Journal of Consulting and Clinical Psychology
2002, Vol. 70, No. 2, 299–310

Copyright 2002 by the American Psychological Association, Inc.
0022-006X/02/$5.00 DOI: 10.1037//0022-006X.70.2.299

# Community Clinic Treatment of Depressed Youth: Benchmarking Usual Care Against CBT Clinical Trials

V. Robin Weersing and John R. Weisz
University of California, Los Angeles

This study used a benchmarking strategy to evaluate the effectiveness of community psychotherapy for depressed youth relative to evidence-based treatment in clinical trials. Symptom trajectories of depressed youth treated in community mental health centers (CMHCs) were compared with trajectories of youth treated with cognitive–behavioral therapy (CBT) in clinical trials. Overall, outcomes of CMHC youth more closely resembled those of control condition youth than youth treated with CBT. Within the CMHC sample, ethnic minority status and low therapy dose were related to worse outcomes. However, when outcomes for Caucasian youth and youth receiving longer term services were examined, the CMHC sample still performed more poorly than youth treated with CBT. The findings support the value of developing, testing, and exporting effective therapies for depressed youth to community clinic settings.

To date, there have been more than 1,500 controlled investigations of the effects of psychotherapy for children and adolescents (Kazdin, 2000). Results of these studies, summarized in meta-analyses, have told a consistent tale: Psychotherapy for youth can produce beneficial effects, on the same order of magnitude as both adult psychotherapy and many medical interventions (Casey & Berman, 1985; Kazdin, Bass, Ayers, & Rodgers, 1990; Weisz & Weersing, 1999; Weisz, Weiss, Alicke, & Klotz, 1987; Weisz, Weiss, Han, Granger, & Morton, 1995). However, this extensive youth treatment literature suffers from two significant limitations: (a) historic overemphasis on the treatment of disruptive behavior problems to the exclusion of mood problems and (b) a potentially tenuous relationship between the positive results of psychotherapy research trials and the actual effectiveness of therapy for youth in community settings.

Comparatively speaking, we have very little information about the treatment of depressed children and adolescents. The vast majority of clinical trials for youth focus on the treatment of disruptive behavior problems, with a mere 15 published studies targeting depressive symptomatology. The small number of clinical trials for youth depression is troubling given the lifetime impairments related to early onset of depression (see, e.g., Rohde, Lewinsohn, & Seeley, 1994) and the high cumulative prevalence of depression by the end of puberty (25.3%; Lewinsohn & Hops, 1993). Furthermore, depression is a potent risk factor for suicide (Gould et al., 1998; Shaffer, Gould, et al., 1996)—the third leading cause of death for school children, adolescents, and young adults (National Institute of Mental Health, 1999).

Historically, the field has also paid little attention to the conditions under which clinical trial research has taken place. Most randomized controlled trials (RCTs) evaluate the effects of psychotherapy under conditions designed to optimize treatment effects and demonstrate the benefits that psychotherapy can offer (Kazdin et al., 1990; Weisz, Weiss, & Donenberg, 1992). In RCTs, treatments are crafted from an explicit theory of intervention (usually cognitive–behavioral, in the case of depression), and the techniques of intervention are well specified, typically in the form of a treatment manual. Individual variation between therapists is minimized by extensive training in the manualized techniques, and treatment adherence is monitored. The conduct of therapy in the real world bears little resemblance to this model, and the services provided to depressed youth in community settings are likely quite eclectic and dependent on individual therapists' preferences and skills (Addis & Krasnow, 2000; Kazdin, Siegel, & Bass, 1990; Weersing, Weisz, & Donenberg, in press). Recent broad-based services research has suggested that the effects of this eclectic community treatment may be much more modest than the effects of therapy in RCTs (see, e.g., Bickman, 1996; Weiss, Catron, Harris, & Phung, 1999).

In addition, the depressed youth included in psychotherapy RCTs may differ in important ways from "real world" depressed youth. For instance, a number of depression clinical trials have screened participants to obtain a sample with minimal comorbid psychiatric problems (e.g., Lewinsohn, Clarke, Hops, & Andrews, 1990). Preliminary investigations of depression treatment response have indicated that comorbidity may negatively affect clinical trial treatment outcome (e.g., Brent et al., 1998); thus, treatment effects for more clinically complex, community samples of depressed

---

V. Robin Weersing and John R. Weisz, Department of Psychology, University of California, Los Angeles (UCLA).

Data were collected for this project as part of the UCLA Clinic Study, a longitudinal investigation of the effectiveness of youth psychotherapy. The current investigation was supported, in part, by National Institute of Mental Health (NIMH) Research Grant R01-MH49522 to John R. Weisz, NIMH National Research Service Award MH11779-02 to V. Robin Weersing, and NIMH Research Scientist Award K05-MH01161 to John R. Weisz. We thank Erin Dabbs and Jennifer Soumakian for assistance with data collection and management. We are grateful also to David A. Armor, Teresa Treat, and David Brent for their helpful comments on an earlier version of this article. Finally, we offer our deepest thanks to the families who agreed to participate in this investigation.

Correspondence concerning this article should be addressed to V. Robin Weersing, who is now at Western Psychiatric Institute and Clinic, University of Pittsburgh Medical School, 3811 O'Hara Street, Suite 112, Pittsburgh, Pennsylvania 15213. E-mail: weersingvr@msx.upmc.edu

youth may not be as positive as clinical trial data would suggest (Hammen, Rudolph, Weisz, Rao, & Burge, 1999).

In combination, these potential limitations leave us with substantial questions about the effectiveness of therapies for depressed youth, delivered under real-world clinical conditions to typical patient samples. To begin addressing this knowledge gap, we report the results of a naturalistic investigation into the effectiveness of outpatient psychotherapy for depressed youth seen for services at community mental health centers (CMHCs). At the time of intake into local CMHCs, we identified youth with depression symptoms comparable to the typical young participant in a depression RCT. We followed these youth for 2 years after intake, gathering information on their clinical profiles, services received, and outcomes achieved. To strengthen our naturalistic design, we then compared the outcomes of CMHC youth with a benchmark based on multiple RCTs.

Traditionally, the benchmarking strategy has been used to assess the transportability of empirically supported treatments (EST) from research to practice settings (e.g., Franklin, Abramowitz, Kozak, Levitt, & Foa, 2000; Wade, Treat, & Stuart, 1998). Treatment outcome data from the EST, as delivered in a practice setting, are compared point-by-point with the "gold standard" of the outcomes for the EST as originally delivered in one or more RCTs. When treatment effects in the practice setting are similar in magnitude to those in the RCT, there is good support for the generalizability of the RCT benchmark results to the clinical context and clientele.

Our use of the benchmarking strategy was somewhat different. As a means of assessing community care for depressed youth, we sought to identify a research standard of care for comparison—creating a best practice benchmark from a review of the entire youth depression treatment literature. After identifying a best practice treatment, we procured all available clinical trials testing its effects. We then (a) assessed the demographic and clinical comparability of the research samples and our sample of depressed CMHC youth and (b) documented treatment differences between RCTs and CMHC. After this step, we compared outcomes of CMHC youth with the average outcomes of depressed youth treated with the best practice RCT treatment.

