant result (favoring CBT). Finally, there were 4 comparisons between the psychodynamic family of treatments and other treatments (described as nonspecific, nonpsychiatric, psychiatric, and other, respectively), and none of these comparisons was significant. Thus, 4 of 5 significant results involved comparisons between cognitive-behavioral treatments and treatments based on other theoretical orientations.

It has been suggested that research allegiance to a particular theoretical orientation may result in delivering the preferred treatment in a more sophisticated and informed manner (Luborsky et al., 1993; Luborsky et al., 1999). Luborsky and colleagues (2002) attempted to control for such effects by averaging the score of three measures of researcher allegiance (ratings of the reprint, ratings by colleagues who know the researcher's work well, and self-ratings of allegiance by the researchers' themselves) and calculating the correlation of this score and the outcome of the treatments compared. The result was an r of .85 for a sample of 29

comparative treatment studies; when applied to their meta-analytic findings, this resulted in a corrected mean d value of 0.14 (in contrast to the original 0.20). There are significant problems when dealing with research allegiance statistically, as this is likely to overcorrect for any researcher allegiance effect that may have had a biasing influence on study results (Hunsley & Di Giulio, 2002).

Focused meta-analyses. Leichsenring et al. (2004) conducted a meta-analysis of 17 randomized studies of short-term psychodynamic psychotherapy (STPP) across a range of patient conditions (social phobia, PTSD, depression, cocaine and opiate dependence, personality disorders, chronic functional dyspepsia, and anorexia and bulimia nervosa). Some of these studies were also included in the meta-analyses used by Luborsky et al. (2002). They included only randomized controlled trials which compared an STPP to another active treatment and required that treatment manuals were used and therapists were experienced in STPP or specifically trained in STPP for the study ($N=15$ studies). STPP yielded a mean d of 1.39 after therapy ($p<.001$) and 1.57 at follow-up ($p<.001$) on target problems. The authors compared the efficacy of STPP with other forms of psychotherapy (mostly CBT, but also group interpersonal psychotherapy, brief supportive therapy, routine primary care, drug counselling, and brief adaptive psychotherapy); only two of the included studies had group sizes of more than 70 participants per treatment group. Leichsenring and colleagues separated treatment outcomes into target problems, general psychiatric symptoms, and social functioning. Within each study, they calculated the effect size difference between the active treatment groups for pre-post and post-follow-up for each of these groups of outcomes. They then averaged these differences for each outcome group and found mean between-group d 's ranging from -0.22 to 0.23,

none of which was statistically significant when analyses were conducted separately for (a) pretherapy-posttherapy effect size differences and then (b) posttherapy-follow-up effect size differences in target symptoms, general psychiatric symptoms, and social functioning. These results are very promising in terms of the efficacy of STPP. However, given that the between-group comparisons were calculated with only 7 to 15 studies and with small sample sizes within each treatment group in all but two of the studies, the power to detect small differences that may exist between treated groups was very low.

In the most comprehensive meta-analysis of sex-offender treatment to date, Lösel and Schmucker (2005) examined controlled treatment outcome evaluations published prior to 2004. Outcome was defined as recidivism; the authors followed a broad definition of recidivism, ranging from incarceration to lapse behavior. Sixty-nine studies containing 80 independent comparisons between treated and untreated offenders were analyzed. Although physical treatments (including physical castration and hormonal treatment) had much higher effects than psychosocial treatments (odds ratio = 7.37 vs. 1.32, respectively, largely due to the extreme effects of physical castration), only CBT ($OR = 1.45$) and behavioral treatments ($OR = 2.19$) had a significant impact on sexual recidivism. With odds ratios close to 1, the other approaches (including therapeutic community, insight-oriented, other and psychosocial treatments) did not significantly influence recidivism.

Mitte (2005) conducted a meta-analysis of (randomized and nonrandomized) behavioral, CBT, and pharmacological treatments for panic disorder with and without agoraphobia. Mitte computed the average of between-treatment (behavioral versus CBT) effect size differences across studies, and found no significant differences for

anxiety symptoms (mean $d=0.09$, effect sizes ranged from -0.07 to 0.24 ; $n=26$ studies), but found significant differences in favor of CBT for associated depressive symptoms (mean $d=0.18$, effect sizes ranged from 0.01 to 0.35 ; $n=22$ studies). As suggested previously, it is hardly surprising that significant differences are not always found (especially within a small sample of studies) when comparing various forms of behavioral and cognitive-behavioral treatments.

