

Flawed Meta-Analyses Comparing Psychotherapy With Pharmacotherapy

Donald F. Klein, M.D.

Objective: The author sought to illustrate the invalidity of meta-analyses that claim to quantitatively compare the benefits of psychotherapy to pharmacotherapy in patients with psychiatric disorders.

Method: Studies included in four meta-analyses were retrieved and their study designs evaluated.

Results: The meta-analyses compared effect sizes from disparate studies that were not uniformly blind, random, controlled, or of high quality. The studies did not directly address the comparative efficacy question or lacked assay sensitivity.

Conclusions: Numerous types of studies exemplify the need for caution in evaluating meta-analytic conclusions without first critically examining the included studies. Estimates of the relative efficacy of different treatments are not well founded when based almost exclusively on indirect, multiply confounded comparisons. Meta-analyses based on flawed studies or studies that lack demonstrated assay sensitivity are also inadequate for the criticism of treatment guidelines. Some bodies of data are inadequate to support a proper meta-analysis.

(*Am J Psychiatry* 2000; 157:1204–1211)

The comparative efficacy of different treatments is a matter of grave public health importance, relevant to important questions illuminating therapeutic mechanisms. Both patients and clinicians wish assurance that prescribed treatments are not just effective but the most effective of available treatments. Comparative information about cost, toxicity, necessary professional training, speed of onset, and likelihood of relapse are also necessary but are not dealt with here. These issues are meaningfully analyzed in APA's "Practice Guideline for Major Depressive Disorder in Adults" (1) and in the Agency for Healthcare Policy and Research's Clinical Practice Guideline: Depression in Primary Care, Volume 2 (2).

A particularly contentious area concerns the relative merits of pharmacotherapy and psychotherapy. This technical question in therapeutic evaluation is confounded by the regulatory fact that of the major providers of mental health care, only medical doctors are legally privileged to prescribe psychotropic medications. The practice of psychotherapy is not limited by any legal restrictions. The economic and ideological consequences of this issue may play a role in claims that concern the relative merits of these guild-related treatments (3). In particular, the aforementioned treatment guidelines have been criticized (4) for going beyond the facts by stating that medication is necessary in the treatment of severe depression.

Strictly speaking, however, motives are beside the point. Data and analyses must be met on their own ground. My concern is to demonstrate that unjustified conclusions about comparative therapeutic merit, derived from the

supposedly objective procedure of meta-analysis, rest on an almost total lack of directly relevant data.

Controversies over meta-analysis have led to an enormous technical literature defining correct practices. This article does not review these critical issues. The reader is referred to Chalmers (5) and Hunter and Schmidt (6) with regard to such essential issues as exhaustive breadth of literature search, quality of original research, sample heterogeneity, and publication bias.

Equivalent treatment samples are required for an unbiased estimate of treatment differences that cannot be attributed to baseline sample differences. If a drug appears more effective than placebo, the alternative hypothesis (i.e., that the drug-treated group had a better prognosis than the placebo group) must be shown to be unlikely. This is accomplished by randomization. Unbiased measurements and appropriate analyses are also necessary to disconfirm experimenter bias.

In contrast, meta-analyses amalgamate studies that may differ in elements such as samples, setting, measures, and therapists. Such differences may be ignored, or adjustments may be attempted by known covariates, such as sample size. This produces a naturalistic, nonexperimental, comparative estimate.

The language of effect sizes does not justify the assumption that multiply differing studies estimate a common causal parameter, despite the claim of Smith et al: "Respect for parsimony and good sense demands an acceptance of the notion that imperfect studies can converge on a true conclusion" (7). An effect size is usually calculated as the difference between the means of the experimental

and comparison groups after treatment, divided by either the pooled or comparison group standard deviation. (If treatment affects the experimental group's standard deviation, this would be obscured by a pooled estimate.) One can calculate effect sizes for each treatment, derived from trials with varying comparison groups, without differentiation. The comparison group may be historical, a wait list, a placebo, or an alternative treatment, and all may be evaluated, with or without randomization, blind or not blind, with different measures of varying reliability, reactivity, or sensitivity to particular treatments. One study may be of subjects at home, the other of psychiatric inpatients. Once an effect size is calculated, it becomes the meta-analytic raw data.

Naive readers assume meta-analyses produce reliable results because of the accumulated large sample size. However, when effect sizes derived from studies of one treatment are compared with effect sizes derived from different studies of the second treatment, there is no assurance of the validity of sample, setting, or treatment comparability.

The importance of quality assessment of the original studies entered into a meta-analysis has been emphasized by Moher et al. (8). However, Juni et al. (9) showed that global quality assessment, as a covariate, is ineffective: "Rather the relevant methodological aspects should be identified, ideally, a priori, and assessed individually."