We anticipated that outcomes for depressed CMHC youth would not be as positive as those for youth treated in the clinical trials. To identify factors that might be implicated, should the effects be weaker, we examined predictors of depression symptom trajectory in the CMHC sample. We turned to three sources to identify appropriate predictor variables. First, we reviewed the small RCT depression treatment response literature (Brent et al., 1998; Clarke et al., 1992; Jayson, Wood, Kroll, Fraser, & Harrington, 1998; Rohde et al., 2001) and identified older youth age and presence of comorbid diagnoses as negative treatment indicators. In the literature, the relationship between comorbidity and depression treatment response appears to be complex, with contradictory findings reported across studies. For example, comorbid disruptive behavior problems have been unrelated to treatment outcome in some samples of depressed youth (Clarke et al., 1992) and predictive of poor treatment response in other studies (Rohde et al., 2001). Similarly, comorbid anxiety has been found to predict poor treatment response (Clarke et al., 1992), good treatment response (Rohde et al., 2001), and superior response to CBT, relative to the outcomes achieved by alternative psychotherapies for depression (Brent et al., 1998). The effects of "double depression"—that is, an episode of major depressive disorder superimposed on preexisting dysthymic disorder—on treatment outcome also have been unclear (e.g., Clarke et al., 1992). In untreated samples of youth, however, diagnosis of double depression has predicted slow recovery (Kovacs et al., 1997). We thought it likely that depressed youth in our CMHC sample would meet criteria for a range of comorbid diagnoses (Hammen et al., 1999; Weisz, Southam-Gerow, & McCarty, 2001). In an attempt to untangle the effects of comorbid diagnoses on treatment response, we chose to investigate four "types" of comorbidity: (a) global comorbidity (i.e., total number of nondepression diagnoses); (b) comorbid anxiety (i.e., presence of one or more anxiety disorder diagnoses); (c) comorbid disruptive behavior (i.e., presence of oppositional defiant disorder or conduct disorder); and (d) double depression.

As a second source of predictor variables, we examined recent studies of broad-based mental health services for youth and identified therapy dose as a potentially important determinant of treatment outcome. Unlike clinical trials, community settings may have great variability in number of therapy sessions provided to youth. Service research to date has yielded inconsistent dose results, with some teams (Angold, Costello, Burns, Erkanli, & Farmer, 2000) asserting that youth in community services have better outcomes with longer term treatment, and other studies finding no differences in outcome between youth who received a full course of treatment and those who received less than a minimum treatment dose (i.e., less than eight sessions; Andrade, Lambert, & Bickman, 2000). We therefore chose to explore whether receipt of a minimal dose of therapy might be implicated in treatment response within our depressed CMHC sample. As our third source of predictors, we included any additional variables that differentiated our CMHC sample and treatments from the RCTs. For instance, we expected that our Los Angeles–based CMHC sample would be more ethnically diverse than the samples of most RCTs (Weisz, Huey, & Weersing, 1998). Given accumulating evidence that minority status affects help seeking, therapy engagement, and treatment outcome for youth (see, e.g., Yeh, Eastman, & Cheung, 1994), we planned to include ethnicity as a predictor variable.

## Method

### *Participants*

Participants were screened from outpatient treatment programs of six Los Angeles–area CMHCs. At the time of intake into the clinics, youth and their families were invited by CMHC staff to participate in a study on the effectiveness of psychotherapy. Families who indicated interest were contacted before their first scheduled treatment session and interviewed by research project staff to (a) obtain formal consent for participation, (b) collect baseline data on youth symptoms and diagnoses, and (c) determine study eligibility.

To obtain a sample of youth with depression symptoms comparable to those of youth in RCTs, we required our participants to meet both diagnostic and dimensional criteria for significant depression. Youth were included in the investigation who presented with (a) *Diagnostic and Statistical Manual of Mental Disorders, third edition, revised* (*DSM–III–R*; American Psychiatric Association, 1987) diagnoses of major depressive disorder and/or dysthymic disorder, according to either parent or youth report, *and* (b) youth self-reported depression in at least the borderline clinical range (Children's Depression Inventory score greater than 12; Kovacs, 1992). We expected that CMHC youth also would meet criteria for a variety of comorbid diagnoses; however, youth were not excluded from the investigation unless they were unable to complete study measures as a result of psychosis or developmental disability.

The resulting sample of depressed youth ($N = 67$) included 37 girls and 30 boys between the ages of 7 and 17 ($M = 12.9$, $SD = 2.6$). The sample was 48% Caucasian and 52% ethnic minority youth, primarily Latino and African American. The ethnic composition of our sample reflected the diversity of Los Angeles (58% ethnic minority; U.S. Census Bureau, 2000) and the State of California (50% ethnic minority; U.S. Census Bureau, 1999) and was representative of the population typically served by area CMHCs. Of the original 67 participants at intake, 19 (28%) were lost to the study after their initial assessment session. These dropouts did not differ from completers on any clinical or demographic variables except gender, with more boys being retained by the study than girls, $\chi^2(1, N = 67) = 6.04$, $p = .01$. Data from the entire sample of 67 are presented in comparisons of CMHC and RCT samples at intake; data from the 48 completer youth are used in analyses relating to treatment outcome.

## Measures

*Children's Depression Inventory (CDI; Kovacs, 1992).* The CDI is a 27-item youth-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. High scores on the measure indicate presence of significant depression (cutoff values of 12 for borderline clinical depression and 19 for clinical depression). The CDI is "the most widely used and researched measure of childhood depression" (Kendall, Cantwell, & Kazdin, 1989, p. 121) and has shown acceptable reliability and validity in numerous investigations (e.g., Craighead, Curry, & Illardi, 1995; Smucker, Craighead, Craighead, & Green, 1986). In the present study, the CDI was used to screen participants and served as the primary measure of outcome.

We selected the CDI as our *primary* outcome measure on the basis of (a) the measure's excellent reliability, validity, and clinical sensitivity; (b) the CDI's status as the most widely used measure of depression in youth psychotherapy research; and (c) psychometric data supporting the use of the CDI across the entire developmental range of our sample, from childhood (e.g., Kovacs, 1992; Smucker et al., 1986) through adolescence (e.g., Craighead et al., 1995; Doerfler, Felner, Rowlinson, Raley, & Evans, 1988). We considered adopting diagnostic status (depressed vs. not depressed) as our primary outcome; however, this approach had several limitations. As discussed in the results, only 6 of the 13 benchmark studies included diagnostic data on their samples at intake, and even fewer studies conducted diagnostic assessments at posttreatment and follow-up. In addition, the *DSM* changed versions twice over the period spanned by the benchmark studies and changed once during the course of our investigation. Accordingly, the CDI appeared to be a more useful measure of outcome in our sample and to provide the best point of comparison between our results and the results of RCTs.

*Diagnostic Interview Schedule for Children (DISC 2.3; Shaffer, Fisher, Piacentini, Schwab-Stone, & Wicks, 1991).* The DISC 2.3 is a structured diagnostic interview administered to parents (DISC–P) and youth (DISC–C) to determine the presence of *DSM–III–R* diagnoses. The interview has been used widely in epidemiological surveys and was written for consumer acceptability and ease of use with a diverse population of parents and youth (see, e.g., Lahey, Flagg, Bird, & Schwab-Stone, 1996). The DISC 2.3 has good evidence of reliability, validity, and clinical sensitivity (e.g., Schwab-Stone et al., 1996; Shaffer, Fisher, Dulcan, & Davies, 1996).