Hoffman et al. (2007) conducted a meta-analysis of psychological interventions for chronic low back pain; across four studies (averaging within-study between-treatment effect sizes differences), CBT was equivalent to self-regulatory treatments (SRT; including hypnosis and behavioral treatments such as biofeedback and relaxation training) at posttest (mean $d = -0.13$, *ns*) for pain intensity, and marginally less effective than SRT across three studies for associated depression (mean $d = -0.41$, $p < .10$). Given the low number of studies in these analyses, it is difficult to conclude what the true treatment difference might be.

Weisz et al. (2006) meta-analyzed 32 studies from the child and adolescent treatment literature that directly compared evidence-based treatments (i.e., treatments included in at least one published list of treatments showing beneficial effects) to usual care (psychotherapy, counselling, or case management provided as part of regular services). Client conditions were largely externalizing problems (conduct problems and delinquency were the focus of all but two of the studies). Averaging across the 32 studies, the authors found the mean d for evidence-based treatments (EBT) versus usual care (UC) was 0.30 , indicating that the average youth treated with an EBT was better off after treatment than 62% of youths who received UC. Follow-up data from 16 of the studies indicated that the mean difference at

follow-up in effect size between the EBT and UC groups was a significant $d = 0.38$. Notably, the superiority of EBTs over UC was not due to the use of homework to facilitate treatment generalization, efforts to ensure treatment integrity in the EBTs, research therapists delivering EBTs, theoretical allegiance of the researchers, evidence from voluntary treatment seekers, or differences in treatment setting. EBT superiority was not reduced by high levels of youth severity, comorbidity, or by inclusion of minority youths as study participants.

Weisz et al. (2006) meta-analyzed 35 randomized studies of psychotherapy for child and adolescent depressive symptomatology (elevated levels of depressive symptoms or formal diagnosis of major depressive disorder or dysthymic disorder). When data from multiple informants (i.e., youth, parents, and teachers) were combined, they found an overall mean d of 0.34. This was significantly less than the mean d of 0.99 found in previous meta-analysis for the treatment of depression and less than the mean effect size typically found for the treatment of youth disorders in general. One element of their analysis involved determining whether treatments that emphasized cognitive change were more effective than treatments that did not. They computed mean d (pre-post) separately for 31 treatments that involved an emphasis on cognition (i.e., cognitive therapy and CBT; mean $d = 0.35, p < .01$) and 13 treatments that did not emphasize cognition (primarily behavioral treatments, but also included attachment-based family training and interpersonal psychotherapy; mean $d = 0.47, p < .01$). None of the included studies were comparative treatment studies. The difference between treatments with a cognitive emphasis and those without was not significant.

As is evident in a number of the meta-analyses just reviewed, researchers sometimes conduct analyses to test for treatment equivalence using a relatively small number of studies, many of which have small sample sizes. It is important for readers of meta-analyses to remember that meta-analytic methods are not a simple statistical remedy for improving upon underpowered studies often found in the clinical treatment literature. Combining data from multiple treatment outcome studies may well provide a better estimate of the “true” impact of the treatment than is possible with one single study. However, comparing different treatments on the basis of such meta-analytic estimates can be problematic. If only a handful of underpowered studies are used to estimate a treatment effect, the accuracy of the estimate may be poor and the meta-analytic comparison may, itself, be underpowered to detect differences between the compared treatments. Similar problems will occur if the treatment comparisons are based on a small number of underpowered comparative treatment studies.

Like the distribution of most psychological data, meta-analytic estimates are distributed around the “true” population mean value of the treatment effect (see Schmidt, 1992). Estimates derived from a small set of studies, involving relatively small sample sizes, are likely to be found across the distribution, not just clustered near the population mean. As with all data, sample estimates are likely to more accurately reflect the population mean if the sample is large and generally representative. This elementary statistical fact is as true for secondary data analysis (i.e., meta-analytic data) as it is for data obtained for primary studies.

