ORIGINAL ARTICLE

# Cognitive Therapy vs Medications in the Treatment of Moderate to Severe Depression

*Robert J. DeRubeis, PhD; Steven D. Hollon, PhD; Jay D. Amsterdam, MD; Richard C. Shelton, MD; Paula R. Young, PhD; Ronald M. Salomon, MD; John P. O'Reardon, MD; Margaret L. Lovett, MEd; Madeline M. Gladis, PhD; Laurel L. Brown, PhD; Robert Gallop, PhD*

**Background:** There is substantial evidence that antidepressant medications treat moderate to severe depression effectively, but there is less data on cognitive therapy's effects in this population.

**Objective:** To compare the efficacy in moderate to severe depression of antidepressant medications with cognitive therapy in a placebo-controlled trial.

**Design:** Random assignment to one of the following: 16 weeks of medications (n=120), 16 weeks of cognitive therapy (n=60), or 8 weeks of pill placebo (n=60).

**Setting:** Research clinics at the University of Pennsylvania, Philadelphia, and Vanderbilt University, Nashville, Tenn.

**Patients:** Two hundred forty outpatients, aged 18 to 70 years, with moderate to severe major depressive disorder.

**Interventions:** Some study subjects received paroxetine, up to 50 mg daily, augmented by lithium carbonate or desipramine hydrochloride if necessary; others received individual cognitive therapy.

**Main Outcome Measure:** The Hamilton Depression Rating Scale provided continuous severity scores and allowed for designations of response and remission.

**Results:** At 8 weeks, response rates in medications (50%) and cognitive therapy (43%) groups were both superior to the placebo (25%) group. Analyses based on continuous scores at 8 weeks indicated an advantage for each of the active treatments over placebo, each with a medium effect size. The advantage was significant for medication relative to placebo, and at the level of a nonsignificant trend for cognitive therapy relative to placebo. At 16 weeks, response rates were 58% in each of the active conditions; remission rates were 46% for medication, 40% for cognitive therapy. Follow-up tests of a site × treatment interaction indicated a significant difference only at Vanderbilt University, where medications were superior to cognitive therapy. Site differences in patient characteristics and in the relative experience levels of the cognitive therapists each appear to have contributed to this interaction.

**Conclusion:** Cognitive therapy can be as effective as medications for the initial treatment of moderate to severe major depression, but this degree of effectiveness may depend on a high level of therapist experience or expertise.

*Arch Gen Psychiatry. 2005;62:409-416*

**Author Affiliations:**
Departments of Psychology (Dr DeRubeis), and Psychiatry (Drs Amsterdam, Young, O'Reardon, and Gladis), University of Pennsylvania, Philadelphia ; Departments of Psychology (Dr Hollon), and Psychiatry (Drs Shelton, Salomon, Lovett, and Brown), Vanderbilt University, Nashville, Tenn; Department of Mathematics and Applied Statistics, West Chester University, West Chester, Pa (Dr Gallop).

ANTIDEPRESSANT MEDICATIONS (ADMs) are the most widely used treatment for major depressive disorder (MDD) in the United States.[1] Evidence from numerous randomized placebo-controlled trials has supported the efficacy of ADMs, particularly among more severely depressed patients.[2]

Cognitive therapy (CT), a type of cognitive behavioral therapy pioneered by Beck et al,[3] has also shown promise in the treatment of MDD.[4] Rush et al[5] initially reported that CT was more effective than ADM in a randomized, comparative trial. However, their ADM dosages were low and the medications were tapered 2 weeks before the final outcome assessment. Despite these shortcomings, these findings generated enthusiasm for CT as an alternative to ADM for the treatment of depression.

*See also page 417*

The Treatment of Depression Collaborative Research Program (TDCRP),[6] initiated by the National Institute of Mental Health, compared CT, ADM, pill placebo, and interpersonal psychotherapy treatments in depressed outpatients. No differences in outcome were observed between CT and ADM among all patients.



©2005 American Medical Association. All rights reserved.

However, in a secondary analysis of more severely depressed patients (intake Hamilton Depression Rating Scale [HDRS] scores of 20 or above), ADM outcomes were superior to both placebo and CT, and the outcomes of CT were not significantly superior to placebo.[7] Because of the size and methodological sophistication of the TDCRP, these findings have had a major impact on the field. For example, the American Psychiatric Association's *Practice Guidelines for Major Depressive Disorder in Adults*[4] recommended the use of ADM, and not CT, as first-line therapy for patients with moderate to severe MDD.

Concerns have been raised about the quality of the CT provided in the TDCRP, leading some to question whether these findings should supersede those of other randomized comparisons of ADM and CT in MDD.[8] Neither Murphy et al[9] nor Hollon et al[10] reported differences in efficacy between ADM and CT in their randomized trials, although both studies, like that of Rush et al,[5] included patients with MDD with a broad range of depression severity. A mega-analysis that focused on the more severely depressed patients from the 4 studies mentioned earlier yielded a finding that did not favor ADM over CT.[11]

To date, no comparison of ADM and CT has contained a large sample of patients with moderate to severe MDD, and only the TDCRP study included a placebo control group. We therefore conducted a large 2-site, placebo-controlled, randomized trial to test the relative efficacy of ADM and CT in outpatients with more severe MDD. We made special efforts to ensure that the provision of both ADM and CT was consistent with "best practices" in the respective treatment modalities.

