Effectiveness of cognitive therapy for depression in a community mental health center: A benchmarking study

By: Kari A. Merrill (Eddington), Valerie E. Tolbert, and Wendy A. Wade

Merrill, K.A., Tolbert, V.A., & Wade, W.A. (2003). Effectiveness of cognitive therapy for depression in a community mental health center: A benchmarking study. *Journal of Consulting and Clinical Psychology, 71*(2), 404-409. PubMed PMID: 12699035.

Abstract:

The current study examined the feasibility and effectiveness of transporting an empirically supported treatment for depression, cognitive therapy (CT), to a community mental health center setting. CT was delivered to 192 adult outpatients with major depression, and a benchmarking strategy compared results with those of 2 randomized controlled trials (RCTs). The 3 samples were largely similar in terms of initial severity of depression, and CT was as effective in reducing depressive symptoms in the current sample as in the RCTs. More favorable outcome was associated with less severe initial depression, more therapy sessions, more years of education, and absence of a comorbid personality disorder. This study demonstrates that an empirically supported treatment can be transported effectively to a clinical setting.

Keywords: cognitive therapy | effectiveness | depression | community mental health center

Article:

Although randomized controlled trials (RCTs) are critical for establishing treatment efficacy, researchers and clinicians have argued that there is a need to follow these trials with studies conducted under conditions that better represent the real settings and populations for which the protocol in question is to be applied (e.g., Chambless & Hollon, 1998; Goldfried & Wolfe, 1998). Termed *effectiveness* studies, these investigations have the goal of increasing external validity. In clinical settings it may be difficult to use the restrictions used in controlled efficacy studies. Careful selection of clients generally may not be possible, and variables such as differential attrition and treatment drift are common (Hoagwood, Hibbs, Brent, & Jensen, 1995). In addition, intensive therapist training and monitoring may not be available to clinicians outside research settings. Furthermore, it may not be feasible to withhold treatment from appropriate clients in a control condition. Thus, transporting empirically supported treatments to naturalistic clinical settings may require a shift from internal validity to favor the maximizing of generalizability.

The current study investigated the effectiveness of Beck's cognitive therapy for depression (CT; Beck, Rush, Shaw, & Emery, 1979) delivered in a community mental health center, the

Center for Behavioral Health (CBH; Bloomington, Indiana). In addition to providing much-needed outcome data for CT on a large, community-based sample, this study provides data collected in a service-oriented setting. Our primary research question was "How effective is CT delivered to patients in an outpatient community-based clinic compared with results obtained in controlled trials?" We adopted a benchmarking strategy (McFall, 1996) used in recent studies on anxiety disorders (e.g., Wade, Treat, & Stuart, 1998) to compare our results with those of two RCTs.

Two RCTs were chosen as benchmarks against which to compare the results from our CBH sample. The first study, by Hollon et al. (1992), involved 107 depressed patients randomly assigned to either CT alone, CT plus pharmacotherapy, or pharmacotherapy. The second is the Treatment of Depression Collaborative Research Program (TDCRP; Elkin, Parloff, Hadley, & Autry, 1985), in which 250 patients were randomly assigned to CT, interpersonal psychotherapy, antidepressant, or medication placebo (for an overview of methods, see Elkin et al., 1985).

An important consideration in treatment outcome research is not only whether the treatment is effective but also for whom it is effective. Several potential predictors of outcome have been identified in previous studies, such as severity of depression, homework compliance, and cognitive functioning (Elkin et al., 1995; Persons, Burns, & Perloff, 1988; Rude & Rehm, 1991). In the current study, potential predictors of outcome were examined first by defining outcome as the posttreatment Beck Depression Inventory (BDI; Beck et al., 1979) score, consistent with previous studies. Second, outcome was defined by categorizing clients according to whether they evinced clinically significant change (as defined by norms inOgles, Lambert, & Sawyer, 1995).

