Effectiveness of Cognitive-Behavioral Therapy for Adolescent Depression: A Benchmarking Investigation

V. Robin Weersing, Yale University School of Medicine

Satish Iyengar, University of Pittsburgh

David J. Kolko, Boris Birmaher, David A. Brent, Western Psychiatric Institute and Clinic

In this study, we examined the effectiveness of cognitive-behavioral therapy (CBT) for adolescent depression. Outcomes of 80 youth treated with CBT in an outpatient depression specialty clinic, the Services for Teens at Risk Center (STAR), were compared to a "gold standard" CBT research benchmark. On average, youths treated with CBT in STAR experienced significantly slower symptom improvement than youths in the CBT benchmark. However, outcomes for STAR teens were more similar to the research benchmark when accounting for differences in referral source (clinical versus advertisement) between the datasets. Results support further efforts to test the effectiveness of CBT in clinically representative community practice settings and samples.

THIRTY YEARS AGO, the existence of youth depression was still a topic of debate. Since that time, epidemiological research has revealed that depression is a serious and relatively common psychiatric

0005-7894/06/036–048$1.001/0

syndrome in youth, with one out of every four youths experiencing a clinically significant mood episode by the end of puberty (Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993). Early onset of depression symptoms substantially impacts current functioning in social and educational roles and may affect future adult functioning (e.g., Rohde, Lewinsohn, & Seeley, 1994). In addition, youth depression is a potent risk factor for suicide (Brent et al., 1993; Gould et al., 1998; Shaffer et al., 1996)–the third leading cause of death for schoolchildren, adolescents, and young adults (National Institute of Mental Health, 1999).

With this growing understanding of the prevalence and impact of youth depression have come efforts to develop effective interventions. At last review, there were 15 randomized controlled trials (RCTs) of psychosocial interventions for depressed children and adolescents (Brent, Gaynor, & Weersing, 2002). Examination of this small body of work quickly reveals that cognitive behavioral therapy (CBT) is the research treatment of choice, with 13 of the 15 RCTs testing its effects. Overall, results of these CBT studies have been quite promising. CBT reliably outperforms wait-list and attention placebo control conditions, and meta-analyses of CBT outcomes yield medium to large effect sizes (1.27, Lewinsohn & Clarke, 1999; 1.06, Reinecke, Ryan, & Dubois, 1998). For depressed adolescents, CBT may be superior to family therapy (Brent et al., 1997), supportive counseling (Brent et al., 1997), and relaxation alone (Wood, Harrington, & Moore, 1996), and equivalent in efficacy to interpersonal psychotherapy (Rosselló & Bernal, 1999). A recent investigation comparing the effects of CBT, antidepressant medication, and their combination has yielded more equivocal results,

although the research team concluded that the combined CBT and medication protocol produced the greatest benefit for depressed teens (Treatment for Adolescents With Depression Study Team [TADS] 2004).

Despite substantial empirical support for cognitive-behavioral techniques, the typical package of services received by depressed youth in community clinic settings is unlikely to include exposure to CBT. Surveys of community practitioners suggest that, in everyday practice, therapists most often provide psychodynamic therapy, family systems interventions, or eclectic combinations of techniques drawn from multiple theoretical orientations (Addis & Krasnow, 2000; Kazdin, Siegel, & Bass, 1990; Weersing, Weisz, & Donenberg, 2002). The effectiveness of these treatments has yet to be demonstrated. In a recent study, Weersing and Weisz (2002) tracked the outcomes of depressed youth treated with psychodynamic-eclectic therapy in community mental health centers (CMHCs). While CMHC youth did improve over the course of the study, the rate of improvement was indistinguishable from the natural remission rate of untreated youth depression reported in epidemiological surveys (e.g., Kovacs, 1996; Kovacs, Obrosky, Gatsonis, & Richards, 1997).

While these poor effects of therapy are provocative, they leave a fundamental question unanswered: Would community clinic youth have fared any better if they had received CBT? Little data exists to address this query. Historically, clinical trial research has occurred under conditions and in samples substantially different from clinical practice, and it is unclear if the positive effects of CBT in RCTs would generalize across these many differences (Hammen, Rudolph, Weisz, Rao, & Burge, 1999; Weisz, Weiss, & Donenberg, 1992). Currently, studies examining predictors of clinical trial treatment response may provide the best available clues as to the likely "robustness" of CBT in practice. Unfortunately, the number of such studies in the youth depression literature is small, and findings have not been clear-cut. For example, the relationship between comorbidity and depression treatment response varies dramatically across youth depression clinical trials, with well-designed studies reporting opposite effects (cf. Brent et al., 1998; Clarke, Hops, Lewinsohn, Andrews, Seeley, & Williams, 1992; Rohde, Clarke, Lewinsohn, Seeley, & Kaufman, 2001).

In short, there is much that remains to be learned about the effects of CBT for youth depression in real-world clinical samples and contexts. To begin addressing this knowledge gap, in this article we report the results of a study investigating the effectiveness of CBT in an outpatient depression specialty clinic. Data for this project were collected from the Services for Teens at Risk (STAR) Center, a working outpatient service based in a large psychiatry department. While the STAR Center functions as an active clinic, it does share many features with clinical trials. As in most RCTs, the STAR Center concentrates on the treatment of a focused clinical problem—depression. STAR uses CBT as its psychosocial intervention model, and, upon joining the Center, therapists are extensively trained and supervised in CBT techniques. In addition, treatment at STAR is fully funded by the state government and is free to teens and their families.

In a number of other respects, however, therapy at STAR is representative of real-world clinical care. While therapists are trained in CBT at the beginning of their employment, they operate autonomously once they are senior clinicians. The length and session-by-session content of treatment is not fixed across patients, and psychotropic medication may be used as deemed medically necessary (see Birmaher, Brent, & Work Group on Quality Issues, 1998; Hughes et al., 1999). Teens and families come to the STAR Center via clinical referral routes, including direct referral from inpatient units in the medical center's psychiatric hospital. Finally, unlike many clinical trials, the STAR Center does not exclude youth from treatment if they meet criteria for serious comorbid diagnoses (e.g., substance abuse) in addition to their primary diagnosis of depression.

Given this blend of clinical trial and clinical practice characteristics, we viewed the STAR Center as a natural laboratory in which to begin examining the generalizability of CBT effects beyond the circumstances and samples typically encountered in therapy research trials. To accomplish this task, we reviewed the medical records of depressed adolescents who presented for intake at STAR between 1995 and 2000. As part of the STAR Center's operation as a CBT clinic, standardized assessments are administered to teens and their parents at intake and during therapy. Historically, these data have been collected for clinical purposes—to guide treatment planning and provide youths and their families feedback about treatment progress. In the current investigation, these data were used to model the rate of improvement in depression symptoms over the course of treatment at STAR. To anchor the magnitude of these effects, we then compared the STAR Center symptom trajectories to a relevant benchmark, the outcomes of CBT for adolescent depression in the clinical trial literature.

