National Institute of Mental Health 
Treatment of Depression Collaborative Research Program

General Effectiveness of Treatments

Irene Elkin, PhD; M. Tracie Shea, PhD; John T. Watkins, PhD; Stanley D. Imber, PhD; Stuart M. Sotsky, MD; Joseph F. Collins, ScD; 
David R. Glass, PhD; Paul A. Pilkonis, PhD; William R. Leber, PhD; John P. Docherty, MD; Susan J. Fiester, MD; Morris B. Parloff, PhD

- We investigated the effectiveness of two brief psychotherapies, interpersonal psychotherapy and cognitive behavior therapy, for the treatment of outpatients with major depressive disorder diagnosed by Research Diagnostic Criteria. Two hundred fifty patients were randomly assigned to one of four 16-week treatment conditions: interpersonal psychotherapy, cognitive behavior therapy, imipramine hydrochloride plus clinical management (as a standard reference treatment), and placebo plus clinical management. Patients in all treatments showed significant reduction in depressive symptoms and improvement in functioning over the course of treatment. There was a consistent 
ordering of treatments at termination, with imipramine plus clinical management generally doing best, placebo plus clinical 
management worst, and the two psychotherapies in between but generally closer to imipramine plus clinical management. In analyses carried out on the total samples without regard to initial severity of illness (the primary analyses), there was no evidence of greater effectiveness of one of the psychotherapies as compared with the other and no evidence that either of the psychotherapies was significantly less effective than the standard reference treatment, imipramine plus clinical management.

Comparing each of the psychotherapies with the placebo plus clinical management condition, there was limited evidence of the specific effectiveness of interpersonal psychotherapy and none for cognitive behavior therapy. Superior recovery rates were found for both interpersonal psychotherapy and imipramine plus clinical management, as compared with placebo plus clinical management. On mean scores, however, there were few significant differences in effectiveness among the four treatments in the primary analyses. Secondary analyses, in which patients were dichotomized on initial level of severity of depressive symptoms and impairment of functioning, helped to explain the relation of significant findings in the primary analyses. Significant differences among treatments were present only for the subgroup of patients who were more severely depressed and functionally impaired; here, there was some evidence of the effectiveness of interpersonal psychotherapy with these patients and strong evidence of the effectiveness of imipramine plus clinical management. In contrast, there were no significant differences among treatments, including placebo plus clinical management, for the less severely depressed and functionally impaired patients. (Arch Gen Psychiatry. 1989;46:971-982)

The National Institute of Mental Health (NIMH) (Rockville, Md) Treatment of Depression Collaborative Research Program was the first multisite coordinated study initiated by the NIMH in the field of psychotherapy. The study had two major aims: (1) to test the feasibility and value of the collaborative clinical trial model (a model frequently used in the field of psychopharmacology) in the area of psychotherapy research and (2) to study, within this research model, the effectiveness of two specific forms of psychotherapy (cognitive behavior therapy [CBT] and interpersonal psychotherapy [IPT]) for treating nonbipolar, nonpsychotic depressed outpatients. The study did prove to be feasible. Future publications...
will discuss the benefits and the problems of using the collaborative clinical trial model in psychotherapy research. The present article reports initial outcome findings for the two psychotherapies, in the areas of depressive symptoms and overall functioning. The general background of this collaborative study has already been described by Elkin et al. and readers are referred to that article for details regarding history, rationale, design, and procedures of the study.

Although initial efficacy findings for both CBT and IPT had been promising, they had not previously been compared directly with each other. The present study constituted the first direct comparison of these two psychotherapeutic approaches for the treatment of depressed outpatients. In addition, it provided more neutral settings for studying these two psychotherapies, each of which had previously been studied largely in settings strongly identified with the respective treatment.

To evaluate the effectiveness of the psychotherapies, they were compared with a reference treatment condition, i.e., a treatment that had already been found to be effective with this patient population. Since there was no specific form of psychotherapy that research had definitively established as efficacious with depressed patients, pharmacotherapy with a tricyclic antidepressant drug, imipramine hydrochloride, served as the reference condition in the current study. Considerable research evidence was available, at the time this study was planned, for the efficacy of imipramine. Since there was variability, however, in both imipramine and placebo response rates, and since most studies had been on inpatient samples, it was necessary to establish the effectiveness of imipramine at these sites and with these outpatient samples. For that reason, a pill-placebo condition was also included in the study. To standardize clinical care in the two pharmacotherapy conditions, a "clinical management" (CM) component was developed.

In comparisons of the four treatment conditions in this study, the two major questions addressed were the following: (1) Is there evidence of the effectiveness of each of the psychotherapies, as compared both with the standard reference treatment of imipramine-CM and with the placebo plus CM (PLA-CM) control condition? (2) Are there any differences in the effectiveness of the two psychotherapies?

The present article will focus on the general effectiveness of these treatments in reducing depressive symptoms and improving patients' functioning. These are major targets of change in any treatment of depression. The results presented, particularly on measures of depressive symptoms, will be comparable with those reported in previous studies comparing CBT or IPT with antidepressant drugs. To test adequately the effectiveness of these treatments, however, it is also necessary to evaluate changes in attitudes and behaviors specifically targeted by each of the psychotherapies. A report by Imber and colleagues addressed this issue of mode-specific effects.

The findings to be presented herein are limited to outcome at termination of treatment. The issue of the temporal pattern of therapeutic effects, particularly the important question of whether any of the treatments studied has more rapid effects, is addressed in a report by Watkins et al. Finally, the possibility that certain patient characteristics may differentially predict improvement in the different treatment conditions is explored in a report by Sotsky et al. All of these aspects of outcome, as well as patient status during follow-up, will have to be considered for a comprehensive picture of the study's findings.

PATIENTS AND METHODS

The research design and methods have been described in detail and will only be summarized here. At each of three research sites, patients were randomly assigned to the four treatment conditions: CBT, IPT, imipramine-CM, and PLA-CM. A total of 250 patients were assigned to treatment: 82 at the University of Pittsburgh (Pa) and 84 each at George Washington University (Washington, DC) and the University of Oklahoma (Oklahoma City).

Patients

The major referral sources for the program were psychiatric outpatient services at each research site, self-referrals, and other mental health facilities. Prospective patients were prescreened to rule out those who were clearly ineligible for the study. The 550 patients who passed prescreening were then interviewed by a clinical evaluator using the Schedule for Affective Disorders and Schizophrenia interview.

To be included in the study, patients had to meet Research Diagnostic Criteria for a current episode of definite major depressive disorder (with the additional criterion that the required symptoms had to be present for at least the previous 2 weeks) and had to have a score of 14 or greater on an amended version of the 17-item Hamilton Rating Scale for Depression (HRSD). Patients who met criteria for hypomania, or had a score of 20 or greater on the Hamilton Rating Scale for Hypomania, were dropped from the study.

