Journal of Consulting and Clinical Psychology
February 1996 Vol. 64, No. 1, 81-87

Preventing Hung Juries About Therapy Studies

Donald F. Klein
Columbia University

ABSTRACT

A series of arguments are presented that emphasize the importance of comparatively evaluating psychotherapies with appropriate pharmacotherapy and pill placebo. The lack of a pill-placebo arm has rendered moot those studies that compared pharmacotherapy directly with psychotherapy because of the lack of an internal sample defining calibration with regard to medication responsivity. The National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program is critically discussed because this program incorporated the recommended design features but still led to substantial controversy. The inclusion of a pill-placebo arm in future therapeutic studies is most desirable. NIMH should initiate a funded program specifically for multisite, pill-placebo-controlled studies of psychotherapy, pharmacotherapy, and their combination, jointly sponsored and supervised by skilled psychopharmacologists and psychotherapists.

This research is based on the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program Public Use Data Tape and Documentation. The NIMH Treatment of Depression Collaborative Research Program is a multisite program initiated and sponsored by the Psychosocial Treatments Research Branch, Division of Extramural Research Programs (now part of the Mood, Anxiety, and Personality Disorders Research Branch, Division of Clinical Research), NIMH. The program was funded by Cooperative Agreements with six participating sites: George Washington University (MH 33762), University of Pittsburgh (MH 33753), University of Oklahoma (MH 33760), Yale University (MH 33827), Clark Institute of Psychiatry (MH 38231), and Rush Presbyterian–St. Luke's Medical Center (MH 35017). The principal NIMH collaborators were Irene Elkin, Coordinator (now at the University of Chicago); M. Tracie Shea, Associate Coordinator (now at Brown University); John P. Docherty, (now at Nashua Brookside Hospital); and Morris B. Parloff, (now at Georgetown University). The principal investigators and project coordinators at the three participating research sites were Stuart M. Sotsky and David Glass, George Washington University; Stanley D. Imber and Paul A. Pilkonis, University of Pittsburgh; and John T. Watkins (now at the Atlanta Center for Cognitive Therapy) and William Leber, University of Oklahoma. The principal investigators and project coordinators at the three sites responsible for training therapists were as follows: Myrna Weissman (now at Columbia University), Eve Chevron, and Bruce J. Rousanville, Yale University; Brian F. Shaw (now at Toronto Hospital, University of Toronto) and T. Michael Vallis (now at Dalhousie University), Clarke Institute of Psychiatry; and Jan A. Fawcett and Phillip Epstein, Rush Presbyterian–St. Luke's Medical Center. Collaborators in the data management and data analysis aspects of the program were C. James Klett, Joseph F. Collins, and Roderic Gillis of the Veterans Administration Cooperative Studies Program, Perry Point, Maryland.
I acknowledge the help of Donald Ross, Robin Garfinkel, and Michael Liebowitz.
Correspondence may be addressed to Donald F. Klein, New York State Psychiatric Institute, Unit 22, Columbia University, 722 West 168th Street, New York, New York, 10032.
Received: June 29, 1995
Revised: August 30, 1995

Accepted: September 6, 1995

In this issue, Jacobson and Hollon state, "It is all too easy to fall into the trap of interpreting data selectively in the service of an a priori position" (1996, p. 74). This is certainly true. As Abelson (1995) stated:

> If a result of a study is contrary to prior beliefs, the strongest holders of those prior beliefs will tend to marshal various criticisms of the study's methodology, come up with alternative interpretations of the results, and spark a possibly long-lasting debate (p. 11).

I hope that this discussion will help the field focus on the data and designs necessary to allow thoughtful, independent assessments of psychotherapy and pharmacotherapy efficacy.

With regard to the Treatment of Depression Collaborative Research Program (TDCRP), Jacobson and Hollon summarized that "Klein was right all along," and that "differences favoring pharmacotherapy over CBT [cognitive—behavioral treatment] were consistent across measures and represented more than random fluctuations in the data" (1996, p. 75). Perhaps I should quit here but there is much more to address, so, not so fast! I remind the reader that the TDCRP study concluded as follows.

