Journal of Consulting and Clinical Psychology
2006, Vol. 74, No. 4, 658–670

0022-006X/06/$12.00 DOI: 10.1037/0022-006X.74.4.658

Randomized Trial of Behavioral Activation, Cognitive Therapy, and Antidepressant Medication in the Acute Treatment of Adults With Major Depression

Sona Dimidjian
University of Washington

Steven D. Hollon
Vanderbilt University

Keith S. Dobson
University of Calgary

Karen B. Schmaling
University of North Carolina at Charlotte

Robert J. Kohlenberg
University of Washington

Michael E. Addis
Clark University

Robert Gallop
West Chester University

Joseph B. McGlinchey
Brown University

David K. Markley
University of Washington

Jackie K. Gollan
University of Chicago

David C. Atkins
Fuller Graduate School of Psychology

David L. Dunner and Neil S. Jacobson
University of Washington

Antidepressant medication is considered the current standard for severe depression, and cognitive therapy is the most widely investigated psychosocial treatment for depression. However, not all patients want to take medication, and cognitive therapy has not demonstrated consistent efficacy across trials. Moreover, dismantling designs have suggested that behavioral components may account for the efficacy of cognitive therapy. The present study tested the efficacy of behavioral activation by comparing it with cognitive

Sona Dimidjian, Robert J. Kohlenberg, David K. Markley, and Neil S. Jacobson, Department of Psychology, University of Washington; Steven D. Hollon, Department of Psychology, Vanderbilt University; Keith S. Dobson, Department of Psychology, University of Calgary, Calgary, Alberta, Canada; Karen B. Schmaling, College of Health and Human Services, University of North Carolina at Charlotte; Michael E. Addis, Department of Psychology, Clark University; Robert Gallop, Department of Mathematics and Applied Statistics, West Chester University; Joseph B. McGlinchey, Department of Psychiatry and Human Behavior, Brown University; Jackie K. Gollan, Department of Psychiatry, University of Chicago; David C. Atkins, Travis Research Institute, Fuller Graduate School of Psychology; David L. Dunner, Department of Psychiatry and Behavioral Sciences, University of Washington.

This article is based in part on Sona Dimidjian's doctoral dissertation at the University of Washington. David L. Dunner has received recent research funding from, is a consultant or on the advisory board for, and serves on the speaker's bureau of a number of pharmaceutical companies, including GlaxoSmithKline. Portions of this article were presented at the 37th Annual Convention of the Association for Advancement of Behavior Therapy, Boston, Massachusetts, November 2003. GlaxoSmithKline provided medications and pill placebos for the trial. The research was supported by National Institute of Mental Health Grant MH55502 (R01) first to Neil S. Jacobson and, after his death, to David L. Dunner.

Neil S. Jacobson passed away before the publication of this article. He was the originator of this study and shaped every aspect of its inception and implementation. His untimely death in 1999 left an irreplaceable gap, but his ideas about behavior therapy and his commitment to empirical inquiry have continued to serve as an inspiration and guide. We also gratefully acknowledge our colleagues for their many contributions to this study. Sandra Coffman and Christopher Martell provided on-site supervision and treatment for the cognitive therapy and behavioral activation conditions, respectively. Steve Sholl and David Kosins provided cognitive therapy. Ruth Herman-Dunn and Tom Linde provided behavioral activation therapy. Linda Cunning, Steven Dager, Kerri Halfant, Helen Hendrickson, and Alan Unis provided pharmacotherapy. Carolyn Bea and Chris Budech coordinated the pharmacotherapy conditions. Peggy Martin completed medical evaluations for the study. Lisa Roberts and Elizabeth Shilling were participant coordinators. Patty Bardina, Evelyn Mercier, Shireen Rizvi, Mandy Steiman, and Dan Yoshimoto were project evaluators. Melissa McElrea, Kim Nomensen, and Eric Gortner provided research support. Marina L. Smith, Jennifer A. Jones, Patricia R. Symons, Sonia Venkatraman, and Melissa Wisler conducted the adherence ratings. Leslie Sokol provided external ratings of the competence of cognitive therapy, and Jan Fawcett provided external ratings of pharmacotherapy. Marsha Linehan provided extremely valuable comments on the results and their implications. Virginia Rutter has been an unwavering supporter of this research, and we are grateful for her ongoing collaboration.

Correspondence concerning this article should be addressed to Sona Dimidjian, who is now at the Department of Psychology, University of Colorado, Boulder, CO 80309-0345. E-mail: sona.dimidjian@colorado.edu

therapy and antidepressant medication in a randomized placebo-controlled design in adults with major depressive disorder ($N = 241$). In addition, it examined the importance of initial severity as a moderator of treatment outcome. Among more severely depressed patients, behavioral activation was comparable to antidepressant medication, and both significantly outperformed cognitive therapy. The implications of these findings for the evaluation of current treatment guidelines and dissemination are discussed.

Keywords: behavioral activation, cognitive therapy, antidepressant medication, major depression

Antidepressant medications (ADMs) are the standard treatment for depression, particularly more severe major depression (American Psychiatric Association, 2000), and represent the most common form of treatment for major depression (Olfson, Marcus, Druss, & Pincus, 2002). However, ADM is not useful for every depressed person, and not all individuals want to take medications, particularly given the side effects that often accompany their use (American Psychiatric Association, 2000).

Of the psychosocial treatments for depression, cognitive therapy (CT) has been the most extensively studied, with numerous outcome studies documenting its efficacy (Hollon, Thase, & Markowitz, 2002). However, in the largest and best known controlled treatment trial, the National Institute of Mental Health Treatment of Depression Collaborative Research Program (TDCRP), CT was less effective than ADM and no more effective than placebo among more severely depressed participants (Elkin et al., 1995). Although this study has had considerable influence on the field, questions have been raised about the adequacy with which the CT was implemented (Jacobson & Hollon, 1996). Furthermore, a subsequent mega-analysis pooling data from the TDCRP and other relevant studies failed to find significant differences between ADM and CT among more severely depressed participants (DeRubeis, Gelfand, Tang, & Simons, 1999). The uncertainty surrounding the relative efficacy of CT and ADM highlights the importance of studies that include controls and ensure that the interventions are adequately implemented.

The emergence of CT over the past 2 decades eclipsed more behavioral approaches; however, findings from a component analysis of CT suggest that the behavioral components alone worked as well as the full package and may hold greater public health relevance (Jacobson et al., 1996). Specifically, the behavioral activation (BA) component alone produced as much change in depressive symptoms as the full CT condition during acute treatment and evidenced no more relapse than CT over a 2-year follow-up (Gortner, Gollan, Dobson, & Jacobson, 1998; Jacobson et al., 1996). Other process-oriented research on CT has similarly highlighted the value of behavioral strategies, suggesting that a focus on creating cognitive changes about interpersonal relationships was associated with worse functioning after CT, whereas a focus on creating actual interpersonal change was associated with improvement (A. M. Hayes, Castonguay, & Goldfried, 1996). Taken together, these findings provided additional support for behaviorists, who had long questioned whether the cognitive interventions in cognitive–behavioral therapies were essential to its success (Jacobson, Martell, & Dimidjian, 2001). These data also revitalized interest in purely behavioral treatments for depression and led to the development of a more fully realized behavioral intervention based on a contextual approach (Martell, Addis, & Jacobson, 2001).

Whereas the earlier model of BA tested in the component analysis study was defined primarily by the proscription of cognitive interventions, the fundamental principle of the expanded BA model is the use of idiographic functional analysis for the understanding of depressive behavior and contextual interventions for its remediation. The BA approach is rooted in the behavioral tradition established by Ferster (1973) and Lewinsohn (1974), both of whom identified the link between avoidant behavior and depression and recommended activation strategies to undermine punishment and increase positive reinforcement from the environment (see also Rehm, 1977). The expansion of the BA component treatment is an attempt to renew focus on the purely behavioral aspects of these traditions, which were largely overlooked in recent decades.

The current study was developed as a replication and extension of both the TDCRP and the component analysis study, addressing the principal criticisms and methodological shortcomings of each. It also paralleled many features of the DeRubeis et al. (2005) study, which compared ADM and CT. The current study compared BA, CT, and ADM in the context of a placebo-controlled trial and included careful steps to ensure the fidelity of the respective treatments. The present study had two primary aims. First, it tested the relative efficacy of BA in the acute treatment of major depression by comparing it both with CT alone and with ADM in the context of a placebo-controlled trial. Second, it tested whether either psychosocial treatment was a viable alternative to ADM in the treatment of moderate to severe depression. Primary predictions specified a significant advantage for ADM over placebo for severely depressed participants and no significant differences between the active treatments. No differences were expected with the less severely depressed participants.