Evidence from Recent Comparative Treatment Studies

Given the limited number of recent meta-analyses examining the psychotherapy equivalence question, we decided to examine the outcome of recent comparative treatment studies. Accordingly, we examined the contents of several journals that typically publish such —studies—*American Journal of Psychiatry*, *Archives of General Psychiatry*, *Journal of the American Medical Association*, *Journal of Clinical Psychology*, and *Journal of Consulting and Clinical Psychology*—for 2006, the most recent complete publication year. Our literature search returned 12 randomized trials (two effectiveness trials and 10 efficacy trials) in which (a) two active treatments were compared or (b) a treatment was compared to whatever treatments were usually offered in the clinical setting (i.e., treatment as usual). In all instances, to be included in our presentation, researchers must have conducted statistical analyses directly comparing the outcomes of patients in the differing treatment conditions.

Table 1 presents a summary of the studies. Every study we found included a variant of CBT, broadly defined, as one of the tested treatments. Inspection of Table 1 reveals that many of the treatments resulted in substantial patient improvement. However, in nine of the studies, one treatment was significantly more efficacious than the other(s). This finding was obtained in both adequately powered and underpowered studies. In three studies, no treatment differences were reported: Christensen, Atkins, Yi, Baucom, and George (2006) compared two forms of behavioral couple therapy (traditional versus integrative), McBride, Atkinson, Quilty, and Bagby (2006) compared CBT and interpersonal psychotherapy for depression, and Strauman et al. (2006) compared cognitive therapy and self-system therapy (which contains some aspects of CBT) for the treatment of depression. Of these

three studies, two were underpowered to detect treatment differences, with only the Christensen et al. (2006) having a sample size close to the recommended 70 participants per treatment condition.

Insert Table 1 about here

One study merits particular attention. In comparing dialectic behavior therapy (DBT) to treatment as usual, Linehan et al. (2006) attended to several methodological suggestions made by treatment researchers to ensure fairness in comparative treatment studies. Specifically, in order to maximize internal validity, they controlled for: availability of treatment, assistance finding and getting to a first appointment with a therapist, hours of individual psychotherapy offered, therapist sex, therapist training, therapist clinical experience, and therapist expertise (with the alternative treatment group therapists having more expertise), availability of group clinical consultation, allegiance to treatment approach, institutional prestige associated with treatment, and general factors associated with receiving psychotherapy. Therapists delivering the alternative treatment (community treatment by experts; CTBE) were nominated by community mental health leaders as experts at treating difficult clients. The content and dosage of therapy was not prescribed by the researchers (i.e., experts could treat clients how they saw fit within the constraints of seeing clients *at least* once per week), the study paid for CTBE at the same rate as for DBT, and no participants were dropped because of failure to pay. Even when controlling for these factors, which have been shown in previous research to be salient to treatment outcome across a variety of treatment modalities, the DBT group was half as likely as the comparison group to attempt suicide during the treatment year, and used crisis services significantly less (1% of

DBT patients went to the emergency department at least once for any type of psychiatric emergency, versus 57.8% of CBTE patients).

Implications for Psychotherapy Research

Taken together, the empirical evidence clearly indicates that, when statistical comparisons have adequate power to detect differences, psychotherapy *nonequivalence* is the rule, not the exception. Across age groups and patient conditions, most researchers have found that some treatments are superior to others. That being said, it also appears that searching for differences among variants of CBT may not always yield statistically significant findings.

From our perspective, there is little to be gained from more research comparing one treatment to another—the Dodo bird verdict is generally not supported in well-designed and adequately powered studies. The only circumstance in which comparative treatment studies can be useful is when a new and promising treatment is compared to a treatment of established efficacy. It could be argued that, if a new treatment is to be tested for a condition in which there is already extensive replicated evidence of treatment efficacy for an established treatment, it would not be ethical to compare the new treatment to a no-treatment control group. Instead of conducting a treatment outcome study, it may be most appropriate to contrast the new treatment to an established treatment in a comparative treatment design, rather than withhold from patients access to a treatment known to be efficacious.