Exclusion criteria included specific additional psychiatric disorders (definite bipolar II and probable or definite bipolar I, panic disorder, alcoholism, drug use disorder, antisocial personality disorder, Briquet's syndrome, and Research Diagnostic Criteria diagnosis of major depressive disorder, psychotic subtype), two or more schizotypal features, history of schizophrenia, organic brain syndrome, mental retardation, concurrent treatment, presence of specific physical illness or other medical contraindications for the use of imipramine, and presence of a chronic state inconsistent with participating in the research protocol, eg, current active suicide potential or need for immediate treatment.

Patients passing clinical screening received a complete medical evaluation. A 7- to 14-day wait and drug washout period followed; at the end of this period, patients returned for a rescreening interview with the clinical evaluator. The 250 patients who passed rescreening and who had given informed consent for all study procedures, including random assignment, were then entered into the program. Assignment of patients to treatment conditions was based on a separate computer-generated random order for each site.

Therapists

A different group of experienced therapists conducted treatment in each of the conditions, with the exception of the two pharmacotherapy conditions, which were carried out double blind by the same therapists. A total of 28 therapists (10 psychologists and 18 psychiatrists) took part in the outcome study, 8 in CBT and 10 each in IPT and pharmacotherapy. The preponderance of psychiatrists (64%) was largely due to the need to include only psychiatrists in the pharmacotherapy conditions. Details regarding selection, training, and certification of therapists are provided elsewhere.

The average age of the therapists was 41.5 years (range, 30 to 60 years), and they had an average of 11.4 years of clinical experience (range, 2 to 27 years). Twenty (71%) were male. Patients were assigned to therapists within treatments according to availability. Each psychotherapist saw between 3 and 11 patients (except for one therapist who moved after seeing 1 patient), and each pharmacotherapist saw between 3 and 9 patients in each of the two pharmacotherapy conditions.

Treatments

A major focus in this study was the careful definition, or standardization, of each of the treatments so that conclusions could be drawn regarding their specific effects. Each of the four treatment conditions was carried out in accord with a detailed manual describing the theoretical underpinning of the approach, the general strategies involved, the major techniques that could be used, and suggestions for dealing with specific problems. During the training/pilot phase of this study, the therapists all received further training in their respective approaches, met competence criteria in carrying out the treatments as described in the protocol, and were monitored throughout the outcome study. Detailed descriptions of each of the treatments can be found in the revised versions of the manuals, which have all been published.
The HRSD used in this study was a modified version of Hamilton's scale. The scores presented in this article are based on the first 17 items, comparable with those reported most frequently in the literature. The HRSD and GAS were rated by a trained clinical evaluator following a semistructured interview, the Schedule for Affective Disorders and Schizophrenia—Change version. In three reliability studies (one just before and two during the outcome study), intraclass correlations were calculated for clinical evaluator ratings of a common set of taped interviews. Agreement indices across raters, ranged from .98 to .96 for the HRSD and from .88 to .85 for the GAS.

**Statistical Analysis**

The major analyses of outcome were 3 x 4 (sites x treatments) analyses of covariance (ANCOVAs) of mean scores on the four measures and \( \chi^2 \) analyses of categorical data (recovered/not recovered by treatment) on the HRSD and BDI. These were followed by analyses of differences between each pair of treatments. Paired t tests were used to assess change from pretreatment to posttreatment within each of the treatment conditions. To test the effect on outcome of the initial severity of the patient's depression, secondary 2 x 4 (level of severity x treatments) ANCOVAs were carried out, with subsequent pairwise comparisons. Analyses of recovery data by \( \chi^2 \) were also carried out within severity subgroups. Univariate, rather than multivariate, ANCOVAs were used so as not to obscure any differences in outcome due to rater perspective, differences frequently observed in studies of psychotherapy. We also wished to ensure the separate analysis of the HRSD and BDI, to facilitate comparisons with previous studies.

Marital status, which was significantly related to outcome, was not distributed evenly across treatment groups and could thus bias the findings. For this reason, it was always used as a covariate in the ANCOVAs. Pretreatment scores on the dependent variable were also included as a covariate, except in those few instances on the HRSD and GAS in which there was significant (P < .05) heterogeneity of regression and the use of a pooled regression for the ANCOVA was not justified. In these instances, the ANCOVAs reported used only marital status, and not pretreatment score, as a covariate. The presence of significant heterogeneity of regression suggests that the treatments used different populations of patients, with which the decision to use low pretreatment scores on the variable in question. This issue is addressed in the secondary analyses, in which patients were divided into two severity groups on the basis of a priori cutting scores on the pretreatment HRSD (≥20 vs <20) and on the pretreatment GAS (≥50 vs >50). These severity analyses will help to clarify the role played by pretreatment levels of these variables.

Outcome data analyses were conducted on three samples of patients:

1. The completor sample included all patients who completed at least 12 sessions and at least 15 weeks of treatment. Although there were 162 completers, data analyses were based on 155 patients, since termination evaluations were not available for seven patients.

2. The end point 204 sample included all patients who received at least 3.5 weeks of treatment (which generally included at least four treatment sessions). The 204 patients include the 162 completers plus 42 patients who dropped out or were withdrawn after 3.5 weeks of treatment. In this sample, the last score obtained, either at interim or early termination evaluation, was used for the patients who dropped out or were withdrawn from the study.

3. The end point 239 sample consisted of all patients who entered treatment. This is also the specific sample, but it includes an additional 35 patients who dropped out or were withdrawn before 3.5 weeks of treatment. Twenty of these patients did not have interim or termination evaluations, and, to include them in the analyses, rescreening or intake scores were used as their end point scores. Examination of our data suggests that these scores represent a reasonable estimate of termination status.

These are clearly not independent samples; each successive sample includes the previous one. Some researchers choose to report findings on one or another of these samples. To get a comprehensive picture of the outcome of the treatments in this study, we analyzed data from all three samples. The completer analysis best reflects treatment effects for those patients who have received a full course of treatment, an important focus in any treatment study. This analysis does not, however, translate into any possible biases due to different attrition in the different treatment conditions. The analysis of end point samples attempts to address this problem by including termina-
tion scores not only for completers but also for patients dropping out or withdrawn from treatment. The end point 204 sample thus reflects terminal status for all patients who had at least minimum exposure to their respective treatment, and the end point 239 sample (sometimes referred to as an “intend to treat” sample) reflects terminal status for all patients who entered treatment. The results of an end point analysis, particularly the end point 239 analysis (since it includes some patients who have barely received any treatment), may be considered an estimate of the overall performance of a treatment program, including its ability to retain patients in treatment. Since there were some differences in the results of analyses of the three samples, we include data on all three to provide the most complete and accurate representation of our findings.

Although our main questions concerned differences between the two psychotherapies and between each of them and imipramine-CM, all six possible comparisons between treatment conditions were of interest. To protect against inflation of the type I error rate associated with multiple comparisons, probability levels for comparisons between pairs of treatments were adjusted for the total number of comparisons, by means of the Bonferroni t test or Brudner’s method of partitioning significance level.