> Thus, there is no evidence in the major analyses that either of the psychotherapies was inferior to the standard reference treatment, at termination of treatment on measures of depressive symptoms or general functioning. These statistical analyses did not, of course, permit the inference that the psychotherapies and the standard reference treatment were "equal" in effectiveness. However, since we had satisfactory power in these analyses for detecting large effect size differences between pairs of treatment (in the total unstratified sample), it is unlikely that very large or important differences were missed. (Elkin et al., 1989, p. 977)

Despite Elkin et al.'s disclaimer that they were not asserting that the treatments were equal, they felt that nothing important had been missed. Therefore, as a practical matter, one could choose one treatment as well as the other. Their statement fostered the interpretation that medication and psychotherapy were equivalent for depression, as evidenced by a (still-unretracted) front page article in the *New York Times* (Psychotherapy is as Good as Drug, 1986). It is these scientifically and socially misleading conclusions that require forthright criticism. However, it was only with the belated release of the final public access data tape by the National Institute of Mental Health (NIMH; in late 1994) that these analyses became possible. Data collection was completed in 1984.

Jacobson and Hollon (1996) cite an article by Elkin et al. (1996) , that reportedly supports the views expressed in this article concerning the superiority of medication over psychotherapy and placebo. However, the concern exists that several previously criticized issues still may not have been adequately addressed. I do not know whether Elkin et al. still require a Bonferroni correction for multiple treatment comparisons. I have argued that a per comparison Type I error rate rather than a per experiment rate is more appropriate for treatment trials, because each comparison is meaningful. If Elkin et al. no longer use a Bonferroni correction, should the change be reconciled with past practices? Elsewhere, it has been argued that the analysis of covariance model is impugned by less stringent criteria for slope heterogeneity than Elkin et al. have used (Klein & Ross, 1993). This critique also deserves comment.

I also wonder if the TDCRP has addressed Jan Fawcett's public statement (cited in Klein & Ross, 1993) that pharmacotherapy and case management were not optimally conducted. It is important to determine

how closely treatments of all types correspond with manualized directives, an issue still unaddressed in the TDCRP to my knowledge. Because the random regression model allows missing data to be modeled on the basis of observable data, how is the possibility of differential dropout addressed in their analysis? Are site interactions included in the new analyses? How are the new findings reconciled regarding medication efficacy, with past claims that the power was sufficient to ensure that no substantial effect was likely to have been missed?

Heuristically, it is important to note that Jacobson and Hollon have incompletely presented the results of the Klein and Ross (1993) reanalysis. They fail to assert the finding that CBT was not only worse than medication but that it was never superior to pill placebo—case management (PLA-CM) at the .05 level. Thus, it is not simply that imipramine is better, faster, and cheaper than CBT, but that the whole basis for the belief that cognitive psychotherapy is doing anything specific has been placed in jeopardy. Furthermore, although interpersonal psychotherapy (IPT) did somewhat better, the TDCRP did not give it a rousing vote of confidence because late substantial effects were found on only one dependent measure, the Hamilton Rating Scale for Depression (HRSD). If a psychotherapy cannot beat pill placebo—case management, it cannot be terrific.

Do Psychotherapies Differ?

Jacobson and Hollon (1996) also do not cite one of the more provocative reanalyses of Klein and Ross (1993) . As is well known, it is remarkably hard to find differences between the outcomes of credible psychotherapies or any evidence that a proposed specific beneficial mechanism of action has anything to do with therapeutic outcome. For instance, Elkin et al. (1989) found no overall differences between CBT and IPT.

However, Klein and Ross (1993) found, by Johnson-Neyman analysis, that for patients whose Beck Depression Inventory (BDI) score exceeded 30, generally considered the boundary between moderate and severe depression, CBT was substantially inferior to IPT. Counterintuitively, in just those patients for whom one might think that CBT was particularly indicated, it seems that one should actually advise an alternative psychotherapy.

That this is not a simple fluke is made less likely by the symmetrical finding by Sotsky et al. (1991) that only among patients with low social dysfunction was IPT significantly better than placebo. However, in the high social dysfunction group, IPT ranked the worst of the four treatments. Sotsky et al. also found CBT better than placebo—case management only in the low cognitive dysfunction group. There is a problem with the Sotsky et al. analysis in that they first cull out promising predictors before entering them into the main analysis. This renders their *p* values questionable, and these authors do "consider these analyses exploratory in nature" (1991, p. 1005).