Method

Participants

The University of Washington Institutional Review Board (IRB) approved the protocol.[1] All participants provided written informed consent prior to enrollment in the study. Participants consisted of 241 individuals between the ages of 18 and 60 years who met criteria for major depression according to the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [*DSM–IV*]; American Psychiatric Association, 1994) and scored 20 or higher on the Beck Depression Inventory (BDI-II; A. T. Beck, Steer, & Brown, 1996) and 14 or greater on the 17-item Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960). All *DSM–IV* diagnoses were made

[1] IRB approval inadvertently lapsed for approximately 6 weeks at the time of the death of Neil S. Jacobson (the original principal investigator); approval for use and publication of data collected during that time was subsequently granted by the IRB.

using the Structured Clinical Interview for the *DSM–IV* Axis I Disorders (SCID-I; First, Spitzer, Gibbon, & Williams, 1997). Recruitment occurred between 1998 and 2001; the majority of participants were recruited from media advertisements ($n = 150$; 62%), a substantial minority by referral from local agencies ($n = 64$; 27%), and the rest by word of mouth or other referral sources ($n = 27$; 11%).

Participants were excluded if they had a lifetime diagnosis of psychosis or bipolar disorder, organic brain syndrome, or mental retardation. Additional exclusion criteria included the following: substantial and imminent suicide risk; a current (e.g., within the past 6 months) or primary diagnosis of alcohol or drug abuse or dependence or a positive toxicology screen; a primary diagnosis of panic disorder, obsessive–compulsive disorder, psychogenic pain disorder, anorexia, or bulimia; or presence of antisocial, borderline, or schizotypal personality disorder. In addition, participants who had not responded favorably within the preceding year to an adequate trial of either CT or paroxetine also were excluded.

Because medications were administered in the trial, individuals also were required to have satisfactory results from a physical examination, laboratory screen (complete blood count, complete metabolic panel, thyroid screen including TSH, T3, T4, and urinalysis), and electrocardiogram (if over 40 years of age). Participants were excluded if they had an unstable medical condition, were using any medication that would complicate the administration of paroxetine, or had a known allergy to paroxetine. Moreover, women were not enrolled if pregnant, lactating, or not using suitable contraception if capable of becoming pregnant.

Procedure

Participants who passed an initial diagnostic telephone screening were scheduled for an on-site clinical evaluation to provide informed consent and ascertain study eligibility. If eligible, participants then completed a medical evaluation to assess possible medical contraindications. Once eligibility was determined, participants were assigned by the participant coordinator to one of four acute treatment conditions using a computer-generated randomization list: BA, CT, ADM, or pill placebo (PLA). Twice as many participants were assigned to the ADM condition to accommodate the design of the continuation phase of the study. Based on the moderating effect of pretreatment severity in the TDCRP (Elkin et al., 1989), severity was used as a stratification variable during randomization. Scores on the pretreatment HRSD were used to form two groups: high severity (HRSD $\geq$ 20) and low severity (HRSD $\leq$ 19). Participants were assigned to therapists within modality based on therapist availability.

Participants completed standard comprehensive outcome assessments, conducted by evaluators blind to treatment assignment, at mid- and post-treatment (approximately 8 and 16 weeks from the start of treatment, respectively) and at nonstandard time points as clinically indicated (e.g., at early termination). The HRSD was also administered to ADM and PLA participants as part of each treatment session by the treating pharmacotherapist, who was blind to whether participants were receiving active medication.

Therapists

BA was provided by two licensed psychologists and a licensed clinical social worker; on average, they had each been in clinical practice for approximately 7 years. Neil S. Jacobson provided initial training in BA. Therapists received individual off-site supervision via telephone from two of the current authors (Michael E. Addis and Keith S. Dobson) and participated in an on-site consultation meeting chaired by Neil S. Jacobson, before his death, and by Christopher Martell thereafter.

CT was provided by three licensed psychologists, who had been in clinical practice for an average of 14 years. Two had extensive training in CT prior to the outset of the trial, including training by the Beck Institute, and had served as cognitive therapists in earlier studies on depression by our group. The third had received specialized training in CT focused on the treatment of anxiety disorders. All were certified by the Academy of Cognitive Therapy during the course of the study. Two study authors (Steven D. Hollon and Keith S. Dobson) oversaw initial training and provided individual supervision off-site via telephone. The therapists also participated in an on-site consultation meeting chaired by Sandra Coffman.

Five pharmacotherapists provided ADM and PLA; all were board certified with an average of approximately 12 years of clinical experience. Training and supervision were provided by one of the authors (David L. Dunner), an experienced pharmacotherapy researcher who has conducted numerous controlled clinical trials.

Treatments

Behavioral activation. The BA treatment condition utilized in the study was an expanded version of the approach used in the component analysis study, which was based exclusively on the behavioral interventions recommended by A. T. Beck, Rush, Shaw, and Emery (1979). The expanded BA model is based on a conceptualization of depression that emphasizes the relationship between activity and mood and the role of contextual changes associated with decreased access to reinforcers that may serve an antidepressant function. The model highlights the centrality of patterns of avoidance and withdrawal (e.g., of interpersonal situations, occupational or daily-life routine demands, distressing thoughts or feelings, and so forth). Because contacting potential antidepressant reinforcers is often initially punishing, avoidance of contact minimizes distress in the short term but is associated with greater long-term difficulty, both by reducing opportunities to contact potentially antidepressant environmental reinforcers and by creating or exacerbating new problems secondary to the decreased activity. Increased activation is presented as a strategy to break this cycle. In general, BA seeks to identify and promote engagement with activities and contexts that are reinforcing and consistent with an individual's long-term goals. Specific behaviorally focused activation strategies include self-monitoring, structuring and scheduling daily activities, rating the degree of pleasure and accomplishment experienced during engagement in specific daily activities, exploring alternative behaviors related to achieving participant goals, and using role-playing to address specific behavioral deficits. In addition, the expanded BA model includes an increased focus on the assessment and treatment of avoidance behaviors, the establishment or maintenance of regularized routines, and behavioral strategies for targeting rumination, including an emphasis on the function of ruminative thinking and on moving attention away from the content of ruminative thoughts toward direct, immediate experience.

Although BA and CT share certain elements (e.g., session structure, emphasis on collaborative relationship with the participant, use of homework, etc.), the use of specific cognitive interventions was clearly proscribed in the BA condition. Information on BA is available in the published treatment manuals (Jacobson et al., 2001; Martell et al., 2001). Participants in the BA condition received a maximum of twenty-four 50-min sessions over 16 weeks, with sessions generally held twice weekly for the first 8 weeks and once weekly for the next 8 weeks.

Cognitive therapy. CT was provided in a manner consistent with standard CT for depression as specified by A. T. Beck et al. (1979) and J. S. Beck (1995). CT therapists used three broad classes of interventions targeting the following areas: (a) behavioral dysfunction, (b) situation-specific negative thinking and cognitive distortions, and (c) underlying dysfunctional beliefs or cognitions assumed to be related to the participant's current depression and risk of future depression. These components were implemented in an integrative fashion, in contrast to the sequential manner used in the component analysis study (Jacobson et al., 1996). CT therapists were able to use the full range of BA strategies outlined in the CT texts cited above but did not utilize the strategies added as part of the expanded BA model previously described. The CT condition followed the same protocol regarding frequency, schedule, and allotment of treatment sessions as did the BA condition.

Pharmacotherapy. Both the ADM and PLA conditions were administered in a triple-blind manner during the first 8 weeks of the study (i.e., participants, pharmacotherapists, and evaluators were kept blind to treatment condition). At 8 weeks, the blind was broken, and PLA participants were offered their choice of treatment at study expense. ADM was administered in a single-blind manner for the final 8 weeks of the acute phase (i.e., participants and therapists were aware that the medication was active, and only evaluators were kept blind as to treatment condition). Paroxetine was selected as the medication because selective serotonin reuptake inhibitors (SSRIs) are the most widely used and best tolerated ADM class.

Both ADM and PLA conditions followed the clinical management protocol developed for the TDCRP, modified for use with an SSRI (Fawcett, Epstein, Fiester, Elkin, & Autry, 1987). Although formal psychotherapy strategies were proscribed, the pharmacotherapists were encouraged to develop therapeutic relationships characterized by support, reassurance, and optimism about the treatment regimen to maximize participant adherence. Toward this end, pharmacotherapists also were encouraged to provide information, help participants develop reasonable expectations regarding treatment, and give limited advice. The typical session consisted of the administration of the HRSD by the pharmacotherapist; inquiry about treatment response, side effects, and nonstudy medication; and further renewal or modification of the participant's pill dosage. Medical evaluation also was conducted as indicated, and a pill count was conducted at each visit to determine participants' compliance with the medication protocol.