## METHODS

The protocol was approved by the respective institutional review boards at the University of Pennsylvania, Philadelphia, and Vanderbilt University, Nashville, Tenn. All participants provided written informed consent prior to any research activity. Subjects were recruited from referrals and from media announcements that described the respective research clinics. Evaluations were conducted blind to treatment condition by interviewers who were trained and supervised by one of the authors (M.M.G.), who herself trained with the Biometrics Research Department (New York State Psychiatric Institute, New York). The Structured Clinical Interviews for *DSM-IV* Diagnosis (Axis I and Axis II) were used to determine diagnostic eligibility for the study.[12,13] In addition, the first 17 items of the 24-item HDRS[14,15] were used to determine whether the depressive symptoms of the subjects were severe enough for inclusion in the trial. There are numerous versions of the HDRS, but most investigators report the standard 17-item version. Selected items were modified to allow patients to be scored for either typical or atypical presentations of symptoms associated with sleep, appetite, and change in weight.[16] The remaining 7 HDRS items, including 3 that emphasize cognitive symptoms, were obtained in the interviews but were not employed in the primary analyses reported in this article. Secondary analyses including these additional items did not alter the basic pattern of the results. Interviewers at both sites (4 at Pennsylvania and 3 at Vanderbilt) rated a subset of these tapes. An intraclass correlation coefficient of 0.96 was obtained for the 17-item total HDRS score (n=24). Assessment of the reliability of the major depressive episode designation yielded a $\kappa$ coefficient of 0.80 (n=12).[17] Diagnoses were confirmed by an experienced research psychiatrist.

Inclusion criteria were: diagnosis of MDD according to *DSM-IV*[18] criteria, age 18 to 70 years, English speaking, and willingness and ability to give informed consent. Consistent with the TDCRP's definition of "more severely depressed," all included patients had scores of 20 or higher on the modified 17-item HDRS at the screen and baseline visits, separated by at least 7 days.

A total of 437 patients were evaluated for the study, 96 of whom either did not meet diagnostic criteria for MDD or did not achieve HDRS scores of 20 or higher at both the screen and baseline study visits. Another 101 patients met the following exclusion criteria: (1) history of bipolar I disorder (n=25); (2) substance abuse or dependence judged to require treatment (n=26); (3) current or past psychosis (n=10); (4) another *DSM-IV* Axis I disorder judged to require treatment in preference to the depression (anxiety disorders, n=5; eating disorders, n=2); (5) 1 of the 3 excluded *DSM-IV* Axis II disorders deemed to be poorly suited to the treatments under investigation (antisocial, n=3; borderline, n=8; schizotypal, n=1); (6) suicide risk requiring immediate hospitalization (n=4); (7) medical condition that contraindicated study medications (n=13); or (8) nonresponse to an adequate trial of paroxetine in the preceding year (n=4).

The 240 patients who met entry criteria were randomly assigned to 1 of 3 treatment conditions: ADM (n=120), pill placebo (n=60), or CT (n=60). The ADM cell was designed to have twice as many patients as the CT and pill placebo cells because responders to ADM at study week 16 were to be randomized a second time for a companion study of subsequent relapse prevention.[19] Antidepressant medication and CT were each provided for 16 weeks. For ethical reasons, the pill placebo condition was terminated after 8 weeks; this duration was sufficient to reveal differences in drug vs placebo.[20] Randomization was implemented after stratifying on sex and the number of prior episodes.

### TREATMENT PROCEDURES

#### Pharmacotherapy and Placebo

Patients in the pharmacotherapy cells were treated for 8 weeks with paroxetine or placebo. All 5 pharmacotherapists (3 at Pennsylvania, 2 at Vanderbilt) were male, board-certified psychiatrists with extensive experience (9-23 years) treating MDD pharmacologically. Patients in pharmacotherapy received weekly treatment sessions for the first 4 weeks, and every other week thereafter. Initial sessions typically lasted 30 to 45 minutes; subsequent sessions lasted about 20 minutes. During the first 8 weeks, patients and pharmacotherapists remained blind as to whether the pills contained paroxetine.

Pharmacotherapy sessions were conducted in accordance with the manual used in the TDCRP.[21] Jan Fawcett, MD, the author of the manual, provided training and consultation in clinical management throughout the study. Pharmacotherapy sessions focused on the following: (1) medication management, which involved education about medications, adjustment of dosage and dosage schedules, and discussions of adverse effects; and (2) clinical management, which involved a review of the patient's functioning in major life spheres, brief supportive counseling, and limited advice giving. Techniques and strategies specific to CT were prohibited.

All patients in the pharmacotherapy conditions began treatment with paroxetine or placebo (10-20 mg daily). This dose was raised in 10- to 20-mg increments as tolerated based on response and the occurrence of dose-limiting adverse effects,



©2005 American Medical Association. All rights reserved.

to a maximum of 50 mg daily by week 6 of treatment, or until a significant reduction in symptoms was seen. The minimum acceptable dose was 20 mg/d; 10 mg/d if 20 mg was not tolerated.

After 8 weeks of pharmacotherapy, the double-blind condition was broken for the patient and pharmacotherapist (but not for the interviewer). Patients who had been given placebo were offered treatment without cost. Those in the ADM cell continued receiving their established paroxetine dose. For patients who did not meet the established response criteria (discussed later) by week 8 of treatment, augmentation with lithium carbonate or desipramine hydrochloride was initiated, unless there was an overriding clinical consideration, such as intolerance of paroxetine, in which case the pharmacotherapist was free to prescribe another ADM.

### Cognitive Therapy

Cognitive therapy was provided by 6 therapists, 3 at each site. One therapist at each site was female. Five therapists were licensed psychologists with PhD degrees; 1 at Vanderbilt was a psychiatric nurse practitioner. Their experience with psychotherapy ranged from 5 to 21 years at the beginning of the trial. All 3 Pennsylvania therapists and 1 of the Vanderbilt therapists had extensive experience conducting CT (7-21 years). The other 2 Vanderbilt therapists each began the trial with 2 years of CT experience. These 2 therapists received training through the Beck Institute for Cognitive Therapy (Bala Cynwyd, Pa) during the trial and were judged to have met criteria for competence as a result of this supplemental training.