To counterbalance this basically conservative approach, we chose a liberal overall probability level of P < .10 for accepting results as significant. For individual comparisons between any two treatment conditions, this means that the probability level obtained must actually be <.017 to be considered significant at an adjusted α level <.10. (We will also indicate “trends” toward significance where the observed probability value in individual comparisons is <.025, corresponding to an adjusted α level of <.15.) This approach left us with satisfactory statistical power (.81 to .86) to detect medium size effects in our primary ANCOVAs and adequately powered (>.80) to detect large size effects in the pretreatment ANCOVAs. Power was adequate (.82 to .95) for detecting medium size interactions of treatment and severity and adequate (.82 to .96) for detecting large size effects in the overall ANCOVAs. Power was frequently less than adequate (.80 to .86), however, for detecting large size effects in the pairwise comparisons within severity groups.

RESULTS

Patient Characteristics

A total of 250 patients (45% of those screened) met study criteria and were randomized into treatment. The primary reason for rejection was failure to meet the major depressive disorder and/or HRSD inclusion and/or HRSD exclusion criteria and/or screening. Eleven patients were dropped out before the first treatment session (3 in CBT, 2 in IPT, and 6 in imipramine-CM). Of the 239 patients who actually entered treatment, 165 (70%) were female and 212 (89%) were white. The average (± SD) age was 35 ± 8.5 years. Sixty-three (26%) were single, 96 (40%) were married or cohabitating, and 80 (33%) were separated, divorced, or widowed. The sample was characterized by a high level of education (120 [50%] had college degrees and 80 [34%] had some college education, and 80 [34%] had high school education or less. The mean (± SD) score at rescreening on the HRSD was 19.5 ± 4.4, and 91 patients (38%) were classified as having “definite endogenous” depression by Research Diagnostic Criteria. There had been no previous episodes of major depressive disorder for 85 (36%), one or two previous episodes for 82 (34%), and three or more for 72 (30%). Mean age at onset of first episode of major depressive disorder was 26.3 ± 10.3 years. The duration of the current episode was 6 months or less for 98 patients (41%), 6 months to 1 year for 57 (24%), and greater than 1 year for 84 (35%). In terms of previous treatment for depression, 36 (15%) had been hospitalized, 50 (21%) had received tricyclic antidepressants, 68 (28%) had received some kind of antidepressant medication, 65 (27%) had received some kind of antianxiety medication, and 136 (58%) had received psychotherapy. A total of 188 patients (76%) had received some form of previous treatment for depression. There were no statistically significant differences among treatment groups on any of these characteristics in the intake sample, although there was some imbalance on marital status.

Attrition

Of the 239 patients entering treatment, 77 (32%) terminated before completion, i.e., had less than 15 weeks and/or 12 sessions of treatment. This includes both patients who dropped out of their own accord and those who were withdrawn by study staff. By treatment group, there were 19 early terminators (32%) in CBT, 14 (25%) in IPT, 19 (33%) in imipramine-CM, and 28 (40%) in PLA-CM. Excluding patients who terminated for clearly external reasons or because they had improved or thought they had improved, there were 59 patients (25%) who terminated for what appeared to be negative treatment-related reasons, including, eg, dissatisfaction with treatment, desiring another treatment, intolerable side effects, and noncompliance. Only 9% of the patients could be classified as clear symptomatic failures. Differences among treatments in these three categories of early termination (all attrition, treatment-related attrition, and symptomatic failure) were evaluated by χ2 and by life table analyses but did not achieve levels of statistical significance. A more detailed report of attrition findings and their implications will be presented separately.

Because systematic or differential dropout can introduce bias in the results and affect the interpretation of findings, we compared those patients who completed treatment with those who did not on the major demographic and clinical variables obtained before treatment. One statistically significant difference (P < .05) emerged from these comparisons. Early terminators (across treatments) were more severely depressed at intake than patients completing treatment (pretreatment HRSD scores, 20.6 v 19.0, respectively; F1,231) = 6.32, P < .02). There was no significant treatment by termination status interaction in the HRSD or in the two items that accounted for more than 50% of the attrition variance. The interaction between early terminators and completers was almost identical within each treatment condition. The complete patients, in general, were fairly similar to the sample of patients entering treatment in terms of the distribution of patient characteristics both within and across treatment conditions.

Comparisons were also made among treatments in regard to the distribution of demographic and clinical variables within the completers sample. The pattern of attrition shifted the proportion of single and married patients (which had already been somewhat unevenly distributed) enough to result in a statistically significant (P < .05) differential distribution of marital status among treatment conditions in the complete sample. Compared with IPT and PLA-CM, CBT completers included more single patients and fewer married patients. As noted earlier, the variable marital status was used as a covariate in analyses of outcome measures.

Outcome Analyses

Pretreatment–Posttreatment Differences.—Paired t tests revealed highly significant differences (at P < .001) between pretreatment and posttreatment means for all groups (including PLA-CM) on all four measures in all three samples.

Table 2.—Trends in Baseline Means, SDs, and the results of ANCOVAs for the four outcome measures, in each of the three samples. There were no significant differences among groups in their pretreatment scores in any of these analyses. Statistically significant differences were found across the four groups at termination of treatment in 4 of the 12 ANCOVAs. These included analyses of the two patient measures (BDI and HSCL-90 R) in the completers sample and the two clinical evaluator measures (HRSD and GAS) in the end point 239 sample. None of the ANCOVAs of the end point 204 sample reached a level of statistical significance.

The general direction of results was similar in all 12 analyses, with PLA-CM always having the highest (most symptomatic) scores, imipramine-CM (and, in one instance, IPT) having the lowest scores, and the psychotherapies generally in between imipramine-CM and PLA-CM. The latter observation is usually explained by the fact that patients entering treatment tended to have lower scores than those who were referred for treatment.

In the completers sample, there was evidence of significant superiority of imipramine-CM over PLA-CM (P = .006) on the HSCL-90 R T. In the end point 239 sample, imipramine-CM was significantly better than PLA-CM (P = .003) and IPT (P = .014). In both samples, there was a trend for patients in both IPT and imipramine-CM groups to do better than the patients who received PLA-CM (P = .018 and .017) on the HRSD. There were no significant differences in the end point 204 sample, although there was one trend (P = .020) for imipramine-CM to do better on the GAS than PLA-CM.