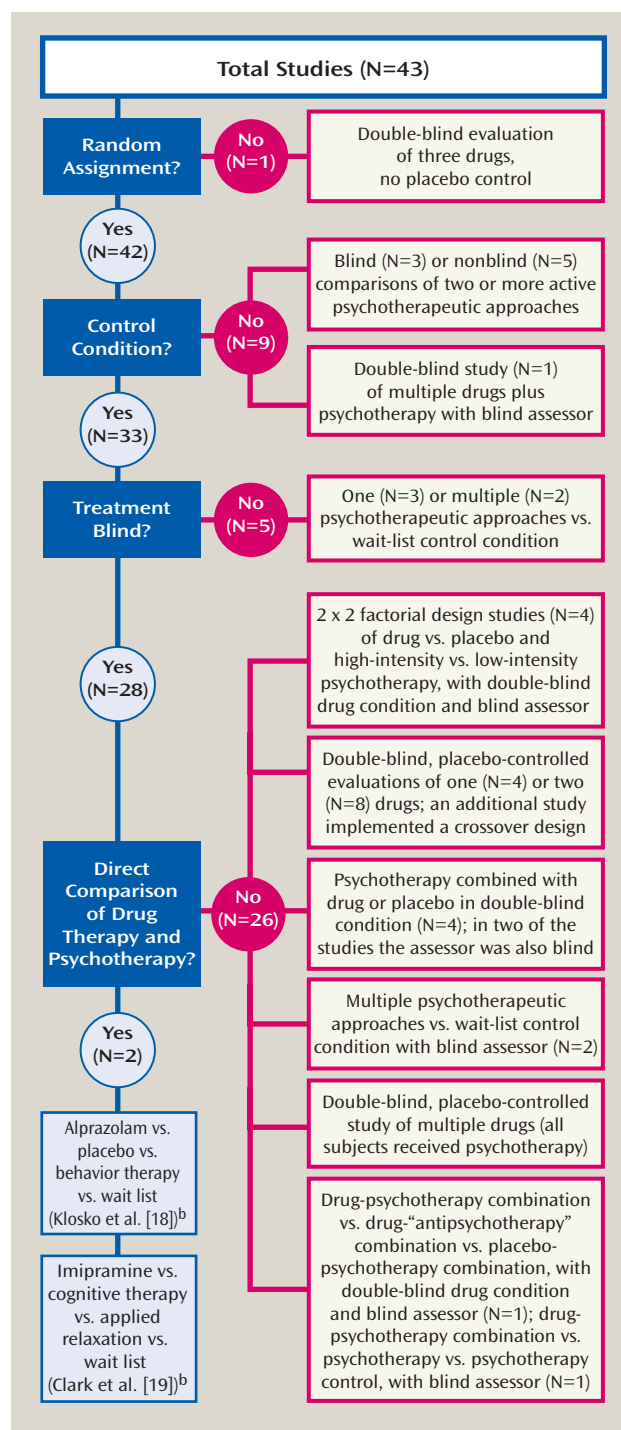
In preparing this article, the meta-analytic literature that compared pharmacotherapy with psychotherapy was reviewed. The examples selected are among the best because they provided the necessary detailed documentation. Their component articles were reviewed for "relevant methodological aspects."

Meta-Analysis of Panic Disorder Treatment: Cognitive Behavior Therapy Equivalent to Drug Therapy?

In 1995, Gould et al. (10) meta-analytically compared the effectiveness of pharmacotherapy, cognitive behavior therapy, and combinations of the two in the treatment of panic disorder. Of 43 cited studies, 14 evaluated only drug treatment, while 15 evaluated only psychotherapy. Twelve studies evaluated psychotherapy combined with pharmacotherapy but did not (and could not) produce a direct contrast. Furthermore, the studies differed greatly in design and sample size (Figure 1).

Gould et al. (10) calculated two effect sizes for each study. They first averaged all dependent measures within each study, thus combining studies that differed in number and kind of dependent variables. The second effect size used panic frequency as the sole outcome variable. Gould et al. concluded that pharmacological and behavior treatments both produce significantly better results than contrast conditions and that cognitive behavior therapy was at least as effective as pharmacotherapy.

FIGURE 1. Flaws in Studies Included in a Meta-Analysis of the Effectiveness of Drug Therapy Versus Cognitive Behavior Therapy for Panic Disorder^a



^a From Gould et al. (10); specific identification of studies available on request.

^b See text for critique of study.

Meta-analysts emphasize the need for complete literature retrieval. To reduce the possibility of a publication bias for positive results, the search should extend to unpublished studies. Gould et al. performed only a MED-

LINE database search that was limited to 20 years, from 1974 to March 1994. Many relevant studies were published before 1974. Furthermore, a 1993 randomized, controlled study of panic disorder by Black et al. (11) that evaluated the effects of fluvoxamine, cognitive therapy, or placebo was missed. The omission of this placebo-controlled, direct comparison study is particularly unfortunate. Black et al. concluded that fluvoxamine was significantly more effective than cognitive therapy, which did not differ significantly from the placebo-treated group on most measures, thus contradicting the findings of Gould et al. (10).

Furthermore, Gould et al. erroneously included several studies that did not meet their inclusion criteria. In 1986, Charney et al. (12) evaluated alprazolam, imipramine, and trazodone. Patients received a placebo first before being nonrandomly switched to a drug on the basis of "neurobiologic investigations of the mechanism of action of anti-panic and antidepressant drugs rather than on specific clinical characteristics" (p. 581). This study was neither randomized nor did it have a concurrent control group. Additionally, Gould et al. (10) included nine studies that lacked any contrast group, making it difficult to understand how a comparative effect size was calculated.

Four studies included by Gould et al. (13–16) developed four-cell designs that allowed estimates of drug (imipramine or alprazolam) versus placebo effects and more intensive versus less intensive psychotherapy. However, these designs do not allow any direct contrast of medication versus psychotherapy by use of a common placebo control. Although Marks et al. (13) found no specific imipramine benefit, a data reanalysis by Raskin (17) showed a substantial, specific, significant benefit of imipramine for panic and anxiety. Gould et al. (10) did not refer to this report.