All therapists followed the procedures outlined in standard texts of cognitive therapy for depression[3,22] and for comorbid personality disorders.[23] Guidelines called for 50-minute sessions to be held twice weekly for the first 4 weeks of treatment, once or twice weekly for the middle 8 weeks, and once weekly for the final 4 weeks. At each site, cognitive therapists met together weekly for 90 minutes to review ongoing cases.

## OUTCOME MEASURES

The modified 17-item HDRS was used as the primary outcome measure.[14-16] All evaluations were videotaped. We will report on categorical indexes that use the HDRS at 8 weeks, when all 3 conditions were compared, and at 16 weeks, when only the 2 active treatments were compared. We employed a response criterion based on absolute levels of symptoms, rather than percentage reduction, to ensure that no patient was classified as a responder who exhibited a high level of symptomatology. At 8 weeks, the criterion for response was an HDRS score of 12 or lower, with the last observation carried forward in the case of study dropouts.[24]

Response criteria for the 16-week comparison were also based on an HDRS score of 12 or less, but they were designed to prevent a transient exacerbation of depressive symptoms from keeping a patient from meeting response criteria. The criteria were: completion of the 16-week treatment phase and one of the following: (1) 16-week HDRS score of 12 or lower and either a 14-week score of 14 or lower or a 10- and 12-week score of 12 or lower; or (2) a 12-, 14-, and an 18-week HDRS score of 12 or lower. The 16-week response designation was also used to determine which patients would be invited to participate in the succeeding phase of the study, an investigation of relapse prevention among responders to acute ADM or CT.[19] All of these response criteria were also applied to determine full remission, with the additional requirement of the final HDRS score having been 7 or less during the acute phase, consistent with the MacArthur recommendation.[25]

## STATISTICAL ANALYSES

Based on numerous placebo-controlled randomized clinical trials, we estimated that the drug-placebo comparison would yield an effect size of a one-half standard deviation on the HDRS (estimated mean difference=3.5, estimated SD=7). Based on these estimates, a cell size of at least 60 was required to achieve a power of 0.8 or more to detect a difference at the 0.05 level, 2-tailed.

Treatment differences in response categories at 8 and 16 weeks were examined using the Cochran Q test.[26] The Cochran Q test approach extends the $\chi^2$ test of association for contingency tables to sets of contingency tables. Within-site comparisons and tests of site × treatment interactions were assessed using logistic regression models based on the Wald $\chi^2$.[27] At 8 weeks, the association for 2 sets of $3 \times 2$ tables was assessed. (Note that the 2 sets correspond to the 2 sites of the study, 3 corresponds to the 3 treatments, and the second 2 corresponds to response/nonresponse.) Under the null hypothesis, there is no difference in response classification for the 3 levels of treatment adjusting for site.[28] Similarly, the 16-week contrasts consisted of assessing the association for the 2 sets of $2 \times 2$ tables. (Note that only 2 treatments are involved in tests of the 16-week data.)

For continuous data, the analytical method was a multilevel model adjusting for the repeated measures with nested random effects. This analysis falls under the heading of random coefficient models, hierarchical linear models (HLM), and multilevel linear models.[29,30] In this approach, each subject's growth curve is characterized by a set of person-specific parameters. The standard HLM involves 2 levels: within-subject (level 1) and between-subject (level 2). At level 1, the outcome varies within subjects over time as a function of a person-specific growth curve. At level 2, the person-specific change parameters are viewed as varying randomly across subjects, as a function of the subject's treatment. The combination of the level 1 and level 2 models results in a mixed linear model with fixed and random coefficients. The person-specific parameters correspond to a random intercept and a random slope for each subject. The random terms are assumed to follow a bivariate normal distribution, which allows the random terms to be correlated. This is commonly referred to as an unstructured covariance structure. Analyses that fall under the heading of HLM provide accurate statistical inferences for data that have a nested or hierarchical structure by modeling the within-subject correlation by the inclusion of random effects.[29,30] Ignoring this within-subject correlation for nested data may result in underestimation of the variance in the model, which would result in significance levels closer to 0. Therefore, more powerful HLMs were used to answer our primary questions that concern continuous outcomes. The HLM analysis examined differences in linear rates of change between ADM, CT, and pill placebo groups in HDRS scores over the course of the first 8 weeks. Similarly, a second HLM examined the difference in linear rates of change between ADM and CT in HDRS scores over the 16-week acute phase. For these models all available data were used, making the HLM application a full intent-to-treat analysis. For dropouts, all and only those data gathered prior to the date of attrition were used in these models. The effects of site and of initial baseline HDRS total score were covaried. Because we collected a second baseline score (at least 1 week after the initial one) on each of the 240 patients, we were able to conduct a full intent-to-treat analysis on these data, even when including the initial baseline as a covariate. The model (performed using SAS version 8.0, PROC MIXED; SAS Institute Inc, Cary, NC) estimated fixed effects for treatment and for site. Population-averaged estimates for the linear trend over time and linear trend over time per treatment are produced by this



©2005 American Medical Association. All rights reserved.