Participants were seen weekly for the first 4 weeks and biweekly thereafter through Week 16 (although PLA participants were terminated at Week 8). The first pharmacotherapy session was approximately 30–45 min, and subsequent sessions lasted up to 30 min. Medications were provided on a flexible schedule designed to bring each participant to a maximally tolerated dose of up to 50 mg per day. All medicated participants were to receive 10 mg/day of paroxetine, with the dosage increased to 20 mg in Week 2, 30 mg in Week 4, 40 mg in Week 6, and the maximum dose of 50 mg in Week 12. If there were significant side effects at any point, the dose could be reduced temporarily and raised again at a later time. All decisions of this nature were made in consultation with the supervising psychiatrist (David L. Dunner).

Measures

Participants completed both standardized clinical interview and self-report measures. All interviewers were trained, certified, and monitored in the assessment techniques by senior project personnel. Interviewers were blind to participants' treatment condition and were supervised weekly to prevent rater drift.

Diagnostic measures. The SCID-I (First et al., 1997) is a semistructured clinical interview that yields judgments with respect to all five axes of *DSM–IV* diagnosis. It served as the primary clinical diagnostic instrument in the study. The Structured Clinical Interview for *DSM–IV* Axis II Personality Disorders (First, Spitzer, Gibbons, Williams, & Benjamin, 1996) also was used to assess for the presence of selected personality disorders (i.e., avoidant, dependent, obsessive–compulsive, depressive, schizotypal, borderline, and antisocial personality disorders). Trained clinical evaluators administered both instruments at pretreatment.

Depression severity measures. Depression severity was assessed using the modified 17-item version of the HRSD (Hamilton, 1960) and the BDI-II (A. T. Beck et al., 1996). The HRSD is the most commonly used interview-based measure of depressive severity and has documented reliability and validity (Williams, 1988). The HRSD was modified to include atypical sleep, appetite, and weight symptoms and was administered by clinical evaluators at pretreatment, midtreatment, posttreatment, and nonstandard assessments as required (e.g., early termination). In addition, the HRSD was administered at each session during the first 8 weeks for ADM and PLA participants by the treating pharmacotherapists, analogous to what typically is done in most pharmacotherapy trials. The BDI-II is a widely used self-report measure of the severity of depressive symptoms with excellent psychometric properties (A. T. Beck et al., 1996). The BDI-II was administered at pretreatment, midtreatment, posttreatment, and nonstandard assessments as required (e.g., early termination); at each time point, participants were asked to rate their symptoms during a one-week period.

Measurement of adherence and competence. Treatment adherence in the current trial was assessed by a team of five undergraduate raters blind to treatment condition and trained to use a version of the Collaborative Study Psychotherapy Rating Scale (Hollon et al., 1988) modified to accommodate the inclusion of BA. The revised instrument contained a total of 66 items, rated on a 0–6 scale, including at least 15 items presumed to be unique to each of the respective conditions. After establishing reliability, raters completed a total of 90 tapes ($n = 36$ each for the CT and BA conditions; $n = 18$ for ADM).

The Cognitive Therapy Scale (CTS) was used to assess the competence with which CT was delivered (Young & Beck, 1980). The CTS is an 11-item instrument designed to measure the quality of treatment delivery for CT therapists, with demonstrated reliability when used by expert raters (Dobson, Shaw, & Vallis, 1985; Vallis, Shaw, & Dobson, 1986). A total score of 40 or greater on the CTS represents the standard threshold of acceptable competence in CT delivery. The off-site CT supervisors (Steven D. Hollon and Keith S. Dobson) rated 36 CT sessions (i.e., 12 per therapist) as part of the ongoing process of quality control, and Leslie Sokol of the Beck Institute also provided expert ratings on a subset of the tapes. Pharmacotherapy tapes were monitored on an ongoing basis by the supervising psychiatrist (David L. Dunner), and an external pharmacotherapy expert, Jan Fawcett, assessed a subset for competence. No comparable measure of competence in BA was available at the time of the study.

Response and remission criteria. Response represents significant symptomatic improvement, whereas remission represents improvement to the point of being asymptomatic within normal range. On the HRSD and BDI, response was defined as at least 50% reduction from baseline. Remission was defined as scores ≤ 7 on the HRSD and ≤ 10 on the BDI.

Reliability of measures. A randomly selected subset of taped clinical interviews ($n = 28$) was rated by a second group of study clinical evaluators to ascertain interrater reliability. Analyses revealed a high level of rater agreement. For the major depressive disorder diagnostic module of the SCID-I, the kappa coefficient was .78. For the HRSD, the intraclass correlation (ICC) was .95 for intake interviews and .99 for follow-up interviews. Experts at Vanderbilt University also rated a sample of taped HRSD interviews ($n = 12$), with a cross-site ICC of .98.

For treatment adherence ratings, after didactic training, raters completed eight randomly selected audiotaped therapy sessions that were also rated by the treatment integrity supervisor (Joseph B. McGlinchey). Average two-way, mixed ICCs (consistency definition) for the group's ratings across classes of items were .83 for the cognitive items, .94 for the behavioral items, and .97 for the pharmacotherapy items. For CT competence ratings, the two CT supervisors exhibited strong concordance, with a reliability of .94 for total CTS scores across 12 sessions. Concordance between either of the CT supervisors and the external expert was more modest, with an average ICC of .47 across 36 sessions. However, concordance was suppressed by a single outlier; when omitted, concordance was .59.

Statistical Analyses

Tests of baseline differences in demographic and clinical characteristics were investigated using analysis of variance (ANOVA) for continuous variables and chi-square tests of independence for categorical variables. In the presence of small or empty cells in the tests of categorical variables, the chi-square test was replaced by Fisher's exact test.

Sex was the only variable on which randomization did not achieve equivalence between conditions. Although sex did not predict response to treatment, it was included as a covariate because the outcomes of BA, which had significantly fewer women, were of primary interest.

Two sets of primary outcome analyses were conducted; the first set sought to determine whether the sample was pharmacologically responsive by comparing change among participants receiving ADM, in contrast to participants receiving PLA, on the HRSD administered at each treatment session over the first half of the acute phase. The second set examined change across the full acute phase and included only the three active treatment conditions. Separate analyses were conducted for each outcome measure within each severity subgroup. Separate analyses for each severity group were implemented because of potential problems with multicollinearity associated with including both the dichotomous severity variable (based on the HRSD) and pretreatment severity (the continuous form of the BDI-II or HRSD) as the first outcome measure in the same analysis. Planned contrasts tested for differences between all possible treatment pairs. Given the primary hypotheses of no difference between active treatments, corrections for multiple comparisons were not used.

Hierarchical linear modeling (HLM), controlling for sex, was used as the primary method to investigate active treatment differences using the intent-to-treat sample (Raudenbush & Bryk, 2001). The standard HLM model involves two levels: within-subject (Level 1) and between-subjects (Level 2). At Level 1, the outcome varies within participants 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 participants, as a function of the participant's treatment. The person-specific parameters correspond to a random intercept and random slope per participant. To determine which person-specific parameters were needed, we used procedures in which the log-likelihoods between the nested models were compared to determine the number of random effects needed (Verbeke & Molenberghs, 2000). Effect size calculations for the HLM models were derived as specified by Raudenbush and Liu (2001) and Verbeke and Molenberghs (2000).

For the modified intent-to-treat comparisons of ADM and PLA on the in-session HRSD, we included random effects estimating the variability in the intercepts and the variability in the slopes between participants. For the comparisons of active treatments, a single random effect for the intercept term was used for the HRSD, and random effects for the intercept and slope were used for the BDI. Homogeneity of random effects across treatment groups existed in all analyses except that of the BDI low-severity group.

Treatment differences in categorical rates of response and remission at posttreatment were examined using Cochran–Mantel–Haenszel (CMH) tests, controlling for sex. Categorical analyses were conducted with the full intent-to-treat sample, using last observation carry forward (LOCF) for participants who failed to complete treatment or were lost to follow-up.

Bioequivalence testing (Rogers, Howard, & Vessey, 1993; Schuirmann, 1987) was used to determine whether treatments were sufficiently close in outcome to be considered statistically equivalent. We chose the margin of noninferiority to correspond to the effect size for the ADM-to-PLA comparison during the acute phase of treatment based on the HRSD ratings conducted at each pharmacotherapy session, derived as specified by Raudenbush and Liu (2001) and Verbeke and Molenberghs (2000). Currently, the Food and Drug Administration standard for bioequivalence is Schuirmann's 2 one-sided t tests (Schuirmann, 1987); these determine if the difference between ADM and BA lies completely within the noninferiority margin, which considers the two to be negligibly different.

To assess whether missing data had a substantive influence on results, we used the pattern-mixture approach (Hedeker & Gibbons, 1997). To determine if the differential trends over time were dependent on completion status, we included a three-way interaction of completion status, time, and treatment group in the HLM analysis. A significant finding for this three-way interaction would suggest that the slope estimates and treatment group comparisons were dependent on completion status; a nonsignificant finding would indicate that the slope estimates and treatment group comparisons were not biased by completion status.