These findings (tentative and unreplicated as they are) are inexplicable on the basis of the therapeutic action theories propounded by the creators of IPT and CBT. However, they are entirely compatible with the hypothesis (championed by Jerome Frank; see Frank & Frank, 1991) that psychotherapies are not doing anything specific; rather, they are nonspecifically beneficial to the final common pathway of demoralization, to the degree that they are effective at all. Presumably, this antidemoralization effect is initiated by having a strong, knowledgeable, professional ally who therapeutically provides the patient with emotional support, usable coping skills, and success experiences and helps reframe life experiences so as to heighten self-esteem.

However, what would happen to the patients with marked interpersonal difficulties if the therapist repeatedly leads them to focus on just the areas where they consistently fail and implicitly suggests that,

unless their interpersonal tactics improve, their depression will persist? Analogously, what about the dysfunctional perfectionist faced with a cognitive therapist who consistently reminds the patient that his or her perfectionism is self-defeating and unwarranted and that if it persists, the future is gloomy indeed? Such patient—therapy mismatches should promote demoralization. There may be specific differences between therapies but unfortunately, in this sample, they seem to be due to a toxicity that sharply limits their range of application to the less ill. There may be much to be said for the old-fashioned, nondirective, Rogerian style, as Shear, Pilkonis, Cloitre, and Leon (1994) showed in a panic disorder study. This is another researchable question.

Replications are the soul of scientific advance. However, there will be no replications if unwelcome, but theoretically meaningful, findings are not even cited, even in such a well-balanced review as that by Jacobson and Hollon.

Is Site Important?

Jacobson and Hollon (1996) cite Elkin et al.:

> ...patients receiving CBT at one site did extremely well and had mean scores very similar to those for patients receiving imipramine-CM, and the same was true for patients receiving IPT at another site ... until we unravel these findings, final judgment must be withheld about the specific effectiveness of the two psychotherapies with more severely depressed and impaired patients. ... (1989, p. 980)

Jacobson and Hollon wish that the TDCRP investigators had pursued the site difference issue more energetically, and that they, as well as Klein and Ross (1993) , should have tempered their conclusions. However, Klein and Ross (1993) performed site analyses using a posttreatment measure as a dependent variable, a pretreatment measure and two marital status measures as covariates, and a Site ($n = 3$) × Treatment ($n = 4$) design. This is the type of analysis used in the original reports. Analyses were conducted with the HRSD, the Hopkins Symptom Checklist—90 (HSCL-90), BDI, and the Global Assessment Scale (GAS). The completers, the sample of 204 and the sample of 239 were analyzed. According to conventional two-way analyses of covariance, there were no Site × Treatment interactions at the .05 level of significance for either adjusted means or heterogeneity of slope. Given that there are no a priori hypotheses with regard to the different, unidentified sites, no further analyses are warranted. I did, however, inspect the data, which suggested that Site 2 was somewhat of an outlier in that CBT did particularly badly both overall and in the severely depressed group, with regard to the BDI and GAS. Nonetheless, the overall null result indicates that Treatment × Site interactions were insufficient to account for the treatment findings.

I believe that the publication and analysis of the data from the TDCRP has been unacceptably dilatory. I agree with Jacobson and Hollon that knowledge of the actual identity of the various sites might be illuminating because, in principle, this would allow for more powerful hypothesis-driven analyses. The decision to censor this information is scientifically and ethically unwarranted.

Jacobson and Hollon state, "The bottom line is that there are reasons to doubt that CBT was adequately implemented in all three sites in the TDCRP" (1996, p. 76). However, there is no more substantive reason for this conclusion than for doubts about the implementation of pharmacotherapy, IPT, or for that matter, placebo—case management. What is needed is a still-missing integrity of treatment analysis.

Pharmacotherapy and Psychotherapy Collaborations

Both pharmacological and psychosocial treatments for anxiety disorders have strong proponents whose views are buttressed by treatment trials of varying degrees of cogency. A major limitation of much research data, however, is that most come from studies conducted at sites with a particular evangelic (often paternal) persuasion in terms of treatment ideology. Medications are studied at sites with psychobiological—psychopharmacologic orientations, whereas psychosocial studies are conducted at cognitive—behavioral centers.

A welcome recent development has been the creation of interdisciplinary multicenter studies comparing medication and psychosocial treatments. Collaborations involve researchers with demonstrated expertise in pharmacological or CBT for particular disorders such as panic disorder, obsessive-compulsive disorder or social phobia. A major question is whether treatment outcome will be as good at a site of the opposite ideological persuasion, where such a treatment is relatively unfamiliar, as at a site where such treatment is routine (the so-called allegiance effect). This is a relatively unstudied, but important issue because (a) it controls for sampling differences between different centers, (b) it controls for investigator bias, and (c) conducting a given treatment simultaneously at sites of different orientations, using a common, mutually agreed-upon protocol and standardized evaluations, forces both centers to expose or abandon outcome-relevant research idiosyncracies.

