Luborsky L., Rosenthal R., Diguer L., Andrusyna T.P., Berman J.S., Levitt J.T., Seligman D.A. & Krause E.D., The Dodo bird verdict is alive and well – mostly. *Clinical Psychology: Science and Practice*, 2002, 9, 1: 2-12 (Commentaries [pp. 13-34]: D.L. Chambless; B.J. Rounsaville & K.M. Carrol; S. Messer & J. Wampold; K.J. Schneider; D.F. Klein; L.E. Beutler).

THE DODO BIRD VERDICT IS ALIVE AND WELL – MOSTLY

Lester Luborsky, Robert Rosenthal, Louis Diguer, Tomasz P. Andrusyna, Jeffrey S. Berman, Jill T. Levitt, David A. Seligman, Elizabeth D. Krause

Lester Luborsky, Department of Psychiatry, University of Pennsylvania; Robert Rosenthal, Department of Psychology, University of California, Riverside, CA; Louis Diguer, Department of Psychology, Laval University; Tomasz P. Andrusyna, Department of Psychiatry, University of Pennsylvania; Jill T. Levitt, Department of Psychology, Boston University; David A. Seligman, Access Measurement Systems, Ashland, MA; Jeffrey S. Berman, Department of Psychology, University of Memphis; Elizabeth D. Krause, Department of Psychology, Duke University, Durham, NC.

Partially supported by the National Institute on Drug Abuse, Research Scientist Award 2K05 DA 00168-23A (to Lester Luborsky), and NIMH Clinical Research Center Grant MH45178 (to Paul Crits-Christoph), and helpful suggestions from Irene Elkin.

Correspondence concerning this article should be addressed to Lester Luborsky, 514 Spruce St., Philadelphia, PA 19106. Electronic mail may be sent via internet to luborsky@landru.cpr.upenn.edu.

Abstract. 17 meta-analyses have been examined of comparisons of active treatments with each other, in contrast to the more usual comparisons of active treatments with controls. These meta-analyses yield a mean uncorrected absolute effect size for Cohen's *d* of .20, which is small and non-significant (an equivalent Pearson's r would be .10). Its smallness confirms Rosenzweig's supposition in 1936 about the likely results of such comparisons. In the present sample, when such differences are then corrected for the therapeutic allegiance of the researchers involved in comparing the different psychotherapies, these differences tend to become even further reduced in size and significance, as shown in Luborsky et al. (1999).

Key Words: comparisons of active psychotherapies, psychotherapy outcomes, correction for researcher's treatment allegiances, empirically validated treatments, Dodo bird verdict.

Saul Rosenzweig's seminal survey of 1936, "Some implicit common factors in diverse methods of psychotherapy," launched the field of psychotherapy's lasting interest in this topic. He supposed that the common factors across psychotherapies were so pervasive that there would be only small differences in the outcomes of comparisons of different forms of psychotherapy. It was a long time in coming, but in 1975 Luborsky, Singer, and Luborsky examined about 100 comparative treatment studies and found that Rosenzweig's hypothesis was essentially right – there was a trend of only relatively small differences from comparisons of outcomes of different treatments. Around that time I, and then others, began to call such small differences by the title from Rosenzweig's quote from Alice in Wonderland: "everybody has won so all shall have prizes" which was the "Dodo bird's verdict" after judging the race. The term "Dodo bird verdict" has since become commonly used and researchers have continued to write articles for or against the existence of, or the meaning of, that trend.

It is the aim of this paper to survey and then to evaluate whether Rosenzweig's (1936) hypothesis is still fitting and still flourishing. We will examine the exact amount of support for this trend, for the task is still very necessary – even expert psychotherapy researchers have different opinions, and even high affect, about the expected results. For a brief sample of these many opinions see: Tschuschke et al., (1998); Crits-Christoph, (1997); Wampold, Mondin, Moody,

Stich, Benson, & Ahn, (1997); King & Ollendick, (1998); Henry, (1998); King, (1997); Howard et al., (1997); Wampold et al., (1997); Cuijpers, (1998); Reid, (1997); Luborsky, Diguer, Luborsky & Schmidt, (1999); Beutler, (1991); Nietzel et al., (1987).

The use of a collection of meta-analyses to check the Dodo bird verdict:

Even before we present results of our collection of meta-analyses of studies on this topic we must review our reasons for using them in the way we did:

First, our collection only relies on meta-analyses because, according to Rosenthal & Rubin (1985) and Rosenthal (1998), meta-analyses ordinarily take into account each study's sample size and the magnitude of "effect size" of the treatments compared in each study. An effect size is a type of measure of the degree of association of two variables. The measures in each of these meta-analyses is a Cohen's d (or a variant of it), a difference between the two treatment means relative to their within group variations.

Second, we have limited ourselves to meta-analyses of the relative efficacy of pairs of different active psychotherapies in comparison with each other. We chose to emphasize such comparisons of active treatments because (a) these give an assessment of the relative efficacy of different active treatments that is of greater interest to clinicians than comparisons of treatment versus controls, and (b) the background variables for the patients in each treatment are likely to be more comparable when there is a direct comparison of two different active types of psychotherapy. We will not deal here with the level of efficacy of these treatments, for there is much evidence already for their mostly good level of efficacy (such as, Lambert & Bergin (1994), Lipsey & Wilson (1993), and Shadish, Matt, Navarro, Siegle, et al (1997)).

Third, we further limited our domain to meta-analyses of studies of common psychiatric diagnoses applied to adults (age 18 or over): depression, anxiety disorders (including obsessive-compulsive disorder and phobia) and mixed neurotics. Our findings do not apply to patients who are psychotic nor do they apply to children.

Fourth, we also limited the scope of our review to meta-analyses of some of the common types of therapies: behavior therapy, cognitive therapy, cognitive-behavior therapy, dynamic therapy, rational-emotive therapy and drug therapy. Drug therapy is included because it is one of the most common comparisons with psychotherapies and psychotherapists tend to be especially interested in the results of this comparison.

The meta-analyses that fit our criteria are briefly discussed below. Our search for these meta-analyses was aided by a computer-based literature search (using the PsychInfo and MedLine databases) and by two large lists of meta-analyses in Lipsey and Wilson (1993) and Chambless et al. (1996). For finding the meta-analyses in the computer sources our search labels included: "comparative treatment studies," "non-significant difference effect," "Dodo bird verdict," and "empirically validated treatments."

Our meta-analyses only include comparisons of effect sizes of active treatments with each other

In 1936 Rosenzweig had reasoned that psychotherapy outcome studies would show that different psychotherapies seem to have major ingredients in common that would lead them to have only small and non-significant outcome differences[2]; one such major ingredient is that they all involve a helping relationship with the therapist – Rosenzweig's conclusion was confirmed by Luborsky, Singer, and Luborsky in 1975, as we noted at the start.

In 1980 Smith, Glass, and Miller supported Luborsky et. al's 1975 conclusion but by a more systematic and much larger review of 475 comparative treatment studies of psychotherapy. They found an average effect size (ES) of treatment versus control studies of psychotherapy of .85. The ES measure used in this study and in the present study are all variants of Cohen's *d*, as described by Rosenthal (1991). A Cohen's *d* of .85 can easily be interpreted as a difference between the two group means of 85% of the standard deviation. Note that Smith, Glass, and Miller (1980) did not present effect sizes as we did from comparisons of active treatments with other active treatments, but rather effect sizes for each type of therapy compared with controls.