**Table. Baseline Characteristics**

| Variable | Whole Sample (n = 240) | University of Pennsylvania (n = 120) | Vanderbilt University (n = 120) |
|---|---|---|---|
| HDRS score, mean ± SD | 23.4 ± 2.9 | 23.4 ± 2.9 | 23.4 ± 2.8 |
| Female, %* | 59 | 52 | 65 |
| White, % | 82 | 75 | 88 |
| Age, mean ± SD, y | 40 ± 12 | 39 ± 12 | 42 ± 11 |
| Education, mean ± SD, y | 15 ± 2 | 15 ± 3 | 15 ± 2 |
| Income, mean ± SD, (in thousands of US $) | 34 ± 33 | 32 ± 33 | 36 ± 34 |
| Married/cohabitating, % | 33 | 33 | 33 |
| Employed, % | 82 | 82 | 82 |
| Chronic or recurrent MDD, % | 90 | 84 | 95 |
| Age of onset, mean ± SD, y* | 22 ± 12 | 25 ± 11.7 | 20 ± 12 |
| No. of prior episodes, mean ± SD | 2.4 ± 2.6 | 2.3 ± 3.0 | 2.6 ± 2.2 |
| Duration of current episode, mean ± SD, mo | 46 ± 75 | 46 ± 88 | 46 ± 61 |
| Dysthymia, %* | 25 | 9 | 41 |
| History of prior ADM therapy, %* | 60 | 47 | 74 |
| History of psychiatric hospitalization, % | 12 | 8 | 15 |
| Met *DSM-IV* specifiers for melancholia, % | 30 | 33 | 28 |
| Met *DSM-IV* specifiers for atypical features, %* | 22 | 15 | 28 |
| Axis I or Axis II comorbidity, % | 83 | 76 | 91 |
| Any Axis I comorbidity, %* | 72 | 63 | 82 |
| Comorbid anxiety disorders, %* | 53 | 41 | 65 |
| Comorbid PTSD, %* | 17 | 5 | 28 |
| Comorbid eating disorders, %* | 17 | 9 | 24 |
| Comorbid substance abuse or dependence, % (primarily by history) | 35 | 34 | 36 |
| Any Axis II comorbidity, %* | 48 | 38 | 57 |
| Cluster A personality disorder, % | 3 | 3 | 3 |
| Cluster B personality disorder, % | 3 | 2 | 5 |
| Cluster C personality disorder, %* | 28 | 19 | 38 |
| Personality disorder NOS, % | 17 | 19 | 14 |

Abbreviations: ADM, antidepressant medication; HDRS, Hamilton Depression Rating Scale; MDD, major depressive disorder; NOS, not otherwise specified; PTSD, posttraumatic stress disorder.
*Indicates significant site difference ($P<.05$).

model. The model tests the linear slope difference between the groups. Differential rates of change per treatment between sites were assessed by the site × time × treatment interaction. If nonsignificant, this interaction term was removed from the model.

## RESULTS

### BASELINE CHARACTERISTICS

A total of 240 patients were randomized to treatments, 120 at each site. Baseline HDRS score means did not differ between conditions or between sites (overall mean ± SD = 23.4 ± 2.9; range = 20-35). Other important baseline variables are presented in the **Table** along with results of *t* tests for continuous variables and $\chi^2$ tests for binary variables. The modal patient in the sample was middle-aged, white, with partial college education and modest income. One third of the sample was married or cohabitating. The Pennsylvania sample, relative to the Vanderbilt sample, was more likely to be male and ethnically and racially diverse.

Overall, but especially at Vanderbilt, the sample was highly chronic or recurrent, with early onsets and a substantial rate of prior hospitalizations. Comorbidity rates were high at Pennsylvania, and even higher at Vanderbilt. Nearly three quarters of the patients met criteria for an Axis I comorbidity, the most common of which were the anxiety disorders. Nearly half the patients met criteria for at least 1 Axis II disorder.

Of all the variables listed in the Table, the rates of substance abuse, Axis I comorbidity, melancholic depression, and atypical depression differed significantly as a function of treatment condition. Because none of these variables predicted response across the treatments, no confounds were identified that would compromise the tests of comparative efficacy. Secondary analyses with these variables as covariates yielded the same pattern of results as those without these covariates. Thus, these variables were not included as covariates in the analyses reported in this article.

### ATTRITION

During the first 8-week treatment period, the overall attrition rate in the sample was 13%. Attrition was 11% in the ADM cell (8 at Pennsylvania, 5 at Vanderbilt); 15% in the CT cell (6 at Pennsylvania, 3 at Vanderbilt); and 13% in the pill placebo cell (4 at each site). One patient in the ADM cell and 2 patients in the pill placebo cell withdrew consent immediately following randomization, while 9 others withdrew consent during treatment without stating a specific reason (5 in the ADM cell and 5 in the CT cell). Only 6 patients dropped out



©2005 American Medical Association. All rights reserved.

owing to adverse effects: 4 (3%) in the ADM cell, and 2 (3%) in the pill placebo cell. Four (7%) patients in the CT cell and 4 (7%) patients in the pill placebo cell dropped out because of dissatisfaction with treatment. Two patients (2%) were withdrawn from ADM treatment owing to worsening of symptoms. One patient in the ADM cell committed suicide during the second week of treatment.

Over the second 8 weeks of the trial, 6 patients in the ADM cell (4 at Pennsylvania and 2 at Vanderbilt; 5% overall) dropped out because of adverse effects (n=4) or because they were no longer interested in treatment (n=2), whereas no patient in the CT cell dropped out during this period. Thus, over the 16-week treatment course, attrition rate was 16% for the ADM cell and 15% for CT. Attrition did not differ significantly between sites or across conditions after 8 weeks or after 16 weeks.