Method

Participants

Participants were 192 adults admitted for treatment through the CBH Depression Treatment Clinic (DTC) between 1994 and 1997 for whom complete assessment and treatment data were available and who consented to participate. All clients admitted to the DTC had a primary diagnosis of major depressive disorder, defined by *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994) criteria, and had no active symptoms of substance dependence, antisocial personality, bipolar disorder, or psychosis. Of 322 clients seeking treatment from the DTC during that period, 98 (30%) did not have a primary diagnosis of major depression and 32 (10%) were excluded from this analysis because of incomplete data. No exclusions were made on the basis of depression severity, medication status, or presence of other psychopathology (except those mentioned above).

Measures

Diagnostic interview. The Structured Clinical Interview for the*DSM-III-R* (SCID I and II; Spitzer, Williams, & Gibbon, 1987; Spitzer, Williams, Gibbon, & First, 1990), administered by DTC therapists, determined Axis I and Axis II diagnoses.

BDI. The BDI was administered at intake and prior to each therapy session.

Demographic information. Information from this intake form includes age, gender, ethnicity, marital status, current income, educational background, and insurance status.

The Hamilton Rating Scale for Depression (HRSD; 17-item version). The HRSD (Hamilton, 1960) is a commonly used clinician rating of depressive symptoms.

Global Assessment of Functioning (GAF; a revision of the Global Assessment Scale). The GAF (Endicott, Spitzer, Fleiss, & Cohen, 1976) is a single rating of overall functioning in the *DSM-III-R* with 10 anchor points and scores ranging from 1 to 100.

Procedure

Clients not directly referred to the DTC were screened by CBH staff to determine their appropriateness for evaluation in the DTC. Clients then completed the intake self-report measures and were assessed by DTC clinicians using the SCID I and II and HRSD.

Clients were treated individually (sessions were held once a week) using Beck's manualized CT protocol (Beck et al., 1979). The standard length of treatment for depression using the CT protocol in most controlled trials is approximately 12–20 weeks. At the CBH, the number of sessions was not predetermined, as is often the case in clinical settings. Unlike participants in research studies, who often get free or substantially discounted treatment, clients in mental health clinics are often restricted by financial considerations and insurance. Therefore, length of therapy was open-ended, and termination decisions were left to the discretion of the therapists on a case-by-case basis.

Therapist Training

Two primary CBH therapists, a doctoral-level psychologist and a master's-level therapist, were trained via an intensive program offered at the Beck Institute for Cognitive Therapy and Research. Those therapists in turn conducted in-house training consisting of a 3-day workshop followed by intensive one-on-one supervision using a preceptor model. Newly trained therapists were provided with close supervision (including feedback from taped sessions) from the primary therapists and weekly case meetings to ensure treatment integrity. The final team of therapists consisted of three doctoral-level psychologists and five master's-level therapists. No differential effects of CT based on individual therapists or therapist education level were found.[1]

Results

Client Characteristics

Tables 1 and 2 show demographic and clinical characteristics of the three samples. We report effect sizes instead of simple *t* tests for paired comparisons of the dependent variables because of

[1] Differential therapist effects on posttreatment BDI scores for clients with at least two sessions of therapy were examined with initial BDI score as a covariate, yielding $F(7, 143) = 0.90$;*ns*. Effects of therapist education (master's versus doctoral level) were tested similarly, yielding $F(1, 143) = 1.03$;*ns*.

the unequal sizes of the three samples and the fact that our null hypothesis was that the samples would be different. A substantial proportion (42%) of the CBH sample had secondary Axis I diagnoses, and 18% had Axis II diagnoses.[2] There were no significant differences between those with or without comorbid diagnoses in initial BDI, $t(190) = -1.26$; HRSD, $t(190) = 1.89$; or GAF, $t(187) = -0.495$, scores. About half (52%) of the CBH clients were on antidepressants at intake. No significant baseline differences based on antidepressant status were found: BDI, $t(188) = -1.02$; HRSD, $t(188) = -0.93$; GAF, $t(165) = 1.42$.