In past research, the benchmarking strategy has been used to assess the transportability of research treatments from laboratory to practice settings (e.g., Wade, Treat, & Stuart, 1998). Outcome

data from the transported treatment are compared, point-by-point, to the "gold standard" outcomes in the original clinical trial, using the same measures and definitions of improvement. If the effects of the intervention in the community replicate these ideal outcomes, logic would dictate that the transported protocol "works" in the new clinical context and clientele (McFall, 1996). A growing number of benchmarking studies support the transportability of research-based interventions for adults with panic disorder (panic control training; Wade et al., 1998), obsessive-compulsive disorder (exposure and response prevention; Franklin, Abramowitz, Kozak, Levitt, & Foa, 2000), bulimia (CBT; Tuschen-Caffier, Pook, & Frank, 2001), depression (cognitive therapy; Merrill, Tolbert, & Wade, 2003), and social phobia (CBT; Lincoln et al., 2003).

Our use of the benchmarking method followed a similar logic. We began by identifying a CBT clinical trial to use as a gold standard of outcome. This task was made easier by the history of the STAR Center. From 1991 to 1995, STAR referred patients to the Brent et al. (1997) clinical trial of CBT for adolescent depression.[1] The Brent study is a major RCT within the youth depression literature, with a substantial sample (N = 107, CBT n = 37), standardized diagnostic and symptom assessment, excellent treatment adherence ratings, and statistically and clinically significant results supporting the efficacy of CBT (for review, see Weersing & Brent, 2003). The Brent trial recruited their sample of depressed youths from clinical sources, such as STAR, as well as through newspaper advertisement. Accordingly, youth seen in STAR between 1995 and 2000 should include teens similar to those who were enrolled in Brent et al. (1997), as well as more clinically complicated youth who would likely have been screened out of the clinical trial. Thus, along a number of dimensions, the Brent RCT seemed a relevant and interesting yardstick against which to measure the effectiveness CBT in the more representative sample and context of the STAR Center.

We adopted a three-step approach to data analysis and benchmarking. The full dataset of the Brent trial was made available for analysis, making it possible for us to statistically test for differences between

STAR and the RCT. As a first step, we compared the demographic and clinical characteristics of the STAR sample to the sample of Brent et al. (1997). Second, we examined differences in treatment between the STAR Center and our benchmark. Third, we assessed the generalizability of CBT effects by comparing the STAR depression symptom trajectory against the outcomes of the Brent RCT. Finally, to unpack any observed differences in treatment outcome, we also planned follow-up analyses to search for predictors of treatment response within and between the two samples.

Method

PARTICIPANTS

Participants for this investigation were drawn from two sources: (a) the patient database and medical records system of the STAR Center, an outpatient clinic providing treatment to depressed and suicidal adolescents, and (b) the research data file of the Brent et al. (1997) clinical trial of CBT for adolescent depression.

STAR Sample. The STAR Center is affiliated with a psychiatric hospital within a large academic medical center. As part of the standard intake procedures of the medical center, patients are asked whether they consent to anonymous use of their medical records data in future clinical research.[2] STAR Center families who provided consent to this use of their data were screened for inclusion into this investigation by an "honest broker." The honest broker was an individual approved by both the hospital and our institutional review board to review patients' medical records and abstract anonymous data for the use of this research project. The broker held no personal, professional, financial, or other interest in the outcomes of the study or the publication of this work. This process was designed to maximally protect patients' privacy, while still allowing for meaningful research to be conducted in a clinical service setting, using real patient data.

In order to obtain a sample with depression symptoms comparable to those of youth in Brent et al. (1997), youths were selected who presented at intake with: (a) *DSM-III-R* or *DSM-IV* (American Psychiatric Association, 1987, 2000) diagnosis of major depressive disorder (MDD); and (b) youth self-reported depression in at least the borderline

[1]STAR and the Brent RCT were based at the same medical center. The STAR clinic is embedded in the Center for Children and Families, a general outpatient service for children and adolescents. The Brent RCT also shared some space with the outpatient service and had access to additional facilities dedicated to research activity and support staff.

[2]Our medical records department released data only for those families who provided consent for archival research. As a result, we were unable to obtain information on the number or characteristics of youths and families who declined to participate.

clinical range (Beck Depression Inventory score of 13 or higher; Beck, Steer, & Garbin, 1988). We expected that STAR clinic youth also would meet criteria for a variety of comorbid diagnoses; however, youth were not excluded from the current study on the basis of comorbidity. The resulting sample of 80 adolescents was predominantly female (77%) and Caucasian (85%). The majority (52%) met criteria for at least one diagnosis in addition to MDD, most often an anxiety disorder (45%) or comorbid dysthymic disorder (34%).

RCT Sample. Depressed youths were enrolled in the Brent RCT between 1991 and 1995, a time span not overlapping with the intake dates covered in our STAR medical record review. During the RCT recruitment period, the STAR Center referred interested teens and families to the clinical trial, and the research team also advertised for participants in community newspapers. All potential subjects were screened for presence of *DSM-III-R* MDD and were required to evidence Beck Depression Inventory (BDI; Beck et al., 1988) scores in at least the borderline clinical range (greater than or equal to 13). Youth were excluded from the study if they met criteria for psychosis, bipolar disorder, obsessive-compulsive disorder, eating disorder, recent substance abuse, current physical or sexual abuse, pregnancy, or chronic physical illness. The sample was predominantly Caucasian (85%) and female (75%) and had moderate rates of comorbid anxiety disorder (32%) and dysthymia (22%). Demographic and clinical characteristics of the sample are described further in the results section.

MEASURES

BDI (Beck, Steer, & Garbin, 1988). The BDI is a 21-item self-report measure of depression. The measure includes items assessing a wide range of depression symptoms including dysphoria, anhedonia, suicidality, and disturbances in sleep, appetite, and cognitive functioning. The BDI is the most widely used dimensional measure of depression, with over 25 years of research on its psychometric characteristics (Beck et al., 1988). The measure was originally developed for use with adults; however, there is a significant body of research supporting the use of the BDI with adolescents and documenting appropriate adolescent norms (e.g., Roberts, Lewinsohn, & Seeley, 1991). In the current investigation, the BDI was used to identify significantly depressed subjects at intake into STAR, and session-by-session BDI scores were used to construct our outcome measure–depression symptom trajectory over the course of treatment.

Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime Versions (K-SADS-PL; Chambers et al., 1985). The K-SADS-PL is a semistructured interview designed to ascertain present episode and lifetime history of psychiatric illness according to *DSM* criteria. Probes and objective criteria are provided to rate individual symptoms. Interrater and test-retest reliability have been established, as well as convergent and discriminant validity (Kaufman et al., 1997). In both STAR and the RCT, the K-SADS-PL was used to identify youth with current MDD, document the level of comorbid psychiatric diagnoses, and assess suicidality.