In this study, the DISC was used to identify youth with significant symptoms of depression (i.e., diagnoses of major depression and/or dysthymia) and to assess the level of comorbidity in the sample. To reduce participant burden, youth were administered only the mood disorder and conduct disorder modules of the DISC–C, whereas parents were given the DISC–P in its entirety. We chose to focus the youth interview on mood and conduct problems, given evidence that youth may be better reporters of their depression symptoms than their parents (Hammen & Rudolph, 1996; Kazdin & Marciano, 1998; Schwartz, Gladstone, & Kaslow, 1998) and because of the surreptitious nature of many conduct symptoms.

*Therapy Procedures Checklist (TPC; Weersing, Weisz, & Donenberg, in press).* The TPC is a 50-item therapist-report measure that assesses use of different child therapy techniques. The measure contains three scales (Psychodynamic, Cognitive, and Behavioral), each scored 1–100. High scores on any theoretical domain indicate extensive use of techniques from that domain. Scales have shown excellent internal consistency (all $\alpha$s $>$ .86) and good test–retest reliability (all $r$s $>$ .75) in multiple samples of youth therapists (Weersing et al., in press). In the present study, we administered the measure to CMHC therapists at posttreatment to assess whether the therapy provided to youth was representative of typical eclectic community practices.

## Assessment Procedure

Youth and their parent(s) were interviewed on four occasions: (a) shortly after intake, before the beginning of treatment; (b) 6 months after intake; (c) 1 year after intake; and (d) 2 years after intake. In addition, consent was obtained to (a) search clinic records for number of treatment sessions and (b) contact the youth's primary therapist to gather information on therapy procedures. Families were paid $50 per completed assessment interview, and therapists were paid $15 to complete the TPC. Youth were given a small age-appropriate gift at each interview (e.g., movie tickets).

## Benchmarking Procedure

*Identification of a research standard of care.* As the first step in our benchmarking procedure, we sought to identify the research standard of care for youth depression. We began by attempting to identify all extant psychotherapy clinical trials for youth depression. Studies were obtained through (a) computer searches of PsycINFO, PsychLit, and Medline;[1] (b) review of reference lists in identified studies; and (c) examination of reference lists in relevant review articles, book chapters, and meta-analyses (e.g., Brent, Gaynor, & Weersing, in press; Weisz, Valeri, McCarty, & Moore, 1999). Through this process we identified 15 clinical trials that (a) treated depressed youth, identified as symptomatic by either standardized diagnostic or dimensional measures; (b) included random assignment to one or more psychosocial treatment conditions; and (c) were published in an English-language, peer-reviewed journal.

Of these 15 RCTs, 13 (87%) tested the effects of cognitive–behavioral therapy (CBT). Two recent meta-analyses of CBT for adolescent depression have yielded very impressive effect sizes (1.27, Lewinsohn & Clarke, 1999; 1.06, Reinecke, Ryan, & Dubois, 1998), and preliminary evidence has suggested that CBT may outperform alternative research treatments for depressed youth (Brent et al., 1997). In addition, at the time of our review, the only two treatments for youth depression identified by the Task Force on Empirically Supported Procedures as "probably efficacious" were CBT programs (no treatment met "well-established" criteria; Kaslow & Thompson, 1998). CBT appeared to be the standard of care in research on treatment of youth depression and thus the most appropriate basis for a best practice benchmark.

Accordingly, we selected for benchmarking the 13 RCTs that included one or more CBT treatments for youth depression (15 CBT conditions across studies). Characteristics of the 13 RCTs are summarized in Table 1.

*Creation of the statistical benchmarks.* Each of the 13 benchmark studies included one or more CBT interventions tested against a variety of

---

[1] Search terms included key words for depression (*depression*, *dysthymia*, *major depression*), psychosocial treatments (*psychotherapy*, *therapy*, *treatment*), and youth (*child*, *adolescent*, *youth*). The search included all possible combinations of terms from these three categories.

Table 1
*Comparison of Community Mental Health Center Sample and Services to Randomized Control Trial Benchmark Studies*

| | | | Demographic characteristics | | | Clinical characteristics | | Treatment characteristics | |
|---|---|---|---|---|---|---|---|---|---|
| Study | Percentage attrition | Primary measure | Mean age | Percentage male | Percentage minority | Depression diagnoses | Other diagnoses | Benchmarking conditions | Sessions |
| Brent et al. (1997) | 10 | BDI | 15.6 | 24 | 17 | 78% MDD<br>22% DD | 32% ANX<br>21% ODD/CD | CBT program | 14 |
| Butler et al. (1980) | 2 | CDI | 11.5 | 63 | — | — | — | Cognitive<br>Attention control<br>No treatment control | 10<br>10 |
| Clarke et al. (1995) | 15 | CES–D | 15.3 | 30 | 7 | Diagnosed depression excluded | 12% ANX<br>2% ODD/CD | Cognitive<br>No treatment control | 15 |
| Clarke et al. (1999) | 22 | BDI | 16.2 | 29 | — | 76% MDD<br>13% DYS<br>11% DD | 24% ANX<br>ODD/CD excluded | CBT program<br>CBT + parents<br>Wait list control | 16<br>25 |
| Kahn et al. (1990) | 0 | CDI | 12.1 | 49 | — | — | — | CBT program<br>Wait list control | 12 |
| Lewinsohn et al. (1990) | 14 | BDI | 16.2 | 39 | — | 49% MDD<br>51% MIN | ANX and ODD/CD excluded | CBT program<br>CBT + parents<br>Wait list control | 14<br>21 |
| Liddle and Spence (1990) | 0 | CDI | 9.2 | 68 | — | — | — | CBT program<br>Attention control<br>Wait list control | 8<br>8 |
| Reynolds and Coats (1986) | 20 | BDI | 15.7 | 37 | 0 | — | — | CBT program<br>Wait list control | 10 |
| Rosselló and Bernal (1999) | 19 | CDI | 14.7 | 46 | 100 | 24% MDD<br>76% DD | — | CBT | 12 |
| Stark et al. (1987) | 21 | CDI | 11.2 | 57 | — | — | — | CBT program<br>Wait list control | 12 |
| Vostanis et al. (1996a, 1996b) | 26 | MFQ–C | 12.7 | 44 | 12 | 30% MDD<br>54% MIN<br>16% DYS | 46% ANX<br>19% ODD/CD | CBT program | 6 |
| Weisz et al. (1997) | 0 | CDI | 9.6 | 51 | 37 | — | — | CBT program<br>No treatment control | 8 |
| Wood et al. (1996) | 9 | MFQ–C | 14.2 | 31 | — | 92% MDD<br>8% MIN | 25% ANX<br>23% ODD/CD | CBT program | 6 |
| CMHC | 28 | CDI | 12.9 | 45 | 52 | 21% MDD<br>24% DYS<br>54% DD | 58% ANX<br>61% ODD/CD | Community treatment | Varied (11) |

*Note.* Dashes indicate that data were not reported. BDI = Beck Depression Inventory; CDI = Children's Depression Inventory; CES–D = Center for Epidemiologic Studies Depression Scale; MFQ–C = Mood and Feelings Questionnaire, Child Version; MDD = major depressive disorder; DYS = dysthymic disorder; DD = comorbid MDD and DYS; MIN = minor and/or intermittent depression; ANX = any anxiety disorder; ODD/CD = oppositional defiant disorder or conduct disorder; CBT = cognitive–behavioral therapy; CMHC = community mental health center.

control conditions or against an alternative treatment. For the purposes of this study, we used treatment outcome data from the CBT conditions and control groups (wait list, no treatment, or attention placebo; see Table 1) to create two composite benchmarks—one indexing the mean effect of CBT across studies (CBT benchmark), the other capturing the mean effect of passage of time and simple attention (control benchmark).