Only two studies (5%) in the Gould et al. meta-analysis directly compared psychotherapy to pharmacotherapy, but even these direct comparisons were problematic. In 1990, Klosko et al. (18) randomly assigned patients—the majority of whom were diagnosed with panic disorder with no more than limited avoidance—to treatment with alprazolam, placebo, or behavior therapy; other patients were assigned to a wait-list control condition. Treatment with alprazolam and placebo was double-blind. An independent assessor was blind to all groups. The wait-list group had none of the hope-engendering effects of being in treatment and therefore substantially differs from a placebo group.

Behavior therapy and alprazolam treatment did not significantly differ in effectiveness, but both were better than the wait-list condition. However, alprazolam treatment was not statistically superior to placebo. Since alprazolam is effective in the treatment of panic disorder, this trial lacked assay sensitivity (i.e., the ability to detect specific treatment effects).

It is likely that this protocol underestimated alprazolam benefit, since after week 13, patients had their alprazolam treatment tapered over the next 2 weeks:

Psychiatrists continued meeting with subjects until they had stopped taking medication completely or, if they were unable to withdraw from medication, until they were restabilized on the study medication once again. At this point, treatment was considered to be over, and the psychiatrist completed the post study termination record....In fact, only one out of sixteen subjects withdrew completely from alprazolam. The remaining fifteen were quickly stabilized at or near their study dosage level.

No evidence is provided that stabilization occurred. Therefore, this study does not parallel clinically usual, uninterrupted alprazolam treatment. Behavior therapy was superior to both placebo and the wait-list condition with regard to producing zero panic but was superior only to the wait-list condition with regard to high-end state functioning.

In the only other study to directly compare drug therapy to psychotherapy, Clark et al. (19) randomly assigned patients with panic disorder to treatment with cognitive therapy, applied relaxation, or imipramine; others were assigned to a wait-list condition in which they remained for 3 months before random assignment to the three active-treatment groups.

Subjects had, at most, moderate agoraphobic avoidance and considered panic their main problem rather than phobic restrictions. Therefore, this group mainly consisted of pure panic patients. The authors justified their use of imipramine as having been repeatedly shown to be more effective than placebo, but that is not true for pure panic patients. In a prior study (20), when pure panic patients were isolated as a diagnostic stratum, neither imipramine nor alprazolam were superior to placebo. Pure panic patients are more prone to spontaneous remission than those complicated by agoraphobia and therefore likely to have a good, but nonspecific, outcome.

In the Clark et al. study (19), all three treatment groups significantly improved compared to the untreated patients in the wait-list condition, who were given no reason to expect improvement. At 3 months, cognitive therapy was superior to applied relaxation and imipramine; at 6 months, however, cognitive therapy and imipramine were indistinguishable but significantly better than the applied relaxation group. Cognitive therapy exhibited a somewhat quicker benefit than imipramine but only on scales that emphasized phobic restriction, not on panic measures. The lack of a placebo comparison group prevented a firm internal calibration that this sample was appropriate for achieving benefit from medication and therefore appropriate for a comparative trial. The sample restriction to largely pure panic disorder patients prevents generalizing even these insecure conclusions to the more usual agoraphobic patients.

Meta-Analysis of Panic Disorder Treatment: Psychotherapy Superior to Pharmacotherapy?

In 1989, Clum (21) meta-analytically compared the effectiveness of psychological interventions to drug therapy for the treatment of panic disorder by surveying 43 studies that differed greatly with regard to design, sample size, and diagnosis (Figure 2). Five of the included studies were presented only as abstracts at professional meetings, and the details could not be obtained. Nine studies evaluated only psychotherapy. Fifteen studies evaluated only drug therapy. Thirteen evaluated psychotherapy combined with drug therapy but did not compare the therapies. Only one study directly compared these treatments. This 1988 study by Klosko et al. (22) was identical to their 1990 study (18), which was criticized earlier in the discussion of the Gould et al. (10) meta-analysis.

Clum focused on improvement rates for panic attacks and symptoms. He computed improvement rates rather than an actual effect size to allow for inclusion of open, uncontrolled trials, although they are unlikely to have similar samples. Significant improvement was defined as either the absence of or a 50% reduction in panic attacks. With these criteria, Clum concluded that while treatment superiority was unclear, generalizations could be made that supported the use of psychotherapy over pharmacotherapy in the treatment of panic disorder.