ASSESSMENT PROCEDURE

STAR sample. Assessment in the STAR sample occurred as part of routine clinical care at the Center. At intake, youth and family demographic characteristics were collected, and STAR youth completed an initial BDI and were administered the K-SADS-PL by a trained intake clinician. If available, a parent or guardian was also interviewed with the K-SADS, and teens were re-interviewed to resolve any diagnostic discrepancies. Youth history of suicidal behaviors as reported on the parent and youth K-SADS also were recorded. Over the course of treatment, youths completed BDIs at each session to monitor their progress and assist in treatment planning. As a result, the number of BDIs varied between participants as a function of the number of sessions attended and the overall length of treatment. Youths who were in therapy for a short span contributed a small number of assessments, clustered in a tight time frame. Youths attending treatment over a period of many months contributed a greater number of BDI scores over a longer time. In addition to the BDI, number and date of sessions were recorded, and use of psychotropic medication noted.

RCT sample. Assessments in the clinical trial were conducted at (a) intake, (b) the sixth treatment session, (c) treatment termination (approximately 3 months post-intake), (d) 3-month intervals in the first year following termination, and (e) 2 years after termination. Youth and family demographic characteristics were collected at intake. Primary outcome measures collected over time included youth-rated depression symptoms on the BDI, diagnostic status on the K-SADS, interviewer-rated depression, and functional status. In addition, information was collected on youths' suicidality, feelings of hopelessness, and cognitive distortions. Parents' current and lifetime psychopathology was assessed, as were family environment variables.

BENCHMARKING PROCEDURE

The Brent RCT and the STAR Center used the same scales for measuring depression (BDI), calculating diagnoses (K-SADS), and assessing suicidality (K-SADS), facilitating comparison of samples and outcomes across the two settings. Furthermore, the raw dataset of the Brent clinical trial was made available, and we were able to directly test for significant differences between the RCT and STAR with a variety of statistical techniques, including *t*-tests for

Table 1
Youths in STAR compared to the CBT benchmark sample (Brent et al., 1997)

	STAR	CBT benchmark
Number of youths	80	37
Demographic characteristics		
Age (mean/SD)	15.56 (1.40)	15.72 (1.36)
Gender composition (percent male)	23	24
Ethnic composition (percent minority)	15	24
Depression profile		
Intake depression severity (median BDI/SD)	24.00 (8.34)	22.00 (8.12)
Endpoint depression severity (median BDI/SD) [a]	9.00 (10.84) [**]	2.00 (8.47) [**]
Double depression [b] (percent)	34 [#]	16 [#]
History of suicide attempt (percent)	50 [**]	14 [**]
Comorbidity at intake		
Comorbid anxiety [c] (percent)	45	41
Comorbid disruptive behavior [d] (percent)	5 [#]	16 [#]
Exclusionary comorbidity [e] (percent)	15 [*]	0 [*]
At least one comorbid diagnosis [f] (percent)	53	51
Total comorbid diagnoses [f] (mean/median)	0.70 (1.00)	0.70 (1.00)

Note. CBT = cognitive-behavioral therapy.
[a] Distribution of endpoint BDI scores was substantially skewed, and the non-parametric Mann-Whitney *U* test was used for descriptive analysis.
[b] Diagnoses of major depression comorbid with dysthymic disorder.
[c] Diagnoses of separation anxiety disorder, overanxious disorder, panic disorder, specific phobia, social phobia, obsessive- compulsive disorder, posttraumatic stress disorder, generalized anxiety disorder, and/or anxiety disorder NOS.
[d] Diagnoses of conduct and/or oppositional defiant disorder.
[e] Diagnoses that would have resulted in exclusion from Brent et al. (1997).
[f] Not including major depression and dysthymia.
[*] *p* < .01.
[**] *p* < .001.
[#] *p* < .10.

the means of continuous variables, chi-square procedures for categorical variables, and hierarchical linear models for symptom change over time.

Results

COMPARISON OF SAMPLES

As can be seen in Table 1, the sample of youth receiving treatment in STAR was similar in many respects to the youth included in the Brent clinical trial. Youth demographic characteristics were virtually identical, with no significant differences between STAR and the RCT in terms of age, gender, or inclusion of minority youth (all $p > .22$). At intake, the level of depression symptoms endorsed on the BDI also was equivalent, although there was a trend toward higher rates of "double depression" (comorbid MDD and dysthymia) in STAR compared to the RCT, $\chi^2(1, N = 117) = 3.84, p = .05$.

Unlike RCT youths, STAR teens often had a history of suicidality. A full half of the STAR sample reported a prior suicide attempt, while only 13% of RCT youth had a history of attempt, $\chi^2(1, N = 109) = 13.87, p < .001$. As the Brent RCT had specific diagnostic exclusionary criteria, we also expected that youth in STAR would evidence higher levels of comorbidity. Fifteen percent of the STAR sample did meet criteria for one or more of the exclusionary diagnoses from the Brent RCT (substance abuse, $n = 6$; obsessive-compulsive disorder [OCD], $n = 5$; eating disorder, $n = 5$; bipolar-spectrum illness, $n = 2$; or depression with psychotic features, $n = 2$), a significant difference from the RCT sample, $\chi^2(1, N = 117) = 6.18, p = .01$. STAR youth did not significantly differ from RCT youth in proportion meeting criteria for comorbid anxiety (excluding OCD) and disruptive behavior disorders ($p > .65; p > .07$).[3] Although the type of comorbidity generally differed between the samples, the mean number of comorbid diagnoses in STAR and RCT was equivalent, and a similar proportion of youth in both samples met criteria for at least one comorbid diagnosis (all $p > .62$).

COMPARISON OF TREATMENTS

Treatment dose and structure. Both STAR and the RCT based their care on the same CBT protocol for adolescent depression. The shared treatment manual was principle-based and provided general instruction in the use of CBT techniques and the logical sequence of introducing new skills and

[3] Given the low rate of disruptive behavior disorders in both the STAR and RCT samples, we analyzed these categorical data with Fisher's Exact Test ($p = .07$) rather than the Pearson chi-square procedure.

material. However, the manual did not include session-by-session scripts for therapist behavior, a specific teen workbook, or mandatory homework assignments. In the Brent RCT, youths were limited to 12 to 16 sessions of the active intervention and were eligible to receive up to 4 additional booster sessions over a 4-month follow-up period. Services in STAR were not fixed in length, and, under these conditions, STAR youth tended to receive a more variable dose of treatment than RCT youth. As can be seen in Table 2, the median number of sessions in STAR was marginally greater than in the clinical trial (Mann-Whitney U; $p = .08$), but there also was a trend for a larger proportion of STAR youth to receive a very low dose of treatment (less than 8 sessions), $\chi^2(1, N = 107) = 3.58$, $p = .06$.