Recovery Analyses.—To address the extent to which patients in the different treatment conditions met a predefined level of clinical recovery, the data were analyzed in terms of the percentage of
Table 1.—Mean Pretreatment and Adjusted Termination Scores for the Four Primary Outcome Measures*  

<table>
<thead>
<tr>
<th></th>
<th>CBT</th>
<th>IPT</th>
<th>IMI-CM</th>
<th>PLA-CM</th>
<th>Significance Level†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean ± SD</td>
<td>N</td>
<td>Mean ± SD</td>
<td>N</td>
</tr>
<tr>
<td><strong>HRSD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prestudy</td>
<td>37</td>
<td>19.2 ± 3.6</td>
<td>47</td>
<td>18.9 ± 3.9</td>
<td>37</td>
</tr>
<tr>
<td>Termination</td>
<td>37</td>
<td>7.6 ± 5.8</td>
<td>47</td>
<td>6.9 ± 5.8</td>
<td>37</td>
</tr>
<tr>
<td><strong>GAS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prestudy</td>
<td>37</td>
<td>52.6 ± 7.2</td>
<td>47</td>
<td>52.6 ± 6.6</td>
<td>37</td>
</tr>
<tr>
<td>Termination</td>
<td>37</td>
<td>69.4 ± 11.0</td>
<td>47</td>
<td>70.7 ± 11.0</td>
<td>37</td>
</tr>
<tr>
<td><strong>BDI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prestudy</td>
<td>37</td>
<td>26.8 ± 8.4</td>
<td>47</td>
<td>25.5 ± 7.7</td>
<td>36</td>
</tr>
<tr>
<td>Termination</td>
<td>37</td>
<td>10.2 ± 8.7</td>
<td>47</td>
<td>7.7 ± 8.6</td>
<td>36</td>
</tr>
<tr>
<td><strong>HSCL-90 T</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prestudy</td>
<td>37</td>
<td>1.38 ± 0.55</td>
<td>47</td>
<td>1.35 ± 0.45</td>
<td>36</td>
</tr>
<tr>
<td>Termination</td>
<td>37</td>
<td>0.47 ± 0.43</td>
<td>47</td>
<td>0.48 ± 0.43</td>
<td>36</td>
</tr>
</tbody>
</table>

*Pretreatment scores are means from a one-way analysis of variance. Adjusted termination scores are treatment main effect least-square mean scores from a treatment by site analysis of covariance with pretreatment scores and marital status used as covariates. Completer patients indicates patients with at least 12 sessions and 15 weeks of treatment; end point 204 patients, patients with at least 3.5 weeks of treatment; end point 239 patients, all patients entering treatment; CBT, cognitive behavior therapy; IPT, interpersonal psychotherapy; IMI-CM, imipramine hydrochloride plus clinical management; PLA-CM, placebo plus clinical management; HRSD, Hamilton Rating Scale for Depression; GAS, Global Assessment Scale; BDI, Beck Depression Inventory; and HSCL-90 T, Hopkins Symptom Checklist-90 Total Score. Higher scores on GAS indicate better functioning.

†Significance level for F test comparing the four treatment groups.
‡Significantly different from PLA-CM (observed P<.01; with Bonferroni correction, P<.10).
§Only marital status was used as covariate because of the lack of equality of slopes between pretreatment and termination scores.
¶Trend for difference from PLA-CM (observed P<.05; with Bonferroni correction, P<.15).
¶Significant (P<.10) treatment by site interaction.

patients who reached a score of 6 or less on the HRSD and also in terms of the percentage who reached a score of 9 or less on the BDI. These cutting scores, used in other studies comparing CBT and tricyclic drugs, reflect the presence of little or no remaining depressive symptoms and are more stringent than most “improvement” criteria.

As can be seen in Table 2, there were significant differences across the four treatment conditions, for all sample sizes, in the percentage of patients reaching a score of 6 or less on the HRSD. Multiple comparisons among pairs of treatments, with the use of the Brunden correction, showed both IPT and imipramine-CM to be significantly superior to PLA-CM in the end point 239 sample (P<.010 and .018, respectively). The percentages of patients entering treatment who reached the HRSD recovery criterion were 43% and 42% for IPT and imipramine-CM, respectively, and 21% for PLA-CM. Patients receiving CBT had a recovery rate of 36%, not significantly worse than that with IPT and imipramine-CM, though also not significantly better than that with PLA-CM. There were also trends for superiority of IPT and imipramine-CM over PLA-CM in the completer (P=.021 and .020) and end point 204 (P=.024 and .018) samples. The percentage of patients reaching the recovery criterion among the patients who completed treatment was 57% for imipramine-CM, 55% for IPT, 51% for CBT, and 29% for PLA-CM. There were no significant differences among treatments on the BDI recovery criterion, at least in part because the patients in the PLA-CM group did so well on this criterion.

In the analyses of both mean scores and recovery rates, no signif-
Table 2.—Patients Recovered at Termination*

<table>
<thead>
<tr>
<th></th>
<th>CBT</th>
<th>IPT</th>
<th>IMI-CM</th>
<th>PLA-CM</th>
<th>Significance Level†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Completer Patients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>37</td>
<td>47</td>
<td>37</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>HRSD=6</td>
<td>19 (51)</td>
<td>26 (55)</td>
<td>21 (57)</td>
<td>10 (29)</td>
<td><strong>.074</strong></td>
</tr>
<tr>
<td>BDI=9</td>
<td>24 (65)</td>
<td>33 (70)</td>
<td>25 (69)</td>
<td>18 (51)</td>
<td><strong>.299</strong></td>
</tr>
<tr>
<td>(N=36)</td>
<td>(N=35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>End Point 204 Patients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>50</td>
<td>55</td>
<td>49</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>HRSD=6</td>
<td>20 (40)</td>
<td>26 (47)</td>
<td>24 (49)</td>
<td>13 (26)</td>
<td><strong>.076</strong></td>
</tr>
<tr>
<td>BDI=9</td>
<td>29 (58)</td>
<td>33 (60)</td>
<td>30 (61)</td>
<td>25 (50)</td>
<td><strong>.668</strong></td>
</tr>
</tbody>
</table>

*Table 1 for explanation of abbreviations.
†Trend for difference from PLA-CM (observed P<.025; with Brunden correction, P<.15).
‡Significantly different from PLA-CM (observed P<.017; with Brunden correction, P<.10).

Secondary Analyses.—The patient sample had included, by design, a fairly wide range of severity of depression. To investigate the influence of pretreatment severity on outcome, two separate sets of analyses were carried out. These analyses must be considered exploratory, since the design did not include stratification on this variable. In the first set of these exploratory analyses, patients’ conditions were considered more severe if they had a score of 20 or greater on their rescreening HRSD. In the second set of analyses, they were considered more severe if they had a score of 50 or less on the GAS. Forty-four percent of the sample met the HRSD criterion for severity; 41%, the GAS criterion; and 25%, both criteria. These cutting scores, defined a priori, reflect a severe level of symptoms (HRSD) or a severe level of both symptoms and impairment of functioning (GAS), within the context of an outpatient sample. The GAS cutting score provided a useful differentiation in a study of recurrent depression.

Analyses of covariance of the four measures in all three samples, with the use of the HRSD severity breakdown, yielded two significant treatment by severity interactions. These were both on the HRSD, in the completer and end point 204 samples. The interaction for the HRSD in the end point 239 sample barely missed our criterion for statistical significance (P=.11), and since there was a significant interaction for the HRSD in the other two samples, the end point 239 sample was also included in the subsequent analyses described below.