All analyses were conducted using SPSS Version 11.5 or SAS Version 9.1.

Results

Participant Enrollment

Of the 388 participants who completed a comprehensive intake assessment, 250 were eligible for randomization (of whom 9 declined participation), resulting in 241 participants randomized to treatment (CT = 45; BA = 43; ADM = 100; PLA = 53). Of the 138 excluded participants, the majority were screened out because of subthreshold major depression or low severity as measured by the BDI or HRSD (n = 89); the remainder were screened out because of medical complications (n = 16), diagnostic considerations (n = 19), substance abuse or dependence (n = 8), acute suicidality (n = 3), or other reasons (n = 3). Figure 1 presents the flow of participants over the course of the study.

Baseline Characteristics

Table 1 presents baseline sample demographic and clinical characteristics. Treatment groups did not differ, with the exception of sex. In the full sample, there was a significant difference between treatments in sex, $\chi^2(3, N = 241) = 9.30$, $p = .026$, with fewer women assigned to BA (47%; n = 20) as compared with CT (73%; n = 33), ADM (68%; n = 68), and PLA (72%; n = 38). This difference was driven by the low-severity subgroup, $\chi^2(3, N = 103) = 9.18$, $p = .027$, in which fewer women were assigned to BA (28%; n = 5) as compared with CT (75%; n = 15), ADM (61%; n = 26), and PLA (59%; n = 13).

Treatment Integrity

Adherence. Therapists in the various conditions were strongly adherent to the respective treatments. The cognitive items on the adherence scale received the greatest ratings in the CT condition sessions (M = 6.07), while receiving negligible endorsement in the BA and ADM sessions (Ms = 0.58 and 0.06, respectively). The behavioral items received the greatest endorsement for the BA condition (M = 8.65), a lesser though substantive degree of endorsement for CT (M = 5.01), and negligible endorsement for ADM (M = 0.22). The pharmacotherapy items received the greatest endorsement in ADM sessions (M = 8.06), while receiving negligible endorsement in the BA and CT conditions (Ms = 0.57 and 0.06, respectively).

CT competence. Study supervisors (Steven D. Hollon and Keith S. Dobson) rated CT therapists as delivering the treatment competently (M = 46.86, SD = 4.05). External ratings from the Beck Institute suggested a more modest level of competence (M = 40.33, SD = 4.17) and were significantly lower than those of study supervisors, $t(35) = 8.08$, $p < .001$.

ADM dosage. Mean dosage did not differ as a function of severity and is therefore presented for the full sample of ADM participants. The mean (± SD) paroxetine dose during the first week of treatment was 10.00 (± 0.00) mg/day. The mean (± SD) daily dose was increased to 19.26 (± 2.64) mg in the second week, 24.52 (± 6.25) mg in the fourth week, 30.00 (± 8.74) mg in the sixth week, and 31.67 (± 11.45) mg in the eighth week. By the 12th week, the mean dosage had been increased to 35.17 (± 13.08) mg/day, which was maintained through the end of the acute phase.

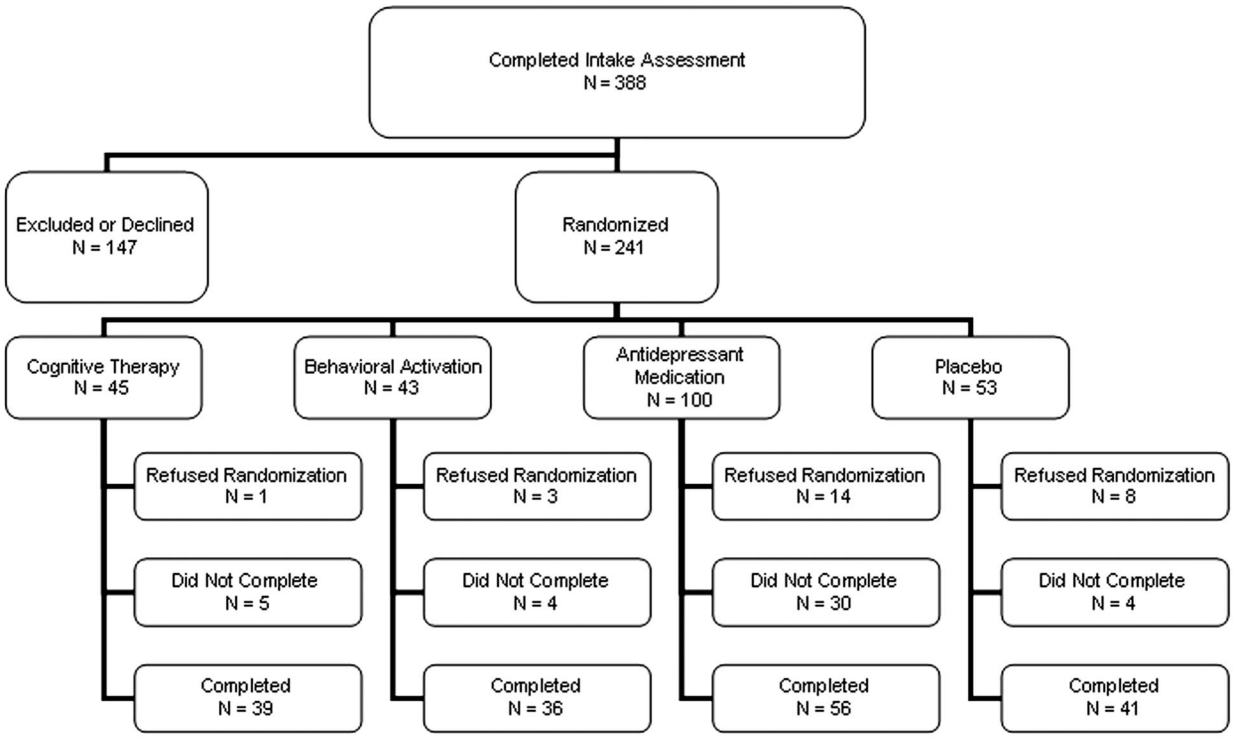

Figure 1. Participant flow.

Attrition

Overall rates of attrition were generally low, with the exception of the ADM condition. Rates of attrition did not differ as a function of severity and are therefore presented only for the full sample. Over the full acute phase, there were significant differences between active treatments in rates of attrition, $\chi^2(2, N = 188) = 19.02$, $p < .001$. The rate of attrition for ADM (44%; $n = 44$) was significantly higher than for either CT (13.3%; $n = 6$), $\chi^2(1, N = 145) = 12.92$, $p < .001$, or BA (16.3%; $n = 7$), $\chi^2(1, N = 143) = 10.07$, $p = .002$.

With respect to the timing of attrition, there were significant differences in the rates of refusal of randomization between patients assigned to pharmacotherapy in contrast to CT or BA, $\chi^2(2, N = 241) = 6.13$, $p = .04$ (Fisher's exact test); 14% of the patients assigned to pharmacotherapy ($n = 22$) failed to attend a single session in contrast to 2% in CT ($n = 1$) and 7% in BA ($n = 3$). Overall, there were significant differences in attrition rates during the first 8 weeks, $\chi^2(3, N = 241) = 17.30$, $p < .001$, with 36% of participants in ADM ($n = 36$) dropping out in contrast to 11% in CT ($n = 5$), 9% in BA ($n = 4$), and 23% in PLA ($n = 12$). Pairwise comparisons between treatments indicated a significantly higher rate of attrition between ADM versus CT, $\chi^2(1, N = 145) = 9.48$, $p = .002$, and ADM versus BA, $\chi^2(1, N = 143) = 10.64$, $p = .001$; there was a trend toward higher attrition in ADM than PLA, $\chi^2(1, N = 153) = 2.87$, $p = .090$. The rates of attrition between the active treatments were not significantly different during the second half of the acute phase, $\chi^2(2, N = 143) = 3.24$, $p = .20$ (Fisher's exact test).

Reasons for the high level of attrition in ADM were diverse. In addition to being significantly more likely not to start treatment if assigned to one of the pill conditions, 9% ($n = 9$) of the ADM participants dropped out because of side effects, 5% ($n = 5$) were withdrawn because of nonadherence to study protocol, 6% ($n = 6$) experienced lack of efficacy or worsening of symptoms (including one participant who died by suicide), 3% ($n = 3$) dropped out because of dissatisfaction with study treatment, 1% ($n = 1$) relocated, 1% ($n = 1$) dropped out because of feeling improved, and 5% ($n = 5$) were lost for reasons unknown. Of PLA participants who dropped out after starting treatment, 2% ($n = 1$) dropped out because of side effects, 2% ($n = 1$) dropped out because of concerns about confidentiality, 2% ($n = 1$) relocated, and 2% ($n = 1$) were lost for reasons unknown. Of CT participants, 2% ($n = 1$) dropped out because of dissatisfaction with study treatment, 2% ($n = 1$) experienced lack of efficacy or worsening of symptoms, 2% ($n = 1$) found the research burdensome, and 4% ($n = 2$) were lost for reasons unknown. Of BA participants, 2% ($n = 1$) dropped out because of dissatisfaction with study treatment, 2% ($n = 1$) experienced lack of efficacy or worsening of symptoms, 2% ($n = 1$) found the research burdensome, and 2% ($n = 1$) relocated.