## MEDICATION DOSAGES

The mean (±SD) daily paroxetine dose during the first week of treatment was 14.0±4.9 mg (Pennsylvania=12.4±4.3; Vanderbilt=15.5±5.0; $P<.001$). The mean daily dosage was increased to 31.6±11.2 by the fourth week (Pennsylvania=30.4±12.4; Vanderbilt=32.8±9.9; $P=.26$), and to 38.8±11.0 at week 8 (Pennsylvania=38.3±11.9; Vanderbilt=39.3±10.2; $P=.66$).

Mean daily paroxetine dosage over the second 8-week treatment period was 37.3±12.4 mg (Pennsylvania=33.5±14.2; Vanderbilt=40.8±9.4; $P=.003$). One patient was switched to buproprion and another to sertraline owing to intolerance of paroxetine; the dosages of these patients were excluded from the calculations of means. The difference in average dosage between sites was primarily due to differential prescribing in the patients with augmented treatment (47 of the 101 patients). Patients with nonaugmented treatments (21 at Pennsylvania, 33 at Vanderbilt) received similar dosages at the 2 sites (Pennsylvania=36.0±12.6; Vanderbilt=39.5±8.2; $P=.28$). However, among the patients with augmented treatments (Pennsylvania n=27, Vanderbilt n=20), mean paroxetine dosages were lower at Pennsylvania than at Vanderbilt (Pennsylvania=31.8±15.2; Vanderbilt=42.9±10.8; $P=.004$). This difference is in contrast to the mean paroxetine dosages of patients with augmented treatment at the end of 8 weeks of treatment, which were similar between sites (Pennsylvania=40.4±10.7, Vanderbilt=44.3±8.7; $P=.16$). The difference in prescribing patterns between Pennsylvania and Vanderbilt among the patients with augmented treatment was not planned or addressed by the protocol. In the absence of specific guidelines, Pennsylvania psychiatrists followed conventional practice more closely in this regard, whereas Vanderbilt psychiatrists followed a more aggressive strategy than is typically practiced. Examination of the data did not suggest that the site differences in outcome in the ADM cell could be explained by this difference in practices between the sites.

Of the 47 patients who received augmentation treatment, 32 (64%) were prescribed lithium, 28 (56%) were prescribed desipramine, and 1 (2%) was treated with venlafaxine. (Percentages add to greater than 100% because some patient treatments were augmented with more than 1 drug, either in combination or in sequence.)

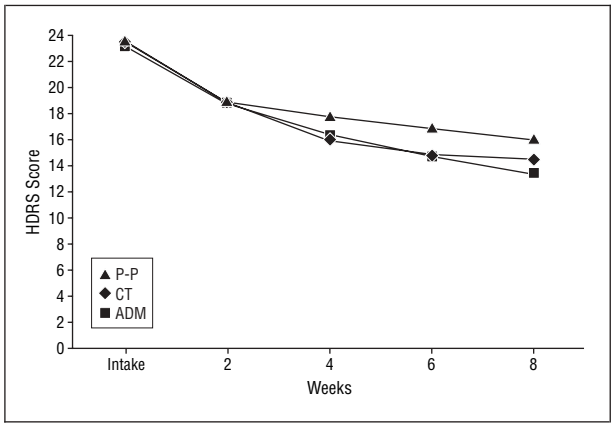



**Figure 1.** Biweekly Hamilton Depression Rating Scale (HDRS) scores during the first 8 weeks of acute treatment. ADM indicates antidepressant medication therapy (n=120); CT, cognitive therapy (n=60); and P-P, pill placebo (n=60).

## OUTCOME AT 8 WEEKS (PLACEBO-CONTROLLED)

### Categorical Analyses at 8 Weeks

Rates of response at 8 weeks (≤12 on the HDRS, with last observation carried forward for dropouts) were 50% in the ADM cell (95% confidence interval (CI), 41%-59%), 43% in the CT cell (95% CI, 31%-56%), and 25% in the pill placebo cell (95% CI, 16%-38%). A Cochran Q test controlling for site indicated a difference in response across the 3 treatments ($\chi^2_2=10.22$, $P=.006$). Pairwise comparisons indicated a significant difference in favor of ADM compared with pill placebo ($\chi^2_1=10.17$, $P=.001$), and a significant difference in favor of CT compared with pill placebo ($\chi^2_1=4.44$, $P=.04$). The pairwise comparison of ADM with CT was nonsignificant ($\chi^2_1=0.71$, $P=.40$). A logistic regression model indicated that response rates across treatment conditions were not differential between the sites (Wald $\chi^2_2=1.59$, $P=.45$).

### Continuous Analyses at 8 Weeks

Across the first 8 weeks of treatment, the test of a site × treatment interaction was not significant ($F_{2,228}=1.65$, $P=.19$). As displayed in **Figure 1**, there was improvement during the first 8 weeks across treatments ($F_{1,231}=395.9$, $P<.001$). The term in the model that tests for differential change on the HDRS as a function of treatment over time was significant ($F_{2,231}=3.81$, $P=.02$). Pairwise contrasts revealed a significant advantage of ADM compared with pill placebo ($F_{1,231}=7.60$ $P=.006$) and a nonsignificant trend in favor of CT compared with pill placebo ($F_{1,231}=2.96$, $P=.09$). Effect sizes, derived from the subject-specific slopes of the HLM model,[31] were 0.60 for ADM compared with pill placebo, and 0.44 for CT relative to pill placebo. The difference between ADM and CT was not significant ($F_{1,231}=0.56$; $P=.46$); the associated effect size estimate was 0.16, in favor of ADM. Effect sizes between 0.5 and 0.8 can be considered "medium" in magnitude, and effect sizes between 0.2 and 0.5 "small."[32]



©2005 American Medical Association. All rights reserved.