Similar ANCOVAs, with the GAS severity criterion, yielded seven significant severity x treatment interactions: for two measures in the completer sample (HRSD and HSLC-90 T), one in the end point 239 sample (HRSD), and all four measures in the end point 204 sample.

Figure 1 illustrates the pattern of the interaction in the end point 204 sample for HRSD scores, with the use of the HRSD severity criterion, and for GAS scores, with the use of the GAS severity criterion. (These two examples were chosen because, in addition to the significant severity by treatment interactions, there was also significant heterogeneity of regression for the GAS and near-significant [P=.06] heterogeneity for the HRSD.) The pattern for the two variables was somewhat different. On the GAS, the interaction was due largely to the marked difference between imipramine-CM and the other groups, particularly PLA-CM. All treatments did about the same for the less severely depressed and impaired patients, but imipramine-CM did much better for the more severely depressed and impaired patients and PLA-CM did much worse for these patients. Cognitive behavior therapy and IPT were similar to each other and did somewhat worse for the more severely depressed and impaired than the less severely depressed and impaired patients. On the HRSD, on the other hand, IPT was more like imipramine-CM, both doing about the same for the more and less severely depressed patients. Significance differences were found either between the two psychotherapies or between either of them and imipramine-CM.

Subsequent ANCOVAs were carried out for the more and less severe subsamples wherever there was a significant severity x treatment interaction. No differences were found among the less severely depressed treatment groups in any of these analyses. Significant differences across treatments were found for the severely depressed patients, in every instance in which there had been a significant severity x treatment interaction. Pairwise comparisons revealed that IPT was significantly more effective than PLA-CM in three instances, all on the HRSD in end point samples, while imipramine-CM was quite consistently significantly more effective than PLA-CM (in 8 of 10 instances), across a number of measures.

Data for more and less severely depressed subgroups were also analyzed in terms of the percentage of patients who reached the recovery criterion of a score of 6 or less on the HRSD. The left-hand side of Fig 2 presents these percentages for the more and less severely depressed groups, as defined by the HRSD, while the right-hand side of Fig 2 presents the percentages for the more and less severely depressed and impaired groups, as defined by the GAS.

The results are similar to those reported for the mean scores. Again, there were no significant differences among treatments for the less severely depressed groups, although the two psychotherapies had slightly higher rates than the two pharmacotherapy conditions for the GAS-defined less severely depressed and impaired patients. For the more severely depressed groups (using both severity criteria), there were consistently significant differences in recovery rates among treatments in all three samples. Further analyses revealed that, for the more severely depressed patients, IPT was significantly superior to PLA-CM in five of six tests and imipramine-CM was significantly superior to PLA-CM in all six tests.

Pairwise comparisons for the more severely depressed and impaired subgroup, as defined by the GAS, also yielded a significant difference between one of the psychotherapies and imipramine-CM, while CBT and PLA-CM did better for the less, than for the more, severely depressed patients. These patterns were fairly representative of most of the other measures and samples within each of the severity breakdowns.
with imipramine-CM being superior to CBT in the end point 294 sample. With the exception of this lone finding, there were no significant differences between either of the psychotherapies and imipramine-CM. As in the analyses of the total sample, there were no significant differences between the two psychotherapies.

**COMMENT**

**Effectiveness of IPT and CBT**

To evaluate the effectiveness of IPT and CBT, they were compared both with a standard reference treatment of tricyclic medication and with a control treatment of pill-placebo, each combined with CM, a minimal supportive therapy. Comparison with a standard reference treatment addresses the question of whether the "experimental" treatments, in this case both psychotherapies, are less (or more) effective than the currently accepted treatment for a disorder. In the major analyses in this study, there was no evidence that either of the psychotherapies was significantly less (or more) effective than imipramine-CM. (Evidence of greater effectiveness of imipramine-CM during the course of treatment is described in the report by Watkins et al. 17)

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 do 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 treatments (in the total unstratified sample), it is unlikely that very large or important differences were missed.

We also compared each of the psychotherapies to the PLA-CM condition. This served as a control for regular contact with an experienced and supportive therapist, as well as the more general support provided by the research setting. Since this condition also included expectations related to the drug administration, it was not an ideal control for the psychotherapies. Nonetheless, it could serve as a useful comparison for evaluating the specific effectiveness of the psychotherapies.

There was limited evidence of the specific effectiveness of IPT and no evidence of the specific effectiveness of CBT, when compared with PLA-CM, in the major outcome analyses. There was some, but not a great deal of, evidence of the superiority of imipramine-CM over PLA-CM in these analyses. The order of the treatments was quite consistent, with imipramine-CM (and in one instance IPT) doing best, PLA-CM doing worst, and the psychotherapies generally in between, but the magnitude of the differences in mean scores was not large. There was some evidence of the superiority of IPT, as well as imipramine-CM, over PLA-CM in the recovery analyses.

The general lack of significant differences between the

![Percentage of patients in more and less severely depressed groups recovered at termination, by treatment condition. Top, Completer patients; middle, end point 204 patients; and bottom, end point 239 patients. For the Hamilton Rating Scale for Depression (HRSD), less severe indicates a prestudy total score of less than 20, and more severe, a prestudy total score of 20 or more. For the Global Assessment Scale (GAS), less severe indicates a prestudy score of greater than 50, and more severe, a prestudy score of 50 or less. CBT indicates cognitive behavior therapy; IPT, interpersonal psychotherapy; IMI-CM, imipramine hydrochloride plus clinical management; and PLA-CM, placebo plus clinical management.](image)
psychotherapies and PLA-CM does not seem to be due to a poor showing by the psychotherapies. The results for both CBT and IPT were, by and large, comparable with those in other studies,3,4,7,31,38 comparing them with antidepressant drugs. (The one exception here was the study by Rush et al.,2 in which their patients receiving CBT showed more improvement than those receiving CBT in the present study. One possible reason for this difference was the more intensive supervision received by the therapists throughout the Rush et al. study; in the present study, intensive supervision in the training phase was replaced by more infrequent consultation during the outcome study.) In addition, highly significant differences (P<.001) were found between pretreatment and posttreatment means for the patients in IPT and CBT; this was true, however, for patients in the PLA-CM condition as well.

The main reason for the general lack of significant findings seems to be due, not to lack of improvement in the psychotherapy groups, but rather to the very good performance of the PLA-CM condition. This is in contrast to the finding of fairly poor performance of waiting-list or delayed-treatment groups used as controls in other studies of brief psychotherapies for depression.29 It is important to remember that, because of the CM component, PLA-CM in this study is not a no-treatment condition or even an “inactive” placebo condition. Patients were seen once a week for 20 to 30 minutes by a well-trained and experienced psychiatrist, who not only administered the medication and reviewed symptoms, side effects, and general functioning, but also offered support and encouragement. This combination of minimal supportive therapy and expectations regarding the medication, along with the generally supportive research and treatment environment, may be sufficient for many patients to achieve a significant reduction in depressive symptoms and improvement in general functioning. This may be true, however, only for that subgroup of patients with major depressive disorder who are less severely depressed and functionally impaired, as will be discussed in the section on “Initial Severity” below.