Side Effects

Side effects were recorded on the basis of pharmacotherapist inquiry and participant report. Overall, side effects reported by participants receiving medication were consistent with the known profile for paroxetine. Results are presented for side effects reported by at least 10% of ADM or PLA participants and are

Table 1
Baseline Characteristics of Participants

Baseline characteristics	Full sample ($N = 241$)	High severity ($n = 138$)	Low severity ($n = 103$)
Sex: *n* (% female)	159 (66.0)	100 (72.5)	59 (57.3)
Race: *n* (% White)	197 (81.7)	118 (85.5)	79 (76.7)
Age (years): *M* (*SD*)	39.90 (10.97)	39.86 (11.50)	39.95 (10.28)
Currently married or cohabiting: *n* (%)	94 (39.0)	53 (38.4)	41 (39.8)
College graduate: *n* (%)	121 (50.21)	65 (47.10)	56 (54.37)
Employed outside home: *n* (%)	171 (71.0)	92 (66.7)	79 (76.7)
BDI: *M* (*SD*)	32.01 (7.48)	35.30 (6.97)	27.60 (5.67)
HRSD: *M* (*SD*)	20.74 (4.12)	23.60 (2.89)	16.90 (1.67)
Severity			
Low (HRSD 14–19): *n* (%)	103 (42.7)		103 (100.0)
High (HRSD ≥ 20): *n* (%)	138 (57.3)	138 (100.0)	
Current episode length (months): *Mdn* (*SD*)	12.00 (71.30)	12.0 (68.77)	11.0 (74.90)
Number of prior episodes: *Mdn* (*SD*)	1.00 (1.44)	1.00 (1.56)	0.00 (1.23)
Depressive subtype			
Melancholic: *n* (%)	73 (30.3)	46 (33.3)	27 (26.2)
Atypical: *n* (%)	42 (17.4)	22 (15.9)	20 (19.4)
Recurrent depression: *n* (%)	139 (57.7)	88 (63.8)	51 (49.5)
Chronic depression (> 2 years): *n* (%)	83 (34.4)	47 (34.1)	36 (35.0)
Age (years) of onset of 1st episode: *M* (*SD*)	27.65 (13.27)	26.22 (13.14)	29.55 (13.27)
Previous psychiatric hospitalization: *n* (%)	23 (9.5)	18 (13.0)	5 (4.9)
Any current Axis I diagnosis: *n* (%)	68 (28.2)	50 (36.2)	18 (17.5)
Any lifetime Axis I diagnosis: *n* (%)	121 (50.2)	79 (57.3)	42 (40.8)
Avoidant, dependent, obsessive–compulsive, or depressive personality disorder: *n* (%)	49 (20.3)	32 (23.2)	17 (16.5)
Any current anxiety diagnosis: *n* (%)	57 (23.7)	43 (31.2)	14 (13.6)
Any lifetime substance abuse/dependence: *n* (%)	102 (42.3)	63 (45.7)	39 (37.9)

Note. Statistics summarized in each cell of the table are given after the variable name. BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression.

presented for the full sample because differences were few as a function of initial severity (high-severity participants reported more nausea and less diarrhea than low-severity participants). Relative to placebo participants, participants receiving paroxetine reported more sexual side effects: anorgasmia, 17% versus 0%, $\chi^2(1, N = 153) = 10.14$, $p = .002$, and decreased libido, 15% versus 0%, $\chi^2(1, N = 153) = 8.81$, $p = .003$; gastrointestinal distress: nausea, 19% versus 6%, $\chi^2(1, N = 153) = 5.01$, $p = .025$; and sleep-related difficulties: insomnia, 25% versus 9%, $\chi^2(1, N = 153) = 5.32$, $p = .021$, somnolence, 38% versus 6%, $\chi^2(1, N = 153) = 18.47$, $p < .001$, and yawning, 12% versus 0%, $\chi^2(1, N = 153) = 6.90$, $p = .009$ (Fisher's exact test). Paroxetine patients also reported more dry mouth, 17% versus 6%, $\chi^2(1, N = 153) = 3.92$, $p = .048$, and excessive sweating, 13% versus 0%, $\chi^2(1, N = 153) = 7.53$, $p = .005$ (Fisher's exact test). Side effects were not assessed in BA and CT but were assumed to parallel those reported by placebo participants.

Pharmacological Responsiveness of Sample

For the high-severity subgroup, there was evidence of differential improvement over time by treatment on the HRSD as conducted in-session by the pharmacotherapists, $F(1, 64) = 5.87$, $p = .018$. High-severity participants receiving ADM improved significantly more per treatment week than did participants receiving PLA; in contrast, for the low-severity subgroup, there was no evidence of differential improvement over time by treatment on the HRSD, $F(1, 49) = 0.98$, $p = .33$. Slope estimates (± *SE*s) for the high-severity subgroup were −1.22 (± 0.16) for ADM and −0.57 (± 0.22) for PLA. Slope estimates (± *SE*s) for the low-severity subgroup were −1.05 (± 0.16) for ADM and −0.77 (± 0.23) for PLA. Associated effect sizes were 0.65 and 0.31 for the high- and low-severity subgroups, respectively.

Analysis of Active Treatment Outcomes

Table 2 provides descriptive statistics on the primary outcome measures as a function of severity. In the high-severity subgroup, there was significant overall improvement by time for all groups on the BDI, $F(1, 83) = 219.86$, $p < .0001$, and on the evaluator-rated HRSD, $F(1, 190) = 443.85$, $p < .0001$. In addition, as shown in Figure 2, significant differences in slopes were found among the treatments on both the BDI, $F(2, 81) = 4.15$, $p = .019$, and the HRSD, $F(2, 188) = 3.12$, $p = .047$. Participants in BA improved significantly more per treatment week than did participants in CT on both the BDI, $t(81) = 2.23$, $p = .029$, and the HRSD, $t(188) = 2.09$, $p = .038$. Similarly, participants in ADM improved significantly more per treatment week than did participants in CT on both the BDI, $t(81) = 2.76$, $p = .007$, and the HRSD, $t(188) = 2.31$, $p = .022$. There were no significant differences in the rates of improvement comparing participants in BA and ADM on the BDI, $t(81) = 0.25$, $p = .80$, or on the HRSD, $t(188) = 0.05$, $p = .96$. BDI slope estimates (± *SE*s) were −1.12 (± 0.20) for CT, −1.76 (± 0.20) for BA, and −1.82 (± 0.15) for ADM. HRSD slope estimates (± *SE*s) were −0.74 (± 0.09) for CT, −0.99 (± 0.08) for BA, and −0.99 (± 0.063) for ADM. Associated effect sizes for BA relative to CT were 0.87 (BDI) and 0.59 (HRSD); for ADM relative to CT, effect sizes were 0.96 (BDI) and 0.51

Table 2
BDI and HRSD Means, Standard Deviations, and Ns by Condition Over Time

Measure and treatment condition	Intake			8 weeks			16 weeks		
	M	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
Low severity									
BDI									
CT	27.30	6.89	20	12.94	10.29	17	9.76	8.15	17
BA	28.72	4.59	18	15.33	10.03	15	11.00	10.08	13
ADM	27.79	5.67	43	13.89	8.61	28	7.91	6.29	22
PLA	26.59	5.43	22	14.68	7.81	19	—	—	—
HRSD									
CT	16.65	1.84	20	10.41	4.05	17	7.19	4.09	16
BA	17.28	1.45	18	12.40	6.58	15	7.92	7.68	13
ADM	16.98	1.60	43	11.57	5.32	28	8.45	5.26	22
PLA	16.68	1.86	22	12.05	5.54	19	—	—	—
High severity									
BDI									
CT	34.12	5.67	25	21.00	14.64	21	17.44	15.57	18
BA	36.68	5.91	25	16.82	8.56	22	8.75	7.96	16
ADM	35.61	7.13	57	14.39	11.00	38	7.78	9.61	27
PLA	34.55	8.36	31	22.50	12.97	22	—	—	—
HRSD									
CT	22.72	2.61	25	12.67	6.96	21	10.33	7.62	18
BA	23.16	2.53	25	12.86	6.93	22	7.56	6.94	16
ADM	23.79	2.60	57	13.13	7.74	38	8.63	7.19	27
PLA	24.32	3.69	31	16.09	7.60	22	—	—	—

Note. The data on PLA are presented for illustrative purposes. Dashes indicate that the PLA condition had no data at the 16-week point. The analyses related to pharmacological responsiveness are based on weekly ratings on the HRSD; these data are not included in this table and are available by request from Sona Dimidjian. BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression; CT = cognitive therapy; BA = behavioral activation; ADM = antidepressant medication; PLA = pill placebo.