Table 1. Demographic Characteristics of the Center for Behavioral Health (CBH), Hollon et al. (1992), and Elkin et al. (1989) Samples

Variable	CBH	Hollon et al. (1992)	Elkin et al. (1989)[a]
Total *n* in CT condition	192	25	59
Gender, % female	70	80	70
Ethnicity, % Caucasian	92	91	89
Age in years, *M* (± *SD*)	32.7 (± 11.2)	32.6 (± 10.8)	35 (± 8.5)
Age range, years	18–66	18–62	21–60
Marital status %[b]			
Married	30	32	40[c]
Never been married	47	26	26
Divorced/separated/widowed	22	42	33
Education (highest level %)			
Less than high school	8	20	25[d]
Completed high school	27	32	35
Some college	29	34	40
College graduate and beyond	36	14	
Employment status %[e]			
Employed	65	62	—
Homemaker	4	13	—
Student	17	0	—
Unemployed	12	25	—

Note. Percentages may add up to < 100% because of missing data or categories. CT = cognitive therapy for depression. Dash indicates data are unavailable.

[a] Demographic statistics are for the total Treatment of Depression Collaborate Research Program sample. [b] CBH and Elkin et al. samples differed significantly in married status, $\chi^2(2, N = 251) = 9.05$, $p = .01$. [c] For Elkin et al., married or cohabitating. [d] High school or less. [e] CBH and Hollon et al. samples differed significantly in employment status, $\chi^2(2, N = 213) = 9.87$, $p = .02$.

Effectiveness of CT

Treatment was terminated prematurely for 68% of the sample, of which 58% left without notice, 2.6% transferred or were referred, and 7.3% moved away. The average number of sessions was 7.8, and 28 clients did not return after intake. Treatment completion was indicated when the therapist determined that no further treatment was required. Posttreatment BDI and/or GAF scores were missing for 12 clients.

Among those receiving at least one session of CT, paired comparisons revealed a significant reduction in BDI scores, $t(159) = 16.3$, $p < .001$, and increase in GAF scores, $t(152) = -15.2$, $p <$

[2] Information on comorbid Axis I diagnoses was not available in the two RCTs. Data were available on comorbid personality disorders in the TDCRP sample. However, the method of assessment of personality disorders in that study was quite different, and the accuracy of any direct comparison would be questionable.

.001, from pre- to posttreatment. The average change in BDI and GAF scores was 14.9 (± 7.9) and 13.2 (± 7.6), respectively. Main effects of antidepressant use and initial severity as well as the interaction between the two also were tested. The main effects of antidepressant use on BDI, $F(1, 150) = 0.02$, and GAF scores, $F(1, 150) = 0.36$, were nonsignificant, as were the effects of the interaction on the BDI, $F(1, 150) = 0.14$, and GAF, $F(1, 150) = 0.34$.

Table 2. Clinical Characteristics and Scores on Intake Measures for the Center for Behavioral Health (CBH), Hollon et al. (1992), and Elkin et al. (1989) Samples

Variable	CBH sample	Hollon et al. (1992)	Elkin et al. (1989)
Concurrent antidepressant use, %	52	0	0
Depressive episodes, %			
No previous episodes	34	27	36[a]
At least one previous episode	65	64	64[a]
Previous psychiatric treatment, %	67	64	70[a]
GAF[b]			
M	56.9	46.3	52.4
SD	7.0	7.6	7.9[c]
BDI			
M	27.2	30.1	27.0
SD	8.9	5.7	7.9[c]
HRSD[d]			
M	21.1	24.1	19.5
SD	6.9	4.3	3.9[c]

Note. Percentages may add up to < 100% because of missing data or categories. BDI = Beck Depression Inventory. [a] Percentage reflects previous treatment for depression in the total Treatment for Depression Collaborative Research Program sample. [b] Global Assessment of Functioning (GAF) scores were higher in the CBH sample compared with Hollon et al. ($d = 1.49$) and Elkin et al. ($d = 0.62$). [c] Means and standard deviations are for cognitive therapy for depression patients only. [d] Hamilton Rating Scale for Depression (HRSD) scores were higher in Hollon et al. compared with CBH ($d = 0.45$).