Standard Reference Treatment.—The fact that there were also relatively few significant differences in the primary analyses, between the PLA-CM and PLACM users, raises the question: Was imipramine-CM an adequate standard reference treatment in this study? Our imipramine-CM condition compared favorably with the tricyclic conditions in the comparative studies cited earlier.2,4,7,31,38 It is difficult to make direct comparisons between our findings and those in the broader psychopharmacologic literature, however, because the studies differ in a number of ways. Our study was longer than most antidepressant drug studies, and this may provide more chance for improvement; on the other hand, our “recovery” criterion was more stringent than the “improvement” criteria used in most of these studies. Also, unlike in many of these studies, we reported on all patients entered into the study, regardless of level of compliance, reason for attrition, etc. Perhaps of greatest importance, our clinician data are based on ratings by independent clinical evaluators, while most of the psychopharmacologic studies report therapist ratings.

To provide information more comparable with that in the psychopharmacologic literature, we carried out several additional analyses, using the end point 204 sample. Eight-week end point analyses for just the imipramine-CM and PLA-CM groups yielded significant differences favoring imipramine-CM over PLA-CM on both the clinical evaluator–rated HRSD (P<.05) and the therapist-rated HRSD (P<.01). Differences at termination of treatment between imipramine-CM and PLA-CM, both in mean scores and in recovery rates on the therapist-rated HRSD, were significantly different in all three samples (ranging from P = .001 to P = .014). Perhaps most striking were the high recovery rates at termination (as judged by therapists) for patients in the imipramine-CM group having at least 3.5 weeks of treatment (78%). Given the stringency of this recovery criterion, this should demonstrate that, by usual psychopharmacology standards, imipramine-CM was an effective treatment in this study and set a high comparison standard for the psychotherapies.

It is important to note that therapists rated patients as more improved than did clinical evaluators, and therapists also differentiated more clearly between patients in active drug and placebo groups. It is these differences between therapist and clinical evaluator ratings that make for an apparent discrepancy between the very favorable findings for imipramine-CM with “usual psychopharmacology standards” and the more modest results presented in this article (for the primary analyses), based on ratings by independent clinical evaluators. (Results using therapist measures, which also demonstrated significant differences between the psychotherapies and PLA-CM, are not presented herein because of the lack of calibration of the different groups of therapists and the possibility of biases that may be associated with different frames of reference of the psychotherapists and pharmacotherapists.)

Alternative Statistical Criteria.—It has been suggested that the relative lack of significant findings reported for our major analyses of clinical evaluator and patient measures may be due to the stringency of our statistical criteria for significance of differences between pairs of treatments. We consider as most appropriate the α levels that were chosen a priori (and reported in the “Results” section), particularly given the need to protect against type I error when findings on several non-independent variables are reported, and the necessary precautions when analyzing subgroups in a clinical trial.30 Nonetheless, in response to questions raised, we have also examined the results of our major data analyses with an alternative approach to significance levels. Using F-protected t tests, with a conventional α level of .05 for the overall F test and an α of .05 for subsequent t tests between pairs of treatments, we found the following:

Only 1 of our 12 ANCOVAs (HSCL-90 T in the completer sample) resulted in an F significant at <.05; in pairwise comparisons, we found that imipramine-CM contributed significantly (P = .049) superior to PLA-CM. If the significance criterion is relaxed to include the two F values at P = .053 (HRSD and GAS in the end point 239 sample), the previous trends for superiority of IPT and imipramine-CM over PLA-CM on the HSCL-90 become significant, and IPT is added to imipramine-CM as being significantly superior to PLA-CM on the GAS. Using this alternative set of criteria would result, then, in an increase in evidence of the superiority of IPT over PLA-CM. There would still be only 3 tests of 12, however, in which IPT and imipramine-CM were significantly superior to PLA-CM, none in which CBT was superior to PLA-CM, and no evidence of differences in effectiveness between the two psychotherapies and imipramine-CM.

Comparison of IPT and CBT

A second major question in this study was whether there were any differences in the effectiveness of IPT and CBT. These two psychological treatments, both developed for treating depression, had never before been compared within the same study. In interpreting the findings for the two psychotherapies (and for pharmacotherapy as well), it is impossible in this study to separate treatments from the therapists carrying out those treatments. We will, in future analyses, explore the outcome differences by individual therapists within each treatment approach. For now, however, it is important to keep in mind that we are not comparing IPT and CBT per se, but rather a “package” of each.
of the treatment approaches with those therapists who chose to and were chosen to carry it out.\cite{4,5}

We found no significant differences between IPT and CBT in any of the major analyses (or, for that matter, in the secondary severity analyses). Furthermore, an examination of mean scores reveals that, with a few exceptions, the actual numerical differences between IPT and CBT and the related effect sizes were very small. This was true even in some cases where IPT was significantly different from PLA-CM, but CBT was not. Thus, there is no evidence in this study of the greater effectiveness of one of the psychotherapies on measures of depressive symptoms and overall functioning (although, again, there was adequate power in the primary analyses for detecting any large effect size differences that might exist). These findings are consistent with many reported in the psychotherapy literature\cite{8} of a general lack of differences in outcome between different forms of psychotherapy. Despite the similar outcome in the analyses reported herein, however, one might still expect to find differences on measures more specifically related to each of the treatment approaches. Even on such measures, however, Imber et al.\cite{1} found few differences. The general lack of differences between the two psychotherapies, together with the good results for the PLA-CM condition, suggests once again the importance of common factors in different types of psychologically mediated treatment.\cite{9}

Factors Influencing Outcome

Initial Severity.—The primary analyses reported in this article were based on all patients within each sample. It was recognized from the outset, however, that a diagnosis of major depressive disorder did not guarantee a homogeneous patient sample.\cite{10} A separate report\cite{11} presents analyses exploring the predictive role of a number of patient demographic, clinical, and personality variables. Pretreatment severity of illness is considered such a crucial variable, however, that its possible role in influencing outcome deserves special attention. Review of both the psychotherapy and pharmacotherapy literature\cite{12,13} had suggested that initial level of symptoms and functioning may influence outcome of treatment. We were especially interested in the possible differential role of severity in the different treatment conditions. For that reason, we reported secondary analyses of the influence on outcome of the initial severity of the patient’s illness.