(HRSD), and for ADM relative to BA, effect sizes were 0.09 (BDI) and 0.01 (HRSD).

Bioequivalence testing (Rogers et al., 1993; Schuirmann, 1987) was used to determine whether ADM and BA were sufficiently similar to each other to be considered statistically equivalent. Using both the BDI and HRSD, ADM and BA lie within the margin of noninferiority, with a probability larger than 99.1%.

As shown in Figure 2, in the low-severity subgroup, there was significant overall improvement by time for all groups on the BDI, $F(1, 62) = 166.10$, $p < .0001$, and on the HRSD, $F(1, 146) = 193.02$, $p < .0001$. However, there was no evidence of differential improvement over time by treatment on the BDI, $F(2, 60) = 0.47$, $p = .63$, or on the HRSD, $F(2, 144) = 0.05$, $p = .95$. Specific pairwise comparisons between treatments also failed to indicate significant differences in slopes, and associated effect sizes were also small.

Categorical rates of response and remission at posttreatment also were calculated for the high- and low-severity subgroups, using LOCF for participants who dropped out of treatment or failed to provide posttreatment data. Because our primary hypotheses concerned the high-severity subgroup, categorical outcomes for this group are presented in Figures 3 and 4.

Among the more severely depressed participants, overall combined rates of response and remission based on the BDI were 48% ($n = 12$) in CT, 76% ($n = 19$) in BA, and 49% ($n = 28$) in ADM. On the basis of the HRSD, overall rates were 56% ($n = 14$) in CT, 60% ($n = 15$) in BA, and 40% ($n = 23$) in ADM. Results indicated a nonsignificant trend on the BDI, $\chi^2(2, N = 107) = 5.64$, $p = .06$, and no significant differences between treatments on the HRSD, $\chi^2(2, N = 107) = 3.62$, $p = .16$. The direction of the differences on the BDI was driven by the superior performance of BA, in which a significantly greater percentage of BA participants met BDI response criteria as compared with both participants receiving CT, $\chi^2(1, N = 50) = 3.92$, $p = .048$, or those receiving ADM, $\chi^2(1, N = 82) = 4.91$, $p = .027$. Rates of remission for the high-severity subgroup based on the BDI were 40% ($n = 10$) in CT, 52% ($n = 13$) in BA, and 42% ($n = 24$) in ADM. On the basis of the HRSD, overall rates of remission were 36% ($n = 9$) in CT, 56% ($n = 14$) in BA, and 23% ($n = 13$) in ADM. There were no significant differences between treatments on the BDI, $\chi^2(2, N = 107) = .99$, $p = .61$. Results indicated significant differences between treatments on the HRSD, $\chi^2(2, N = 107) = 8.88$, $p = .012$, with a significantly greater percentage of BA participants reaching remission as compared with ADM participants, $\chi^2(1, N = 82) = 9.82$, $p = .002$.

The poor performance of CT relative to BA and ADM on the continuous measures was in part a consequence of a subset of extreme nonresponders based on observed posttreatment assessments. Specifically, considering all high-severity patients, 28% ($n = 7$) of CT participants endorsed scores of greater than 30 on the

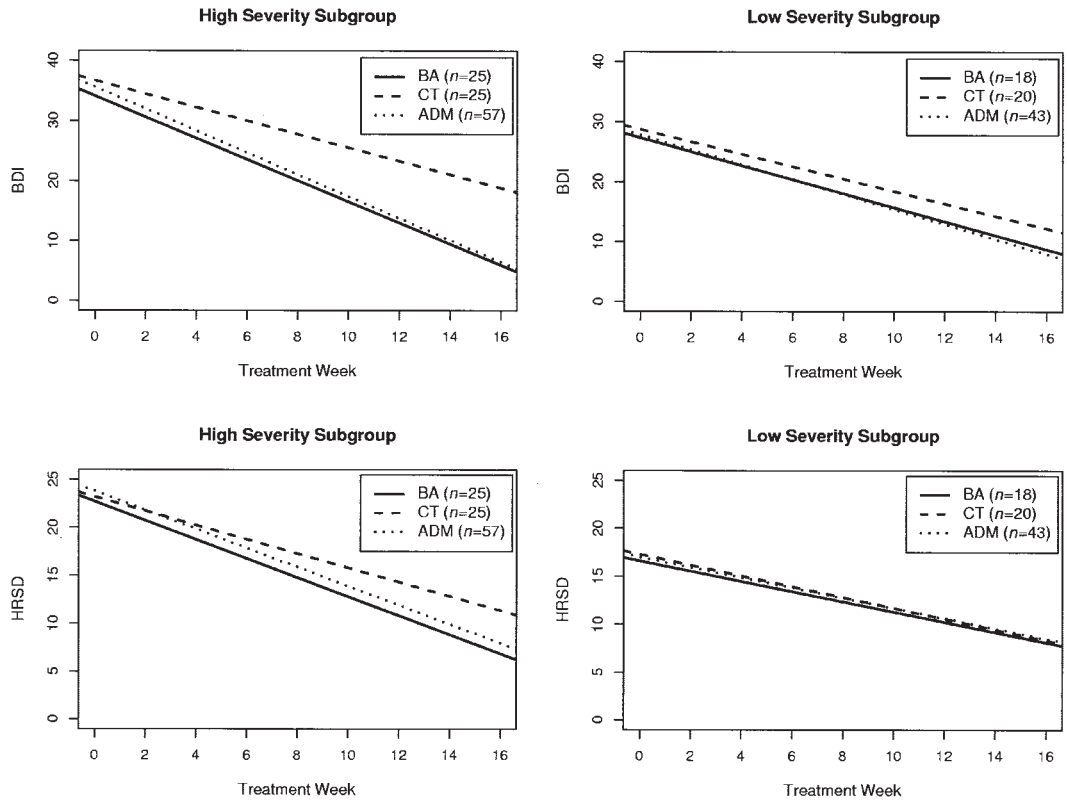

Figure 2. BDI and HRSD slope trajectories for active treatments during the full acute phase. BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression; CT = cognitive therapy; BA = behavioral activation; ADM = antidepressant medication.

BDI in contrast to only 2% ($n = 1$) of ADM and 0% ($n = 0$) of BA participants. On the HRSD, 8% ($n = 2$) of CT participants endorsed scores of greater than 20 in contrast to only 5% ($n = 3$) of ADM and 4% ($n = 1$) of BA participants.

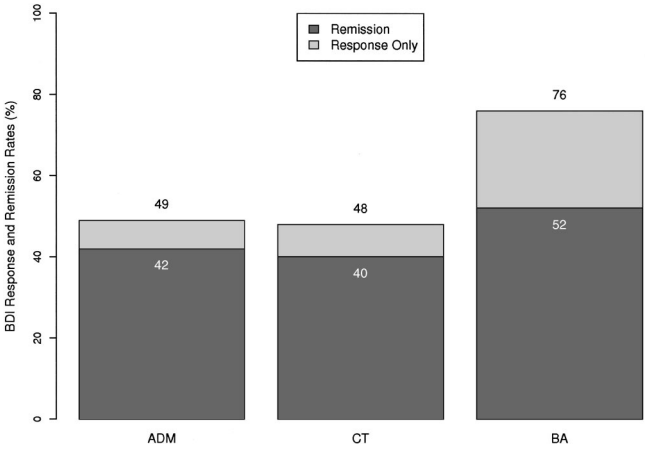

Figure 3. Response and remission rates at posttreatment based on the Beck Depression Inventory (BDI) for the high-severity subgroup for antidepressant medication (ADM), cognitive therapy (CT), and behavioral activation (BA). Total bar represents response; lower bar represents remission.

Among the less severely depressed participants, overall rates of response based on the BDI were 65% ($n = 13$) in CT, 50% ($n = 9$) in BA, and 56% ($n = 24$) in ADM. On the basis of the HRSD, overall response rates were 60% ($n = 12$) in CT, 39% ($n = 7$) in BA, and 47% ($n = 20$) in ADM. Results indicated no significant differences between treatments on the BDI, $\chi^2(2, N = 81) = 0.25$, $p = .88$, or on the HRSD, $\chi^2(2, N = 81) = 1.02$, $p = .60$. Rates of remission based on the BDI were 55% ($n = 11$) in CT, 44% ($n = 8$) in BA, and 42% ($n = 18$) in ADM. On the basis of the HRSD, overall rates of remission were 50% ($n = 10$) in CT, 39% ($n = 7$) in BA, and 33% ($n = 14$) in ADM. Results indicated no significant differences between treatments on the BDI, $\chi^2(2, N = 81) = 0.77$, $p = .68$, or on the HRSD, $\chi^2(2, N = 81) = 1.59$, $p = .45$.