Nevertheless, many comparative treatment studies not involving new treatments will undoubtedly continue to be conducted. Based on available evidence, and assuming they have sufficient power to detect group differences, most such studies will continue to find that treatments are not equivalent in their clinical

effects. Knowing full well that such studies are to be conducted, we join with a growing number of researchers in suggesting that these studies should be designed to also provide information on both mechanisms and mediators of efficacious treatment (e.g., Jensen, Weersing, Hoagwood, & Goldman, 2005; Kazdin, in press). We need to know much more about how and why treatments work or fail to work, not just that one treatment is better than another. This type of information is especially important because, as shown repeatedly in the meta-analytic literature, treatments that fail to demonstrate their superiority in comparative trials still, nonetheless, demonstrate efficacy with respect to some conditions for some patients. Do all therapies exert their influence through the same mechanisms? If some therapies work via different mechanisms, is it possible to develop a treatment that optimally combines these differing mechanisms? These are the types of questions that need to be answered in order to truly advance our knowledge about the effects of psychological treatments.

Implications for Evidence-Based Psychological Practice

For some clinical conditions, the inescapable conclusion based on many hundreds of treatment studies is that some specific forms of psychological treatment should be viewed as first line options for clinicians. An increasing number of practice guidelines are now available that encourage attention to such findings. These include guidelines available from the Agency for Healthcare Research and Quality (<http://www.ahrq.gov/>), National Institute for Health and Clinical Excellence (<http://www.nice.org.uk/>), the American Psychiatric Association (http://www.psych.org/psych_pract/treatg/pg/prac_guide.cfm), and the American Academy of Child and Adolescent Psychiatry

(<http://www.aacap.org/page.wv?section=Practice+Parameters&name=Practice+Parameters>). It is also important to recognize that there are some conditions for which there may be multiple treatment options that work relatively well, including adult depression and couple conflict (see Hunsley & Lee, 2006).

Unfortunately, the emphasis within the field on establishing psychotherapy equivalence or treatment superiority has resulted in a rather substantial blind spot for many psychotherapy researchers and, possibly, clinicians. Some treatments are better than others but, as stated above, that does not mean that the less efficacious treatments are worthless. It is very important to know that a treatment is likely to be most beneficial to a client, but it is also important to know that, if the treatment fails to work for a specific client, there is another viable to consider, even if this alternative treatment has been found to be somewhat less efficacious in clinical trials. The movement to promote EBTs in clinical practice is precisely about encouraging the use of *all* treatments that have been shown to work in sound empirical investigations (e.g., Hunsley, 2007, in press).

Consider what is known about the impact of psychotherapy as routinely delivered in clinical settings. Hansen, Lambert, and Forman (2002) analyzed data from over 6,000 adult patients seen in a range of clinical settings (e.g., employee assistance programs, university counseling centers, community mental health clinics, and health maintenance organizations). In this large data set only 35% of clients met criteria for improvement or recovery. Very similar outcome results (29% of patients met criteria for improvement) were recently reported by Wampold and Brown (2005) in their sample of over 6,100 adult patients who received therapy services through a managed care organization. In a meta-analysis of 2,500 clients in

“real world” clinical practice, Lambert et al. (2003) found that less than a quarter of patients usually make substantial gains in treatment. Furthermore, a meta-analysis of studies of usual clinical care for children and adolescents indicated that obtained effect sizes averaged about zero (mean $d = 0.01$; Weisz, 2004; Weisz, Donenberg, Han, & Weiss, 1995).

In contrast to these findings, across 28 studies of EBTs, involving over 2,100 adult patients, Hansen et al. (2002) found that 57% of patients met criteria for recovery by the end of treatment, and fully two-thirds met criteria for improvement or recovery. Is there evidence that EBTs can work in real world clinical settings? Hunsley and Lee (2007) reviewed 35 treatment effectiveness studies for adult ($N=21$) and child/adolescent disorders ($N=14$). They included only studies that were designed to test an efficacious treatment (i.e., that had been previously tested in at least one efficacy study) in a routine clinical setting. They reported that the treatments provided in these effectiveness studies typically obtained outcome results comparable to those found in meta-analytic summaries of the efficacy studies on the same treatments. These findings suggest that treatments with established efficacy can be transported to clinical settings without any substantial loss of effectiveness. When combined with the growing evidence base showing that EBTs are superior to usual clinical services (Addis et al., 2003; Linehan et al., 2006; Mufson et al., 2004; Weisz et al., 2006), the need for dissemination and utilization of *all*, not just the best, EBTs is obvious.