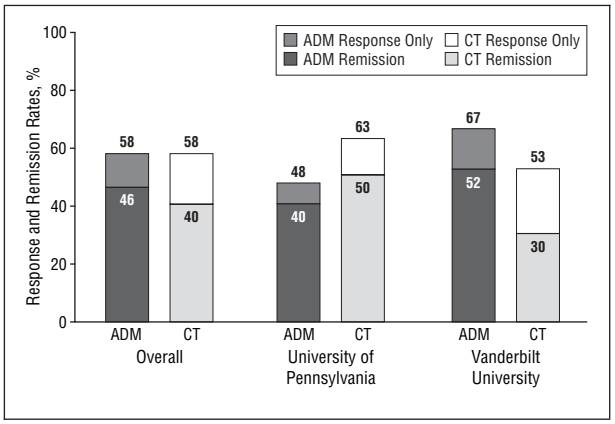



**Figure 2.** Response and remission rates at 16 weeks for antidepressant medication therapy (ADM) (n=120, 60 per site) and cognitive therapy (CT) (n=60, 30 per site).

## OUTCOME AT 16 WEEKS (ACTIVE TREATMENTS ONLY)

### Categorical Analyses at 16 Weeks

As displayed in **Figure 2**, 58% of the patients treated with ADM (95% CI, 48%-66%) and 58% of the patients in the CT cell (95% CI, 45%-70%) met the response criteria at 16 weeks. This difference was not significant ($\chi^2_1$=0.01, *P*=.92). The test of a site × treatment interaction yielded a nonsignificant trend (Wald $\chi^2_1$=3.31, *P*=.07). The difference in response rates was not significant at either Pennsylvania (Wald $\chi 2_1$=1.83, *P*=.18) or Vanderbilt (Wald $\chi^2_1$=1.50, *P*=.22).

Remission rates were 46% in the ADM cell (95% CI, 37%-55%) vs 40% in the CT cell (95% CI, 28%-53%). The test of site × treatment interaction was significant (Wald $\chi^2_1$=4.15, *P*=.04). At Pennsylvania, the remission rates were not significantly different (Wald $\chi^2_1$=0.81, *P*=.37), whereas at Vanderbilt, the remission rate in ADM was higher than the remission rate in CT at the level of a nonsignificant trend (Wald $\chi^2_1$=3.80, *P*=.05).

### Continuous Analyses at 16 Weeks

After 16 weeks, there was no significant main effect for treatment in the HLM analysis, but there was a significant site × treatment interaction ($F_{1,173}$=5.51, *P*=.02; effect size=0.36). As a result, we report analyses separately for each site (**Figure 3**). At Pennsylvania the difference between the groups was not significant ($F_{1,173}$=1.67, *P*=.2, effect size=0.37), whereas at Vanderbilt, change in the ADM cell was significantly greater than change in the CT cell ($F_{1,173}$=4.19, *P*=.04, effect size=0.57).

To explore differences between the populations at the 2 sites that might explain the site × treatment interaction, we first identified pretreatment variables that predicted responses within the respective treatments (at *P*<.10, using the subject-specific slopes as estimated through the HLM model).

In the ADM cell, patients with Axis I comorbidity fared better than patients without Axis I comorbidity (ES=0.17, *P*=.07). Of the patients with Axis I comorbidity, presence of generalized anxiety disorder met this criterion (ES=0.30, *P*=.02). Also in the ADM cell, patients fared worse if they had chronic depression (ES=-0.24, *P*=.004) or were unemployed (ES=-0.25, *P*=.04). Of the 4 variables that predicted response in the ADM cell, only Axis I comorbidity was differentially prevalent in the ADM conditions at the 2 sites (Pennsylvania=52%, Vanderbilt 87%; $\chi^2$=17.23, *P*<.001). Patients with Axis I comorbidity fared as well in the ADM cell at Pennsylvania as they did at Vanderbilt (slope estimate at Pennsylvania=-0.82, SE=0.08; slope estimate at Vanderbilt=-0.82, SE=0.06). Thus, the higher rate of Axis I comorbidity at Vanderbilt contributed to the superior performance of patients in the ADM cell at Vanderbilt, relative to Pennsylvania.

In the CT cells, patients fared better if they met criteria for the melancholic subtype of MDD (ES=0.40, *P*=.02); they fared worse if they were comorbid for social phobia (ES=-0.26, *P*=.06). Neither of these variables was associated with differential prevalence in the CT conditions at the 2 sites.

## COMMENT

Both pharmacotherapy and CT treatments outperformed placebo in moderately to severely depressed outpatients. The finding of superiority of an established medication relative to placebo is often regarded as essential in any comparison between the established treatment and a newer treatment.[33]

Comparisons between the 2 active treatments in this trial must take into account a site × treatment interaction. Such interactions reflect either differences in patients sampled or differences in treatment procedures. There is evidence of a contribution from each of these kinds of factors in the present study.

With respect to differences in the samples, Axis I comorbidity was differentially prevalent in the ADM conditions at the 2 sites. Patients with Axis I comorbidity, who were primarily comorbid with anxiety disorders, fared particularly well on ADM, perhaps because paroxetine, the primary medication in this study, has anxiolytic effects. These patients with comorbidity responded as well at Pennsylvania as they did at Vanderbilt, but the higher prevalence of these patients at Vanderbilt led to a better overall response to ADM at Vanderbilt.