Calculation of the composite benchmarks involved two steps: (a) transformation of depression scores across studies to the same metric and (b) aggregation of transformed scores across studies to form a mean CBT benchmark and a mean control benchmark. We began by identifying the primary dimensional depression measure for each study (see Table 1) and obtaining published normative data for the measure in nonclinical, community samples of youth.[2] Next, we used the mean and standard deviation of the measure in the normal sample of youth to compute normative $z$ scores for CBT ($z_{nt}$) and control ($z_{nc}$) groups in each study at each assessment point. These computations took the general form $z_{nt} = (\bar{x}_t - \mu)/(\sigma)$ and $z_{nc} = (\bar{x}_c - \mu)/(\sigma)$, where $\bar{x}_t$ was the CBT group mean, $\bar{x}_c$ was the control group mean, $\mu$ was the normal population mean, and $\sigma$ was the normal population standard deviation for the depression measure (Kendall & Grove, 1988). By using normative data to construct our *z* scores, we were able to place all studies on the same metric for comparison, and we gave this metric clinically significant meaning.

---

[2] Study measures and normative data were as follows: Children's Depression Inventory ($\mu$ = 9.09, $\sigma$ = 7.04; Smucker et al., 1986); Beck Depression Inventory ($\mu$ = 7.17, $\sigma$ = 7.50; Roberts, Lewinsohn, & Seeley, 1991); Center for Epidemiologic Studies Depression Scale ($\mu$ = 16.98, $\sigma$ = 10.65; Roberts et al., 1991); Mood and Feelings Questionnaire, child version ($\mu$ = 27.05, $\sigma$ = 13.73; Kent, Vostanis, & Feehan, 1997). For the Mood and Feelings Questionnaire (MFQ–C), we had to rely on psychometric data collected in samples of child and adolescent outpatients. These outpatient samples are likely to have higher means on the MFQ–C than would a sample of nonclinic, community youth. Thus, $z$ scores for studies using the MFQ–C (Vostanis, Feehan, Grattan, & Bickerton, 1996a, 1996b; Wood, Harrington, & Moore, 1996) may be artificially lowered in comparison with studies using other depression symptom measures. However, removal of studies including the MFQ–C from benchmarking outcome analyses did not alter our conclusions.

As the final step in the creation of the two benchmarks, we aggregated scores within condition (CBT and control) across studies at each assessment point. For the CBT benchmark, studies were available to compute a mean benchmark at intake ($n = 15$), posttreatment ($n = 15$), 1- to 3-month follow-up ($n = 9$), 4- to 6-month follow-up ($n = 5$), 7- to 9-month follow-up ($n = 3$), and 10- to 12-month follow-up ($n = 4$) assessments. Because many of the control conditions in the RCTs were wait lists, benchmarking data were available to create control group means only at intake ($n = 11$), posttreatment ($n = 11$), and 1- to 3-month follow-up ($n = 4$) assessments. Data from single control conditions were available at 4- to 6-month ($n = 1$), 7- to 9-month ($n = 1$), and 10- to 12-month ($n = 1$) follow-up assessments.

We aggregated scores for each time point with an unweighted mean. Use of the unweighted mean permitted direct computation of standard errors and confidence intervals for CBT and control benchmark means at each assessment point. In addition, only 8 of the 80 normative *z* scores calculated for CBT and control groups were based on studies with less than 10 participants per cell (a common cut point for use of sample-weighted aggregates in meta-analysis; see Hedges & Olkin, 1985), suggesting that use of the unweighted mean would not unduly influence results.[3]

# Results

We adopted a four-step approach to data analysis and benchmarking. First, we assessed the demographic and clinical comparability of the CMHC sample and the samples of the 13 RCTs providing benchmarking data. Second, we examined differences in treatment between the CMHC and the benchmark studies. Third, we compared the mean CMHC depression symptom trajectory against the outcomes of our CBT and control benchmarks. Fourth, we examined the impact of key demographic, clinical, and treatment variables on depression symptom trajectory within the CMHC sample and compared results for subgroups of CMHC youth with the RCT benchmarks.

## *Comparison of CMHC Sample With RCT Benchmark Samples*

Demographic and clinical characteristics of the CMHC sample and 13 benchmark study samples are provided in Table 1. Overall, youth treated in the CMHC appeared to be similar in many respects to youth in the benchmark studies. The mean age of CMHC youth was 12.9 years, compared with 13.4 years for benchmark youth. Gender proportions were also similar (45% vs. 44% male). It was difficult to assess the ethnic comparability of CMHC and benchmark youth, as the majority of RCTs did not provide ethnicity data for their samples. As can be seen in Table 1, it appeared that the CMHC sample contained a substantially higher proportion of ethnic minority youth than all of the benchmark studies, save one conducted in Puerto Rico (Rosselló & Bernal, 1999).

As is discussed in the benchmarking outcome results, CMHC, CBT, and control youth had comparable levels of depression symptoms at intake (for CMHC, $\bar{z}_{nt} = 1.66$; for CBT, $\bar{z}_{nt} = 1.71$; for control, $\bar{z}_{nc} = 1.79$). Despite this similarity in depression symptoms across CMHC and benchmark samples, CMHC youth did appear to have higher rates of comorbid clinical problems. However, as with ethnicity, many benchmark studies did not report complete diagnostic data for their samples, leaving the true rates of comorbidity in the RCTs unknown. Over 70% of CMHC youth met criteria for at least one comorbid diagnosis (median comorbid diagnoses = 1.00).

## *Comparison of CMHC Therapy With RCT Benchmark Treatments*

For CMHC youth, therapy was delivered as part of the normal service provision of the clinics and was not manipulated in any fashion. Number of treatment sessions ranged from 1 to over 90 (*Mdn* = 11), with 35% of youth receiving less than a minimum dose of treatment (i.e., less than eight sessions). In session, we expected that CMHC therapists would favor eclectic and psychodynamic approaches over cognitive and behavioral approaches, in line with national surveys of community practitioners (e.g., Silver & Silver, 1983; Weersing, Weisz, & Donenberg, in press). Therapists' scores on the TPC were consistent with this hypothesis (74% TPC response rate). CMHC therapists reported using a mix of psychodynamic, cognitive, and behavioral techniques with their cases but endorsed significantly more psychodynamic techniques than either cognitive, $t(28) = 2.43$, $p = .02$, or behavioral techniques, $t(28) = 6.14$, $p < .001$.