Treatment content. As described previously, the psychosocial treatment in STAR from 1995 to 2000 was based on the CBT manual from the Brent clinical trial. Accordingly, we describe the content of the manual without specific reference to use in STAR or Brent et al. (1997). At the beginning of treatment, parents and youth were provided psychoeducation about the nature and seriousness of depressive illness (see Brent, Poling, McKain, & Baugher, 1993). Following this initial family orientation, the content of individual CBT sessions focused primarily on altering the irrational, overly negative cognitions viewed to be at the root of depressive symptomatology. Youth were taught to identify their automatic thoughts, accurately label thoughts as distorted or overly pessimistic, and challenge their depressive thinking about themselves and the world. Treatment also targeted difficulties in affect regulation and impulsivity, particularly as related to self-injurious and suicidal behaviors. Youths learned to identify their feelings, use behavioral activities and distraction to regulate mood, and solve problems in a calm and logical manner.

This CBT treatment protocol represents a developmental adaptation of the classic cognitive therapy model developed by Beck and colleagues (Beck, Rush, Shaw, & Emery, 1979). As teens frequently do not complete detailed thought logs, in vivo experiences, such as monitoring cognitions associated with in-session affective shifts, were used to illustrate the cognitive model. In order to match the more concrete cognitive style of younger adolescents, the CBT therapists were instructed to summarize, or have the adolescent summarize, session content frequently. Throughout treatment, therapists were guided to keep the level of abstraction to a minimum, and concrete examples, linked to youths' personal experience, were used whenever possible. For a detailed case example from the clinical trial, see Brent et al. (1996).

Therapist training, supervision, and treatment adherence. Before participating in the Brent RCT, each therapist received 6 months of intensive training in CBT, supervised by external experts from the Beck research team. To be certified for participation in the clinical trial, therapists were required to treat two test cases to criterion. During the RCT, therapists were provided with 1 hour of individual and 1 hour of group supervision per week. As a check on treatment integrity, a random 25% of session tapes were rated by external consultants, and analyses of these ratings indicated that 90% of sessions were of "acceptable" quality or better.

At the beginning of employment in the STAR Center, therapists were provided with training in CBT, including 2 days of background in cognitive theory and 2 additional days of technique-focused training and role-play. During their first 3 months at the Center, clinicians received 2 to 3 hours supervision per week from on-site CBT experts, including former therapists from the Brent RCT.[4] Selected sessions from therapists' first five cases were taped and reviewed in supervision. After this period of training and enhanced supervision, veteran

Table 2
Treatment in STAR compared to treatment in the CBT benchmark
(Brent et al., 1997)

	STAR	CBT benchmark
Number of youths	80	37
Treatment content		
Theoretical orientation	CBT	CBT
Manual available	Yes	Yes
Adherence rated regularly	No	Yes
Treatment length		
Number of weeks in treatment (median)	21.93	26.29
Number of sessions (median)	19.50[#]	15.00[#]
Less than eight sessions (percent)	30[#]	14[#]
Use of psychotropic medication (percent)		
No use of medication during active treatment	35[**]	100[**]
One prescription	44	0
Two or more prescriptions	21	0

Note. CBT = cognitive-behavioral therapy.
[#] $p < .10$.
[**] $p < .001$.

---

[4] Information on therapist identity was not consistently available from the STAR medical records review. Some of the therapists from the Brent RCT were employed by the STAR Center following the conclusion of the trial; these therapists generally assumed supervisory roles.

therapists operated much more independently. Senior therapists attended a 1-hour treatment team meeting and were provided 1 hour of individual supervision per week in which to discuss any problems with their cases (typically 20 patients). Formal treatment adherence ratings and standardized tape review were not regularly conducted with senior staff, although the opportunity for periodic tape review was available if requested by therapists or judged necessary by STAR administrators.

Use of medication. Use of psychotropic medication was grounds for exclusion from the Brent et al. (1997) RCT. However, in the 1-year follow-up period following termination from the study, 19% of youth sought pharmacotherapy, most often treatment with an antidepressant. Youth in STAR were not prohibited from receiving adjunctive psychotropic medication, and the majority did so (65%), a significantly higher proportion than in the RCT, $\chi^2(1, N = 100) = 34.14, p < .001$. These STAR prescriptions were primarily for selective-serotonin reuptake inhibitors (SSRIs; 51%) or other antidepressants (36%).

OUTCOME ANALYSES

STAR symptom trajectory and predictors of treatment response. In order to accurately model outcome in the STAR sample, we estimated BDI depression symptom trajectories for youths using hierarchical linear modeling techniques (HLM 5.05; Raudenbush, Bryk, & Congdon, 2001). Youths seen for services in STAR completed the BDI every session as a part of their regular clinical care. Accordingly, STAR youths varied in both the number and spacing of assessment points, and data were not available from youths after they terminated from the clinic (which occurred at different times for each participant). In these circumstances, the *slope* is widely considered the best estimate of change over time and, in our case, of treatment outcome (see Willett, 1988, for review). In addition to providing the most accurate fit to our data structure, HLM is also robust to missing data points (Bryk & Raudenbush, 1992), and the technique allowed us to model symptom slopes for STAR youths with three or more BDI scores (all youths in our sample).

Conceptually, HLM can be thought of as a series of nested linear regressions. In the STAR sample, we began by computing a model with no predictors of outcome (i.e., an unrestricted base model) in order to obtain estimates of the within-subject and between-subject variance in BDI scores. At the second stage of analysis, the linear effects of time were used to predict within-subject BDI scores, and symptom trajectories were estimated for each individual participant.[5] Results indicated that a linear time model accounted for a majority (68%) of within-subject variance in BDI scores. Overall, youth in STAR significantly improved over the course of treatment, $t(79) = 6.82, p < .001$, with the STAR sample as a whole improving a mean of .62 BDI points per week.

These results appeared to be robust across a wide variety of variables that could potentially impact treatment response. In a series of exploratory HLM analyses ($\alpha = .05$),[6] we found that STAR symptom slopes were not significantly affected by youth age ($p = .10$), gender ($p = .09$), history of suicide attempt ($p = .88$), medication use ($p = .43$), global level of comorbidity (total number of nonmood diagnoses, $p = .84$), presence of double depression ($p = .16$), or presence of a comorbid diagnosis that would have resulted in exclusion from the Brent RCT ($p = .75$). Ethnic minority youth did have flatter slopes (i.e., improved more slowly) than Caucasian teens, $t(78) = 3.14, p = .002$; however, interpretation of these results was complicated by small sample size and substantial collinearity between ethnicity and gender. The current STAR sample contained only 12 African-American youths, all of whom were female.[7] Of note, minority youths did not differ from

---

[5] The Level 1 and Level 2 equations for this basic, linear time model read as follows:

$$\text{Level 1}: (\text{BDI})_{ti} = \pi_{0i} + \pi_{1i}(\text{time})_{ti} + e_{ti}$$

$$\text{Level 2}: \pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}$$

In the Level 1 equation, the outcome variable $(\text{BDI})_{ti}$ is the BDI score at time t for subject i. This score is a function of youths' BDIs at intake (π_{0i}) and their rates of change in BDI scores over the course of therapy (π_{1i}) plus error (e_{ti}). These effects are *random,* in that they are allowed to vary across subjects, and each youth may have his or her own unique value for BDI intercept (π_{0i}) and slope (π_{1i}). In this base time model, there are no specific predictors of youths' slopes at Level 2 of the model. Symptom slopes are simply predicted by the mean slope for the sample (b_{10}) and error (r_{1i}).