With respect to differences in the treatments at the 2 sites, the superior performance of CT at Pennsylvania, relative to Vanderbilt, was likely related to therapist experience. The more experienced cognitive therapists at Pennsylvania produced outcomes at least comparable to those produced by ADM, whereas Vanderbilt's less experienced cognitive therapists produced outcomes that were inferior to those produced by ADM at that site. This calls to mind the site differences observed in the TDCRP.[7] Among their more severely depressed patients, differences favoring ADM relative to CT were large at the 2 sites that had less experience with CT, and were negligible at the one site that had greater experience with CT.[34]

On the whole, these findings do not support the current American Psychiatric Association guideline,[4] based



©2005 American Medical Association. All rights reserved.

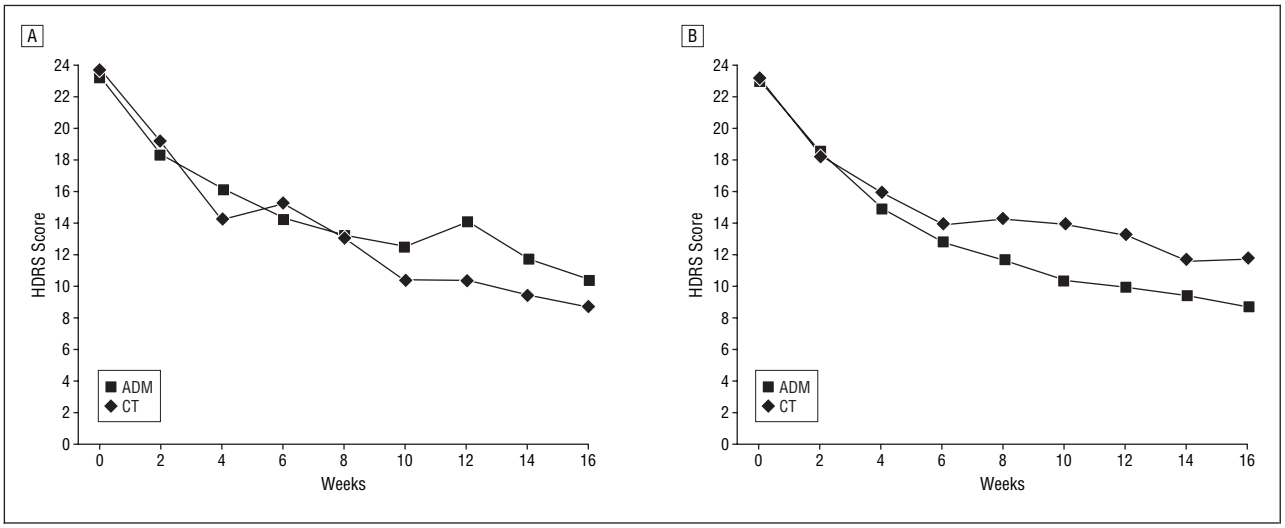



**Figure 3.** Biweekly Hamilton Depression Rating Scale (HDRS) scores for University of Pennsylvania (A) and Vanderbilt University (B) during the 16 weeks of acute treatment. ADM indicates antidepressant medication therapy (n=120, 60 per site); CT, cognitive therapy (n=60, 30 per site).

on the TDCRP,[7] that "most (moderately and severely depressed) patients will require medications." It appears that cognitive therapy can be as effective as medications, even among more severely depressed outpatients, at least when provided by experienced cognitive therapists.

**Submitted for Publication:** August 12, 2003; final revision received August 12, 2004; accepted September 9, 2004.
**Correspondence:** Robert J. DeRubeis, PhD, Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104-6196 (derubeis@psych.upenn.edu).
**Funding/Support:** This study was supported by grants MH50129 (R10) (Dr DeRubeis) and MH55875 (R10) and MH01697 (K02) (Dr Hollon) from the National Institute of Mental Health, Bethesda, Md. GlaxoSmithKline, Brentford, Middlesex, United Kingdom, provided medications and pill placebos for the trial.
**Previous Presentation:** Presented at the 155th Annual Convention of the American Psychiatric Association, May 23, 2002; Philadelphia, Pa.
**Acknowledgment:** Gratitude is expressed to our colleagues for contributing to this research. Robert J. DeRubeis, PhD, and Steven D. Hollon, PhD, were the principal investigators and oversaw the implementation of cognitive therapy at the respective sites. Jay D. Amsterdam, MD, and Richard C. Shelton, MD, were the coprincipal investigators and supervised the implementation of medication treatment. Edward Schweizer, MD, provided important consultation about study design and implementation, especially early in the trial. Paula R. Young, PhD, and Margaret L. Lovett, MEd, served as the study coordinators. John P. O'Reardon, MD, Ronald M. Salomon, MD, and the late Martin Szuba, MD, served as study pharmacotherapists (along with Drs Amsterdam and Shelton). Cory P. Newman, PhD, Karl N. Jannasch, PhD, Frances Shusman, PhD, and Sandra Seidel, MSN, served as the cognitive therapists (along with Drs DeRubeis and Hollon). Jan Fawcett, MD, provided consultation with regard to the implementation of clinical management pharmacotherapy. Aaron T. Beck, MD, Judith Beck, PhD, Christine Johnson, PhD, and Leslie Sokol, PhD, provided consultation with respect to the implementation of cognitive therapy. Madeline M. Gladis, PhD, and Kirsten L. Haman, PhD, oversaw the training of the clinical interviewers, and David Appelbaum, PsyD, Laurel L. Brown, PhD, Richard C. Carson, PhD, Barrie Franklin, PhD, Nana A. Landenberger, PhD, Jessica Londa-Jacobs, PhD, Julie L. Pickholtz, PhD, Pamela Fawcett-Pressman, MEd, Sabine Schmid, MA, Ellen D. Stoddard, PhD, Michael Suminski, PhD, and Dorothy Tucker, PhD, served as project interviewers. Robert Gallop, PhD, and Andrew J. Tomarken, PhD, provided statistical consultation. Joyce L. Bell, BA, Brent B. Freeman, BA, Cara C. Grugan, BA, Nathaniel R. Herr, BA, Mary B. Hooper, MS, Miriam Hundert, BSN, Veni Linos, MSc, and Tynya Patton, MA, provided research support. Kelly Bemis Vitousek, PhD, provided helpful comments on an earlier draft of the manuscript.