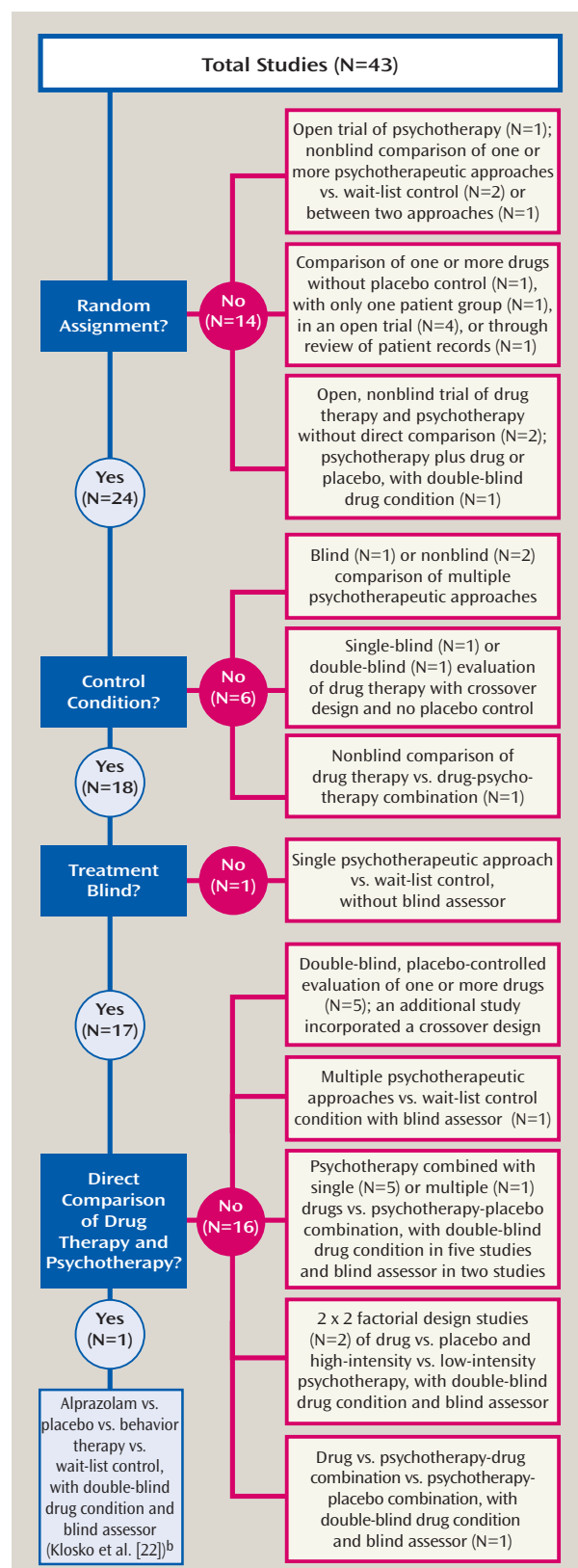
In 1993, Clum et al. (23) improved upon the meta-analysis from 1989 by not accepting open trials and requiring the presence of a control group in each study. They also calculated effect sizes rather than improvement rates. The 29 studies included in Clum et al. (23) were included in the previously described Gould et al. analysis (10). Although this meta-analysis was methodologically improved, nonetheless, seven of the included studies did not employ a control group. Only the aforementioned study of Klosko et al. (18) directly compared psychotherapy to pharmacotherapy.

Meta-Analysis of OCD Treatment: Behavior Therapy Superior to Antidepressants?

In 1994, van Balkom et al. (24) meta-analytically surveyed 86 studies of antidepressants and cognitive and behavior therapies in the treatment of obsessive-compulsive disorder (OCD) (Figure 3). Fifty-five studies evaluated only drug therapy, while 25 studies evaluated only psychotherapy. Five studies evaluated psychotherapy combined with drug therapy but without a direct comparison. Only one study directly compared psychotherapy with drug therapy. Studies within each treatment group differed on design, sample size (which ranged from eight to 255 subjects), specific treatment, and diagnosis.

For their meta-analysis, van Balkom et al. (24) used a non-comparative effect size. Within each treatment, posttest re-

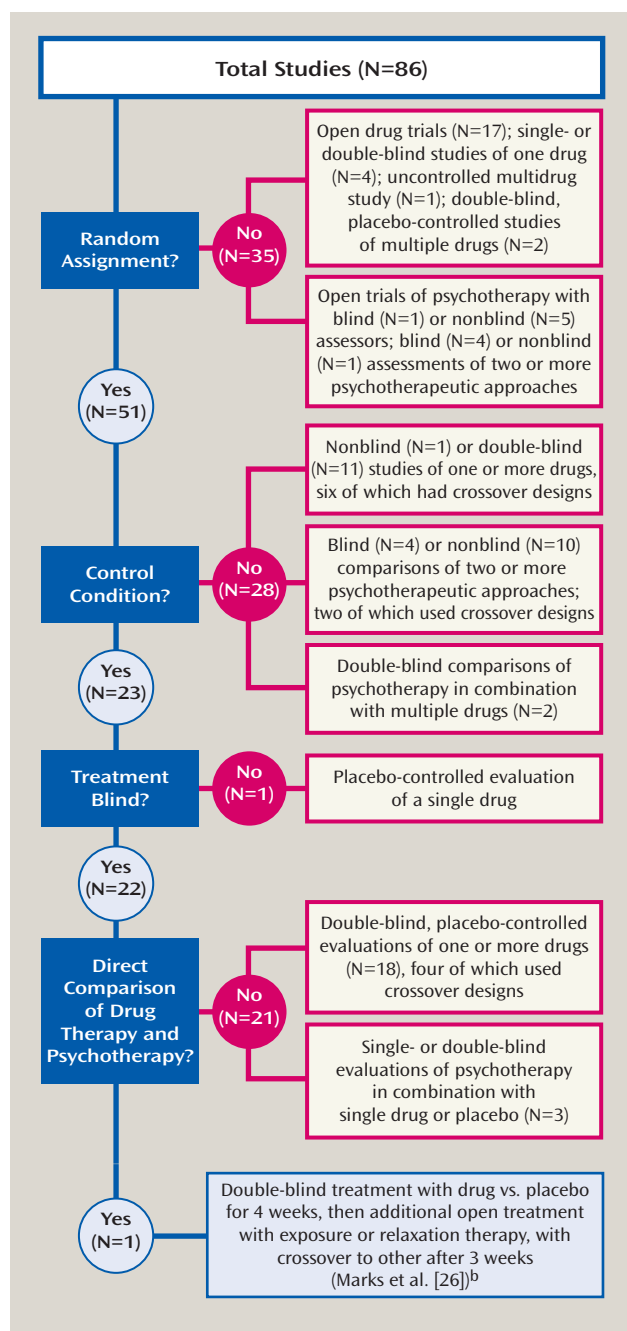
FIGURE 2. Flaws in Studies Included in a Meta-Analysis of the Effectiveness of Psychological Interventions Versus Drug Therapy for Panic Disorder^a