By design, therapists in the benchmark studies used CBT. The CBT programs focused on teaching youth (a) cognitive techniques to identify and modify irrational and depressogenic thought patterns; (b) behavioral strategies to regulate mood (e.g., pleasant activity scheduling, relaxation techniques); and (c) problem-solving skills to reduce interpersonal conflict and stress. In most cases, authors indicated that detailed treatment manuals were available with session-by-session instructions for implementing the CBT program. Number of sessions in the CBT programs ranged from 6 to 25 (see Table 1), with a median of 12.

## *Comparison of CMHC Outcomes With RCT Benchmark Outcomes*

To compare CMHC results with RCT benchmarks, we transformed each CMHC participant's score on the CDI to normative *z*-score format, that is, for CMHC, $z_{nt} = (\bar{x}_t - \mu)/(\sigma)$ (see Method section). We then took the mean of these CMHC *z* scores for all assessment interviews within the range of the benchmark follow-up period: intake ($n = 67$), 6-month ($n = 37$), and 1 year ($n = 35$) follow-up assessment. Note that because the number of therapy sessions was allowed to vary naturally in the community clinics, CMHC youth did not have a posttreatment assessment separate from follow-up assessments.

Figure 1 depicts mean *z* scores plotted against time for CMHC, CBT, and control youth. CBT and control benchmark means are displayed with 95% confidence intervals. High *z* scores indicate greater depression severity. As can be seen in the figure, all three

---

[3] Estimation of the standard error of the weighted mean of the normative *z* scores raised several statistical issues. As the normative *z* scores were calculated for different measures and on different normal samples of youth, the underlying population variance for the entire set of normative *z* scores was unknown. We are unaware of any accepted statistical estimation techniques for the standard error of the weighted mean in this case.

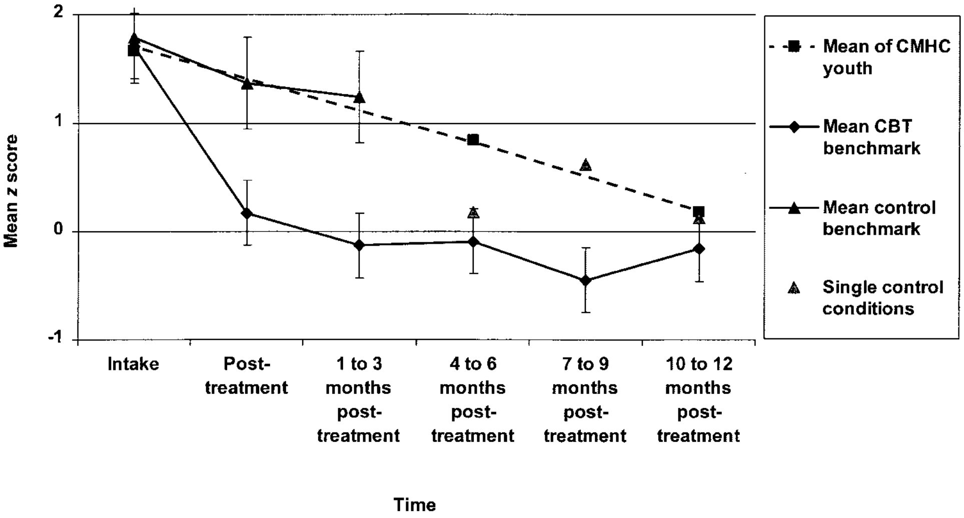

*Figure 1.* Comparison of community mental health center (CMHC) means to benchmark means and confidence intervals. CBT = cognitive–behavioral therapy.

groups began with comparable depression symptoms at intake, and all improved over the course of time. However, the depression symptom slopes, or rates of improvement, were quite different in the three groups. Youth treated with CBT improved sharply between intake and posttreatment assessment, with the CBT posttreatment mean falling within two tenths of a standard deviation unit of the mean for normative samples of community youth. These CBT gains were maintained from posttreatment ($M = 0.16$) through follow-up assessments ($-0.13$; $-0.09$; $-0.49$; $-0.16$). In contrast, control condition youth improved more slowly, with control posttreatment and 1- to 3-month follow-up means more than a full standard deviation higher than the means of community youth. Unsurprisingly, CBT benchmark means were significantly lower than control benchmark means at posttreatment, $t(24) = 5.15$, $p < .001$, and 1- to 3-month follow-up, $t(11) = 5.06$, $p < .001$. Insufficient data were available to construct control benchmark means for comparison at later follow-up assessments.

For each time point, we next evaluated whether the CMHC *z*-score mean fell within the 95% confidence interval for either the CBT benchmark mean or the control benchmark mean. This analytic strategy was analogous to a series of one sample *t* tests ($\alpha = .05$) in which the distribution of either CBT or control means was compared against the value of the CMHC mean. We adopted this method over other statistical techniques, such as independent sample *t* tests between CMHC and benchmark means, as it seemed to best capture our view of the benchmarks as estimates of population values. These confidence interval tests also avoided any complications associated with pooling standard deviations from a sample of individual participants (CMHC) with standard deviations from samples of means (CBT and control).

Overall, CMHC youth appeared to improve at a fairly slow rate. Youth seen at the CMHC were not assessed at posttreatment or at 3-month follow-up, limiting our ability to make direct comparisons between the CMHC and the control benchmark. However, when we examined the CMHC symptom slope between intake and 6-month assessment, it appeared that CMHC youth followed a trajectory similar to that of control youth in the benchmark studies. Although this relationship was quite clear from an examination of the means and confidence intervals in Figure 1, as a check on our visual inspection, we used the hierarchical linear modeling results from our treatment response analyses to compute an estimated 3-month mean for CMHC youth.[4] The estimated 3-month CMHC mean (in normative *z*-score format) was 1.23, nearly identical to the control benchmark mean of 1.24 and markedly higher than the CBT mean of $-0.13$. In addition, the CMHC mean at 6 months fell within the 95% confidence interval of the control benchmark mean at three months—a conservative comparison, as all groups were improving over time.

Although the trajectory of CMHC youth appeared quite similar to the control benchmark, it bore little immediate resemblance to the CBT benchmark. The CMHC mean at 6 months (0.84) fell outside the confidence intervals for the CBT benchmark mean at posttreatment (0.46 to $-0.12$), 1- to 3-month follow-up (0.21 to $-0.47$), and 4- to 6-month follow-up (0.28 to $-0.46$). CMHC youth did eventually achieve a reduction in depression symptoms similar to that of youth treated with CBT, but it took a year for CMHC youth to reach this mark (for CMHC, $M = 0.18$; for CBT, confidence interval = 0.18 to $-0.50$).

These data could be viewed as indicating that community therapy for depressed youth produced changes similar to those achieved by CBT, albeit much more slowly than CBT. However, it is unclear whether the eventual improvement in the CMHC sample should be attributed to the effects of CMHC therapy or rather to natural remission of depressive symptoms. As discussed previously, data were insufficient to construct a formal composite benchmark for control groups after the 3-month assessment point. This said, the trajectory of control youth in the RCTs appeared to be heading toward the CMHC mean at 1-year follow-up, and data from control conditions from individual RCTs (see Figure 1) at

---

[4] These analyses were based on data from 48 youth who completed from two to four assessment interviews over a period of 2 years.