Fortunately, in the early phase of the present review, we also surmised a likely weakness of the method of relying on a treatment versus a control as compared with the method of an active treatment versus another active treatment. It is, in part, that the active treatment versus the control treatment tended to deal less well with the match of the background factors in the patients in the treatments compared. As an example, in Smith, Glass, and Miller (1980), the patients in the sample of studies in cognitive therapy may well have been less psychiatrically severe in their disorders than those who were in dynamic treatment. An active treatment versus another active treatment comparison might have equalized the groups of patients by the probable effects of randomization into each treatment[3]. For this reason also we decided to restrict our sample of meta-analyses to only those that relied on the comparison of two active treatments. We have located 17 of such meta-analyses in 6 reports (Table 1). Each of these 6 meta-analytic reports will be briefly described:

(1) Berman, Miller, Massman (1985) with a larger and more inclusive sample of studies than Miller and Berman (1993) also found small and non-significant differences between cognitive therapy and desensitization. (N = 20 studies). (A positive Cohen's *d* means only that the first treatment is more effective than the other treatment)

(2) Robinson et al. (1990) reported 6 meta-analyses with 4 of them significant using uncorrected effect sizes. (N = 53 studies). These imply that behavioral treatment is less effective than cognitive-behavioral (-.24), that cognitive behavioral is more effective than general verbal (.37); that cognitive therapy is more effective than general verbal (.47) and behavioral is more effective than verbal (.27). But when the effect sizes are corrected for the researchers' allegiance (by a method to be described below) they become lower and non-significant.

(3) Svartberg and Stiles (1991) continued the search for relatively efficacious therapies by meta-analyses of treatment comparisons, one of which reported a significant difference between dynamic versus cognitive-behavioral with a significant correlation of -.47 (14 studies).

(4) Crits-Christoph (1992) found non-significant differences in effect sizes of comparisons of active treatments for dynamic versus other psychotherapies. (11 studies).

(5) Luborsky, Diguer, Luborsky, Singer, and Dickter (1993), in a sample of three studies, again showed non-significant effect sizes for dynamic versus other psychotherapies (3 studies) (Note: to avoid duplication, the studies that were the same as those in Luborsky et al. (1999) were omitted here).

(6) Luborsky, Diguer, Seligman, Rosenthal, et al. (1999) found non-significantly different effect sizes in comparisons of cognitive versus behavioral, dynamic vs. behavioral, dynamic vs. cognitive, and pharmacotherapy vs. psychotherapy. (29 studies). The last of these comparisons was included because the pair is often considered by clinicians as a usable option either singly or in combination.

The main trends implied by the meta-analyses:

The mean effect sizes in the 17 meta-analyses showed low and non-significant differences.

The mean absolute value of the uncorrected Cohen's *d* effect size for the 17 meta-analyses listed in Table 1 was .20 which is not large and is non-significant, given that it is the mean of absolute values. To calculate each of the effect sizes, we first converted the three that used Pearson's r into Cohen's *d* so that they were all expressed in terms of Cohen's *d*.

The effect sizes were further reduced after corrections for the researcher's allegiance and for other factors.

There is another major influence that can alter the typically modest and non-significant difference effect – it is the researcher's allegiance effect. This effect is the association of measures of the researcher's allegiance to each of the treatments compared with measures of the outcomes of the treatments. There had been hints of this effect for many years, as first noted in Luborsky et al. (1975). Now there is a really exhaustive review of the topic (Luborsky et al., 1999) that shows a well-established researcher's allegiance effect – the correlation between the mean of 3 measures of the researcher's allegiance and the outcome of the treatments compared was a huge Pearson's r of .85 for a sample of 29 comparative treatment studies! The 3 measures, described in Luborsky et al (1999), are ratings of the reprint, ratings by colleagues who know the researcher's work well, and self-ratings of allegiance by the researcher' themselves.

This high correlation of the mean of the 3 allegiance measures with the outcomes of the treatments compared implies that the usual comparison of psychotherapies has a limited validity because so far it is not easy to rule out the presence of the large researcher allegiance effect. To make matters worse, it is not clear at all how the allegiance effect comes about. A variety of methods have been suggested by Luborsky et al. (1999) for reducing the intrusion of the researcher's allegiance, but, even when they are implemented, the impact of such methods are likely to remain ambiguous in the precise amount of correction to be applied. Among the recommended precautionary steps: it might be valuable a) to include researchers with a variety of allegiances in the

research group carrying out the study and b) to choose as a comparison to the preferred treatment, a treatment that is equally likely to be judged as credible. (Berman and Luborsky, in preparation; Berman and Weaver 1997).

A sample of the effects of corrections are noted as follows: When the uncorrected correlations in Robinson et al. (1990) were corrected for researchers' allegiance by the mean of their three corrected allegiance scores (the most common type of correction used here) (Luborsky et al, 1999), the correlations become lower and non-significant. The data from Smith, Glass, and Miller (1980) was corrected for reactivity (meaning, influencable by therapist or by researcher) and Luborsky et al. (1993) was corrected for the quality of the research design (Luborsky et al 1999). The more exact changes can be seen by a comparison of the uncorrected with the corrected effect sizes in Table 1. For example, in Luborsky, Diguer, Luborsky, Singer, and Dickter (1993), the uncorrected comparison of two active treatments effect size was .00 (non-significant) and the effect size after correcting for research quality was similar in size: -.01 (non-significant).

To summarize these results, we compared the mean of the effect sizes of corrected comparisons of active treatments from 11 meta-analyses in Table 1 – the 11 were all those for which we had data to compute corrections – with the mean of the corresponding uncorrected effect sizes We first converted all these effect sizes into Cohen's *d* (Cohen, 1977) and then took the mean of the absolute value of the effect sizes. The mean uncorrected effect size with Cohen's *d* was .20 but the mean corrected Cohen's *d* effect size was only .12; the reductions of the corrected effect sizes meant they were no longer significant. Also, the median uncorrected effect size was .21 as compared to a corrected median effect size of .14; the reductions also meant they were no longer significant.

Comparison of effect sizes of meta-analyses for Cognitive and Cognitive-Behavioral vs. Dynamic and other treatments yields small differences.

A few of the common types of comparisons among the 17 meta-analyses warrant an even more focused review. But first, to make comparisons easier, two related subclasses can be combined, that is, the cognitive and the cognitive-behavioral. These two may be reasonable to combine because their comparison yielded only a very small effect size of -.03 in one meta-analysis with 4 studies (Robinson et al., 1990). An example of a comparison of corrected effect sizes that is easily shown in Table 1 compares Cognitive or Cognitive-Behavioral treatment with other treatments. The mean effect size of these six comparisons of Cohen's *d* is .14 (Cognitive vs. Behavioral, .12, Cognitive vs. General Verbal, -.15, Behavioral vs. Cognitive-Behavioral, -.16, Cognitive-Behavioral vs. General Verbal, .09, Cognitive vs. Behavioral, .22, and Dynamic vs. Cognitive, .08). This mean of .14 is not significantly different from means of the other comparisons listed so that such a finding is quite in synchrony with the other findings in the study.