Table 3. Posttreatment Data for the Center for Behavioral Health (CBH), Hollon et al. (1992), and Elkin et al. (1989) Participants Receiving at Least One Therapy Session

Variable	CBH (*n* = 160)	Hollon et al. (1992) (*n* = 25)	Elkin et al. (1989) (*n* = 59)
BDI			
M ± *SD*	14.6 ± 10.3	13.3 ± 12.0	13.4 ± 10.6
d		0.12	0.12
GAF			
M ± *SD*	67.6 ± 9.9	—	64.4 ± 12.4
d		—	0.27

Note. BDI = Beck Depression Inventory; GAF = Global Assessment of Functioning. Dash indicate that data are not available.

Table 3 shows results from the benchmarking comparisons of the CBH sample with the two RCTs using posttreatment data for CT participants with at least one treatment session. To provide a more stringent comparison of our sample with the TDCRP data, we selected CBH clients with BDI ≥ 17, HRSD ≥ 14, and at least four CT sessions ($n = 100$) to compare with Elkin et al.'s (1989) sample of clients with 3.5 or more weeks of treatment ($n = 50$).[3] The effect size for the difference in mean pretreatment BDI scores (29.5 ± 7.7 for CBH; 27.5 ± 8.1 for Elkin et al., 1989) was small ($d = 0.26$) and for GAF scores (55.6 ± 6.9 for CBH; 52.1 ± 7.9 for Elkin et al.,

[3] A comparable sample of clients who had started, but not necessarily completed, treatment was not reported in Hollon et al. (1992).

1989) was moderate (d = 0.48). Mean pretreatment HRSD scores were higher in the CBH group (23.13 ± 5.34 for CBH; 19.6 ± 3.7 for Elkin et al., 1989; d = 0.73). Posttreatment data for the restricted samples are shown in Table 4.

Table 4. Posttreatment Data for the Center for Behavioral Health (CBH) and Elkin et al. (1989) Restricted Samples

Variable	CBH (n = 100)	Elkin et al. (1989) (n = 50)
BDI		
M ± *SD*	13.05 ± 9.7	11.5 ± 9.7
d		0.16
GAF		
M ± *SD*	68.82 ± 9.5	66.5 ± 12.6
d		0.21

Note. CBH sample with pretreatment Hamilton Rating Scale for Depression Score ≥ 14, Beck Depression Inventory (BDI) ≥ 17, and number of sessions ≥ 4; Elkin et al. sample with ≥ 3.5 weeks of therapy. GAF = Global Assessment of Functioing.

Clinically significant change was defined using norms from a nondistressed, general population sample (see Ogles et al., 1995). Forty-eight percent of the total sample and 57% of the restricted sample showed clinically significant change using a liberal criterion (BDI < 13.5; within two standard deviations of the nondistressed sample). A more conservative criterion (BDI ≤ 9; within one standard deviation) yielded 35% of the total sample and 55% of the restricted sample.

Table 5. Variables Associated With Outcome (Posttreatment Beck Depression Inventory [BDI] Scores) for Treatment Starters Using Linear Regression (n = 163)

Pretreatment variable	*B*	*SE B*	*β*
Constant	27.36	4.54	
Composite (BDI and HRSD)	3.75	0.84	.32***
Age (in years)	0.00	0.06	.01
Gender (1 = male; 0 = female)	–2.15	1.50	–.10
Income (1–16 rating scale)	–0.30	0.14	–.16*
Education (in years)	–0.47	0.27	–.12
Previous tx (1= previous psychiatric tx; 0 = no previous tx)	0.90	1.58	.04
Medication (1 = antidepressant use; 0 = no use)	–0.70	1.51	–.03
Total no. of therapy sessions	–0.40	0.12	–.26**
Axis II (1 = Axis II diagnosis; 0 = no Axis II)	4.28	1.89	.16*

Note. HRSD = Hamilton Rating Scale for Depression; Previous tx = any previous psychiatric treatment; antidepressant = concurrent antidepressant use; Axis II = any Axis II diagnosis.
* p < .05. ** p < .01. *** p < .001.