[6] As an example, equations for a model examining the effects of youth demographic characteristics on symptom slopes read as follows:

$$\text{Level 1}: (\text{BDI})_{ti} = \pi_{0i} + \pi_{1i}(\text{time})_{ti} + e_{ti}$$

$$\text{Level 2}: \pi_{0i} = \beta_{00} + \beta_{01}(\text{age}) + \beta_{02}(\text{gender}) + \beta_{03}(\text{minority status}) + r_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{age}) + \beta_{12}(\text{gender}) + \beta_{13}(\text{minority status}) + r_{1i}$$

The first level of this model is identical to the base time model described in footnote 6. However, in the Level 2 equations, the random intercept (π_{0i}) and slope (π_{1i}, youth symptom trajectory) of Level 1 is predicted by youth age, gender, and minority status (ethnic minority vs. Caucasian). The Level 2 regression coefficients (β_{11}, β_{12}, β_{13}) capture the strength of relationship between these demographic predictors and symptom change in STAR.

[7] Based in part on these findings, we have increased clinic outreach efforts in minority communities in the greater Pittsburgh area. We hope to provide more complete information on the effectiveness of CBT for ethnic minority populations in future work.
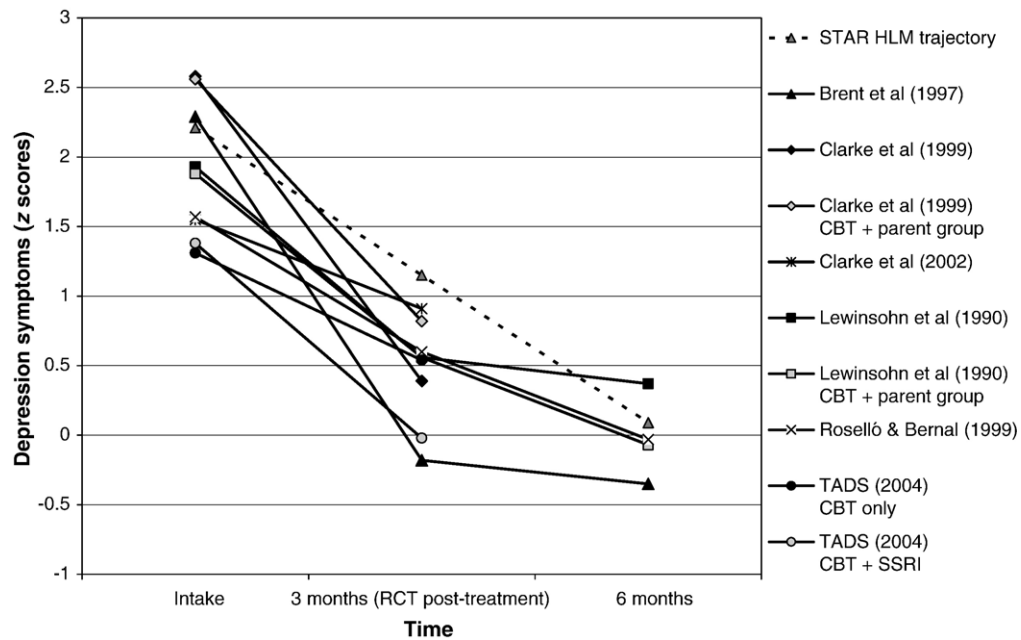
FIGURE 1 Mean HLM symptom trajectory for STAR youths compared to the outcomes of CBT conditions from clinical trials of depressed adolescents.

Caucasian teens in use of medication or number of CBT sessions (all $p > .17$).

Comparison of STAR trajectory to benchmark. In our next set of analyses, we compared the STAR symptom trajectory to the outcomes of youth in Brent et al. (1997). Once again, we used HLM procedures to model outcome, in this instance for the combined sample of STAR ($n = 80$) and the CBT cell of the Brent RCT ($n = 37$; combined $N = 117$).[8] As in our previous HLM analyses, we estimated an unrestricted base model of variance in BDI scores, determined that there was variability to be modeled, and computed individual symptom slopes for each participant in the combined data file. We next predicted between-subject differences in these slopes as a function of enrollment in STAR versus the RCT. As anticipated, youth provided CBT in STAR improved significantly more slowly than youth in the clinical trial, $t(115) = 4.86$, $p < .001$, with this effect accounting for 19% of between-subject variance in symptom slopes.

To place these results in context, in Figure 1 we plotted the mean symptom slope of STAR youths

against the outcomes of the Brent RCT and of several other depression clinical trials (Clarke et al., 1999, 2002; Lewinsohn et al., 1990; Rosselló & Bernal, 1999; TADS, 2004). We selected RCTs that (a) provided treatment to adolescents age 13 to 18 with a primary diagnosis of MDD and (b) assessed depression symptom outcomes with a standardized symptom scale. To compare results across studies, we converted these dimensional symptom scores to a common metric by computing normative z scores (Kendall & Grove, 1988) for the CBT cells of each study at each assessment point. These computations took the form $z_{nt} = (\overline{x}_t - \mu)/(\sigma)$, where $\overline{x}_t$ was the mean score on a depression measure at a given RCT assessment point, μ was the normal population mean, and σ was the normal population standard deviation for the depression measure (see Weersing, 2005). The z score calculated with this formula can be interpreted as an index of depression severity; a intake z score of 2.0 indicates that the mean level of symptoms is two standard deviation units above the community mean for depression, and a normative z score of 0 is equivalent to "normal" level of depression (the community mean). As can be seen in Figure 1, STAR youths began treatment with symptoms comparable to youths in the clinical trials. Youths receiving CBT in the RCTs experienced sharp improvements by posttreatment (3 to 4 months after intake). Symptom change in STAR occurred more slowly, although effects were not dramatically outside the range of other CBT clinical trials.

---

[8] In the STAR data, all youth were in active treatment during the course of assessment. Accordingly, we used BDI scores from the treatment phase of the Brent clinical trial in these HLM analyses. One teen in the RCT attended only one session of active treatment and, thus, contributed only one BDI score for analysis. It was possible to include this participant in the HLM estimation of unrestricted base models and linear time models; however, there were insufficient data to include the youth in models predicting symptom slope.

Comparison of STAR to RCT Subgroups. We next sought to unpack the difference in outcome between STAR and the RCT. The sample and treatment in STAR varied in a number of respects from that of the Brent clinical trial, including presence of exclusionary comorbid conditions (e.g., substance abuse), history of suicide attempt, and use of medication. However, exploratory analyses indicated that these factors were unrelated to symptom slope in the STAR Center (all $p > .42$), suggesting that these differences, while noteworthy, were not likely to be the explanation for the worse outcomes of CBT in STAR.