My department has recently participated in three major collaborative multisite pill-placebo-controlled studies of psychotherapy and pharmacotherapy. In Michael Liebowitz (Columbia University, New York City) and Richard Heimberg's (Albany, New York) joint study of cognitive—behavioral group therapy (CBGT), a psychotherapy placebo, phenelzine, and pill placebo, preliminary analyses showed no site effects for either psychotherapy or pharmacotherapy, despite the fact that the New York site was a psychopharmacological research center and the Albany site was a cognitive—behavioral psychotherapy research center. In parallel fashion, Michael Liebowitz and Edna Foa (Philadelphia) are jointly conducting pharmacotherapy—psychotherapy trials of obsessive-compulsive disorder treatment without apparent site effects. (M. R. Liebowitz, personal communication, April 12, 1995) In both trials, the specific benefit of the active agents was demonstrated.

This was probably due to the close collaboration of Liebowitz with Heimberg and Foa, in which they assumed responsibility for supervision of their respective therapies at their collaborative sites. It may also be due to the fact that while these investigators do advocate a particular therapy, in the sense that they have previously demonstrated its utility, that they are not advocates of any parochial superiority. Rather they were in the scientific business of trying to find out what is what, which affords the field a valuable model.

The present joint-controlled study by David Barlow (Albany, New York), Jack Gorman (Columbia University, New York), Katherine Shear (University of Pittsburgh) and Scott Woods (Yale University) of imipramine, panic control therapy, and pill/placebo has not progressed far enough for definitive site comparisons.

All of these studies demonstrate both design superiority and feasibility and are also models for the field.

Track Record

Jacobson and Hollon (1996) emphasize with regard to the TDCRP that no other study "has ever suggested any such advantage for pharmacotherapy over cognitive therapy" (p. 77). I agree, but I draw a radically different methodological and substantive conclusion. If one is trying to realistically compare the relative merits of pharmacotherapy and psychotherapy, one must have a pill-placebo arm, using participating scientists skilled in the sort of evaluations that pill placebo requires.

Jacobson and Hollon (1996) state that Klein "dismisses these studies" (p. 75), but a more accurate statement would be that I consider these studies moot because they had no internal calibration affirming that this sample was indeed a medication-responsive sample. The key problem is that clinical trials should be conducted on random samples of the defined population of concern, thus allowing proper generalizations. Needless to say this is never done. Instead, the reliance is on samples of convenience and unsystematic replications to make the case. Because of this uncomfortable fact, all conclusions from clinical trials should be in the format–these sample findings generalize to a population that conforms to the entry criteria (as modified by dropouts). However, this sample may differ sharply from other similarly defined samples. If we attempt to generalize to a population characterized by a certain treatment effect, we are only supported in this if we have demonstrated that particular treatment effect in this sample. Therefore, generalization to medication-responsive populations from samples that have not been demonstrated to be medication-responsive is unwarranted. Of course, the failure to demonstrate specific medication responsivity may be a Type II error rather than caused by real sample differences, but firm conclusions are still impossible.

The studies cited by Jacobson and Hollon do imply that one can select apparently well-diagnosed patients for whom psychotherapy is as good as pharmacotherapy or even superior. However, without a pill placebo, it is not known whether this is because in this sample neither therapy worked or even because medication was toxic. That there are so many still debatable, inconclusive studies confirms that the "plain vanilla" sample of convenience, pharmacotherapy-versus-psychotherapy design is inherently inadequate. I am pleased to find that Jacobson and Hollon agree that the field has advanced to the point that a pill-placebo control group is highly desirable, even if they offer somewhat different reasons.