These exploratory severity analyses suggested that initial severity was, in fact, a very important variable in this study. A number of significant severity \times treatment interactions were found, especially for the GAS breakdown. The treatment \times severity interactions help to explain the relative lack of significant findings in the total, unstratified sample (particularly in the end point 204 sample). Looking separately at the more and less severely depressed subgroups, results were quite consistent: there were no differences among treatments (including PLA-CM) in the less severely depressed patient group, while, for the more severely depressed patients, quite consistent differences emerged between the active treatments and PLA-CM. For these patients, especially those defined by the GAS as more severely depressed and functionally impaired, IPT was superior to PLA-CM on both the HRSD mean scores and recovery criterion, while imipramine-CM was quite consistently superior to PLA-CM on a broader range of measures. Cognitive behavior therapy, though not very different from IPT in this sample, did not do well enough, across the board, to demonstrate significant superiority to PLA-CM.

The effectiveness of imipramine-CM for the more severely depressed and functionally impaired patients was especially marked in the end point 204 sample, where it was significantly superior to PLA-CM on all four outcome measures and on the recovery criterion. Interpersonal psychotherapy was superior to PLA-CM only on the HRSD measures, both mean scores and recovery criterion. The end point 204 sample, which includes all patients who had at least 3.5 weeks of treatment, seems most likely to favor the active drug treatment; following 3.5 weeks of treatment, patients might be expected to show considerable improvement if they had received tricyclic medication but would probably not be expected to do so if they had received psychotherapy. This points to one of the difficulties inherent in comparing pharmacotherapy and psychotherapy\cite{14} and should be taken into account in interpreting results of the end point 204 sample.

These secondary analyses (particularly for the GAS breakdown) led to two observations of potential clinical significance.

1. The PLA-CM treatment did not do well for the more severely depressed patients, particularly those with more functional impairment, and rarely brought them to a level of recovery. More specific targeted forms of treatment are indicated for these patients. Imipramine-CM was especially effective for this subgroup, both in reducing depressive symptoms and in improving general functioning, and there was also some evidence of the specific effectiveness of IPT in reducing depressive symptoms.

2. For the less severely ill patients, there were no significant differences across treatments on any of the measures, although there was sufficient statistical power for detecting large effect size differences. Thus, there was no evidence of the superiority of any of the treatments (including imipramine-CM) over PLA-CM in this subgroup. These findings are in accord with an observation by Downing and Rickels,\cite{15} who, after reviewing the drug literature, concluded that “a more intense psychotherapy relationship may tend to obscure differences between response to active drug and response to placebo.” Our pharmacotherapy-treated patients did not receive “intense” psychotherapy; however, they probably did receive more contact and support than is present in many drug studies. It is difficult to compare our PLA-CM results with those of traditional placebo conditions in the psychopharmacology literature, due to many methodologic differences. Where it is possible to focus, however, on what appear to be comparable samples or subgroups, PLA-CM seems more effective than drug study placebos.\cite{16}

It must be stressed that the present findings do not mean that the less severely depressed and functionally impaired patients do not need any treatment, especially since other studies report significant superiority of brief psychotherapies for depression over waiting-list or delayed-treatment control groups,\cite{17} which evidence little change. Further work will be directed toward understanding the therapeutic properties of the PLA-CM condition in this study. The present findings do raise the possibility that some type of minimal supportive therapy in the hands of an experienced practitioner may be sufficient to bring about a significant reduction of depressive symptoms (at least in the short run) for the less severely depressed patients. The findings, if replicated, also raise questions regarding the need for antidepressant medication or for highly specified forms of psychotherapy for the resolution of the depressive episode in these patients. It is of interest that the patients in this subgroup constituted almost 60% of our sample and are probably seen with some frequency in many outpatient settings.

Since we did not include severity as a stratifying variable in our original design, these findings must be considered exploratory. We find the results based on our poststudy stratification provocative, however, especially as they help to explain the relative lack of significant findings in our total unstratified sample. Severity of initial illness has been a fairly consistent...
predictor of outcome in the psychotherapy literature, and some recent studies in both psychopharmacology and psychotherapy have also found that poststudy stratification on severity revealed interesting differential treatment outcomes.57,60

Although clinicians and researchers alike acknowledge the possible importance of severity in influencing the outcome of treatments for depression, there is little systematic preplanned research on this subject. We hope that these results will encourage other investigators to try to replicate the findings, to prestratify on severity, or to focus on specific levels of severity in their research. A later article will explore in more detail the severity findings in the present study and their implications. The article will include further exploration of the regression data and of the differences found with the two different severity criteria.

**Research Site Differences.**—There were two significant differences in outcome due to research site; in both, two of the research sites generally did better than the third. What we were particularly concerned about, however, was the possible interaction of treatments and sites, which would indicate a differential effectiveness of the treatments depending on setting. In the major analyses, there was only one significant treatment by site interaction. This interaction, on the BDI in the end point 239 sample, seemed to be due to a particularly poor outcome on this measure for PLA-CM in one site and for CBT in another site. (Since there was no treatment effect on this variable, the interaction did not alter the findings reported in this article.)

More consistent interactions of treatment and site seem to be present, however, in the more severely depressed and impaired subsample, as defined by the GAS. Site had not been included in the severity analyses because there was an insufficient number of severe cases at one of the sites. Exploratory analyses of the data for the two sites with sufficient numbers of patients revealed consistently significant treatment by site interactions for the more severely depressed and functionally impaired patients. 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. Outcome for imipramine-CM was more consistent at the two sites. Further analyses will explore the extent to which these interactions may be due to the distribution of particular patient characteristics, therapist performance variables, or some other aspects of the treatment settings. 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.

**Patient Samples and Perspectives.**—There were also some differences in findings, depending on whether one looked at completers or at one or another of the end point samples. We have already commented on the particularly striking performance of imipramine-CM in the GAS-defined severely depressed and impaired group in the end point 204 sample. Some of the other differences in findings for particular samples are confounded with the perspective from which ratings were made. For example, significant differences between pairs of treatments on clinical evaluator-rated measures are found largely in the end point samples, but not in the completer sample, although the pattern of results is similar for the completers. This finding is partly due to the increased number of patients in the end point samples, but, at least in the case of IPT, it also seems due in part to the lower attrition from this treatment as compared with PLA-CM. Since early terminators generally had poorer scores at termination, more of these poor scores were included for the PLA-CM group (which had a 40% attrition rate) than for the IPT group (which had a 23% attrition rate), increasing the difference between PLA-CM and IPT means in the end point samples. Thus, the relative superiority of IPT over PLA-CM in the end point samples for the clinical evaluator measures is probably due in part to the ability of IPT to retain patients long enough for them to benefit from the treatment.

The significant findings for patient measures, on the other hand, occurred mainly in the completer sample (and for GAS-defined severely ill patients in the end point 204 sample). The general lack of significant findings in the end point samples seems to be due, at least in part, to the fact that some patients who received imipramine-CM who terminated very early rated themselves as poorly as those in PLA-CM. Further work will be carried out to explore the reasons for some of these perspective/sample differences.