STAR and the RCT also differed, in part, by the referral source of the youth provided treatment. All youth seen in STAR were referred from clinical sources and were actively seeking services, whereas, in the clinical trial, a third of the sample was recruited into the study via newspaper advertisement. In a follow-up report to the 1997 clinical trial, Brent et al. (1998) found that these non-referred youth were significantly more likely to recover from MDD than clinically referred teens. Thus, to better index the effects of treatment in STAR, we compared the outcomes of STAR youth to these two subgroups of youth from Brent et al. (1997)—teens clinically referred to the RCT and those recruited via advertisement.

We ran three HLM models testing the difference in slope between (a) STAR and RCT clinic youth (combined $N = 106$), (b) STAR and RCT advertisement youth (combined $N = 91$), and (c) RCT clinic and RCT advertisement youth (combined $N = 37$). All comparisons were significant, with the three groups evidencing different symptom trajectories. Overall, youth in STAR improved the most slowly (.62 BDI points/week), followed by clinically referred youth from the clinical trial (.66 BDI points/week). Youth recruited via advertisement had the fastest rates of improvement (.90 BDI points/week). As discussed previously, analysis of slope effects is the preferred technique for modeling outcome in these data. However, for descriptive purposes, we also examined endpoint BDI scores for these three groups; this produced the same ordering of outcomes as the HLM slope analysis (STAR median = 9.00, RCT clinic median = 5.00, RCT advertisement median = 1.00). The three HLM symptom trajectories are plotted in Figure 2.

Discussion

In this investigation, we sought to address a basic, but unanswered, question about the treatment of depressed adolescents: Can CBT, the research standard of care for youth depression, produce positive effects in the samples and settings of real-world clinical service? We examined the outcomes of 80 depressed adolescents treated with CBT in an outpatient depression specialty clinic, the STAR Center. To anchor our evaluation of youth outcomes, we then compared the effects of treatment in STAR to a relevant benchmark–a gold-standard CBT clinical trial (Brent et al., 1997).
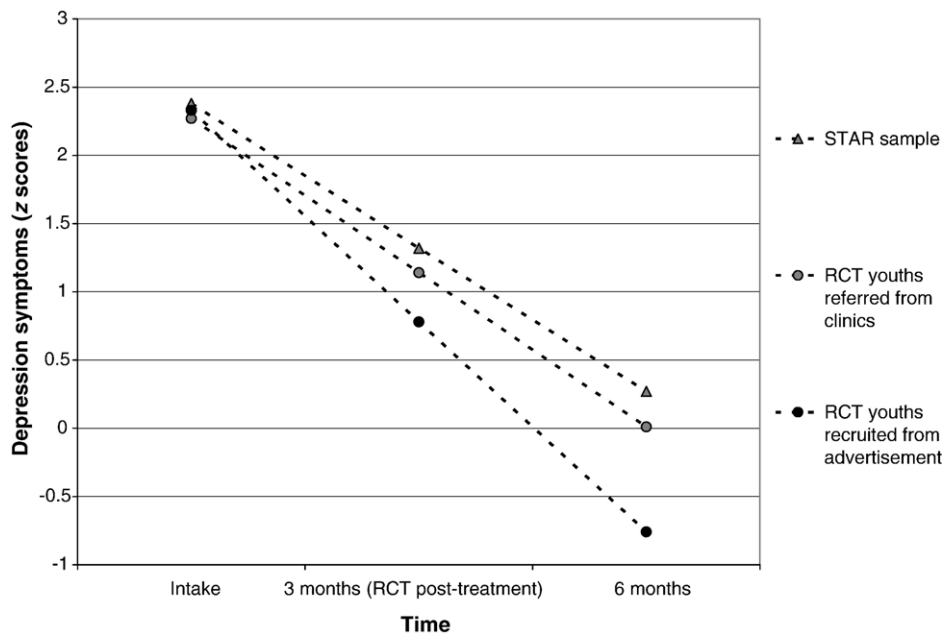

FIGURE 2 Mean STAR HLM symptom trajectory compared to trajectories for Brent RCT youth referred from clinical sources or recruited via newspaper advertisement.

Overall, results appear promising. As can be seen in Figure 1, STAR youth reported depression symptoms at intake comparable to youths in the Brent clinical trial and in the adolescent depression literature at large. Over the course of treatment, STAR youths achieved significant symptom change, improving at a rate of .62 BDI points per week. Given the intake BDI scores in the STAR Center, this rate of improvement maps onto a return to "normal" levels of depression symptoms approximately 6 months after intake (see Roberts et al., 1991). This rate of change was significantly slower than that in the Brent RCT; however, STAR results may compare favorably with the outcomes of community care that does not include exposure to CBT. As reported in Weersing and Weisz (2002), depressed teens treated with eclectic-dynamic therapy in community clinics took nearly a year for their depression symptoms to return to the normal range. This very gradual CMHC improvement mapped almost exactly onto the natural remission rate of untreated youth depression in epidemiological reports (see Kovacs, 1996; Kovacs et al., 1997).

It appears that CBT in STAR may have produced improvement above and beyond the effects of time and natural remission. However, it is also clear that the effects of CBT were attenuated in the STAR Center. To better understand this gap in outcome between STAR and RCT, we looked at differences in sample and treatment characteristics. The two settings differed significantly in youth suicidality, presence of certain psychiatric comorbidities, and use of adjunctive psychotropic medication. However, none of these noteworthy differences between STAR and RCT predicted symptom slope within the STAR Center, suggesting the outcome gap was not likely due to these variables. The failure to find significant clinical predictors of effects is in accord with findings from the Brent clinical trial indicating that CBT was a robust intervention across a variety of clinically complicating factors, including suicidality (Barbe, Bridge, Birmaher, Kolko, & Brent, 2004) and comorbid diagnoses (Brent et al., 1998).

The lack of correspondence between medication use and symptom slope is more puzzling, especially given the recent findings of the TADS investigation of adolescent depression, in which CBT in combination with an SSRI was significantly superior to the effects of CBT alone (TADS, 2004). Youths in STAR were not randomly assigned to receive antidepressants, and it is possible that treating psychiatrists selectively provided medication to the most severe youths in the sample. If providers were very well calibrated to youths' medication needs, this selection bias could erase significant correlations between med-

ication use and outcome. We were unable to assess this hypothesis within this dataset, and it is possible the effects of CBT alone for the full, unmedicated sample of STAR youths would be less impressive than those reported here.

One factor that did appear to account, in part, for the gap in outcome between STAR and RCT was participant referral source. All youths in STAR came to the Center via clinical referral routes— transfer from other mental health providers, referral from schools, discharge from inpatient hospitalization. In contrast, a third of the Brent RCT sample was recruited into the trial via newspaper advertisement. As reported in Brent et al. (1998), CBT was particularly efficacious for these youths, and referral source was a significant predictor of diagnostic status at treatment termination (with recruited youth less likely to meet criteria for MDD). Unsurprisingly, in our analyses, STAR symptom trajectories more closely resembled those of clinically referred RCT youth. Indeed, while the difference between these symptom slopes was statistically significant, at the end of 6 months, the gap in outcome between STAR youth and clinically referred RCT youth would amount, on average, to less than a two-point difference on the BDI.