Particularly compelling is the meta-analysis of the psychotherapy of depression by Robinson, Berman, and Neimeyer (1990) . The average effect size derived from 29 studies that compared psychotherapy with waiting-list controls was 0.84 ($p < .05$); however, the average effect size derived from the nine studies that compared psychotherapy with placebo was 0.28, insignificant both clinically and statistically. In the peculiarly convoluted fashion that seems standard in dealing with uncomfortable facts, Robinson et al. (1990) state the following:

> As our analyses demonstrate, clinical research has firmly established the efficacy of psychological interventions for depression ... It remains unclear, though, which aspects of psychotherapeutic treatment were responsible for producing this improvement. When the effects of psychotherapy were compared with those of placebo treatments, no reliable differences emerged. (p. 40)

Follow-Up Effects

Jacobson and Hollon (1996) point out that long-term follow-up differences among imipramine plus CM (IMI-CM), CBT, IPT, and PLA-CM were minimal, but they miss the crucial point. No one has claimed that pharmacotherapy has a continuing beneficial effect once treatment ceases, whereas this is the very claim that lent psychotherapy its fascinating charm. Starting with psychoanalysis, each school of psychotherapy has argued that they (alone!) specifically remediate the very basis of illness, by releasing the repressed, extinguishing the conditioned reflex, correcting the dysfunctional attitude, diminishing catastrophizing cognitions, improving interpersonal tactics, and so forth. Therefore, there is always an implicit, and often a quite explicit, claim that psychotherapy produces superior long-term effects, or even cures, compared with the transient symptomatic benefits of mere pill pushing. These claims have not been made on the basis of proper comparative trials but rather on the usual self-serving, incomplete set of anecdotal reports. That happy result, however, was not found in the prospective and comparative TDCRP reports.

Because it turned out that those who did well on CBT or IPT had no better prognosis than those who did well on placebo, shouldn't Jacobson and Hollon reflect whether the putative specific psychotherapeutic ingredient even exists? It doesn't bother me that during follow-up pharmacotherapy did no better than placebo, because I never thought it would. In fact, I thought pharmacotherapy would do worse, because some medication responders presumably require the continued presence of medication to maintain their equilibrium. Informed psychopharmacologists seem to be moving to the judgment that among severe, recurrent, medication-responsive patients with depression, the likelihood of relapse would be strongly decreased if medication was maintained indefinitely and that interrupting medication is ill advised. The follow-up data was just included in the final NIMH data release, so it has not been reanalyzed it as yet.

Jacobson and Hollon (1996) state that relapse rates following CBT were higher in the TDCRP than in similar studies, and they argue that this indicates a less-than-adequate implementation of CBT. Perhaps so, but perhaps it is due to varying sample composition. If those other studies had included a pill placebo, prognostic comparability across studies could be better evaluated.

Is Psychotherapy a Good Choice for Severe Depression?

Jacobson and Hollon (1996) also state that there is no basis for the belief "that CBT is contraindicated for severe depression" (p. 76), despite the complete lack of significant CBT superiority to placebo in the TDCRP. Furthermore, our comparison with IPT suggests CBT is contraindicated for just those severe depressions where it is frequently invoked. Worse, where is the positive evidence that it is a good idea? With regard to long-term pharmacotherapy, positive-controlled, comparative evidence for its distinctive superior usefulness exists. (Frank et al., 1990) But for CBT? Or IPT? Or any other psychotherapy?

I plainly disagree with Jacobson and Hollon (1996) with regard to their recommendation that CBT is "a viable alternative to pharmacotherapy in the treatment of even severe outpatient depression" (p. 79). I believe this is bad clinical advice. Is this simply a question of differing opinions that cannot be rationally distinguished? I don't think so. The bottom line is that if the Food and Drug Administration (FDA) was responsible for the evaluation of psychotherapy, then no current psychotherapy would be approvable, whereas particular medications are clearly approvable.

Is Pill Placebo a Necessary Comparison Group?

I am delighted that Jacobson and Hollon find "considerable merit to Klein's (1990) position" (1996, p. 77) . Nonetheless, they think I go too far. "Modified rapture," as Nanki Poo says in *The Mikado.*

"We do not agree that the absence of those controls renders drug—psychotherapy comparisons uninterpretable. The quality of the pharmacotherapy provided can (and must) be evaluated independent of whether drug and placebo differences are evident" (Jacobson & Hollon, 1996, p. 77). That the quality of the pharmacotherapy provided must be independently evaluated is not an argument against the use of pill placebo but rather an argument for precautions in addition to the use of pill placebo. Jacobson and Hollon list many appropriate requirements about medication implementation that are quite beside the point of whether the sample is a medication-responsive one. Every single desirable feature could have been carried out, analyzed, and presented and one still would not have a clue whether the sample was a medication-responsive one.