The main explanations for the "small" effect sizes for differences in outcomes of active treatments

The effect sizes for comparisons of active treatments, both corrected and uncorrected, for the 17 meta-analyses, were usually relatively "small" and non-significant. The adjective "small" is in quotes and preceded by the ambiguous qualifier "relatively" because the choice of a corresponding effect size level varies among the writers on the topic. Cohen (1977) for example,

would call a *d* of .20 "small" (equivalent to a Pearson's r of only .10) but Rosenthal (1990, 1995) would call it greater than small because the designation is somewhat dependent on the requirements of the situation, for example, if only 4 of 100 persons having a heart attack are saved by taking aspirin, that is not a small percentage if you are one of the 4 people![4] Now we are more ready to consider some probable explanations for this relatively small and non-significant relationship:

Explanation 1: The types of treatments do not differ much in their main effective ingredients and therefore "small" differences with non-significant effects are the rule.

The treatment components that are in-common between the treatments compared may be the most influential basis for explaining the small and non-significant difference effect. This was the explanation offered by Rosenzweig (1936) and later restated by Frank and Frank (1991), Luborsky et al. (1975), Strupp and Hadley (1979), and Lambert and Bergin (1994). The last especially stressed the role of common factors across different psychotherapies in explaining the trend toward non-significant differences among the outcomes of different forms of psychotherapy. Elkin, et al., (1989) and Imber, et al. (1990) also considered the common factors across interpersonal and cognitive-behavioral psychotherapy in their explanations for the non-significant differences between different treatments in the NIMH Treatment for Depression Collaborative Research Program. This explanation emphasizes that the common components of different treatments may be so large and so much more potent than specific ingredients, that the comparisons result in small and non-significant differences. Other components have also been suggested as common across treatments: the helping relationship with the therapist, the opportunity to express one's thoughts (sometimes called abreaction), and the gains in self-understanding.

Explanation 2: The researcher's allegiance to each type of treatment compared differs, sometimes favoring one treatment and sometimes favoring the other.

The researcher's allegiance to each of the treatments in comparative treatment studies appears to influence the small effect sizes of each treatment outcome in the expected direction, as shown in the comprehensive evaluation by Luborsky et al. (1999). To explain this more concretely: Treatment A in a meta-analysis may be favored by the researcher's positive allegiance in one study while in another study treatment A may suffer from a researcher's negative allegiance.

Explanation 3: Clinical and procedural difficulties in comparative treatment studies may contribute to the non-significant differences trends.

There have been a series of rebuttals trying to explain the methodological problems that lead to the Dodo bird trend – among these are Beutler (1991), Elliott, Stiles and Shapiro (1993), Norcross (1995), and Shadish and Sweeney (1991). These discussions tend to agree that although research shows that the "small" and non-significant difference effect exists, the effects of different treatments may appear in ways that have not yet been studied. Kazdin (1986), Kazdin and Bass (1989), Wampold (1997), and Howard et al. (1997) further explain that non-significant differences between treatments may reflect procedural and design limitations in comparative treatment outcome studies. These limitations include the representativeness of the measures of treatment process and outcome and the statistical power of the findings. Howard et al. (1997) further suggests doing

separate meta-analyses for each contrasting pair of types of treatments, such as we have done for Cognitive and Cognitive-Behavioral vs. Dynamic and other treatments (on p. 11).

Explanation 4: Interactions between certain patient qualities and treatment types, if not taken into account, may contribute to the non-significant difference effects.

Several studies, such as those by Beutler et al. (1991) and Blatt (1992); Blatt and Folsen (1993); Blatt and Ford (1994), have shown that the match of the patient's personality with different treatments can then succeed in producing significant effects; when such matches are not taken into account, they may contribute to the non-significant difference effects.[5]

The status of the empirically validated treatment movement

Much of the most recent comparative treatment research has been done as part of the increasingly fashionable empirically validated treatment movement (Luborsky, in press). One list of such studies in Chambless et al. (1996) might be thought by some to belong in our review, but it actually belongs in a separate category because it was not supposed to be a meta-analysis of the relative effectiveness of different active treatments. We mention this review just because it is commonly mistaken to be a list which reports the relative efficacy of different treatments. However, it is not – the Task Force itself clearly stated that its focus is only on compiling a list of treatments that had been "empirically validated."

Conclusions and Discussion

The available evidence has been summarized here from 17 meta-analyses of comparisons of active treatments with each other, in contrast to the more usual comparisons of active treatments with control treatments. The studies reviewed mainly included patients with the common diagnoses of depression, anxiety disorders (including obsessive-compulsive and phobic disorders), and mixed neurosis but not patients who are psychotic or children. Also, the sample of studies included only those where patients were treated by these usual treatments: behavior therapy, cognitive therapy, cognitive-behavior therapy, dynamic therapy, rational-emotive therapy, and drug therapy.

Comparisons of active treatments with each other tend to have "small" and non-significant differences: For our sample of 17 of such meta-analyses in 6 reports of meta-analyses of comparisons of active treatments with each other, there is a mean uncorrected absolute effect size of .20 by Cohen's d. (Table 1). This is impressive because of its smallness as well as the fact that the 6 reports include meta-analyses with many studies. Another large-scale review of studies of treatment comparisons of active treatments also happened to find a similar level of effect sizes: .19 by a Pearson r (Wampold et al., 1997). The mean effect size in our review supports our impression that a majority of comparisons of an active treatment vs. an active treatment have relatively "small" effect sizes and non-significant differences between different psychotherapies, especially after corrections for the researcher's allegiances, thus re-affirming the original Dodo bird verdict -- "Everybody has won and all must have prizes." (as in the illustration from Alice's Adventures in Wonderland in Figure 1)

We also calculated medians of effect sizes (Table 1) in order to show that no one meta-analysis method skewed our overall mean effect size. Looking at Table 1, we see that the mean and median effect sizes, both weighted by the size of the sample of each study and unweighted, are almost identical for both the corrected and uncorrected effect sizes. Thus, our sample of meta-analyses has a good distribution, with no one meta-analysis method unduly affecting the overall mean uncorrected Cohen's *d* effect size of .20.

To describe further what the overall uncorrected mean effect size actually represents, we must explain that it is a very conservative estimate. When calculating our mean effect size, we took the absolute value of each of the 17 effect sizes before summing them. This inflates our mean effect size because, if we had kept the signs and summed in that manner, certain effects would cancel each other, resulting in a lower mean effect size. Even then, our mean Cohen's *d* of .20 is equivalent to a Pearson r of only .10. By Rosenthal and Rubin's Binomial Effect Size Display (BESD) method (1991), an r of .10 means that on average there is a 10% difference in success rate between psychotherapies (e.g. a change from 45% to 55%). Though a Cohen's *d* of .20 may not be "small" according to Rosenthal (1990, 1995), Cohen (1977) does see it as small, and this average 10% difference in success rate is the most conservative estimate of the overall mean effect size due to our absolute values method of combining the Cohen's *d*'s.

Our general conclusion, therefore, is that Rosenzweig's clinically-based hypothesis of 1936 has held up – the outcomes of quantitative comparisons of different active treatments with each other, because of their similar major components, are likely to show mostly "small" and non-significant differences from each other.

Comparisons of active treatments with each other often need a correction: The re-examination of 29 mostly newer studies by Luborsky et al. (1999) showed that a correction to the effect sizes is typically needed because researcher's allegiance to each of the therapies compared is highly correlated with treatment outcomes – the correlation was a Pearson's r of .85! Researcher's allegiance is therefore a reasonable basis for correcting effect sizes. After corrections for researchers' allegiance were applied, the effect sizes were usually reduced and non-significant.