6-month, 9-month, and 1-year follow-up paralleled or exceeded CMHC improvements.

### *Predictors of CMHC Symptom Trajectory and Outcomes for Subgroups of CMHC Youth*

As we had anticipated, depressed CMHC youth appeared to fare more poorly than youth treated with CBT in clinical trials. In our next set of analyses, we attempted to determine whether CMHC youth had uniform outcomes or whether a subset of youth within the CMHC sample achieved results more akin to CBT benchmark youth. As a first step, we sought to identify significant predictors of treatment response in the CMHC sample.

*Predictors of CMHC symptom trajectory.* As described previously, we identified older youth age, comorbidity, and receipt of very short-term services (i.e., less than eight sessions) as negative treatment indicators in the literature. To identify additional candidate variables, we examined our results comparing CMHC with RCT samples and treatments. As expected, our Los Angeles–based CMHC sample included more ethnic minority youth (52% of the sample) than the typical depression RCT. We thus included ethnic minority status on our list of treatment predictors.

Hierarchical linear modeling (HLM; version 4.04, Bryk, Raudenbush, & Congdon, 1998) was used to test associations between depression symptom trajectory and our final set of seven predictors: age, ethnicity, total number of comorbid diagnoses, comorbid anxiety disorder, comorbid disruptive behavior disorder, diagnosis of double depression, and therapy dose. Selecting HLM as our analytic strategy offered a number of benefits. Most important, HLM is robust to missing data (Bryk & Raudenbush, 1992), and the technique allowed us to estimate symptom slopes for all CMHC youth with at least two assessment interviews. For HLM analyses involving demographic and clinical predictors of symptom slope, 48 youth met this minimum interview criterion. However, we only were able to obtain CMHC medical records, and, therefore, therapy dose, for 34 youth with the requisite number of interviews.[5] HLM analyses including therapy dose as a predictor are based on this smaller sample.

We computed four sets of HLM analyses: a set of demographic analyses, testing the effects of youth age and ethnicity on symptom slope ($n = 48$); a set of clinical analyses, testing the effects of our four comorbidity predictors on symptom slope ($n = 48$); a set of treatment analyses, testing the effects of therapy dose on symptom slope ($n = 34$); and a set of combined analyses including all significant demographic, clinical, and treatment predictors ($n = 34$). Each set of HLM analyses involved computing three models. First, an unrestricted base model, with no predictors included, was computed to obtain estimates of the total variance in CDI scores, both between and within subjects. Second, the linear effects of time were used to predict within-subject variance in CDI scores, and individual symptom trajectories were estimated for each participant. Third, the appropriate set of demographic, clinical, and treatment variables was used to predict between-subject variance in these CDI symptom slopes. We used a Bonferroni-corrected alpha level to test for significant effects of our predictors (seven predictors; Bonferroni $\alpha = .007$).

In our first set of analyses, which focused on demographic characteristics, the linear effects of time accounted for a high proportion (60%) of within-subject variance in CDI scores. Youth symptom trajectories were quite uniform, with only marginal variance in symptom slopes between subjects, $\chi^2(46, N = 47) = 58.95$, $p = .09$. As planned, we tested for significant predictors of this between-subject variance in slope. Ethnic minority status emerged as significant, $t(45) = 3.81$, $p = .001$, with depressed ethnic minority youth improving more slowly than their Caucasian peers. Figure 2 displays mean CDI scores and standard errors for Caucasian and ethnic minority youth over the four assessment interviews. As can be seen in the figure, the negative impact of minority status on depression symptoms was most apparent at later follow-up assessments, after the majority of youth had completed therapy.

In our second set of analyses, focusing on clinical predictors, we obtained similar results at the first stage of modeling. Overall, time accounted for the majority of within-subject variance (60%), and there was little variability between subjects in symptom slope, $\chi^2(47, N = 48) = 59.01$, $p = .11$. As we had planned, we did attempt to predict between-subject differences in slope using total number of comorbid diagnoses, presence of comorbid anxiety disorder, presence of comorbid disruptive behavior disorder, and diagnosis of double depression; however, when entered simultaneously, none of these clinical predictors produced significant effects (all $p$s $> .23$). Given the likely collinearity between our four comorbidity variables, we also ran our modeling analyses entering each predictor independently. Again, none of the comorbidity predictors were significant (all $p$s $> .12$).

In our third set of analyses, focusing on treatment dose, time again accounted for the majority of within-subject variance (60%), and subject symptom slopes were generally uniform, $\chi^2(32, N = 33) = 42.93$, $p = .09$. Therapy dose was a significant predictor of between-subject differences in symptom slope, $t(32) = 3.05$, $p = .005$. Youth receiving less than eight sessions of therapy fared more poorly over time than those receiving eight sessions or more. Figure 3 displays mean CDI scores and standard errors for the four assessment interviews by therapy dose. As with ethnicity, the effects of therapy dose appeared to be most pronounced at one- and two-year follow-up.

When minority status and therapy dose were both included as predictors in our combined analyses, neither variable cleared our Bonferroni-corrected alpha level for statistical significance (for minority status, $p = .01$; for therapy dose, $p = .19$). This may have been due to the reduced sample size ($n = 34$) and power for tests involving therapy dose. It seemed more likely, however, that the nonsignificant results were due to substantial collinearity between ethnicity and length of services in the CMHC. Fifty-six percent of ethnic minority youth received less than eight sessions of CMHC psychotherapy, whereas only 17% of Caucasian youth attended fewer than eight sessions, $\chi^2(1, N = 34) = 5.81$, $p = .02$. The very similar patterns of results depicted in Figures 2 and 3 highlighted this collinear relationship. Examining the figures, ethnicity and therapy dose appeared to produce nearly identical sets of means,

---

[5] Missing medical records most often had been shredded by the clinic. Youth with missing records did not differ from youth with complete medical record data by demographic characteristics, clinical characteristics, type of treatment (TPC scores), or number of completed research assessment interviews.

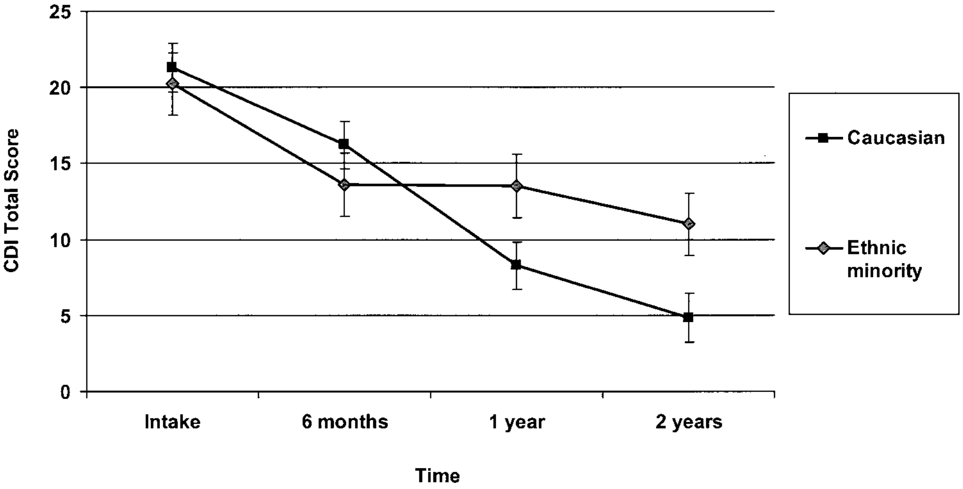

*Figure 2.* Children's Depression Inventory (CDI) total score by ethnicity.

suggesting that the two variables divided the CMHC sample into the same two subgroups of participants.