^a From Clum (21); specific identification of studies available on request.

^b See text for critique of study.

FIGURE 3. Flaws in a Meta-Analysis of the Effectiveness of Antidepressants Versus Cognitive and Behavior Therapy for Obsessive-Compulsive Disorder^a



^a From van Balkom et al. (24); specific identification of studies available on request.

^b See text for critique of study.

sults were subtracted from pretest results and divided by the pooled initial and final standard deviations. An average effect size was derived from within four clinical areas (obsessive-compulsive symptoms, depression, anxiety, and social adjustment), which produced four mean effect sizes per treatment condition. Such noncomparative effect sizes have been criticized as being vulnerable to several artifacts (25).

In the one study that attempted a comparison of pharmacotherapy and psychotherapy, Marks et al. (26) randomly assigned patients to treatment with clomipramine or placebo in a double-blind trial. After 4 weeks, patients were hospitalized for a second study phase, where in addition to their clomipramine/placebo treatment they received either exposure therapy or relaxation therapy for 3 weeks. After 3 weeks, patients who received relaxation therapy were switched to exposure therapy for an additional 3 weeks. This design prevents a simple contrast of pharmacotherapy and psychotherapy, since all drug patients also received psychotherapy. An interaction analysis allowed for some contextually dependent comparative estimates, but these were not significant.

Despite the lack of any significant interaction, Marks et al. (26) contrasted two particular cells to support the claim that exposure therapy improved rituals more than clomipramine treatment. This was not an a priori hypothesis and occurred in only two of eight comparisons. Therefore, these post hoc contrasts cannot be considered definitive (or persuasive). This data analysis illustrates the fallacy of averaging effect sizes, since after 7 weeks clomipramine appeared more effective for mood and social adjustment than exposure therapy, whereas exposure therapy seemed particularly effective for behavioral avoidance. These clinically useful distinctions are meta-analytically obscured by using the supposed common metric of a single effect size.

The conclusion of van Balkom et al. (24) was that behavior therapy is more effective for treatment of OCD than antidepressants, a finding that was based on self ratings but not on assessor ratings. Since only one of the 86 studies directly compared psychological with pharmacological interventions and showed no significant a priori distinctions between the two treatments, this conclusion goes far beyond the data.

“Mega-Analysis” of Depression Treatment: Practice Guidelines Promoting Unproven Practices?

DeRubeis et al. (4) addressed the particular issue that the treatment guidelines from APA and the Agency for Health Care Policy and Research recommend the use of antidepressant medication and not cognitive behavior therapy for more severe cases of depression. They argued that the narrow conclusions from these guidelines were based on one study and that “a meta-analysis that combines outcome data from additional sources would advance the debate.” Therefore, they cited three other studies that compared antidepressant medication to cognitive behavior therapy. DeRubeis et al. obtained each study’s raw data, allowing a “mega-analysis” of the subset of more severely depressed patients.

They concluded, “Until findings emerge from current or future comparative trials, antidepressant medication should not be considered, on the basis of empirical evi-

dence, to be superior to cognitive behavior therapy for the acute treatment of severely depressed outpatients.”

However, it is the experimental design that determines the meaningfulness of treatment comparisons. Assay sensitivity (i.e., the demonstration that a particular trial is able to detect a specific treatment effect, given such elements as the investigators, setting, measurement techniques, particular patient sample, and sample size) must be demonstrated. In a properly designed trial, the finding of treatment superiority to placebo fulfills this requirement.

Stating that treatments were not statistically different from each other in a given trial is a far cry from asserting equivalent benefit. This is the elementary statistical error known as affirming the null hypothesis, which has been thoroughly discussed in the context of those who wish to do away with placebo controls by simply comparing new agents with a standard agent. The problem is that in some samples, the standard agent is not efficacious. Indeed, many antidepressant trials, for differing reasons, cannot distinguish a so-called standard treatment from placebo. Therefore, the Food and Drug Administration, particularly in the relatively ill-defined area of depression, refuses to accept non-placebo-controlled trials as definitive (27, 28).

The National Institute of Mental Health Treatment of Depression Collaborative Research Program (29, 30) is the only study cited by DeRubeis et al. (4) that actually demonstrated assay sensitivity (31, 32) by the clear superiority of imipramine to placebo in the more severely depressed patients within this study. That imipramine was superior to cognitive therapy, coupled with the inability to distinguish cognitive therapy from placebo, furnishes the basis for stating that for patients with severe depression, imipramine has demonstrated specific effectiveness, whereas cognitive behavior therapy has not. The other three studies lacked a placebo control.