Conclusion

Based on decades of research, it is clear that all psychotherapies are decidedly not equivalent in their clinical impact. Even among efficacious treatments, the mean

difference between treatments is frequently estimated to be approximately $d = .20$ (Luborsky et al., 2003; Wampold et al., 1997). If $d = .20$ is the best estimate of differences between efficacious treatments, and assuming a normal distribution of individual study results, it is entirely expected that there will be some instances of treatment equivalence in the literature. Whether these instances are informative or meaningful is, however, a separate issue. Based on evidence to date, a small number of these instances of treatment equivalence may be very interesting and clinically useful (e.g., finding that two very different types of treatment yield comparable results), some are only relatively informative (e.g., finding that different forms of CBT yield comparable results), and some, frankly, are misleading and irrelevant (e.g., finding that two treatments yield comparable results in studies without adequate power to detect group differences).

In this era of enhanced professional accountability and evidence-based health care, it is unlikely that evaluators, including policymakers, third party payers, and prospective clients, will be as benign and generous as the Dodo bird was in declaring all therapies to be “winners” (Winter, 2006). However, in promoting the fact that some treatments are better than others, we must not throw the proverbial “baby” out with the “bathwater.” If, despite persistent attempts by a clinician skilled in the provision of the first line treatment, insufficient progress is made in therapy, the responsible step is to consider alternative treatment options that have some supporting empirical evidence. Turning to second and third line treatments is routinely done in psychiatry and in other areas of medicine, and there is no reason that psychotherapy patients should expect any less attention from clinicians to the full range of available evidence-based psychotherapy options.

References

Addis, M. E., Hatgis, C., Krasnow, A. D., Jacob, K., Bourne, L., & Mansfield, A. (2004).

Effectiveness of cognitive-behavioral treatment for panic disorder versus treatment as usual in a managed care setting. *Journal of Consulting and Clinical Psychology, 72*, 625-635.

Bellack, A.S., Bennett, M.E., Gearon, J.S., Brown, C.H., & Yang, Y. (2006). A randomized clinical trial of a new behavioral treatment for drug abuse in people with severe and persistent mental illness. *Archives of General Psychiatry, 63*, 426-432.

Beutler, L.E. (1991). Have all won and must all have prizes? Revisiting Luborsky et al.'s verdict. *Journal of Consulting and Clinical Psychology, 59*, 226-232.

Bohart, A. C., O'Hara, M., & Leitner, L. M. (1998). Empirically violated treatments:

Disenfranchisement of humanistic and other psychotherapies. *Psychotherapy Research, 8*, 141-157.

Borkovec, T.D., & Costello, E. (1993). Efficacy of applied relaxation and cognitive-behavioral therapy in the treatment of generalized anxiety disorder. *Journal of Consulting and Clinical Psychology, 61*, 611-619.

Bradley, R., Greene, J., Russ, E., Dutra, L., & Westen, D. (2005). A multidimensional meta-analysis of psychotherapy for PTSD. *American Journal of Psychiatry, 162*, 214-227.

Carpentier, M.Y., Silovsky, J.F., & Chaffin, M. (2006). Randomized trial of treatment for children with sexual behavior problems: Ten-year follow-up. *Journal of Consulting and Clinical Psychology, 74*, 482-488.

Christensen, A., Atkins, D.C., Yi, J., Baucom, D.H. & George, W.H. (2006). Couple and individual adjustment for 2 years following a randomized clinical trial comparing traditional versus

integrative behavioral couple therapy. *Journal of Consulting and Clinical Psychology, 74*, 1180-1191.

Clark, D.M., Ehlers, A., Hackmann, A., McManus, F., Fennell, M., Grey, N., et al. (2006).

Cognitive therapy versus exposure and applied relaxation in social phobia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74*, 568-578.

Crits-Christoph, P. (1997). Limitations of the Dodo bird verdict and the role of clinical trials in psychotherapy research: Comment on Wampold et al. (1997). *Psychological Bulletin, 122*, 216-220.