*Benchmarking analyses for subgroups of CMHC youth.* It appeared that outcomes for youth treated in the CMHC varied as a function of minority status (Caucasian vs. ethnic minority) and therapy dose (at least eight sessions vs. less than eight sessions). As the next step in our analytic plan, we compared outcomes for these four subgroups of CMHC youth to our CBT and control benchmarks. With these analyses, we sought to determine whether our previous benchmarking results held true for all portions of the CMHC sample.

Over the short term, benchmarking results were similar for the four subgroups of CMHC youth. All four groups began with depression symptoms comparable to youth in the benchmark studies. Between intake and 6-month follow-up, youth in the CMHC subgroups appeared to improve slowly, with the rate of symptom improvement more akin to control youth than CBT benchmark youth. At 12-month follow-up, however, differences between the subgroups emerged. Means for Caucasian youth (−0.12) and youth receiving at least eight sessions of therapy (−0.26) had moved within the range of the CBT benchmark confidence interval (0.18 to −0.50), whereas means for ethnic minority youth (0.63) and youth receiving a very low dose of therapy (0.63) were still well outside the CBT confidence interval.

These results were generally consistent with our main benchmarking analyses, in which the mean for entire CMHC sample lagged behind the mean for CBT youth until 1-year follow-up assessment. However, this pattern may represent the best case scenario in the CMHC. None of the subgroups of CMHC youth achieved results comparable to the CBT benchmark prior to the 1-year follow-up assessment, and the ethnic minority and low therapy dose subgroups never achieved outcomes comparable to CBT youth.

*Effects of CBT for ethnic minority youth enrolled in RCTs.* In our sample, depressed minority youth appeared (a) to receive less benefit from CMHC services than their Caucasian peers and (b) to fare poorly in comparison with youth treated with CBT in the benchmark studies. These findings could be interpreted as evidence that ethnicity is an important moderator of depression treatment response. However, minority status and therapy dose were highly collinear in our sample, limiting our ability to draw conclusions about the effects of ethnicity per se on depression treatment outcome. In addition, the effects of minority status on depression symptoms were evident only at long-term follow up, well after the majority of youth had completed treatment.

In an attempt to clarify the relationship between ethnicity and treatment response, we examined outcomes of ethnic minority youth within the sample of benchmark studies. One benchmark study, Rosselló and Bernal (1999), tested the effects of CBT with a depressed ethnic minority sample.[6] We reconstructed our CBT and control benchmarks to exclude data from Rosselló and Bernal and then compared results of their study with the new benchmark means. Overall, outcomes for the Latino sample of Rosselló and Bernal were similar to results for the new CBT benchmark. At intake, the mean $z$ score for the Rosselló and Bernal sample was comparable to the new CBT and control benchmark $z$ scores. The immediate posttreatment mean for Rosselló and Bernal (0.60) fell between the 95% confidence intervals for the CBT benchmark (−0.17 to 0.44) and control benchmark (0.92 to 1.82) means. By 3-month follow-up, however, the Rosselló and Bernal mean (−0.03) had moved well within the confidence interval of the CBT benchmark (−0.54 to 0.25) and outside of the control benchmark confidence interval (0.50 to 1.99). Although based on the results of a single study, these data suggested that depressed ethnic minority youth could obtain outcomes similar to those of Caucasian youth, when minority youth were treated with a full dose of an active psychosocial intervention. It is also worth noting that the depressed minority youth treated with CBT in Rosselló and Bernal achieved these positive outcomes despite a very high rate (76%) of double depression in the sample—a rate higher even than that in our CMHC sample (54%).

---

[6] While the CBT package in Rosselló and Bernal (1999) appeared comparable to the treatments used in the other RCT benchmark studies, the authors did report that treatment was "culturally adapted" and that therapy was delivered by ethnically similar therapists.

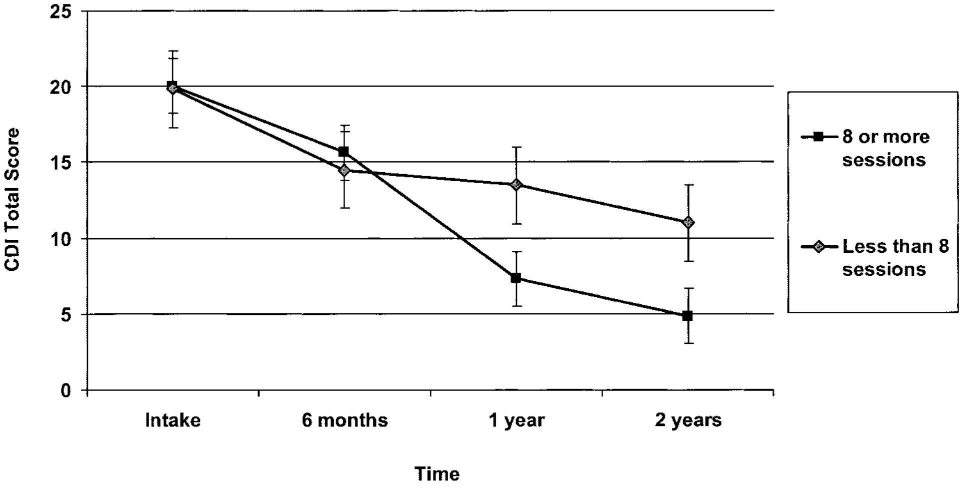

*Figure 3.* Children's Depression Inventory (CDI) total score by therapy dose.

## Discussion

In this article, we have reported the results of a naturalistic investigation of the effectiveness of community psychotherapy for depressed youth. To anchor our results, we compared the outcomes of depressed youth treated in CMHCs with the outcomes of youth treated with CBT in depression clinical trials. The results are not encouraging with regard to the effects of community treatment.

At intake, CMHC youth exhibited depression symptoms comparable to youth enrolled in clinical trials. Yet despite this similarity in starting point, there were substantial differences in depression recovery between CMHC youth and youth treated with CBT in the benchmark studies. Youth treated with CBT showed steep declines in their depression symptoms within 3 months, and these improvements were maintained over the follow-up period. Depressed youth treated in the CMHC had much shallower symptom trajectories, and it took a year for CMHC youth to achieve the level of symptom improvement obtained by youth treated with CBT. Indeed, over short-term follow-up, the mean symptom slope for community treatment more closely resembled the clinical trial control conditions than the CBT benchmark. Six months after intake into the CMHC, depressed youth were as symptomatic as control youth at their 3-month assessment. Given that all youth (CBT, control, and CMHC) appeared to be improving over time, we viewed this comparison as conservative: CMHC youth took twice as long as control youth to achieve a similar level of symptom relief.