In the STAR dataset, we were unable to unpack the psychological factors bound up in "referral source" that may account for differences in treatment outcome. Several possible mechanisms seem viable for future investigation. In the Brent RCT sample, clinically referred youths did not differ in depression severity, comorbidity, or functional status from youths recruited via advertisement. However, clinically referred teens did endorse higher levels of hopelessness, and this difference significantly mediated the effect of referral source on outcome (Brent et al., 1998). Hopelessness was not measured in the STAR Center, but it seems likely that STAR youths would report feeling quite hopeless—in addition to the high level of depression symptoms self-reported by STAR teens, a full half of STAR youth had a history of suicide attempt. In addition, referral source may be associated with important familial characteristics of depressed youths. In the Brent clinical trial, parents of clinically referred and recruited youths also did not differ in demographic or clinical characteristics (e.g., socioeconomic status, parental depression). However, parents who answered clinical trial newspaper advertisements may have differed on unmeasured dimensions, such as greater organizational skills or motivation to support their youths' involvement in therapy. Future research on the effectiveness of CBT may benefit from the assessment of hopelessness, treatment motivation, and

other psychological variables that may be associated with clinical referral status.

We also see value in additional research testing the effectiveness of CBT for depressed adolescents who are members of ethnic minority groups. In the STAR sample, African-American teens had a slower rate of improvement than Caucasian youths. Given the relatively small number of minority youths in our sample, and substantial collinearity between ethnicity and gender, we were unable to tease apart possible psychological or contextual variables underlying this difference in outcome. It is worth noting that STAR services provided to ethnic minority youth were comparable in length and number of sessions to those of Caucasian youth, and minority group membership was not significantly related to use of adjunctive psychotropic medication. Interestingly, these service use findings stand in contrast to data from Weersing and Weisz (2002). In the very diverse Weersing and Weisz CMHC sample, ethnic minority youths also had worse outcomes than Caucasian youths (although trajectories for both groups still closely resembled natural remission). In the CMHC, however, minority group members attended significantly fewer therapy sessions, raising the possibility that differences in therapy attrition might be related to the ethnic difference in outcome. The source of the minority outcome gap in our sample is less clear. Only two RCTs of CBT have been conducted in samples with substantial ethnic minority representation (Rosselló & Bernal, 1999; Weisz, Thurber, Sweeney, Proffitt, & LeGagnoux, 1997); in both these investigations, CBT produced positive effects.

In addition, we see benefit in future transportability studies that take into consideration a broader range of therapist, system, and outcome variables than was possible in the current investigation. While we feel that the STAR Center was a useful context in which to *begin* testing the effectiveness of CBT in practice, the majority of depressed teens receive mental health care from less specialized practitioners, in less specialized settings. Therapists in STAR were able to behave as community providers in many respects. Therapy session content and length were not fixed across youths; psychotropic medications were permitted; sessions were unobserved; and supervision faded as therapist seniority increased. However, STAR therapists' caseloads were not those of general community practitioners. Therapists in STAR were able to focus on, and presumably become expert in, a specific clinical problem (depression) and a particular therapeutic approach (CBT). As CBT protocols are disseminated and tested in more general clinical contexts, issues of therapist special-

ization and expertise such as these will likely become more salient.

Other service system issues also may affect the effectiveness of CBT, if the intervention is transported to and tested in a more general outpatient setting. For example, therapy in the Center was provided free to youths as part of a state-funded mental health initiative. As a result, therapists were not unduly burdened by insurance demands or paperwork, and issues of therapist "productivity" were less pronounced than in other service systems (e.g., managed mental health care). As a matter of policy, new STAR therapists also receive thorough training in the CBT model at the beginning of their employment. The cost associated with therapist training activities may not be easily absorbed by a local child guidance center, and attention should be paid to the health economics of implementing and maintaining empirically supported practice in the community (see Kendall & Southam-Gerow, 1995, for further discussion). In addition to the cost of interventions, real-world payors, therapists, and families likely will desire a broader accounting of the effectiveness of interventions (see Hoagwood, Jensen, Petti, & Burns, 1996), particularly the impact of treatment on youth functioning.

References

Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology*, 68, 331–339.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.) Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Barbe, R., Bridge, J., Birmaher, B., Kolko, D., & Brent, D. (2004). Suicidality and its relationship to treatment outcome in depressed adolescents. *Suicide and Life-Threatening Behavior*, 34, 44–55.

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

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8, 77–100.

Birmaher, B., & Brent, D. A. Work Group on Quality Issues. (1998). Practice parameters for the assessment and treatment of children and adolescents with depressive disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 63S–83S.

Brent, D. A., Perper, J. A., Moritz, G., Allman, C., Roth, C., Schweers, J., Balach, L., & Baugher, M. (1993a). Psychiatric risk factors of adolescent suicide: A case control study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 521–529.

Brent, D. A., Poling, K., McKain, B., & Baugher, M. (1993b). A psychoeducational program for families of affectively ill

children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 770–774.

Brent, D. A., Roth, C. M., Holder, D. P., Kolko, D. J., Birmaher, B., Johnson, B. A., & Schweers, J. A. (1996). Psychosocial interventions for treating adolescent suicidal depression: A comparison of three psychosocial interventions. In & E. D. Hibbs & P. S. Jensen (Ed.), *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* Washington, DC: American Psychological Association.

Brent, D. A., Holder, D., Kolko, D., Birmaher, B., Baugher, M., Roth, C., & Johnson, B. (1997). A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive treatments. *Archives of General Psychiatry*, 54, 877–885.

Brent, D. A., Kolko, D., Birmaher, B., Baugher, M., Bridge, J., Roth, C., & Holder, D. (1998). Predictors of treatment efficacy in a clinical trial of three psychosocial treatments for adolescent depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 906–914.

Brent, D. A., Gaynor, S. T., & Weersing, V. R. (2002). Cognitive-behavioral approaches to the treatment of depression and anxiety. In M. Rutter, & E. Taylor (Eds.), *Child and adolescent psychiatry: Modern approaches* (pp. 921–937). London: Blackwell Scientific Publications.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* London: Sage.

Chambers, W. J., Puig-Antich, J., Hirsch, M., Paez, P., Ambrosini, P. J., Tabrizi, M. A., & Davies, M. (1985). The assessment of affective disorders in children and adolescents by semistructured interview: Test-retest reliability of the Schedule for Affective Disorders and Schizophrenia for School-Age Children, Present Episode Version. *Archives of General Psychiatry*, 42, 696–702.

Clarke, G., Hops, H., Lewinsohn, P. M., Andrews, J., Seeley, J. R., & Williams, J. (1992). Cognitive-behavioral group treatment of adolescent depression: Prediction of outcome. *Behavior Therapy*, 23, 341–354.