## REFERENCES

1. Olfson M, Klerman GL. Trends in the prescription of antidepressants by office-based psychiatrists. *Am J Psychiatry*. 1993;150:571-577.
2. Thase ME, Kupfer DJ. Recent developments in the pharmacotherapy of mood disorders. *J Consult Clin Psychol*. 1996;64:646-659.
3. Beck AT, Rush AJ, Shaw BF, Emery G. *Cognitive Therapy of Depression*. New York, NY: Guilford Press; 1979.
4. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder (revision). *Am J Psychiatry*. 2000;157 (suppl 4):1-45.
5. Rush AJ, Beck AT, Kovacs M, Hollon S. Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed patients. *Cognit Ther Res*. 1977;1:17-37.
6. Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB. NIMH Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch Gen Psychiatry*. 1989;46:971-982.
7. Elkin I, Gibbons RD, Shea MT, Sotsky SM, Watkins JT, Pilkonis PA, Hedeker D. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol*. 1995;63:841-847.
8. Jacobson NS, Hollon SD. Cognitive-behavior therapy versus pharmacotherapy:



©2005 American Medical Association. All rights reserved.

now that the jury's returned its verdict, it's time to present the rest of the evidence. *J Consult Clin Psychol*. 1996;64:74-80.
9. Murphy GE, Simons AD, Wetzel RD, Lustman PJ. Cognitive therapy and pharmacotherapy. *Arch Gen Psychiatry*. 1984;41:33-41.
10. Hollon SD, DeRubeis RJ, Evans MD, Wiemer MJ, Garvey MJ, Grove WM, Tuason VB. Cognitive therapy and pharmacotherapy for depression: singly and in combination. *Arch Gen Psychiatry*. 1992;49:774-781.
11. DeRubeis RJ, Gelfand LA, Tang TZ, Simons AD. Medications versus cognitive behavioral therapy for severely depressed outpatients: mega-analysis of four randomized comparisons. *Am J Psychiatry*. 1999;156:1007-1013.
12. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN)*. New York: Biometrics Research, New York State Psychiatric Institute; 2001.
13. Spitzer RL, Williams JBW, Gibbon M, First MB. *Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II, Version 1.0)*. Washington, DC: American Psychiatric Press; 1990.
14. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960; 23:56-62.
15. Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. *Arch Gen Psychiatry*. 1988;45:742-747.
16. Reimherr FW, Amsterdam JD, Quitkin FM, Rosenbaum JF, Fava M, Zajecka J, Beasley CM, Michelson D, Roback P, Sundell K. Optimal length of continuation therapy in depression: a prospective assessment during long-term fluoxetine treatment. *Am J Psychiatry*. 1998;155:1247-1253.
17. Fleiss JL, Cohen J. *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons; 1973.
18. American Psychological Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychological Association; 1994.
19. Hollon SD, DeRubeis RJ, Shelton RC, Amsterdam JD, Salomon RM, O'Reardon JP, Lovett ML, Young PR, Haman KL, Freeman BB, Gallop R. Prevention of relapse following cognitive therapy vs medications in moderate to severe depression. *Arch Gen Psychiatry*. 2005;62:417-422.
20. Hollon SD, Thase ME, Markowitz JC. Treatment and prevention of depression. *Psychological Science in the Public Interest*. 2002;3:39-77.
21. Fawcett J, Epstein P, Fiester SJ, Elkin I, Autry JH. Clinical management-imipramine/placebo administration manual: NIMH Treatment of Depression Collaborative Research Program. *Psychopharmacol Bull*. 1987;23:309-324.
22. Beck JS. *Cognitive Therapy: Basics and Beyond*. New York, NY: Guilford Press; 1995.
23. Beck AT, Freeman A. *Cognitive Therapy of Personality Disorders*. New York, NY: Guilford Press; 1990.
24. Amsterdam JD, Brunswick DJ, Gilbertini M. Sustained efficacy of gepirone-IR in major depressive disorder: a double-blind placebo substitution trial. *J Psychiatr Res*. 2004;38:259-265.
25. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, Rush AJ, Weissman MM. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry*. 1991;48:851-855.
26. Kuritz SJ, Landis JR, Koch GG. A General overview of Mantel-Haenszel methods: applications and recent developments. *Annu Rev Public Health*. 1988; 9:123-160.
27. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: Wiley; 1989.
28. Stokes ME, Davis CS, Koch GG. *Categorical Data Analysis Using the SAS System*. Cary, NC: SAS Institute Inc; 1995.
29. Bryk A, Raudenbush S. *Hierarchical Linear Modeling: Applications and Data Analysis Methods*. Newbury Park, Calif: Sage Publishing; 1996.
30. Goldstein H. *Models in Educational and Social Research*. New York, NY: Oxford University Press; 1987.
31. Raudenbush SW, Xiao-Feng L. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychol Methods*. 2001;6:387-401.
32. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
33. Klein DF. Preventing hung juries about therapy studies. *J Consult Clin Psychol*. 1996;64:74-80.
34. Jacobson NS, Hollon SD. Prospects for future comparisons between drugs and psychotherapy: lessons from the CBT-versus-pharmacotherapy exchange. *J Consult Clin Psychol*. 1996;64:104-108.



©2005 American Medical Association. All rights reserved.