Jacobson and Hollon correctly, but irrelevantly, argue that "a pill placebo is not needed in the design to describe the characteristics of the sample" (1996, p. 77). For instance, Sotsky and Simmens (in press) have shown that about 25% of the TDCRP patients met the Columbia criteria for atypical depression, whom we have shown respond poorly to imipramine. However, only because there was a pill placebo

were they able to determine that, indeed, this subgroup of patients had no specific imipramine benefit whatsoever. This implies that the superiority of imipramine over psychotherapy was actually even greater for the appropriately medicated patients with depression.

Furthermore, because patients with atypical depression usually have low scores on the HRSD, what appears to be a severity effect (i.e., that imipramine is no better than placebo in patients with less severe depression) may be entirely an artifact of diagnostic composition. Jacobson and Hollon are correct that a sample requires refined description, but without pill placebo, one ends up with suspicions rather than facts.

Jacobson and Hollon (1996) argue that whether a sample is drug responsive is only one relevant consideration. One should also be interested in drug-unresponsive samples. Such samples would allow for interesting studies, such as estimating the value of an as yet untested drug, the evaluation of different psychotherapies, and so forth, but they are not appropriate samples for a comparison of a specific established pharmacotherapy versus psychotherapy, and that was the issue here.

I cannot follow the Jacobson and Hollon argument that a drug effect could mean that the sample is unusually placebo unresponsive. Are they saying that this is a sampling or Type I error? That concern is best met by replication. If they are saying that the sample is unusually placebo unresponsive, as with patients with chronic schizophrenia in comparison with those with panic disorder, then the unresponsiveness is a true characteristic of the sample and no problem is evident.

TDCRP Pharmacotherapy

I am currently analyzing the TDCRP pharmacotherapy and finding deficiencies. These data were not available until the recent release of the final NIMH data tape. Of course, analyses for compliance and dosage effects should have been among the very first analyses presented by Elkin et al. (1989) . Patient compliance by pill count could not be evaluated because unfortunately, this was not sufficiently standardized to allow accurate assessment. Therefore, I used a report from the therapist after each weekly session assessing whether the patient had taken their medication. This was rated on a scale corresponding to *full dose or over, more than 75% of a dose,* and *less than 75% of the dose.* Because each patient had multiple scores, we created a summary variable that defined inadequate drug compliance by the patient who used less than 75% of the dose for at least 33% of the sessions.

Using this stringent definition, only 9 of the 119 (7.6%) patients were noncompliant; of these nine, all but one had fewer than 8 sessions of a possible 16. Of the drug-treated completers, only 1 (1.3%) had inadequate drug compliance. Of those who dropped out, 8 of the 38 (18.2%) were rated as inadequately compliant. The Fisher's exact test (two-tailed, $p = .001$) indicates that lack of compliance was associated with dropouts. Similar analyses using other summary definitions of compliance (and higher proportions of noncompliers) yielded the same finding (unpublished data).

These preliminary analyses raise questions about the adequacy of the pharmacotherapy. Conceivably, a more attuned approach to perceived noncompliance would have resulted in higher treatment retention.

I also note that imipramine is not all of pharmacotherapy and that psychotherapy comparisons with the Specific Serotonin Reuptake Inhibitors, which are broader in spectrum and have fewer side effects, are way overdue.

Should Psychotherapy Researchers Panic?

Having agreed with me about the desirability of pill placebo control groups, Jacobson and Hollon then cite several panic disorder studies that have shown "very high success rates of CBTs" (1996, p. 78), apparently forgetting concern for manifold design deficiencies (e.g., waiting-list controls; see also my recent discussion [Klein, 1995]). The authors state, "in all comparisons with pharmacotherapy, CBT did at least as well, and usually better". I am not sure what the authors refer to because they do not cite references. I know of no comparative study of pharmacotherapy versus psychotherapy in panic disorder where in that sample medication was shown to be statistically superior to placebo, except for the Black, Wesner, Bowen, and Gabel (1993) study where CBT did little better than placebo and much worse than fluvoxamine. Jacobson and Hollon have fallen prey to our mutual concern about partisan, uncritical literature reviews. Furthermore, they miscategorize Shear et al. (1994) as favoring a theoretical approach other than CBT. They might discuss that with Shear et al. who appear quite agnostic.