DeRubeis et al. (4) argued that “cognitive therapy has fared as well as antidepressant medication,” but this is illogical, since in the three studies that fail to detect a difference, all supposed treatment benefits may have been simply placebo or spontaneous remission effects. Apparent treatment equivalence may be an illusory artifact of lack of specific effect.

DeRubeis et al. (4) state,

The absence of placebo conditions in the studies by Rush et al., Murphy et al., and Hollon et al. leaves open the possibility that their groups were not responsive to antidepressant medication....But, just as differences between drug and placebo were obtained in the more severely depressed subgroup of the Treatment of Depression Collaborative Research Program, it is reasonable to expect that Rush et al., Murphy et al., and Hollon et al. would have obtained them as well, although this cannot be known with certainty.

However, this lack of certainty does not dissuade DeRubeis et al. (4) from rejecting these treatment guidelines.

Hollon (33), a co-author of two studies cited by DeRubeis et al. (4), presented critiques of Rush et al. (34) and Blackburn et al. (35) that DeRubeis et al. did not cite:

Early studies that found cognitive therapy (CT) superior to drugs in the treatment of depression have been criticized by myself and others for failing adequately to implement pharmacotherapy (Hollon et al. 1991; Meterissian and Bradwejn 1989). In the first of these studies, Rush and colleagues found CT superior to imipramine pharmacotherapy in terms of acute response (Rush et al. 1977). However, in that study, pharmacotherapy was provided by inexperienced psychiatric residents, medication doses were less than optimal (no more than 250 mg/day), no plasma monitoring was conducted to check on compliance and absorption, and drug withdrawal was begun two weeks before the end of treatment. This was hardly adequate pharmacotherapy.

During this 12-week study the last 2 weeks were used to taper off and then discontinue medication before evaluation. In a footnote, Rush et al. (34) pointed out that by week 10, the superiority of cognitive therapy, measured by the Beck Depression Inventory (BDI), was only significant at the 0.15 level. These authors clearly state “between weeks 10 and 12 the mean BDI scores of the pharmacotherapy group increased by 2.43, while the mean BDI score of the cognitive therapy group decreased by 2.17....A question may be raised as to whether the level of improvement in some of the pharmacotherapy patients reflects the reduction of drug dosage....Future studies should maintain the pharmacotherapy patients on full dosage until the study is completed.” Also, the reactivity of the self-rated, attitude-heavy Beck Depression Inventory to the procedures of cognitive therapy is not considered. The more medication-relevant Hamilton scale is not discussed in this footnote. Nonetheless, Rush et al. (34) concluded that “cognitive therapy results in significantly greater improvement than did pharmacotherapy.” This supposed differential benefit of cognitive therapy was therefore included in the DeRubeis et al. “mega-analysis.”

Hollon (33) stated,

Even to an advocate like myself, it is obvious that neither of these studies appeared to provide consistently adequate pharmacotherapy, something that was evident despite the fact that neither incorporated a pill-placebo control....I think Klein (1996) goes too far in dismissing these studies out of hand, but I think he is correct when he suggests that inclusion of pill-placebo controls would have made their findings far easier to interpret.

DeRubeis et al. (4) did not cite any studies that have shown cognitive therapy to be superior to a credible placebo in the treatment of severe depression, much less studies, calibrated by demonstration of a superior effect of a standard medication to placebo, that also show a similar superiority for cognitive therapy. Those who promulgate a treatment have the responsibility for substantiating its specific worth.

DeRubeis et al. (4) did not take into account the study by Stewart et al. (36), which showed that within the Treatment of Depression Collaborative Research Program sample was a subsample with atypical depressive features. Such patients respond poorly to imipramine (although well to monoamine oxidase inhibitors). In this atypical depressive subset, cognitive therapy significantly outperformed imipramine but did not outperform placebo. The reason is that imipramine actually had a numerically worse outcome than placebo in this atypical depressed subsample. On the other hand, among the nonatypical patients, the specific positive benefit of imipramine was enhanced.

Even apparently well-done outpatient studies that compared a tricyclic antidepressant to cognitive therapy (e.g., Murphy et al. [37] and Hollon et al. [38]) may have included enough subjects with atypical depression to have sabotaged tricyclic effectiveness. Without a placebo control, this unexpected problem is undetectable. Therefore, it cannot be asserted, as DeRubeis et al. (4) have done, that it is "reasonable" to expect equivalent specific drug benefit.

In their "mega-analysis," DeRubeis et al. (4) did not note the flawed nature of the cited data, the existence of relevant reanalyses, or the irrelevance of placebo-free experimental designs to claims for equivalent efficacy. Their conclusions are incorrect and harmful.

Discussion

Meta-analysis is useful for summing small effect sizes derived from low-power—but high quality—studies that have documented equivalence in samples, diagnostic procedures, comorbidity, randomized treatment assignment, reliable and valid measures with equivalent reactivity, comparable settings, skilled therapists, an internally calibrating placebo control demonstrating assay sensitivity, skilled independent assessors, detection of protocol deviation, equivalent investigator allegiances, and lack of differential attrition. Such studies may cumulatively indicate a significant benefit that cannot be established by each small, nonpowerful study.