Clarke, G. N., Lewinsohn, P. M., Rohde, P., Hops, H., & Seeley, J. R. (1999). Cognitive-behavioral group treatment of adolescent depression: Efficacy of acute group treatment and booster sessions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 272–279.

Clarke, G. N., Hornbrook, M., Lynch, F., Polen, M., Gale, J., O'Connor, E., Seeley, J. R., & Debar, L. (2002). Group cognitive-behavioral treatment for depressed adolescent offspring of depressed parents in a health maintenance organization. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 305–313.

Franklin, M., Abramowitz, J., Kozak, M., Levitt, J., & Foa, E. (2000). Effectiveness of exposure and ritual prevention for obsessive-compulsive disorder: Randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology*, 68, 594–602.

Gould, M. S., King, R., Greenwald, S., Fisher, P., Schwab-Stone, M., Kramer, R., Flisher, A. J., Goodman, S., Canino, G., & Shaffer, D. (1998). Psychopathology associated with suicidal ideation and attempts among children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 915–923.

Hammen, C., Rudolph, K., Weisz, J., Rao, U., & Burge, D. (1999). The context of depression in clinic-referred youth: Neglected areas in treatment. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 64–71.

Hoagwood, K., Jensen, P. S., Petti, T., & Burns, B. J. (1996). Outcomes of mental health care for children and adoles-cents: I. A comprehensive conceptual model. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 1055–1063.

Hughes, C. W., Emslie, G. J., Crismon, L., Wagner, K. D., Birmaher, B., Geller, B., Pliszka, S. R., Ryan, N. D., Strober, M., Trivedi, M. H., Toprac, M. G., Sedillo, A., Llana, M. E., Lopez, M., & Rush, A. J.The Texas Consensus Conference Panel on Medication Treatment of Childhood Major Depressive Disorder. (1999). The Texas children's medication algorithm project: Report of the Texas consensus conference panel on medication treatment of childhood major depressive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 1442–1454.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 980–988.

Kazdin, A. E., Siegel, T. C., & Bass, D. (1990). Drawing on clinical practice to inform research on child and adolescent psychotherapy: Survey of practitioners. *Professional Psychology: Research and Practice*, 21, 189–198.

Kendall, P. C., & Grove, W. M. (1988). Normative comparisons in therapy outcome. *Behavioral Assessment*, 10, 147–158.

Kendall, P. C., & Southam-Gerow, M. A. (1995). Issues in the transportability of treatment: The case of anxiety disorders for youth. *Journal of Consulting and Clinical Psychology*, 63, 702–708.

Kovacs, M. (1996). Presentation and course of major depressive disorder during childhood and later years of the life span. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 705–715.

Kovacs, M., Obrosky, D. S., Gatsonis, C., & Richards, C. (1997). First-episode major depressive and dysthymic disorder in childhood: Clinical and sociodemographic factors in recovery. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 777–784.

Lewinsohn, P. M., & Clarke, G. N. (1999). Psychosocial treatments for adolescent depression. *Clinical Psychology Review*, 19, 329–342.

Lewinsohn, P. M., Clarke, G. N., Hops, H., & Andrews, J. (1990). Cognitive-behavioral treatment for depressed ado-lescents. *Behavior Therapy*, 21, 385–401.

Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology, I: Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *Journal of Abnormal Psychology*, 102, 133–144.

Lincoln, T., Rief, W., Hahlweg, K., Frank, M., von Witzleben, I., Schroeder, B., & Fiegenbaum, W. (2003). Effectiveness of an empirically supported treatment for social phobia in the field. *Behavior Research and Therapy*, 41, 1251–1269.

McFall, R. M. (1996, July 1). *Consumer satisfaction as a way of evaluating psychotherapy: Ecological validity and all that versus the good old randomized trial.* Panel discussion presented at the sixth annual Convention of the American Association of Applied and Preventative Psychology, San Francisco.

Merrill, K. A., Tolbert, V. E., & 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, 404–409.

National Institute of Mental Health. (1999, September). *In harm's way: Suicide in America.* Retrieved January 10, 2002, from http://www.nimh.nih.gov/publicat/harmaway.cfm.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2001). *HLM 5.04 for Windows [Computer software].* Chicago: Scientific Software International.

Reinecke, M. A., Ryan, N. E., & DuBois, D. L. (1998). Cognitive-behavioral therapy of depression and depressive symptoms during adolescence: A review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*, 26–34.

Roberts, R. E., Lewinsohn, P. M., & Seeley, J. R. (1991). Screening for adolescent depression: A comparison of depression scales. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*, 58–66.

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1994). Are adolescents changed by an episode of major depression? *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 1289–1298.

Rohde, P., Clarke, G. N., Lewinsohn, P. M., Seeley, J. R., & Kaufman, N. K. (2001). Impact of comorbidity on a cognitive-behavioral group treatment for adolescent depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 795–802.

Rosselló, J., & Bernal, G. (1999). The efficacy of cognitive-behavioral and interpersonal treatments for depression in Puerto Rican adolescents. *Journal of Consulting and Clinical Psychology*, *67*, 734–745.

Shaffer, D., Gould, M. S., Fisher, P., Trautman, P., Moreau, D., Kleinman, M., & Flory, M. (1996). Psychiatric diagnosis in child and adolescent suicide. *Archives of General Psychiatry*, *53*, 339–348.

Treatment for Adolescents with Depression Study Team. (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA*, *292*, 807–820.

Tuschen-Caffier, B., Pook, M., & Frank, M. (2001). Evaluation of manual-based cognitive-behavioral therapy for bulimia nervosa in a service setting. *Behavior Research and Therapy*, *39*, 299–308.

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.

Weersing, V. R. (2005). Benchmarking the effectiveness of psychotherapy: Program evaluation as a component of evidence-based practice. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*, 1058–1062.

Weersing, V. R., & Brent, D. A. (2003). Cognitive-behavioral therapy for adolescent depression: Comparative efficacy, mediation, moderation, and effectiveness. In A. E. Kazdin, & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 135–147). New York: The Guilford Press.

Weersing, V. R., & Weisz, J. R. (2002). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology*, *70*, 299–310.

Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child and Adolescent Psychology*, *31*, 168–180.

Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist*, *47*, 1578–1585.

Weisz, J. R., Thurber, C. A., Sweeney, L., Proffitt, V. D., & LeGagnoux, G. L. (1997). Brief treatment of mild-to-moderate child depression using primary and secondary control enhancement training. *Journal of Consulting and Clinical Psychology*, *65*, 703–707.

Willett, J. B. (1988). Question and answers in the measurement of change. In E. Z. Rothkopf (Ed.), *Review of research in education*, *Volume 15* (pp. 345-422). Washington, DC: American Educational Research Association.

Wood, A., Harrington, R., & Moore, A. (1996). Controlled trial of a brief cognitive-behavioural intervention in adolescent patients with depressive disorders. *Journal of Child Psychology and Psychiatry*, *37*, 737–746.