Jacobson and Hollon (1996) criticize an undue reliance on the as yet unanalyzed multisite CBT, imipramine, and pill-placebo study of panic disorder, given past problems with the TDCRP. This seems unjust, because these investigators are well informed about the analytical deficiencies of the TDCRP. Site interactions, if found, will modify the conclusions they can draw, but why should Jacobson and Hollon assume that this will be the case? Perhaps their pessimism concerning this multisite trial is due to the unpalatable TDCRP conclusions. My conclusion is that more such multisite trials are required.

I caution that investigators should not repeat the appalling history of assuming wonderful psychotherapy benefits from poorly controlled trials. In the most recent report (unpublished data, February 3, 1995, presented at the Psychiatric Research Society) from the important Barlow, Gorman, Shear, and Wood collaboration, about 74% of the panic control therapy patients completed treatment and of them, about 73% were considered responders. This yields an approximately 54% intent-to-treat response rate, which should give pause. Because the drug—placebo blind of this study is not broken, nothing can be said regarding specific or comparative efficacies. In the recent Roche-sponsored study of panic disorder (unpublished data, moclobemide versus placebo; R. Buller, presented ECNP, 1994), there was a 70% placebo response rate! Pill-placebo internal calibration is an overriding necessity.

Jacobson and Hollon (1996) state that in those pill-placebo-controlled trials for CBT, "the quality of CBT was most suspect on independent grounds" (p. 79). This is not documented in their review. Just what were these independent grounds?

We agree that adequate pill-placebo-controlled medication trials as well as psychotherapy trials be conducted by those competent to do so. We also agree that psychotherapy evaluative trials that do not include a pill placebo are behind the times. This automatically implies that comparative studies of psychotherapy and pharmacotherapy are similarly obliged. If this discussion moves the field to that point, it would have been well worthwhile.

Criteria for Thoughtful Evaluations of Therapeutic Research

Jacobson and Hollon (1996) correctly argue that, even when a particular verdict is in, one has to look at the rest of the evidence. Replication, in the context of good design, is the sine qua non of science. To pursue the goal of critically integrating studies that claim relevance to treatment evaluation, I prepared a checklist, to clarify whether a study contributes to or detracts from understanding. Because of space limitations, it is not presented here but is available on request.

Conclusion

Jacobson and Hollon (1996) agree that including a pill-placebo comparison group in the design of

psychotherapy evaluative studies is a good idea. However, they argue that I have gone too far and that studies that do not include such a comparison group are still interpretable.

I assert that they have not gone far enough, because they have not answered my critique that studies that compare psychotherapy with waiting lists are uninterpretably biased and that studies that only compare psychotherapy with medication are inherently moot. They do not directly address my point that because such studies, based on samples of convenience, lack internal calibration, there can be no clear grasp of sample composition or generalizability. Treatment equivalence may mean that both treatments are not working and psychotherapy superiority may indicate drug toxicity for this specific sample. Their other arguments as to refinements with regard to sample and treatment description are true but obscure the point. These supplement the need for a pill-placebo arm, but they do not substitute for it.

One might guess that Jacobson and Hollon are loath to agree with me that most psychotherapeutic research has been a waste of time and money, but where are the logical grounds for disagreement with my stand? I applaud Jacobson and Hollon for rebutting this article, and I hope that they focus in narrowly on this topic. Just what legitimate inferences can be drawn from studies of convenience samples that do not include a pill-placebo arm?

Jacobson and Hollon (1996) imply that multisite studies are problematic because of potential site effects, but I do not see that as a problem, rather as a positive gain in knowledge. If such site effects, in well-conducted studies, were substantial and common, that would afford a whole new avenue of inquiry into the mechanisms of therapeutic benefit. If anything, Jacobson and Hollon should join me in supporting the appropriate multisite studies that my colleagues have pioneered, as the only answer to allegiance effects.

Furthermore, the logical demand that psychotherapeutic research include a pill-placebo arm parallels the FDA demand with regard to medication approval. It is only because of FDA requirements, which stem from legislation demanding the demonstration of both safety and efficacy before marketing, that placebo-controlled studies of pharmacotherapy are common.

Where does this leave investigators? Almost surely, there will never be an FDA for psychotherapy (although I would like it). Multisite, pill-placebo control studies are essential for evaluating, psychotherapy, pharmacotherapy, and their combination. Only NIMH provides funds adequate for such studies. Such studies are few and far between and totally inadequate to meet the social and scientific needs.