Cujipers, P. (1998). Minimising interventions in the treatment and prevention of depression:

Taking the consequences of the Dodo bird verdict. *Journal of Mental Health, 7*, 335-365.

Dimidjian, S., Hollon, S.D., Dobson, K.S., Schmaling, K.B., Kohlenberg, R.J., Addis, M.E. et al.

(2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*, 658-670.

Giesen-Bloo, J., Van Dyck, R., Spinhoven, P., Van Tilburg, W., Dirksen, C., Van Asselt, T., et al.

(2006). Outpatient psychotherapy for borderline personality disorder: Randomized trial of schema-focused therapy versus transference-focused psychotherapy. *Archives of General Psychiatry, 63*, 649-658.

Gosselin, P., Ladoucer, R., Morin, C.M., Dugas, M.J., & Baillargeon, L. (2006). Benzodiazepine

discontinuation among adults with GAD: A randomized trial of cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology, 74*, 908-919.

Greenberger, D., & Padesky, C.A. (1995). *Mind over mood: A cognitive therapy treatment manual for clients*. New York: Guilford Press.

Hansen, N.B., Lambert, M.J., & Forman, E.M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*, 329-343.

Howard, K. I., Krause, M. S., Saunders, S. M., & Kopta, S. M. (1997). Trials and tribulations in the meta-analysis of treatment differences: Comment on Wampold et al. (1997). *Psychological Bulletin, 122*, 221-225.

Hsu, L. M. (2000). Effects of directionality of significance tests on the bias of accessible effect sizes. *Psychological Methods, 5*, 333-342.

Hoffman, B.M., Papas, R.K., Chatkoff, D.K., & Kerns, R.D. (2007). Meta-analysis of psychological interventions for chronic low back pain. *Health Psychology, 26*, 1-9.

Hunsley, J. (in press). Addressing key challenges in evidence-based practice in psychology. *Professional Psychology: Research and Practice*.

Hunsley, J. (2007). Training psychologists for evidence-based practice. *Canadian Psychology, 48*, 32-42.

Hunsley, J., & Di Giulio, G. (2002). Dodo bird, phoenix, or urban legend? The question of psychotherapy equivalence. *The Scientific Review of Mental Health Practice, 1*, 11-22.

Hunsley, J., & Lee, C. M. (2006). *Introduction to clinical psychology: An evidence-based approach*. Toronto: John Wiley & Sons.

Hunsley, J., & Lee, C.M. (2007). Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice, 38*, 21-33.

Jensen, P.S., Weersing, R., Hoagwood, K.E., Goldman, E. (2005). What is the evidence for evidence-based treatments? A hard look at our soft underbelly. *Mental Health Services Research, 7*, 53-74.

Kazdin, A. E. (in press). Mediators and mechanisms of change in psychotherapy research.

Annual Review of Clinical Psychology.

Kazdin, A. E., & Bass, D. (1980). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57*, 138-147.

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry, 59*, 990-996.

Lambert, M.J., Whipple, J.L., Hawkins, E.J., Vermeersch, D., Nielsen, S.L., & Smart, D.W.

(2003). Is it time to track patient outcome on a routine basis? A meta-analysis. *Clinical Psychology: Science and Practice, 10*, 288-301.

Leichsenring, F., Rabung, S., & Leibling, E. (2004). The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders. *Archives of General Psychiatry, 61*, 1208-1216.

Linehan, M.M., Comtois, K.A., Murray, A.M., Brown, M.Z., Gallop, R.J., Heard, H.L., et al.

(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.

Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology, 1*, 117-146.

Luborsky, L., Rosenthal, R., Diguier, L., Andrusyna, T.P., Berman, J.S., Levitt, J., et al. (2002).

The dodo bird verdict is alive and well - mostly. *Clinical Psychology: Science and Practice, 9*, 2-12.

Luborsky, L., Rosenthal, R., Diguier, L., Andrusyna, T.P., Levitt, J.T., Seligman, D.A., et al.

(2003). Are some psychotherapies much more effective than others? *Journal of Applied Psychoanalytic Studies*, 5, 455-460.