A few of the comparisons of active treatments with each other have larger and more significant differences: When considered one by one, a few of the correlations are moderate and reach the conventional level for significance. Such correlations are infrequent as part of the entire set of meta-analyses, but the presence of occasional significant differences in treatment outcomes perhaps should be taken seriously, as Lambert and Bergin (1994) tentatively suggest. Looking at the array of results for the meta-analyses that are surveyed in Table 1, one is struck by the variability of the effect sizes in which a few of them rise above the designation of a small and non-significant level to at least a moderate size. Wampold et al. (1997) noticed the same variability in their results but tended to view these as chance results in their large distribution of results. Crits-Christoph (1997, p. 217) considered these more seriously, just as we are inclined to do, that these exceptions may reflect more than chance. Wampold et al. (1997) list in their Table 1, 14 of such exceptions in the 114 studies (p. 218). These suggest that something more than a Dodo bird verdict may be operating. Among the 114 studies in Wampold et al. (1997), Crits-Christoph (1997) identified only 29 studies with a noncollege student sample that involved comparison of noncognitive-behavioral treatments with each other (e.g. treatments other than cognitive therapy, desensitization, exposure,

relaxation, skills training, and assertion training). Of these 29 studies, only 14 showed some significant difference between the treatment conditions, suggesting that the Dodo bird verdict may not apply as well in all cases.

The basic issue in this discussion is whether a few differences that were more than small and better than non-significant a) should be attributed to chance factors or b) should be pursued as illustrations of more than chance effects. There are arguments in favor of each alternative.

Comparisons of active treatment with controls appear to be less valuable for our main aim: "Controls" were used here to refer to (a) comparison groups that were purposely lacking in a component and (b) non-psychotherapy treatments such as clinical management or wait-for-treatment groups. The type of comparative treatment study that is based on an active treatment vs. a control naturally tends to give higher effect sizes (Grisson, 1996) so that the results from this type of effect size measure cannot justifiably be combined with the results of comparisons of active treatments. It is for this reason that we did not include the many studies using mainly treatment vs. control comparisons, such as Shapiro and Shapiro (1982), Engels, Garnefski, and Dickstra (1993), Van Balkom et al. (1994), Feske and Chambless (1995), and Grawe et al. (1994). Furthermore, the treatment versus control type of comparison tends to be not as revealing of the relative potency of a treatment as are comparisons of active treatments with each other.

Other important questions remain to be examined: The meta-analyses comparing different active treatments with each other suggest further research: (a) The studies of comparative treatments still may not be sufficiently representative of the common diagnoses and the common types of psychotherapies. They may, for example, suffer from an unrepresentative selection of cases, as in Wampold et al. (1997) – according to Crits-Christoph (1997), Wampold et al. (1997) have about half of their 114 studies involving the treatment of various forms of anxiety but too little of severe degrees of the usual diagnoses. Also, the studies in the sample may over-represent behavioral and cognitive-behavioral treatments and under-represent dynamic treatments. Therefore, more clinical trials are needed that correct for these distortions.(Crits-Christoph, 1997) Our sample of studies partially corrects for such limitations. It is, impressive, however that our sample of studies, which overlaps only in part with Wampold et al. (1997), also shows the small and non-significant difference effect that we call the Dodo bird verdict. (b) It will be valuable to have the research quality of each study that is included judged by independent judges whose therapy allegiances to each form of psychotherapy are known. So far, there is basically no correlation between the quality of the research study and the size of the outcomes in psychotherapy (Smith and Glass, 1977 pg. 758 Table 4). We also found the same in Luborsky et al. (1999) where we had two judges rate each study for the quality of the research on 12 dimensions and correlated their mean score with the outcome of the treatment. (c) More investigations are needed of the degree of fit of each of the possible explanations of the small and non-significant difference trend. For example, a comparison is needed for the degree to which the small and non-significant differences trend is best explained by common elements between the two treatments or by other factors as suggested by Crits-Christoph (1997). (d) We may ultimately find that research on the match of the type of patient in relation to the type of treatment will offer more information than the usual comparative treatment research design with its focus on the comparison of different treatment types across patient types. After conducting more of these studies we may find that such match

designs reveal more effective treatments for certain kinds of patients than the usual focus (Beutler, 1991; Blatt, 1992). Barber and Muenz (1996) found that in the TDCRP data (Elkins et al., 1989), although CT and IPT did similarly in that study, if one looks at the patients who are more obsessive the IPT is better than CT, and if one looks at the patients who are more avoidant, then CT is better than IPT – in other words, subgroups of patients might do better or worse with a specific treatment so that what is needed are hypotheses about those subgroups of patients. (e) Similar meta-analyses should be done with long term treatments. We may find that long term treatments have a long term build up of improvement, that may ultimately lead to more benefit (Luborsky, in press). (f) It may be worth examining symptom outcomes along with other kinds of outcomes of psychotherapy. Small and non-significant outcomes do not mean that the treatments compared have the same effects on all patients. Specific outcome measures such as depression and anxiety, may tempt us to forget that there may be other differences between treatments. Two patients, for example, after participating in different treatments, may feel better and not currently depressed, but one of them also may have made other important gains—one currently non-depressed man said that he achieved a better understanding of his relationship with his wife and was able to make helpful changes. We and others therefore should score other aspects of the patients' changes as well, even beyond the symptom measures.

REFERENCES

Barber, J. P., & Muenz, L. R. (1996). The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: Empirical findings from the Treatment for Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 64*, 951-958.

Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of Consulting and Clinical Psychology, 64*, 619-622.

Berman, J. S., Miller, C. R., & Massman, P. J. (1985). Cognitive therapy versus systematic desensitization: Is one treatment superior? *Psychological Bulletin, 97*, 451-461.

Berman, J. S., & Luborsky, L. (2000). Researcher allegiance, treatment credibility, and the outcome of comparisons between psychotherapies. Manuscript in preparation.

Berman, J. S., & Weaver, J. F. (1997, December). Treatment credibility and the outcome of psychotherapy. Paper presented at the meeting of the North American Society for Psychotherapy Research, Tucson, AZ.

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.

Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., & Merry, W. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *Journal of Consulting and Clinical Psychology, 59*, 333-340.

Blatt, S. J., & Felsen, I. (1993). Different kinds of folks may need different kinds of strokes: The effect of patients' characteristics on therapeutic process and outcome. *Psychotherapy Research, 3*, 245-259.

Blatt, S. J., & Ford, R. (1994). *Therapeutic change: An object relations perspective*. New York: Plenum.

Blatt, S. J. (1992). The differential effect of psychotherapy and psychoanalysis on anaclitic and introjective patients. The Menninger psychotherapy research project revisited. *Journal of the American Psychoanalytic Association, 40*, 691-724.

Chambless, D. et al. (1996). An update on empirically validated therapies. *The Clinical Psychologist, 49*, 5-18.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Crits-Christoph, P. (1992). The efficacy of brief dynamic psychotherapy: A meta-analysis. *The American Journal of Psychiatry, 149*, 151-158.

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.

Cuijpers, P. (1998). Minimizing interventions in the treatment and prevention of depression: Taking the consequences of the "Dodo Bird Verdict." *Journal of Mental Health (Uk), 7*, 355-365.