If there is going to be substantial progress, NIMH must initiate a funded program specifically for multisite, pill-placebo-controlled, studies of psychotherapy, pharmacotherapy, and their combination, jointly sponsored and jointly supervised by skilled psychopharmacologists and psychotherapists.

In the past, NIMH has raised false hopes by putting out Requests For Proposals, where no money was actually allocated. Such a practice can only be condemned. This does not mean that set-aside programmatic money must be spent; rather that an amount of money be dedicated from which high-quality grants are assured of funding, once peer reviewed and approved.

Without a conscious decision to budget for and support pill-placebo-controlled, person-oriented therapeutic research by NIMH, the National Institute for Drug Abuse, and the National Institute for Alcohol Abuse and Alcoholism, the field will continue to make, at best, stumbling advances. Also, it should be brought forcibly to the attention of the review groups of federal science institutes and private foundations that studies that purport to evaluate psychotherapy or pharmacotherapy without a pill-

placebo-comparison arm need skeptical evaluation.

There will be practical difficulties because the word will get out that psychotherapy trials with an apparent drug arm are likely to be offering only a pill placebo. Such a presumption will diminish expectancy and hope effects, making the pill-placebo arm a less-than-credible comparison group, thus creating the false appearance of psychotherapeutic superiority. Therefore, studies of psychotherapy must be combined with studies of active pharmacotherapy. Because there is an overriding public need to understand the relative short-term and long-term merits of pharmacotherapy, psychotherapy, and their combination, such joint multisite, multitreatment studies are all the more desirable.

References

Abelson, R. P. (1995). *Statistics as principled argument.* (Hillsdale, NJ: Erlbaum)
Black, D. W., Wesner, R., Bowen, W. & Gabel, J. (1993). A comparison of fluvoxamine, cognitive therapy, and placebo in the treatment of panic disorder.(*Archives of General Psychiatry, 50,* 44—50.)
Elkin, I., Gibbons, R. D., Shea, M. T. & Shaw, B. F. (1996). Science is not a trial (but it can sometimes be a tribulation).(*Journal of Consulting and Clinical Psychology, 64,* 92—103.)
Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S. J. & Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General Effectiveness of Treatments. (*Archives of General Psychiatry, 46,* 971—982.)
Frank, E., Kupfer, D. J., Perel, J. M., Cornes, C., Jarrett, D. B., Mallinger, A. G., Thase, M. E., McEachran, A. B. & Grochocinski, V. J. (1990). Three-year outcomes for maintenance therapies in recurrent depression.(*Archives of General Psychiatry, 47,* 1093—1099.)
Frank, J. D. & Frank, J. B. (1991). *Persuasion and healing: A comparison to psychotherapy.* (Baltimore: Johns Hopkins University Press)
Jacobson, N. S. & Hollon, S. D. (1996). Cognitive—behavior therapy versus pharmacotherapy: Now that the jury's returned its verdict, it's time to present the rest of the evidence.(*Journal of Consulting and Clinical Psychology, 64,* 74—80.)
Klein, D. F. (1995). Response to critique of suffocation alarm theory.(*Anxiety, 1,* 145—148.)
Klein, D. F. & Ross, D. C. (1993). Reanalysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program General Effectiveness Report.(*Neuropsychopharmacology, 8,* 241—251.)
, Psychotherapy is as good as drug in curing depression, study finds. *The New York Times,* p.(A1)
Robinson, L. A., Berman, J. S. & Neimeyer, R. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research.(*Psychological Bulletin, 108,* 30—49.)
Shear, M. K., Pilkonis, P. A., Cloitre, M. & Leon, A. C. (1994). Cognitive—behavioral treatment compared with nonprescriptive treatment of panic disorder.(*Archives of General Psychiatry, 51,* 395—401.)
Sotsky, S. M., Glass, D. R., Shea, M. T., Pilkonis, P. A., Collins, J. F., Elkin, I., Watkins, J. T., Imber, S. D., Leber, W. R., Moyer, J. & Oliveri, M. E. (1991). Patient predictors of response to psychotherapy and pharmacotherapy: Findings in the NIMH Treatment of Depression Collaborative Research Program. (*American Journal of Psychiatry, 148,* 997—1008.)
Sotsky, S. M. & Simmens, S. (in press). *Pharmacology response in atypical depression: Findings from the NIMH Treatment of Depression Collaborative Research Program* ((Report for American Psychiatric Association *DSM-IV* Task Force and sourcebook).)