Predictors of Outcome

Linear regression predicted posttreatment BDI scores in clients with at least one treatment session using several pretreatment variables, shown in Table 5. Variables were entered in one block and missing data replaced with the sample mean. The resulting model had an adjustedR^2 of .28 (standard error of the estimate = 8.6) and was significant, $F(9, 153) = 7.90$, $p < .001$. Four variables had beta coefficients significant at p < .05: initial severity (a composite of pretreatment

BDI and HRSD),[4] number of sessions, presence of Axis II diagnosis, and income. Years of education approached significance at $p < .10$.

Logistic regression tested predictors of the dichotomous dependent variable, clinically significant change, which was defined as a posttreatment BDI score less than 13.5 (see Ogles et al., 1995). Twelve clients with initial BDI scores equal to or below 14 were excluded. The initial regression model yielded $\chi^2(9,N = 150) = 40.33$, $p < .001$; Cox and Snell $R^2 = .26$; classification accuracy rate was 76.7%. Variables, associated beta coefficients, and odds ratios are shown in Table 6. A second model was tested retaining only the predictors with significant ($p < .05$) Wald statistics (income, severity composite, number of sessions, and Axis II). The difference in fit was not significant, $\chi^2(5,N = 150)$ difference = 4.28. The second model yielded, $\chi^2(4,N = 150) = 36.05$, $p < .001$; Cox and Snell $R^2 = .23$, and classification accuracy rate was 74.4%.

Table 6. Summary of Logistic Regression Analysis for Variables Predicting Clinically Significant Change for Treatment Starters (n = 151)

Pretreatment variable	*β*	OR	95% CI	*p*
Model 1				
Gender (1 = male; 0 = female)	–0.11	0.90	0.38–2.14	.809
Age (in years)	–0.02	0.98	0.95–1.02	.414
Income (1–16 rating scale)	0.10	1.10	1.02–1.20	.019
Education (in years)	0.15	1.16	0.98–1.38	.079
Composite (BDI and HRSD)	–0.78	0.46	0.26–0.80	.001
Medications (1 = antidepressant use; 0 = no use)	0.22	1.24	0.50–3.06	.640
Previous tx (1 = previous psychiatric treatment; 0 = no previous treatment)	–0.89	0.41	0.16–1.05	.076
Total no. of therapy sessions	0.10	1.11	1.03–1.19	.006
Axis II (1 = Axis II diagnosis; 0 = no Axis II)	–1.25	0.29	0.08–0.98	.046
Model 2				
Income	0.09	1.09	1.01–1.17	.022
Composite	–0.78	0.46	0.28–0.76	.002
Total no. of therapy sessions	0.11	1.12	1.04–1.19	.001
Axis II	–1.30	0.27	0.09–0.86	.027

Note. Previous tx = any previous psychiatric treatment; antidepressant = concurrent antidepressant use; Axis II = any Axis II diagnosis. OR = odds ratio; CI = confidence interval; BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression.

Discussion

The present study evaluated the effectiveness of CT for depression delivered in a naturalistic clinical setting by a benchmarking strategy. Examination of demographic and clinical characteristics suggest that the CBH sample was similar to the two RCT samples in terms of age, race, gender, previous depressive episodes and treatment, and pretreatment BDI scores. Pretreatment HRSD scores in the CBH full sample were lower compared with Hollon et al. (1992) but, in the restricted sample, were actually higher than Elkin et al. (1989).

[4] This composite was obtained as in Hollon et al. (1992) by first normalizing BDI and HRSD scores and then computing an average of the two normalized scores for each subject. However, note that Hollon et al. used two additional measures in their composite.

An important consideration was the allowance of uncontrolled antidepressant use in the CBH sample. Over half of the CBH clients were on antidepressant medications at intake. However, there were no differences between antidepressant users versus nonusers on pre- or posttreatment measures of depression severity or global functioning.