Elkin, I., Shea, T., Watkins, J., Imber, S., Sotsky, S., Colins, J., Glass, D., Pilkonis, P., Leber, W., Docherty, J., Fiester, S. & Parloff, M. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program – General effectiveness of treatments. *Archives of General Psychiatry, 46*, 971-982.

Elliott, R., Stiles, W. B, & Shapiro, D. A. (1993). Are some psychotherapies more equivalent than others? In T.R. Giles (Ed.), *Handbook of Effective Psychotherapy*. New York: Plenum Press.

Engels, G. I., Garnefski, N., & Diekstra, R. F. W. (1993). Efficacy of rational emotive therapy: A quantitative analysis. *Journal of Consulting and Clinical Psychology, 61*, 1083-1090.

Feske, U., & Chambless, D. L. (1995). Cognitive behavioral versus exposure only treatment for social phobia: A meta-analysis. *Behavior Therapy, 26*, 695-720.

Frank, J. D., & Frank J. (1991). *Persuasion and healing: A comparative study of psychotherapy*. (3rd ed.) Baltimore: John Hopkins University Press.

Grawe, K., Donati, R., & Bernauer, F. (1994). *Psychotherapie im wandel: Von der confession zur profession*. Göttingen: Hogrefe.

Grissom, R. J. (1996). The magical number .7 +/- .2: Meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control. *Journal of Consulting and Clinical Psychology, 64*, 973-982.

Henry, W. P. (1998). Science, politics, and the politics of science: The use and misuse of empirically validated treatment research. *Psychotherapy Research, 8*, 126-140.

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.

Imber, S. D., Pilkonis, P. A., Sotsky, S. M, Elkin, I., et al. (1990). Mode-specific effects among three treatments for depression. *Journal of Consulting and Clinical Psychology, 58*, 352-359.

Kazdin, A. (1986). Comparative Outcome Studies of Psychotherapy: Methodological Issues and Strategies. *Journal of Consulting and Clinical Psychology, 54*, 95-105.

Kazdin, A., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology, 57*, 138-147.

King, N. J. (1997). Empirically validated treatments and AACBT. *Behaviour Change, 14*, 2-5.

King, N. J., & Ollendick, T. H. (1998). Empirically validated treatments in clinical psychology. *Australian Psychologist, 33*, 89-95.

Lambert, M. S., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of Psychotherapy and Behavior Change* (pp. 143-189). New York: Wiley.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181-1209.

Luborsky, L. (in preparation). Clinical observer measures of psychological health-sickness.

Luborsky, L. (in press). The meaning of the new empirically supported treatment studies for psychoanalysis. In *Psychoanalytic Dialogues*. Special Editor, Jeremy Safran.

Luborsky, L. (1995). Are common factors across different psychotherapies the main explanation for the Dodo Bird verdict that "Everyone has won so all shall have prizes?" *Clinical Psychology: Science and Practice, 2*, 106-109.

Luborsky, L., & Barber, J.P. (1993). Benefits of adherence to psychotherapy manuals – and where to get them. In N. Miller L. Luborsky, J.P. Barber, & J. Docherty (Eds.), *Psychodynamic treatment research: A handbook for clinical practice* (pp. 211-226). New York: Basic Books.

Luborsky, L., Diguer, L., Luborsky, E., & Schmidt, K. A. (1999). The efficacy of dynamic versus other psychotherapies: Is it true that "everyone has won and all must have prizes?" – an update. In D. S. Janowsky (Ed.), *Psychotherapy Indications and Outcomes* (pp. 3-22). Washington, D.C.: American Psychiatric Press, Inc.

Luborsky, L., Diguer, L., Luborsky, E., Singer, B., & Dickter, D. (1993). The efficacy of dynamic psychotherapies. Is it true that everyone has won so all shall have prizes?' In N. Miller, L. Luborsky, J.P. Barber, & J. Docherty (Eds.), *Psychodynamic treatment research: A handbook for clinical practice* (pp. 447-514). New York: Basic Books.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Johnson, S., Halperin, G., Bishop, M., & Schweizer, E. (1999). The researcher's own therapeutic allegiances – A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*, 95-132.

Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that "Everyone has won and all must have prizes"? *Archives of General Psychiatry, 32*, 995-1008.

Luborsky, L., Woody, G. E., McLellan, A. T., O'Brien, C. P., & Rosenzweig, J. (1982). Can independent judges recognize different psychotherapies? An experience with manual-guided therapies. *Journal of Consulting and Clinical Psychology, 50*, 49-62.

Miller, R. C., & Berman, J. S. (1983). The efficacy of cognitive-behavior therapies: A quantitative review of the research evidence. *Psychological Bulletin, 94*, 39-53.

Nietzel, M. T., Russell, R. L., Hemmings, K. A., & Gretter, M. L. (1987). Clinical significance of psychotherapy for unipolar depression: A meta-analytic approach to social comparison. *Journal of Consulting and Clinical Psychology, 55*, 156-161.

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.

Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin, 108*, 30-49.

Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin, 113*, 553-565.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45*, 775-777.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park: SAGE Publications, Inc.

Rosenthal, R. (1995). Progress in clinical psychology: Is there any? *Clinical Psychology: Science and Practice, 2*, 133-150.

Rosenthal, R. (1998). Meta-analysis: Concepts, corollaries and controversies. In John Adair and David Belanger (Eds.), *Advances in psychological science Vol. 1: Social, personal and cultural aspects* (pp. 371-384). Hove, England, UK: Psychology Press/Erlbaum (UK) Taylor and Francis, xx, 580 pp.

Rosenthal, R., & DiMatteo, M. R. (in press). Meta-Analysis. In J. Wixted, (Ed.), *Stevens' Handbook of Experimental Psychology*, Volume 4 (Methodology), 3rd Edition, Wiley.

Rosenthal, R., Hiller, J. B., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (in press). Meta-analytic methods, the Rorschach, and the MMPI. *Psychological Assessment*.

Rosenthal, R., & Rubin, D. (1985). Statistical analysis: Summarizing evidence versus establishing facts. *Psychological Bulletin, 97*, 527-529.

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

Shadish, W., Matt, G., Navarro, A., Siegle, G., Crits-Christoph, P., Hazelrigg, M., Jorm, A., Lyons, L., Nietzel, M., Prout, H., Robinson, L., Smith, M. L., Svartberg, M., & Weiss, B. (1997). Evidence that psychotherapy works in clinically representative conditions. *Journal of Consulting and Clinical Psychology, 65*, 355-365.

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

Shapiro, D.A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin, 92*, 581-604.

Smith, M., & Glass, G. V. (1977). Meta-analyses of psychotherapy outcome studies. *American Psychologist, 32*, 752-760.

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

Strupp, H., & Hadley, S. (1979). Specific vs. non-specific factors in psychotherapy. *Archives of General Psychiatry, 36*, 1125-1136.

Svartberg, M., & Stiles, T. C. (1991). Comparative effects of short-term
psychodynamic psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology, 59*, 704-714.

Tschuschke, V., Banninger-Huber, E., Faller, H., Fikentscher, E., Fischer, G., Frohburg, I., Hager, W., Schiffler, A., Lamprecht, F., Leichsenring, F., Leuzinger-Bohleber, M., Rudolph, G., & Kachele, H. (1998). Psychotherapy research – how it should (not) be done. An expert reanalysis of comparative studies by Grawe et al. (1994). *Psychotherapie, Psychosomatik, Medizinische Psychologie, 48*, 430-444.