But if one pools, precautions are needed against being fooled. The meta-analyses presented here show that elementary precautions have been ignored. I have not made the general case that all meta-analyses regarding comparative therapies for the mentally ill are irreparably compromised, although I have earnestly and fruitlessly looked for a model example. These flawed examples should be kept in mind so that sweeping meta-analytic conclusions are greeted skeptically.

It is an enormous problem that peer reviewers of large meta-analyses rarely, if ever, take the time and trouble to recapture the original documents for reassessment. This is extremely tedious, time-consuming work that far exceeds the capacity and tolerance of most voluntary, pro bono peer reviewers. However, DeRubeis et al. (4) reviewed only four studies, making it apparent that their peer reviewers did not understand the relevant methodological issues.

Also, when there are a large number of cited studies, critiquing a random subsample would make the burden manageable. Scholarly journals that wish to publish large meta-analyses should do this preliminary random selection and conceptually relevant tabulation of included studies so as to make real peer review actually possible rather than an illusion.

The thrust of this critique is that judgments about comparative therapeutic efficacy require direct experimental contrasts and cannot be manufactured by statistical legerdemain. It is beyond this paper to provide comprehensive guidelines for meta-analyses. However, one necessary criterion is that there should be no attempt to develop comparative therapeutic evaluations unless the therapies have actually been validly compared, within assay-sensitive randomized studies (27, 28). This elementary criterion would prevent the publication of such flawed meta-analytic efforts without the need for retrieving original studies. The Cochrane Collaboration (39) fosters evidence-based medicine by a meta-analytic program based on controlled, randomized, comparative trials. Unfortunately, the currently available data simply do not allow for a proper meta-analysis that can address the relative merits of pharmacotherapy and psychotherapy.

In the area of therapeutics, partial and biased readings of the evidence are common. The popularity of meta-analysis poses a danger to comprehending the necessity for randomized, controlled experimental research, if valid comparative judgments are supposedly attainable by non-experimental methods. Confounding treatment effects with sample and other local characteristics is the overriding, likely possibility that led to recognition of the importance of randomization and direct, properly controlled comparisons in the first place.

It would advance the field if articles favoring particular treatments were published in debate format (40). Further, Hollon (33) and I agree that comparative treatment studies, conducted jointly by the collaboration of experts in each technique, at each site, are necessary for solid therapeutic evidence. Unfortunately, few such comparative studies are on the horizon (41).

Received Feb. 8, 1999; revisions received Aug. 19, 1999, and Jan. 19, 2000; accepted Jan. 27, 2000. From the Department of Therapeutics, New York State Psychiatric Institute. Address reprint requests to Dr. Klein, Department of Therapeutics, New York State Psychiatric Institute, Unit 22, 1051 Riverside Dr., New York, NY 10032; donaldk737@aol.com (e-mail).

Supported in part by an NIMH Clinical Research Center grant (MH-30906) to the New York State Psychiatric Institute.

The author thanks Cynthia Steppling Lah for her help in the research and preparation of this manuscript.

References

1. American Psychiatric Association: Practice Guideline for Major Depressive Disorder in Adults. *Am J Psychiatry* 1993; 150(April suppl)