McBride, C., Atkinson, L., Quilty, L.C., & Bagby, R.M. (2006). Attachment as a moderator of treatment outcome in major depression: A randomized control trial of interpersonal psychotherapy versus cognitive behavior therapy. *Journal of Consulting and Clinical Psychology*, 74, 1041-1054.

Mitte, K. (2005). A meta-analysis of the efficacy of psycho- and pharmacotherapy in panic disorder with and without agoraphobia. *Journal of Affective Disorders*, 88, 27-45.

Mufson, L., Dorta, K. P., Wickramaratne, P., Nomura, Y., Olfson, M., & Weissman, M. M. (2004). A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry*, 61, 577-584.

Nock, M. K., & Ferriter, C. (2005). Parent management of attendance and adherence in child and adolescent therapy: A conceptual and empirical review. *Clinical Child and Family Psychology Review*, 8, 149-166.

Norcross, J. C. (1995). Dispelling the Dodo bird verdict and the exclusivity myth in psychotherapy. *Psychotherapy*, 32, 500-504.

Reid, W. J. (1997). Evaluating the Dodo's verdict: Do all interventions have equivalent outcomes? *Social Work Research*, 21, 5-16.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry*, 6, 412-415.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47, 1173-1181.

Shadish, W. R., Matt, G. E., Navarro, A. M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin, 126*, 512-529.

Shadish, W. R., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason why we don't let Dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology, 59*, 883-893.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.

Stiles, W.B., Shapiro, D.A., & Elliot, R. (1986). Are all psychotherapies equivalent? *American Psychologist, 41*, 165-180.

Strauman, T.J., Vieth, A.Z., Merrill, K.A., Kolden, G.G., Woods, T.E., Klein, M.H. et al. (2006). Self-system therapy as an intervention for self-regulatory dysfunction in depression: A randomized comparison with cognitive therapy. *Journal of Consulting and Clinical Psychology, 74*, 367-376.

Tang, N.K.Y. & Harvey, A.G. (2006). Altering misperception of sleep in insomnia: Behavioral experiment versus verbal feedback. *Journal of Consulting and Clinical Psychology, 74*, 767-776.

Vieth, A., Strauman, T.J., Kolden, G., Woods, T., Michels, J. & Klein, M.H. (2003). Self-system therapy: A theory-based psychotherapy for depression. *Clinical Psychology: Science and Practice, 10*, 245-268.

Wampold, B.E., & Brown, G.S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73*, 914-923.

- Wampold, B. E., Mondin, G. W., Moody, M., & Ahn, H. (1997). The flat earth as a metaphor for the evidence for uniform efficacy of bona fide psychotherapies: Reply to Crits-Christoph (1997) and Howard et al. (1997). *Psychological Bulletin*, *122*, 226-230.
- Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "All must have prizes." *Psychological Bulletin*, *122*, 203-215.
- Weissman, M.M., Markowitz, J.C., & Klerman, G.L. (2000). *Comprehensive guide to interpersonal psychotherapy*. New York: Basic Books.
- Weiss, B., & Weisz, J. R. (1995). Relative effectiveness of behavioral versus nonbehavioral child psychotherapy. *Journal of Clinical and Consulting Psychology*, *63*, 317-320.
- Weisz, J.R., (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples*. Cambridge, UK: Cambridge University Press.
- Weisz, J.R., Donenberg, G.R., Han, S.S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*, 688-701.
- Weisz, J.R., Jensen-Doss, A., & Hawley, K.M. (2006). Evidence-based youth psychotherapies versus usual clinical care. *American Psychologist*, *61*, 671-689.
- Weisz, J.R., Valeri, S.M., & McCarty, C.A. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin*, *132*, 132-149.
- Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*, *55*, 542-549.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, *117*, 450-468.

Winter, D. (2006). Avoiding the fate of the Dodo bird: the challenge of evidence-based practice. In D. Loewenthal, & D. Winter (Eds.), *What is psychotherapy research?* (pp. 41-46). London: Karnac Books.

Zinbarg, R.E. (2000). Comment on "Role of emotion in cognitive-behavior therapy": Some quibbles, a call for greater attention to patient motivation for change, and implications of adopting a hierarchical model of emotion. *Clinical Psychology: Science and Practice*, *7*, 394-399.