Tschuschke, V., & Kächele, H. (1996). What do psychotherapies achieve? A contribution to the debate centered around differential effects of different treatment concepts. In U. Esser, H. Pabst, G. W. Sperirer (Eds.) (p. 159-181). *The power of the person-centered approach*. GWG-Verlag Köln.

Van Balkom, A. J. L. M., Van Oppen, P., Vermeulen, A. W. A., & Van Dyck, R. (1994). A meta-analysis on the treatment of obsessive-compulsive disorder: A comparison of antidepressants, behavior, and cognitive therapy. *Clinical Psychology Review, 14*, 359-381.

Wampold, B. E. (1997). Methodological problems in identifying efficacious psychotherapies, *Psychotherapy Research, 7*, 21-43.

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*, 203-25.

Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist, 47*, 1578-1585.

Table 1
Meta-analyses of comparisons of effect sizes of active treatments (by Cohen's *d*)

Reports	Meta-analyses (n=17)	N of Studies	Effect Sizes (ES)	
			Uncorr.	Corrected[e]
1. Berman, Miller, & Massman, 1985[h]	Cog. vs. Desen.	20	.06	
2. Robinson, Berman et al., 1990	Cog. vs. Behav.	12	.12	.12
	Cog. vs. C-B	4	-.03	-.03
	Cog. vs. General Verbal[d]	7	.47*	-.15
	Behav. vs. C-B	8	-.24*	-.16
	Behav. vs. General Verbal	14	.27*	.15
	C-B vs. General Verbal	8	.37*	.09
3. Svartberg & Stiles[#], 1991	Dynamic[a] vs. C-B	6	-.47*	
	Dynamic[a] vs. Behav.	5	-.10	
	Dynamic[a] vs. Nonspecific	3	.29	
4. Crits-Christoph, 1992	Dynamic[b] vs. Non-psychiatric treatment	5	.32	
	Dynamic[b] vs. Psychiatric treatment	6	-.05	
5. Luborsky, Diguer, et al.[#], 1993	Dynamic[c] vs. Other	3	.00	-.01
6. Luborsky et al.[#], (1999)	Cog. vs. Behav.	9	.21	.22
	Dynamic[c] vs. Behav.	7	-.03	.14
	Dynamic[c] vs. Cog.	4	.02	.08
	Pharm. vs. Psychotherapy	9	-.41	-.20
Mean Effect Size (absolute value)			.20 (n=17) weighted[f]: .21	.12 (n=11) weighted: .14
Median Effect Size (absolute value)			.21 weighted[g]: .21	.14 weighted: .15

* $p < .05$

[#] The original Cohen's r for these studies was converted to Cohen's d (Cohen, 1977)

[a] Short-term psychodynamic psychotherapy (STPP)

[b] Brief dynamic psychotherapy (BDP)

[c] A variety of dynamic treatments

[d] General verbal therapy is comprised of treatments such as psychodynamic, client-centered, and other forms of interpersonal therapy. These treatments have in common a relatively greater emphasis on insight rather than on the acquisition of a set of specific skills.

[e] Corrected for researcher's allegiance (by a mean of the 3 measures, Luborsky et al., 1999); Study 5 was corrected for quality of research design.

[f] ES's weighted by sample size of each corresponding study.

[g] Weighted, as described by Rosenthal & DiMatteo (in press) and Rosenthal, Hiller, et al. (in press).

[h] Includes Miller and Berman 1983.

FOOTNOTES

[1] Given at a celebration on May 26, 2000 in honor of Saul Rosenzweig's many contributions, at Washington University's Department of Psychology, St. Louis, Missouri.

[2] We have used the usual wording "non-significant difference" rather than "equivalent" because the usual wording is usually more fitting. But, there are times when it is possible to imply an equivalence between two compared groups by following the method suggested by Rogers, Howard, & Vessey (1993). That suggested method reveals whether two groups are sufficiently similar to each other to be thought of as equivalent.

[3] Consider a study comparing T_1 vs. $Control_1$ that finds $d = .80$ and a study comparing T_2 vs. $Control_2$ that finds $d = .30$. We conclude T_1 is better than T_2 because a d of .80 is larger than a d of .30. However, if the study of T_2 had employed a much sicker population of patients, the smaller d is *not* due to a difference between treatments but to a difference between clienteles. A head-to-head comparison of T_1 vs. T_2 for a sample of patients for which both T_1 and T_2 would be appropriate might find no difference at all.

[4] To provide a more exact calibration of the largeness-smallness of comparison of active treatment versus active treatment, these methods can be used: 1) The measure can be compared with the overall effect of the treatment e.g. d=.85. 2) A second method employs the coefficient of robustness (mean d / S_d), an index of the clarity of the directionality of the results in relation to their homogeneity (Rosenthal, 1991).

[5] The use of the term "efficacious" will remind many readers of the becoming-popular distinction between "efficacy" and "effectiveness"-- this is essentially the supposed distinction between a research-context comparison of treatments (efficacy) with a clinic-context comparison of treatments (effectiveness). In the last six or seven years especially, the opinion has spread over the land that clinic-based treatment tends to be less effective than research-based treatment. The idea became even more prevalent after Weisz, et al, (1992) on the differential effectiveness of child psychotherapy under the two conditions reported that "clinic therapy" was far less effective than "research therapy." But, on the contrary, a much larger review, although still not very large, (Shadish, 1996) showed that "clinic therapy" performed reasonably well compared with "research therapy" and the same conclusion was reported in Shadish, Matt, Navarro, Siele, et al (1997).

Clinical Psychology: Science and Practice, 2002, 9, 1.

Let's Face Facts: Common Factors are More Potent than Specific Therapy Ingredients

Stanley B. Messer (*Rutgers University*) & Bruce E. Wampold (*University of Wisconsin-Madison*)

Address correspondence to Stanley B. Messer, GSAPP, Rutgers University, 152 Frelinghuysen Rd., Piscataway, NJ 08854, or to Bruce E. Wampold, Dept. of Counseling Psychology, School of Education, University of Wisconsin-Madison, 321 Education Building, 1000 Bascom Mall, Madison, Wisconsin, 53706.
Address email correspondence to smesser@rci.rutgers.edu or to wampold@macc.wisc.edu
Messer's Telephone: 732-445-2323, FAX: 732-445-4888

Abstract: Luborsky at al.'s findings of a non-significant effect size between the outcome of different therapies reinforces earlier meta-analyses demonstrating equivalence of *bonafide* treatments. Such results cast doubt on the power of the medical model of psychotherapy, which posits specific treatment effects for patients with specific diagnoses. Furthermore, studies of other features of this model--such as component (dismantling) approaches, adherence to a manual, or theoretically relevant interaction effects--have shown little support for it. The preponderance of evidence points to the widespread operation of common factors such as therapist-client alliance, therapist allegiance to a theoretical orientation and other therapist effects in determining treatment outcome. The commentary draws out the implications of these findings for psychotherapy research, practice and policy.
Key Words: common factors, specific ingredients, medical model

The effect size resulting from a meta-analysis that compares different therapies gives us greater confidence in a true difference existing between them than does any single study. By the same token, all the greater is our confidence in the results of a meta-analysis of meta-analyses, particularly if the results converge with previous estimates of the relative efficacy of various therapies. Luborsky *et al*. (in press) have compared active treatments and found a non-significant effect size of .20 based on 17 meta-analyses, which shrank further to .12 when corrected for researcher allegiance. It is of considerable interest that these results closely parallel those of Grissom (1996) who meta-analyzed 32 meta-analyses of comparative treatments, reporting an effect size of .23. In a recent meta-analysis also comparing active treatments, Wampold et al. (1997) found an effect size identical to that of Luborsky et al., namely, .20.