Analysis of Missing Data

To determine if the differential trends over time were dependent on completion status, we included a three-way interaction of completion status, time, and treatment group in the HLM analysis to assess the impact of missing data. On the BDI, results indicated a nonsignificant effect for the Completion Status × Time × Treatment Group interaction for the high-severity subgroup, $F(2, 64) = 0.23$, $p = .80$, and low-severity subgroup, $F(1, 52) = 2.79$, $p = .10$. Similarly, on the HRSD, results indicated a nonsignificant effect for the Completion Status × Time × Treatment Group

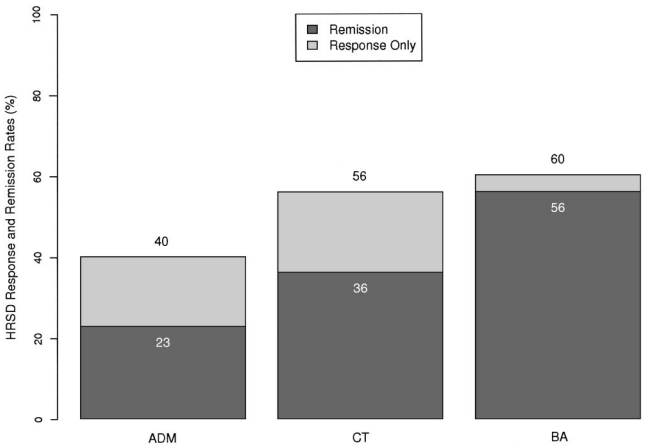

Figure 4. Response and remission rates at posttreatment based on the Hamilton Rating Scale for Depression (HRSD) for the high-severity subgroup for antidepressant medication (ADM), cognitive therapy (CT), and behavioral activation (BA). Total bar represents response; lower bar represents remission.

interaction for the high-severity subgroup, $F(2, 185) = 0.16$, $p = .86$, and low-severity subgroup, $F(1, 142) = 1.18$, $p = .28$. These results suggest that the parameter estimates generated by the original HLM models are valid and are not biased by missing data. Additionally, two-way ANOVA models were also used to test for differences in baseline HRSD and BDI by completion status and treatment condition for the high- and low-severity subgroups. For the BDI, all p values were greater than .34; for the HRSD, all p values for the two-way interaction of completion status and treatment condition were greater than .26.

Discussion

The results of this study indicate that BA is comparable in efficacy to ADM, the current standard, and more efficacious than CT, one of the best supported psychotherapies, among more severely depressed participants. The results also provide further confirmation of the importance of initial severity in the analysis of treatment outcome; differential treatment effects were observed only among those patients who were more severely depressed.

In any comparison between psychotherapy and medication, it is important to examine whether the sample was responsive to medications and whether pharmacotherapy was adequately implemented. Among more severely depressed participants in this trial, ADM significantly outperformed placebo through 8 weeks of treatment. There were no significant differences in outcome between ADM and placebo for the less severely depressed participants, consistent with findings from numerous other studies (Hollon et al., 2002). In the absence of a demonstrated drug effect for such patients, there may be little justification for prescribing psychoactive medications when there are comparably effective psychosocial alternatives free of side effects.

Across the full acute phase, for the more severely depressed participants, BA and ADM were comparable on both self-report and clinical ratings; moreover, BA brought a significantly greater percentage of participants to remission and retained a greater percentage of participants in treatment. The performance of BA with respect to ADM challenges current treatment guidelines, which state that moderately and severely depressed participants require medication (American Psychiatric Association, 2000). The availability of viable alternatives to ADM is particularly important given that not all participants want to take medication, particularly given typical side effect profiles (Hollon et al., 2002).

The findings of this study with respect to CT are at odds with other recent studies in which CT has been comparable to ADM (DeRubeis et al., 2005). However, this pattern of findings is consistent with the TDCRP (Elkin et al., 1989), in which CT was not significantly different from placebo and was significantly outperformed by ADM. Although the quality of CT in the TDCRP has been criticized, it is not clear that these same concerns apply in the present trial. Moreover, the outcomes of CT in this study are comparable to other recent trials; specifically, the remission rate of 36% among CT patients in this study compares favorably to the 40% remission rate recently reported by DeRubeis et al. (2005). Thus, it appears that the superiority of BA was not due to poorly implemented CT but rather to the greater efficacy of BA.

The results of this study build on earlier behavioral approaches to depression (Ferster, 1973; Lewinsohn, 1974) and replicate and extend the findings of the earlier component analysis study (Jacobson et al., 1996). The results underscore the value of sustained use of simple behavioral strategies, such as goal setting, self-monitoring, activity scheduling, problem solving, and graded task assignment, in the treatment of depression. Although the long-term prevention effects of this approach relative to CT are still to be determined, the short-term outcomes in this study are consistent with more recent activation-oriented interventions for depression (e.g., Blumenthal et al., 1999; Hopko, Lejuez, LePage, Hopko, & McNeil, 2003) and with the findings of studies across multiple diagnostic categories suggesting that the cognitive components of CT may add little incremental benefit over purely behavioral interventions (e.g., Borkovec, Newman, Pincus, & Lytle, 2002; Foa, Rothbaum, & Furr, 2003; Gloaguen, Cottraux, Cucherat, & Blackburn, 1998).

This growing body of research raises questions about the necessity of directly targeting negative thinking to achieve treatment response. In this regard, it is important to note that A. T. Beck and colleagues (1979) have long suggested that therapists focus on behavioral strategies early in treatment when patients are more depressed and return to that emphasis later if patients start to worsen. Although the current data do not specifically address whether change in cognition is a mediator of symptom change, they provide strong evidence that behavioral methods are sufficient to produce symptom change irrespective of whether improvement is mediated by cognitive change or not (cf. Bandura, 1977). Future analyses should more directly address the underlying mechanisms of change.

Additionally, the results of this trial suggest that the expanded BA model may have unique advantages over the behavioral strategies tested in the component analysis study (Jacobson et al., 1996). Although further research is necessary to identify specific processes of change, the added elements of the expanded BA model may account for its stronger performance relative to CT in

the current trial. In particular, targeting avoidance behaviors, in accordance with earlier behavioral theory (Ferster, 1973), may be an important innovation. Addressing avoidance is standard in treatments for anxiety, and recent models propose that avoidance may be a fundamental element underlying multiple psychopathologies and that blocking avoidance may be a critical element of treatment (Barlow, Allen, & Choate, 2004). However, treatments for depression have heretofore not specifically emphasized targeting avoidance, with the exception of the use of opposite action for sadness within dialectical behavior therapy (Linehan, 1993) and early investigations of acceptance and commitment therapy with depressed patients (S. C. Hayes, Strosahl, & Wilson, 1999; Zettle & Rains, 1989).

Avoidance minimizes immediate distress at the cost of both diminishing opportunities for reinforcement and exacerbating ongoing stressors. BA explicitly targets the reduction of avoidance behaviors related to both intrapersonal and interpersonal difficulties. For example, suppose a patient responds to interpersonal conflict with a coworker by avoiding work for multiple days. Although this avoidance minimizes aversive interactions with her coworker, the patient also misses the experience of accomplishing tasks at work, which has served an antidepressant function for her in the past. Staying home also creates new problems, such as earning less money and engendering frustration on the part of her supervisor, while doing nothing to address the original problem with the coworker. To interrupt this cycle, the BA model uses focused activation strategies to explicitly target such avoidance patterns and associated functional consequences. In essence, in BA, patients learn to identify patterns of avoidance and to respond with activation; this basic principle is applied repeatedly across multiple situations in therapy.

Moreover, the BA model utilizes a fundamentally different approach to negative and ruminative thinking than used in CT. First, behavioral interventions address the function of negative or ruminative thinking, in contrast to CT's emphasis on thought content. BA encourages attention to the consequences of ruminating (avoidance and withdrawal) and the use of activation strategies as alternatives. In this regard, BA shares important elements with other contemporary behavioral therapies that emphasize function rather than topography of behavior (e.g., S. C. Hayes et al., 1999; Jacobson & Christensen, 1996; Linehan, 1993). BA also overlaps in this way with strategies in CT that explore the utility (as opposed to the validity) of thoughts (J. S. Beck, 1995). It is possible that an emphasis on the utility or function of thinking has a particularly important role in the treatment of depression. Second, BA specifies attention-to-experience interventions to counter ruminative thinking by attending to direct sensations. Similar to recent mindfulness-based treatments (e.g., Segal, Williams, & Teasdale, 2002), these interventions provide a method for addressing rumination that does not engage the content of thoughts. Patients are encouraged to notice when they are ruminating and to move their attention away from the content of ruminative thoughts toward direct and immediate experience; for instance, a patient may be asked to experiment with attending to the sights, smells, or sounds around her when she notices that she is ruminating.