2. Agency for Health Care Policy and Research: Clinical Practice Guideline: Depression in Primary Care, vol 2: Treatment of Major Depression: AHCPR Publication 93-0551. Washington, DC, US Government Printing Office, 1993
3. DeNelsky GY: The case against prescription privileges for psychologists. *Am Psychol* 1996; 51:207-212
4. DeRubeis RJ, Gelfand LA, Tang TZ, Simons AD: Medications versus cognitive behavior therapy for severely depressed outpatients: mega-analysis of four randomized comparisons. *Am J Psychiatry* 1999; 156:1007-1013
5. Chalmers TC: Problems induced by meta-analyses. *Stat Med* 1991; 10:971-980
6. Hunter JE, Schmidt FL: *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, Calif, Sage Publications, 1990
7. Smith ML, Glass JV, Miller TI: *The Benefits of Psychotherapy*. Baltimore, Johns Hopkins University Press, 1980
8. Moher D, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Claussen TP: Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352:609-613
9. Juni P, Witschi A, Bloch R, Egger M: The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282:1054-1060
10. Gould RA, Otto MW, Pollack MH: A meta-analysis of treatment outcome for panic disorder. *Clin Psychol Rev* 1995; 15:819-844
11. Black DW, Wesner R, Bowers W, Gabel J: A comparison of fluvoxamine, cognitive therapy, and placebo in the treatment of panic disorder. *Arch Gen Psychiatry* 1993; 50:44-50
12. Charney DS, Woods SW, Goodman WK, Rifkin B, Kinch M, Aiken B, Quadri LM, Heninger GR: Drug treatment of panic disorder: the comparative efficacy of imipramine, alprazolam, and trazodone. *J Clin Psychiatry* 1986; 47:580-586
13. Marks IM, Gray S, Cohen D, Hill R, Mawson D, Ramm E, Stern RS: Imipramine and brief therapist-aided exposure in agoraphobics having self-exposure homework. *Arch Gen Psychiatry* 1983; 40:153-162
14. Marks IM, Swinson RP, Basoglu M, Kuch K, Noshivani H, O'Sullivan G, Lelliot PT, Kirby M, McNamee G, Sengun S, Wickwire K: Alprazolam and exposure alone and combined in panic disorder with agoraphobia. *Br J Psychiatry* 1993; 162:776-787
15. Mavissakalian M, Michelson L: Self-directed in vivo exposure practice in behavioral and pharmacological treatments of agoraphobia. *Behavior Therapy* 1983; 14:506-519
16. Mavissakalian M, Michelson L: Agoraphobia: relative and combined effectiveness of therapist-assisted in vivo exposure and imipramine. *J Clin Psychiatry* 1986; 47:117-122
17. Raskin A: Role of depression in the antipanic effects of antidepressant drugs, in *Clinical Aspects of Panic Disorder*. Edited by Ballenger JC. New York, Wiley-Liss, 1990, pp 169-180
18. Klosko JS, Barlow DH, Tassinari R, Cherny JA: A comparison of alprazolam and behavior therapy in the treatment of panic disorder. *J Consult Clin Psychol* 1990; 58:77-84
19. Clark DM, Salkovskis PM, Hackmann A, Middleton H, Pavlos A, Gelder M: A comparison of cognitive therapy, applied relaxation and imipramine in the treatment of panic disorder. *Br J Psychiatry* 1994; 164:759-769
20. Cross-National Collaborative Panic Study: Drug treatment of panic: comparative efficacy of alprazolam, imipramine, and placebo. *Br J Psychiatry* 1992; 160:191-202
21. Clum GA: Psychological interventions vs drugs in the treatment of panic. *Behavior Therapy* 1989; 20:429-457
22. Klosko JS, Barlow DH, Tassinari R, Cherny JA: Comparison of alprazolam and cognitive behavior therapy in the treatment of panic disorder: a preliminary report, in *Panic and Phobias: Treatment and Variables Affecting Course and Outcome*. Edited by Hand I, Wittchen HU. Berlin, Springer-Verlag, 1988, pp 54-65
23. Clum GA, Clum GA, Surls R: A meta-analysis for panic disorder. *J Consult Clin Psychol* 1993; 61:317-332
24. van Balkom AJLM, van Oppen P, Vermeulen AWA, van Dyck R, Nauta MCE, Vorst HCM: A meta-analysis on the treatment of obsessive compulsive disorder: a comparison of antidepressants, behavior, and cognitive therapy. *Clin Psychol Rev* 1994; 14:359-381
25. Klein DF: Listening to meta-analysis but hearing bias. *Prevention & Treatment*, vol 1, article 0006c. American Psychological Association, posted June 26, 1998. <http://journals.apa.org/prevention/volume1/pre0010006c.html>
26. Marks IM, Stern RS, Mawson D, Cobb J, McDonald R: Clomipramine and exposure for obsessive-compulsive rituals, I. *Br J Psychiatry* 1980; 136:1-25
27. Laska EM, Klein DF, Lavori PW, Levine J, Robinson DS: Design issues for the clinical evaluation of psychotropic drugs, in *Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines*. Edited by Prien RF, Robinson DS. New York, Raven Press, 1994, pp 29-67
28. Klein DF: Response to Rothman and Michels on placebo-controlled clinical trials. *Psychiatr Annals* 1995; 25:401-403
29. Elkin I, Gibbons RD, Shea MT, Sotsky SM, Watkins HT, Pilkonis PA, Hedeker D: Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol* 1995; 63:841-847
30. Klein DF, Ross DC: Reanalysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program General Effectiveness Report. *Neuropsychopharmacology* 1993; 8:241-252
31. Leber P: Is there an alternative to the randomized controlled trial? *Psychopharmacol Bull* 1991; 27:3-8
32. Temple RJ: When are clinical trial of a given agent vs placebo no longer appropriate or feasible? *Control Clin Trials* 1997; 18:613-620; discussion, 18:661-666
33. Hollon SD: The relative efficacy of drugs and psychotherapy; methodological considerations, in *Psychotherapy Indications and Outcomes*. Edited by Janowsky DS. Washington, DC, American Psychiatric Press, 1999, pp 343-365
34. Rush AJ, Beck AT, Kovacs M, Hollon S: Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Therapy and Res* 1977; 1:17-37
35. Blackburn IM, Bishop S, Glen AIM, Whalley LJ, Christie JE: The efficacy of cognitive therapy in depression: a treatment trial using cognitive therapy and pharmacotherapy, each alone and in combination. *Br J Psychiatry* 1981; 139:181-189
36. Stewart JW, Garfinkel R, Nunes EV, Donovan S, Klein DF: Atypical features and treatment response in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Clin Psychopharmacol* 1998; 18:429-434
37. Murphy G, Simons AD, Wetzel RD, Lustman PJ: Cognitive therapy and pharmacotherapy. *Arch Gen Psychiatry* 1984; 44:33-41
38. Hollon SD, DeRubeis RJ, Evans MD, Wiemer MJ, Garvey MJ, Grove WM, Tuason VB: Cognitive therapy and pharmacotherapy for depression: singly and in combination. *Arch Gen Psychiatry* 1992; 49:774-781
39. The Cochrane Collaboration. <http://hiru.mcmaster.ca/cochrane/default.htm>
40. Klein DF: Critiquing McNally's reply. *Behav Res Ther* 1996; 34:859-863
41. Klein DF: Panic and phobic anxiety: phenotypes, endophenotypes, and genotypes (editorial). *Am J Psychiatry* 1998; 155:1147-1149