Despite their being small, non-significant effects, it should be noted that they are upper estimates or even overestimates of what is likely to be the true difference between pairs of active therapies. We believe this to be so for two reasons: First, these estimates are based on the absolute values of the differences between pairs of therapy, although some of the effect sizes are actually negative. To illustrate, consider therapies A and B whose true efficacy is identical. Suppose one study shows, due to sampling error, that A has a slight edge (say a nonsignificant difference equivalent to an effect size of d = .20) and another study finds, due to sampling error, that B has a slight edge (again, nonsignificant and an effect size d = -.20). Averaging the *absolute* values produces an effect size of .20, when the true effect size is zero. Absolute values do not take account of the sampling error that typically produces these small but opposite effects, which one would expect to occur even if the true difference were zero.

A second reason we regard the Luborsky et al. figure of .20 as an overestimate is that many of their comparisons are of cognitive or behavioral therapies with treatments described vaguely as "verbal therapies," "non-specific therapies," or "non-psychiatric treatment". Many of these treatments used as comparisons for cognitive or behavioral treatments were not meant to be therapeutic in the sense of an active treatment backed by theory, research, or clinical experience. That is, these treatments are not *bonafide*. At best, they were used to control for common factors and at worst, they were foils to establish the efficacy of a particular treatment. Wampold, Minami, Baskin, and Tierney (in press) meta-analyzed therapies for depression and found CBT to be superior to the non-cognitive and non-behavioral therapies until they separated these therapies into two groups: those that were *bonafide* treatments (i.e., treatments

supported by psychological theory and with a specified protocol) and those that were not (such as "supportive counseling" with no theoretical framework). It was shown that the superiority of CBT to these other therapies was an artifact of including non-*bonafide* therapies in the comparisons: CBT was not significantly more beneficial than non-cognitive and non-behavioral treatments that were intended to be therapeutic. Similarly, we take issue with the treatment groups presumably controlling for "common factors" in the studies of comparative psychotherapy effects meta-analyzed by Stevens, Hynan, and Allen (2000).

Study after study, meta-analysis after meta-analysis, and Luborsky et al.'s meta-meta-analysis have produced the same small or non-existent difference among therapies. Despite this record going back almost 25 years to the meta-analysis of therapy outcomes performed by Smith and Glass (1977), Luborsky et al. (in press), Crits-Christoph (1997) and others continue to hold out the possibility or even the likelihood of finding significant differences among the therapies. In a similar vein, Howard, Krause, Saunders and Kopta (1997) suggested that therapies should be ordered along an efficacy continuum. Luborsky et al.'s (in press) meta-analysis places us no closer to either goal than have the previous meta-analyses, precisely because the evidence points to all active therapies being equally beneficial.

Our argument is that there is much greater support in the literature for the efficacy of common factors as conceptualized by Rosenzweig (1936), Garfield (1995), and especially Jerome Frank (Frank & Frank, 1991) than there is for specific treatment effects on which, for example, the Empirically Supported Treatment (EST) movement depends. In the remainder of this commentary, we will briefly summarize the findings that, in our opinion, are a strong endorsement of the common factors view and constitute an indictment of the specific ingredients approach. For more comprehensive coverage of this evidence, see Wampold (2001) and for the argument against over-reliance on EST's, see Messer (2001).

Specific Effects

The medical model (on which ESTs are based) proposes that the specific ingredients characteristic of a theoretical approach are, in and of themselves, the important sources of psychotherapeutic effects. Although a common factors or "contextual" approach (Frank & Frank, 1991; Wampold, 2001) also regards specific ingredients as necessary to the conduct of therapy, the purpose of such interventions is viewed quite differently within this model. In the common factor or contextual model, the purpose of specific ingredients is to construct a coherent treatment in which therapists believe and that provides a convincing rationale to clients. Furthermore, these ingredients cannot be studied independently of the healing context and atmosphere in which they occur.

Component studies. Within the medical model, component studies are considered to be one of the most scientific designs for isolating factors that are critical to the success of psychotherapy. Ahn and Wampold (in press) conducted a meta-analysis of component studies that appeared in the literature between 1970 and 1998. An effect size was calculated by comparing the outcomes of treatment *with the purported active component* versus treatment *without the component*t. The aggregate effect size across the 27 studies was -. 20, indicating a trend in favor of the treatment without the component, but which was not statistically different from zero. That is, presumably effective components were not needed to produce the benefit of psychotherapeutic treatments, as the medical model would predict.

Adherence to a manual. Within the medical model, adherence to the manual is crucial because it assures that the specific ingredients described in the manual, and which are purportedly critical to the success of the treatment, are being delivered faithfully. Consequently, one should expect treatment delivered with manuals to be more beneficial to clients than treatments delivered without them. The meta-analytic evidence suggests that the use of manuals does not increase the benefits of psychotherapy. Moreover, treatments administered in clinically representative contexts are not inferior to treatments delivered in strictly controlled trials where adherence to treatment protocols is expected. Although the evidence regarding adherence to treatment protocols and outcome is mixed,

it appears that it is the structuring aspect of adherence rather than adherence to core theoretical ingredients that predicted outcome (Wampold, 2001). Moreover, the slavish adherence to treatment protocols appear to result in deterioration of the therapeutic relationship (e.g., Henry, Strupp, Butler, Schacht, & Binder, 1993).

Interaction effects. Interactions between treatments and characteristics of the clients that support the specificity of treatments have long been a cornerstone of the medical model of psychotherapy. However, in his review Wampold (2001) could not find one interaction effect theoretically derived from hypothesized client deficits (such as a biologically versus psychologically caused depression), casting further doubt on the specificity hypothesis. Although some interaction effects have been found in psychotherapy (e.g., Beutler & Clarkin, 1990), they are related to general personality or demographic features--not interactions that would be predicted by the specific ingredients of the treatment (e.g., medication versus cognitive therapy). From a common factors viewpoint, interactions are more likely to occur based on clients' belief in the rationale of treatment or from the consistency of clients' culture or worldview with those represented in the treatment. It would predict, for example, that a client who believes that unconscious factors play a large role in human behavior is more likely to do well in psychoanalytic therapy than someone who views social influence as paramount.

We now turn to the evidence related to the question of whether commonalties of treatment are responsible for outcomes.

General Effects

The Alliance-Outcome Relationship

The therapist-client alliance is probably the most familiar of the common factors and is truly pan-theoretical. There have been two meta-analyses of the alliance-outcome relationship. In the first, Horvath and Symonds (1991) reviewed 20 studies published between 1978 and 1990 and found a statistically significant correlation of .26 between alliance and outcome. This Pearson correlation is equivalent to a Cohen's d of .54, considered a medium-sized effect. Thus, 7% of the valiance in outcome is associated with alliance (compared, for example, to 1% [d of .20] for differences among treatments). In the second meta-analysis, Martin, Garske and Davis (2000) found a correlation of .22, equivalent to a d of .45. This is also a medium-sized effect, accounting for 5% of the variance in outcomes. Clearly, the relationship accounts for dramatically more variability in outcome than specific ingredients, (which account for roughly 0%).