A number of limitations should be noted. First, given that the expanded BA model was developed at the University of Washington and CT supervisors were off-site, allegiance effects may have influenced the findings. These concerns were mitigated, however, because investigators with allegiance to their respective treatments were responsible for overseeing those treatments and supervision by these experts was provided throughout the trial. Future trials are essential in which BA is implemented in other venues and comparison treatments have the benefit of on-site expertise. Moreover, such tests of the generalizability of findings are additionally important given the study exclusion criteria and particular sample characteristics (e.g., differences in sex between conditions).

Second, the lack of competency ratings for our BA therapists is a limitation. Despite the positive outcomes of BA in this study, the development and validation of independent competency assessments of BA remain an important issue for future research.

Third, the rate of attrition in ADM was higher than that reported in other trials using paroxetine (e.g., DeRubeis et al., 2005). We cannot rule out the possibilities that patients in our trial were unrepresentative in their unwillingness to accept or inability to tolerate medication or that the treatment implementation contributed to the high rate of attrition. For example, ADM may have demonstrated better retention had the protocol allowed for a more aggressive dosage schedule and greater flexibility in treatment delivery (e.g., augmenting or switching medications). Moreover, the greater attrition in ADM complicated interpretation of the results. There is no evidence that attrition biased the findings based on HLM analyses; these analyses, which took attrition into account, indicate that ADM was as efficacious as BA and superior to CT among more severely depressed patients. At the same time, categorical analyses, which considered only whether patients actually met criteria for response or remission, indicate that ADM was no better than CT and less efficacious than BA. The difference is that the HLM analyses essentially estimate what likely would have happened if the medication dropouts had remained in treatment (they should have done as well as patients in BA), not how they actually did (fewer of them actually benefited from treatment). This distinction should not be overlooked when evaluating the relative advantages of the respective interventions.

In summary, BA did particularly well in this study. It was at least as efficacious as ADM, even among more severely depressed participants, and retained a greater proportion of patients long enough for them to benefit from treatment. This suggests that BA may be a viable alternative to ADM, challenging current treatment guidelines. BA also was more efficacious than CT among more severely depressed participants. These results challenge the assumption that directly modifying negative beliefs is essential for change and raise the possibility that elements of the expanded BA model may offer more robust interventions for depression. Finally, interest in BA was based in part on the notion that it would be a more exportable treatment that is easier to implement and train than CT or other more complex interventions. If this is the case, the public health advantages could be significant. Such questions await future study.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). Practice guidelines for the treatment of patients with major depressive disorder (revision). *American Journal of Psychiatry, 157,* 1–45.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 181–215.

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy, 35,* 205–230.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression.* New York: Guilford Press.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the BDI-II.* San Antonio, TX: Psychological Corporation.

Beck, J. S. (1995). *Cognitive therapy: Basics and beyond.* New York: Guilford Press.

Blumenthal, J. A., Babyak, M. A., Moore, K. A., Craighead, W. E., Herman, S., Khatri, P., et al. (1999). Effects of exercise training on older patients with major depression. *Archives of Internal Medicine, 159,* 2349–2356.

Borkovec, T. D., Newman, M. G., Pincus, A. L., & Lytle, R. (2002). A component analysis of cognitive–behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *Journal of Consulting and Clinical Psychology, 70,* 288–298.

DeRubeis, R. J., Gelfand, L. A., Tang, T. Z., & Simons, A. D. (1999). Medications versus cognitive behavior therapy for severely depressed outpatients: Mega-analysis of four randomized comparisons. *American Journal of Psychiatry, 156,* 1007–1013.

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., et al. (2005). Cognitive therapy vs. medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62,* 409–416.

Dobson, K. S., Shaw, B. F., & Vallis, T. M. (1985). Reliability of a measure of the quality of cognitive therapy. *British Journal of Clinical Psychology, 24,* 295–300.

Elkin, I., Gibbons, R. D., Shea, M. T., Sotsky, S. M., Watkins, J. T., Pilkonis, P. A., & Hedeker, D. (1995). Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 63,* 841–847.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., et al. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46,* 971–982.

Fawcett, J., Epstein, P., Fiester, S. J., Elkin, I., & Autry, J. H. (1987). Clinical management: Imipramine/placebo administration manual. *Psychopharmacological Bulletin, 23,* 309–324.

Ferster, C. B. (1973). A functional analysis of depression. *American Psychologist, 28,* 857–870.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *User's guide for the Structured Clinical Interview for DSM–IV Axis I Disorders.* Washington, DC: American Psychiatric Press.

First, M. B., Spitzer, R. L., Gibbons, M., Williams, J. B. W., & Benjamin, L. (1996). *User's guide for the Structured Clinical Interview for DSM–IV Axis II Personality Disorders (SCID-II).* New York: New York State Psychiatric Institute, Biometrics Research Department.

Foa, E. B., Rothbaum, B. O., & Furr, J. M. (2003). Augmenting exposure therapy with other CBT procedures. *Psychiatric Annals, 33,* 47–53.

Gloaguen, V., Cottraux, J., Cucherat, M., & Blackburn, I. M. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders, 49,* 59–72.

Gortner, E. T., Gollan, J. K., Dobson, K. S., & Jacobson, N. S. (1998). Cognitive–behavioral treatment for depression: Relapse prevention. *Journal of Consulting and Clinical Psychology, 66,* 377–384.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23,* 56–61.

Hayes, A. M., Castonguay, L. G., & Goldfried, M. R. (1996). Effectiveness of targeting the vulnerability factors of depression in cognitive therapy. *Journal of Consulting and Clinical Psychology, 64,* 623–627.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change.* New York: Guilford Press.

Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2,* 64–78.

Hollon, S. D., Evans, M. D., Auerbach, A., DeRubeis, R. J., Elkin, I., Lowery, H. A., et al. (1988). *Development of a system for rating therapies for depression.* Unpublished manuscript, University of Minnesota, Minneapolis.

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest, 3,* 1–39.

Hopko, D. R., Lejuez, C. W., LePage, J. P., Hopko, S. D., & McNeil, D. W. (2003). A brief behavioral activation treatment for depression: A randomized pilot trial within an inpatient psychiatric hospital. *Behavior Modification, 27,* 458–469.

Jacobson, N. S., & Christensen, A. (1996*). Integrative couple therapy: Promoting acceptance and chan*ge. New York: Norton.

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., et al. (1996). A component analysis of cognitive–behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64,* 295–304.

Jacobson, N. S., & Hollon, S. D. (1996). Cognitive-behavior therapy versus pharmacotherapy: Now that the jury's returned its verdict, it's time to present the rest of the evidence. *Journal of Consulting and Clinical Psychology, 64,* 74–80.

Jacobson, N. S., Martell, C. R., & Dimidjian, S. (2001). Behavioral activation treatment for depression: Returning to contextual roots. *Clinical Psychology: Science and Practice, 8,* 255–270.

Lewinsohn, P. M. (1974). A behavioral approach to depression. In R. J. Friedman & Katz, M. (Eds.), *The psychology of depression: Contemporary theory and research* (pp. 157–178). Oxford, England: Wiley.

Linehan, M. M. (1993). *Cognitive–behavioral treatment of borderline personality disorder.* New York: Guilford Press.

Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). *Depression in context: Strategies for guided action.* New York: Norton.

Olfson, M., Marcus, S. C., Druss, B., & Pincus, H. A. (2002). National trends in the use of outpatient psychotherapy. *American Journal of Psychiatry, 159,* 1914–1920.

Raudenbush, S., & Bryk, A. (2001). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Raudenbush, S. W., & Liu, X. F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6,* 387–401.

Rehm, L. (1977). A self-control model of depression. *Behavior Therapy, 8,* 787–804.

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

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics, 15,* 657–681.

Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness-based cognitive therapy for depression.* New York: Guilford Press.

Vallis, T. M., Shaw, B. F., & Dobson, K. S. (1986). The Cognitive Therapy Scale: Psychometric properties. *Journal of Consulting and Clinical Psychology, 54,* 381–385.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Spinger-Verlag.

Williams, J. B. (1988). A structured interview guide for the Hamilton Depression Rating Scale. *Archives of General Psychiatry, 45,* 742–747.

Young, J., & Beck, A. T. (1980). *Cognitive Therapy Scale: Rating manual.* Unpublished manuscript, University of Pennsylvania, Philadelphia.

Zettle, R. D., & Rains, J. C. (1989). Group cognitive and contextual therapies in treatment of depression. *Journal of Clinical Psychology, 45,* 436–445.

Received January 3, 2005
Revision received February 17, 2006
Accepted February 17, 2006 ■