Overall, the outcome data from the CBH clients compare favorably to the two RCT studies. CT was effective in reducing self-reported depressive symptoms and in increasing global functioning. Are these statistical differences clinically significant? A substantial percentage of the CBH clients reported posttreatment BDI scores that fell within the normative range, depending on the cutoff used. These effectiveness results are particularly striking in light of the few prerequisites for client participation used in the study. Because minimum BDI and/or HRSD scores were not required for participation, the sample included clients whose self- or clinician-rated depressive symptoms initially were somewhat mild, providing a conservative test of pre–post comparisons. Effectiveness results were comparable for analyses that either included or excluded the more mild cases. Although there was not a large enough sample of such cases to make direct comparisons of the effectiveness of CT in mild versus moderate to severe cases, these results indicate that the addition of these clients to the study did not have an adverse impact on the effectiveness of CT.

A second goal of study was to explore the association of clinical and demographic client characteristics and treatment-related variables with treatment outcome, information that could have implications for clinical decision-making. Consistent with previous studies (e.g., Frank, Kupfer, Jacob, & Jarrett, 1987; Thompson, Gallagher, & Czirr, 1988), results indicated that more favorable outcome was associated with lower pretreatment severity, higher income, absence of Axis II pathology, and greater number of sessions.

In general, these findings suggest, not surprisingly, that individuals at a higher level of functioning before treatment continue to be less symptomatic by the end of treatment. The data also indicate that a greater number of therapy sessions are associated with a more favorable outcome. This finding is in contrast to other studies (Gordon & Wedge, 1998; Shapiro et al., 1994), although such studies often have used a restricted range of sessions. It is clear, however, that improvement in psychotherapy follows a log-linear model, with rates of improvement in symptoms slowing as the number of sessions mount (Howard, Lueger, Martinovich, & Lutz, 1999). However, if therapy is a sequential process, perhaps changes made later in therapy are qualitatively and quantitatively different than those made earlier.

Most CBH clients attended substantially fewer sessions than clients in the RCTs, yet still showed similar levels of improvement. For example, Elkin et al. (1989) provided clients with a standard maximum of 20 sessions over a 15-week period, as recommended by Beck et al. (1979). Hollon et al. (1992) reported that clients were allowed up to 20 sessions in a 12-week period, with the average client receiving 14.5 sessions over 11.5 weeks. In contrast, the mean number of sessions received by CBH clients (who completed at least one session of CT) was 9.21 (± 6.48, range 1–24).[5] Therefore, at least in the current sample, CT clients achieved similar levels of improvement with fewer sessions than typically provided in RCTs.

[5] Most clients tended to space their appointments more liberally than in the RCTs, resulting in a larger variation of treatment length.

There are several important limitations in the current study, the most obvious of which is the absence of a control group. However, the use of a benchmarking strategy allows us to compare the degree of change in the CBH sample to that in RCTs and to conclude, albeit with caution, that the results support the effectiveness of CT. A second threat to internal validity is the prevalence of antidepressant use in the current study. Although the data demonstrated that antidepressant use did not impact outcome, further research is needed to examine the question of relative effectiveness of medications versus therapy in a community population.

The reliance on the BDI as the primary index of treatment outcome also presents a limitation. Although the BDI is widely used in outcome research, a more comprehensive assessment of outcome would have been preferable. As discussed by Chambless and Hollon (1998), symptom measures alone do not adequately capture quality of life. The limited resources of the present study, however, prohibited compensation of clients for data collection. The advantages of the BDI are its routine, weekly administration as part of the CT protocol and its use in previous studies.

Finally, the current study lacked therapy adherence ratings and, despite the intensive training provided in CT, the degree of adherence is unknown. As discussed by Garfield (1996), whereas research settings emphasize training and treatment integrity, practitioners are much more variable in terms of their use of manuals. Although it would have been desirable to examine adherence as an independent variable, the conditions under which the current study was conducted prohibited the use of the expensive, time-consuming process of gathering such data.