Therapist and Researcher Allegiance

Allegiance refers to the degree to which the therapist delivering the treatment or the researcher studying it believes that the therapy is efficacious. Within a medical model, allegiance should not matter since the potency of the techniques is paramount, but it is central to the common factors model espoused by Frank (Frank & Frank, 1991). Compared to the upper bound of d = .20 for specific effects, Wampold (2001) found that therapist allegiance effects ranged up to d = .65 and Luborsky et al. (1999) found even larger researcher allegiance effects. Indeed, the latter authors found that almost 70% of the variability in effect sizes of treatment comparisons was due to allegiance. How odd it is then that we continue to examine the effect of different treatments (accounting for less than 1% of the variance) when a factor such as the allegiance of the researcher accounts for nearly 70% of the variance!

Therapist Effects

Because the medical model posits that specific ingredients are critical to the outcome of therapy, whether ingredients are administered to and received by the client are considered more important than the therapists who deliver them. By contrast, the contextual or common factors model puts more emphasis on variability in the manner in which therapies are delivered due to therapist skill and personality differences. In effect, the medical model says, "Seek the best treatment for your condition," whereas the contextual model advises, "Seek a good therapist who uses an approach you find compatible." As it turns out, therapists within a given treatment account for a fairly large proportion of the outcome variance (6 to 9%), lending support to the common factors model. (On the importance of the therapist, see also Bergin (1997) and Luborsky, McClellan, Diguer,Woody, & Seligman, 1997).

To summarize, common factors and therapist variability far outweigh specific ingredients in accounting for the benefits of psychotherapy. The proportion of variance contributed by common factors such as placebo effects, working alliance, therapist allegiance and competence are much greater than the variance stemming from specific ingredients or effects. The findings presented by Luborsky et al. (in press) have contributed in an important way to the evidence supporting the common factors versus specific ingredients model. We turn now to what follows from this conclusion.

Recommendations

Research

1. Limit clinical trials comparing *bonafide* therapies since such trials have largely have run their course. We know what the outcomes will be.
2. Focus on aspects of treatment that can explain the general effects or the unexplained variance in outcomes. For example, APA Division 17 (Counseling Psychology) is in the process of studying interventions that work that are not tied to a specific diagnosis (Wampold, Lichtenberg, & Waehler, in press), and Division 29 (Psychotherapy) has a task force reviewing therapist-client relationship qualities and therapist stances that move therapy forward (Norcross, 2000).
3. Decrease the emphasis on specific ingredients in manuals and increase the stress on common factors. Most helpful are psychotherapy books that present both general principles of practice and specific examples, rather than manualized techniques that must be adhered to rigidly in clinical trials.

Practice

1. Cease the unwarranted emphasis on ESTs. They are based on the medical model which has been found wanting, and wrongly leads to the discrediting of experiential, dynamic, family and other such treatments (Messer, 2001). The latter are not likely to differ in outcome from the approved EST's (most of which are behavioral or cognitive-behavioral).
2. Therapists should realize that specific ingredients are necessary but active only insofar as they are a component of a larger healing context of therapy. It is the meaning that the client gives to the experience of therapy that is important.
3. Because more variance is due to therapists than the nature of treatment, clients should seek the most competent therapist possible (which is often well known within a local community of practitioners), whose theoretical orientation is compatible with their own outlook, rather than choose a therapist strictly by expertise in ESTs.

Policy

As psychologists, we should try to influence public policy to support psychotherapy outside the medical system. In many respects the kinds of problems we treat and what constitutes healing or growth do not fit well within a medical model. Luborsky et al.'s (in press) results are a reminder of this fact. We do deal with suffering, and plenty of it--be it over loss, failing marriages, disturbed children, addictions, confused identities, and interpersonal or intrapsychic conflicts. Society has a large stake in terms of both public expenditures and quality-of-life in supporting psychotherapy which has a proven track record of alleviating human suffering but needs a new institutional framework within which to do its job.

References

Ahn, H., & Wampold, B. E. (in press). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. Journal of Counseling Psychology.

Bergin, A. E. (1997). Neglect of the therapist and human dimensions of change: A commentary. Clinical Psychology: Science and Practice, 4, 83-89.

Beutler, L. E., & Clarkin, J. (1990). Differential treatment selection: Toward targeted therapeutic interventions. New York: Brunner/Mazel.

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.

Frank, J. D., & Frank, J. B. (1991). Persuasion and healing: A comparative study of psychotherapy (3rd ed.). Baltimore: Johns Hopkins University Press.

Garfield, S. L. (1995). Psychotherapy: An eclectic-integrative approach. New York: Wiley & Sons.

Grissom, R. J. (1996). The magic number .7+-.2: Meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control. Journal of Consulting and Clinical Psychology, 64, 973-982.

Henry, W. P., Strupp, H. H., Butler, S. F., Schacht, T. E., & Binder, J. (1993). Effects of training in time-limited dynamic psychotherapy: Changes in therapist behavior. Journal of Consulting and Clinical Psychology, 61, 434-440.

Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. Journal of Counseling Psychology, 38, 139-149.

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

Luborsky, L., McClellan, A. T., Diguer, L., Woody, G., & Seligman, D. A. (1997). Clinical Psychology: Science and Practice, 4, 53-65.

Luborsky, L., Rosenthal, R., Diguer, Andrusyna, T. P., Berman, J. S., Levitt, J. T., Seligman, D. A., & Krause, E. D. (in press). The Dodo Bird verdict is alive and well--mostly. Clinical Psychology: Science and Practice.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic relation with outcome and other variables: A meta-analytic review. Journal of Consulting and Clinical Psychology, 68, 438-450.

Messer, S. B. (2001). Empirically supported treatments: What's a non-behaviorist to do? In Slife, B. D., Williams, R. N. & Barlow, S. H. (Eds.), Critical issues in psychotherapy: Translating new ideas into practice. Thousand Oaks, CA: Sage.

Norcross, J. C. (2000). Empirically supported therapeutic relationships: A Division 29 task force. Psychotherapy Bulletin, 35 (2), 2-4.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy: "At last the Dodo said, 'Everybody has won and all must have prizes.'" American Journal of Orthopsychiatry, 6, 412-415.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. American Psychologist, 32, 752-760.

Stevens, S. E., Hynan, M. T., & Allen, M. (2000). A meta-analysis of common factor and specific treatment effects across the outcome domains of the phase model of psychotherapy. Clinical Psychology: Science and Practice, 7, 273-290.

Wampold, B. E. (2001). The great psychotherapy debate: Models, methods and findings. Mahwah, NJ: Lawrence Erlbaum.

Wampold, B. E., Lichtenberg, J. W., & Waehler, C. A. (in press). Principles of empirically supported interventions in counseling psychology. The Counseling Psychologist.

Wampold, B. E., Minami, T., Baskin, T. W., and Tierney, S. C. (in press). A meta-(re) analysis of the effects of cognitive therapy versus "other therapies" for depression. Journal of Affective Disorders.

Wampold, B. E., Mondin, Moody, M., Stich, F., Benson, K, & Ahn, H. (1997). A meta-analysis of outcome studies comparing *bonafide* psychotherapies: Empirically, "all must have prizes." Psychological Bulletin, 122, 203-215.