**CONCLUSIONS**

In the set of findings for the total unstratified sample of patients in our primary analyses, it is clear that there is no evidence of greater effectiveness of one of the psychotherapies as compared with the other and no evidence that either of the psychotherapies was significantly less effective than the standard reference treatment. All treatment conditions (including PLA-CM) evidenced significant change from pre-treatment to posttreatment. Comparing each of the psychotherapies with the placebo plus minimal supportive therapy condition, there was limited evidence of the specific effectiveness of IPT and none of the specific effectiveness of CBT. In general, the results for the two psychotherapies fell between those for imipramine-CM and PLA-CM, being neither significantly less effective than imipramine-CM nor (with some exceptions for IPT) significantly more effective than PLA-CM. They were in general, however, closer to imipramine-CM than to PLA-CM.

The secondary analyses dichotomizing the patient sample according to two severity criteria suggest that a very important role was played by this variable. For the less severely depressed patients, there was no evidence of the specific effectiveness of any of the treatments over the placebo combined with minimal supportive therapy. There was evidence, however, of the specific effectiveness of the active treatments for the more severely depressed and functionally impaired patients; here, the imipramine-CM condition did extremely well, the PLA-CM condition did poorly, and the two psychotherapies were in between. Although IPT was not significantly more effective than CBT, it increased the percentage of patients taking antidepressants and was significantly more effective than CBT in reducing depressive symptoms. The most striking findings, here, were on the recovery data, where both IPT and imipramine-CM were consistently superior to PLA-CM. These “severity” findings (as well as the results of the primary analyses) should not be generalized beyond the current patient sample, i.e., nonpsychotic, nonbipolar outpatients meeting criteria for a major depressive disorder.

Although this study was designed as a psychotherapy investigation, the inclusion of the standard reference and control conditions provides information about the effectiveness of imipramine. The present results may serve to reaffirm the value of imipramine in treating depressed outpatients, at least those who are more severely ill. Imipramine-CM was more frequently superior to PLA-CM than were the psychotherapies and showed significant effects over a broader range of measures. This was not true, however, for the less severely ill patients, for whom imipramine-CM was not significantly superior to PLA-CM. The secondary analyses suggest a means of identifying those more severely ill patients who are especially responsive to the drug. The results reported by Watkins et al also demonstrate that, although all of the groups improved over time, the effects of imipramine-CM were more rapid.

Several factors should be taken into account in interpreting
the findings reported for the two psychotherapies. The first relates to the specific effectiveness of IPT. Interpersonal psychotherapy was consistently superior to PLA-CM, although the differences reached a level of statistical significance only on the HRSD. It should be noted that the HRSD is probably the most established instrument in studies of the treatment of depression. The fact that IPT did so well on the stringent recovery criterion, and especially with the more severely depressed and impaired patients, is of clinical significance.

The additional exploratory analyses suggested that, while the effects of imipramine-CM were fairly consistent across research settings for the more severely depressed and functionally impaired patients, IPT showed a similar degree of effectiveness at one research site and CBT at another. The possible reasons for these site differences must be explored and hypotheses generated and tested to begin to determine the conditions under which each of the psychotherapies may be effective in treating severely depressed outpatients.

In drawing conclusions from the comparisons of the two psychotherapies with the standard reference treatment, it should be borne in mind that imipramine-CM was a very effective treatment, especially for the GAS-defined severely depressed subgroup (in which 76% of the imipramine-CM completers reached the recovery criterion, compared with 18% of the patients in PLA-CM), and thus provided a stringent "standard" for comparison with the psychotherapies. Since the imipramine condition, as well as the placebo condition, included the CM component, the success of this condition with the more severely ill patients may have been due to the combination of drug and minimal supportive therapy. This study did not include a treatment condition combining pharmacotherapy with a specific form of psychotherapy. It is possible that the effects of such a combination would have been superior to that of any of the individual treatments, although results in the literature have been inconsistent.46

The current findings address only the question of the short-term effectiveness of these treatments. To evaluate fully the outcome for the psychotherapies, however, it will be necessary to consider their possible role in the maintenance of improvement and in the prevention of relapse. Future publications reporting analyses of data obtained at 6-, 12-, and 18-month follow-up evaluations will address these issues.

The NIMH Treatment of Depression Collaborative Research Program is a multi-site 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 to six participating sites (George Washington University [MH 20766], University of Pittsburgh [MH 37057], University of Oklahoma [MH 33769], Yale University, New Haven, Conn [MH 33897], Clara Institute of Psychiatry, Toronto, Ontario [MH 38251], and Rush Presbyterian–St Luke's Medical Center, Chicago, Ill [MH 35017]). The principal NIMH collaborators are Irene Elkin, PhD, coordinator; Alice Klerman, MD, associate coordinator (formerly at George Washington University), John P. Docherty, MD (now at Nashua Brookside Hospital, Nashua, NH), and Morris B. Parloff, PhD (now at The American University, Washington, DC). The principal investigators and project coordinators at the three participating research sites are as follows: George Washington University, Stuart M. Sotocay, MD, and David R. Glass, PhD; University of Pittsburgh, Stanley D. Imber, PhD, and Paul A. Pilkonis, PhD; and University of Oklahoma, John T. Watkins, PhD, and William R. Leber, PhD. The principal investigators and project coordinators at the three sites responsible for training therapists for the program: Carlow: Yale University, Almyna Weissman, PhD; George Washington University (University, New York, NY), Eve Chevron, MS, and Bruce J. Rounsaville, MD; Clara Institute of Psychiatry, Brian F. Shaw, PhD, and T. Michael Vallis, PhD; and Rush Presbyterian–St Luke's Medical Center, Jan A. Fawcett, MD, and Philip Epstein, MD. Collaborators in the data management and data analysis aspects of the programs are C. James Klett, PhD, Joseph F. Collins, ScD, and Roderic Gilles of the Veterans Administration Cooperative Studies Program, Perry Point, Md.

We would like to thank members of the Advisory Group to the Collaborative Study (Paul Chadoff, MD, Sol L. Garfield, PhD, Martin M. Katz, PhD, Donald F. Klein, MD, Perry London, PhD, Morris B. Parloff, PhD, Jeanne S. Phillips, PhD, Hans H. Strupp, PhD, and Eberhard H. Uhlenhuth, MD) for their important contributions throughout the study. We would also like to thank H. A. Ali, MA, for various helpful contributions to the study. Robert F. Prien, PhD, for his valuable comments on an earlier draft of the manuscript. Regina T. Dolan, MA, Deborah Dowdall, and Doreen Grant for their assistance in preparing the manuscript, and Robert M. A. Hirschfeld, MD, of the Mount Sinai and Personality Disorders Research Branch, for his advice and support. We are grateful for the many contributions of our statistical consultant, C. James Klett; we also profited from discussions with Jacob Cohen and John Bartko.

References


26. Hill CE. Adherence in cognitive-behavioral therapy, interpersonal therapy, and clinical management. Read before the annual meeting of the Society for Psychotherapy Research June 15, 1986; Santa Fe, NM.