The current study demonstrated that a treatment program involving CT for depression can be implemented successfully in a clinical service setting, and the findings support the effectiveness of CT in this context. Although studies of this type cannot replace controlled trials, they can provide valuable information on transporting empirically supported treatments as well as effectiveness data for a given clinic.

References

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

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

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*, 7–18.

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., Parloff, M. B., Hadley, S. W., & Autry, J. H. (1985). NIMH Treatment of Depression Collaborative Research Program. Background and research plan. *Archives of General Psychiatry*, *42*, 305–316.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. T., Sotsky, S. M., & Collins, J. F. (1989). NIMH Treatment of Depression Collaborative Research Program: I. General effectiveness of treatments. *Archives of General Psychiatry*, *46*, 971–982.

Endicott, J., Spitzer, R. L., Fleiss, J., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry*, *33*, 766–771.

Frank, E., Kupfer, D. J., Jacob, M., & Jarrett, D. (1987). Personality features and response to acute treatment in recurrent depression. *Journal of Personality Disorders*, *1*, 14–26.

Garfield, S. L. (1996). Some problems associated with "validated" forms of psychotherapy. *Clinical Psychology: Science and Practice*, *3*, 218–229.

Goldfried, M. R., & Wolfe, B. E. (1998). Toward a more clinically valid approach to therapy research. *Journal of Consulting and Clinical Psychology*, *66*, 143–150.

Gordon, K., & Wedge, G. (1998). Counseling in primary care: A 2-year follow-up of outcome and client perceptions. *Journal of Mental Health-United Kingdom*, *7*, 631–636.

Hamilton, M. (1960). A Rating Scale for Depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56–61.

Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*, 683–687.

Hollon, S. D., DeRubeis, R. J., Evans, M. D., Wiemer, M. J., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1992). Cognitive therapy and pharmacotherapy for depression: Singly and in combination. *Archives of General Psychiatry*, *49*, 774–781.

Howard, K. I., Lueger, R. J., Martinovich, Z., & Lutz, W. (1999). The cost-effectiveness of therapy: Dose–response and phase models. In N. E. Miller & K. M. Magruder (Eds.), *Cost-effectiveness of psychotherapy: A guide for practitioners, researchers, and policymakers* (pp. 143–152). New York: Oxford University Press.

McFall, R. M. (1996, July1). *Consumer satisfaction as a way of evaluating psychotherapy: Ecological validity and all that versus the good old randomized trial* (panel discussion). 6th Annual Convention of the American Association of Applied and Preventative Psychology, San Francisco.

Ogles, B. M., Lambert, M. J., & Sawyer, J. D. (1995). Clinical significance of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Journal of Consulting and Clinical Psychology*, *63*, 321–326.

Persons, J. B., Burns, D. D., & Perloff, J. M. (1988). Predictors of dropout and outcome in cognitive therapy for depression in a private practice setting. *Cognitive Therapy and Research*, *12*, 557–575.

Rude, S. S., & Rehm, L. P. (1991). Response to treatments for depression: The role of initial status on targeted cognitive and behavioral skills. *Clinical Psychology Review*, *11*, 493–514.

Shapiro, D. A., Barkam, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic interpersonal therapy. *Journal of Consulting and Clinical Psychology*, *62*, 522–534.

Spitzer, R. L., Williams, J. B., & Gibbon, M. (1987). *Instruction manual for the Structured Clinical Interview for DSM–II–R*. (Available from the Biometrics Research Department, New York State Psychiatric Institute, 722 W. 168 St., NY, NY 10032)

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1990). *User's guide for the Structured Clinical Interview for the DSM–III–R: SCID*. Washington, DC: American Psychiatric Press.

Thompson, L. W., Gallagher, D., & Czirr, R. (1988). Personality disorder and outcome in the treatment of late-life depression. *Journal of Geriatric Psychiatry*, *21*, 133–153.

Wade, W. A., Treat, T. A., & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology*, *66*, 231–239.