The eventual improvement of CMHC youth to CBT benchmark levels is consistent with the improvement seen in the control groups of the benchmark studies that reported long-term follow-up results. The CMHC results also bear a strong resemblance to the pattern seen in studies of the natural course of youth depression. In youth, the median length of an episode of major depressive disorder has been estimated to be 9 months (Kovacs, 1996), with a 94% cumulative probability of recovery by 1 year after the onset of the episode (Kovacs, Obrosky, Gatsonis, & Richards, 1997). Taken together, our results and these data on natural remission are consistent with the proposition that community therapy for youth depression, in our sample, did not improve on the symptom reduction that occurs naturally with the passage of time.

Why did youth in the CMHC fare so poorly? Clinical differences between the CMHC and clinical trial samples did not appear to be a viable explanation. Youth had similarly severe depression at intake, and although CMHC youth had higher rates of comorbid disorders, none of our comorbidity predictors were significantly related to symptom slope. It is possible that demographic and contextual differences may have played a role in producing worse CMHC outcomes. Ethnic minority youth appeared to receive particularly little benefit from CMHC treatment, as did youth receiving very short-term services (less than eight sessions of therapy). However, even when examining Caucasian youth and youth receiving longer treatments, these subgroups still lagged a year behind the gains of CBT youth. In addition, when we examined the outcomes of ethnic minority youth in the sample of benchmark studies, minority youth treated with CBT obtained outcomes similar to benchmark youth as a whole.

Although this investigation did not compare CMHC therapy and CBT head-on in a randomized trial, differences in therapy type may be a reasonable explanation for the differences in CMHC and RCT outcome: According to therapist report, CMHC services were predominantly psychodynamic, whereas therapists in clinical trials provided a pure dose of CBT. We originally selected CBT as our benchmark treatment because of the consistent empirical support for the method, including preliminary evidence that CBT may produce outcomes superior to those of alternative therapies for youth depression (Brent et al., 1997). Our results seem to provide additional support for this finding. In addition, our results are consistent with meta-analytic evidence suggesting that for a wide range of child problems, behavioral treatments produce outcomes superior to those of the nonbehavioral therapies traditionally used in community settings (Weisz, Donenberg, Han, & Weiss, 1995; Weiss & Weisz, 1995).

To build on the results of the current investigation, we envision two broad directions for future research: (a) work probing the effectiveness of CBT for depressed youth and (b) additional research examining outcomes associated with community therapy. First, given the possibility that treatment type may have accounted for differences in depression outcome between CMHC and RCT, we see a great deal of value in studies testing the generalizability of positive CBT effects to clinically representative settings and samples. These investigations could take a number of different forms, from studies manipulating a few specific aspects of clinical representativeness (e.g., leaving CBT dose to therapist discretion; Weersing, Brent, & Gaynor, 2001) to studies transporting full CBT protocols into CMHC environments and comparing CBT and community therapy effects directly (Weisz, 2001). Ideally, we see results from this work as feeding back into treatment development efforts and driving CBT models and practices toward greater ecological validity. This clinic-based treatment development model has been explicated in detail elsewhere (Weisz, 2000; Weisz & Weersing, 1999).

In addition, future transportability research could be designed to address specific methodological limitations of the current investigation. By using a benchmarking strategy in this study, we were able to assess the comparability of CMHC and RCT samples, treatments, and outcomes. Our benchmarking design also allowed us to test the effects of CMHC therapy under conditions high in external validity. For example, we did not manipulate the content or length of usual community care, impose on typical CMHC supervisory practices, nor require CMHC families to agree to random assignment to treatment as a precondition of participation. However, our strength in external validity is coupled with weakness, from the standpoint of internal validity. With our benchmarking design, we cannot rule out the possibility that unmeasured differences between the CMHC and RCT samples accounted for our findings. Similarly, it is possible that treatment characteristics other than therapy type may have produced the superior outcomes in CBT (e.g., differences in the amount and structure of supervision between research and practice; Kendall & Southam-Gerow, 1995). Direct, experimental comparison of CBT and CMHC therapy in the same setting, staff, and sample of youth would provide a clear and well-controlled test of the superiority of CBT for youth depression. This type of transportability study also would assess outcomes for CBT and CMHC therapy using the same battery of measures and, thus, address another possible limitation of the current investigation—our reliance on normative $z$ scores to equate the variety of depression measures used in the RCTs. By including data from multiple RCTs, we sought to obtain (a) an unbiased estimate of CBT effects across different teams of investigators and (b) the largest and most diverse sample of participants for comparison ($N = 898$, across studies). However, collapsing across the different sample, treatment, and measurement characteristics of the 13 benchmark studies may have introduced additional sources of error into our analyses.

As a second broad area of future research, we see value in additional studies of the effects of community psychotherapy for depressed youth. Although this investigation was a useful first step in examining the care provided to depressed children and adolescents, future investigations would benefit from an increased range of outcome measures, especially functional outcomes (Hoagwood, Jensen, Petti, & Burns, 1996). It may be that community therapists are less symptom focused than their research counterparts and spend more therapeutic energy on tasks such as maintaining youth in school than on depression symptom reduction per se. These positive outcomes would not be captured by our current assessment scheme. Although youth are putatively the best informed about their own feelings of depression, it also may be useful to gather outcome data from a wider range of informants, including parents and teachers. In addition, therapists may be of assistance in identifying therapy dropouts, attrition being a useful outcome variable in its own right.

We suspect that ethnic differences in attrition may have been a key factor underlying the collinear relationship between ethnicity and therapy dose in our CMHC sample. Certainly, further research seems warranted to untangle the associations we observed among minority status, low therapy dose, and poor recovery from depression. Additional research may help to uncover psychological variables, rather than simple demographic markers, implicated in producing poorer outcomes for ethnic minority youth. For example, it may be helpful for future studies to have information about family life stress, including socioeconomic stress.

It may also be useful for future investigations to further probe the relationship between comorbidity and depression treatment response. In our sample, presence of comorbid diagnoses was not related to depression recovery. However, our assessment of comorbidity was based in large part on parent report, and it is possible that youth report would have uncovered additional diagnoses, such as posttraumatic stress disorder, that may have been important predictors of recovery. There also may be value in moving beyond a purely diagnostic definition of comorbidity and including assessment of significant subclinical psychiatric symptoms (e.g., Cerel & Fristad, 2001). Future work also could expand the assessment of comorbid conditions to include *familial* comorbidity. In epidemiological studies, families of depressed youth evidence high rates of maternal depression and paternal substance abuse (Hammen et al., 1999); these factors may well play an important role in recovery from depression for youth treated in community clinic settings.

It is our hope that through further investigations such as these we may, as a field, expand our understanding of treatment outcome moderators in real-world treatment contexts and use this understanding to develop empirically based treatments that are robust to the conditions of community practice.

## References

References marked with a single asterisk indicate studies that provided pretreatment, posttreatment, or follow-up data for the composite RCT benchmarks. References marked with a double asterisk indicate sources of psychometric data for calculation of normative $z$ scores.

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.

Andrade, A. R., Lambert, W., & Bickman, L. (2000). Dose effect in child psychotherapy: Outcomes associated with negligible treatment. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 161–168.

Angold, A., Costello, J. E., Burns, B. J., Erkanli, A., & Farmer, E. M. Z. (2000). The effectiveness of non-residential specialty mental health services for children and adolescents in the "real world." *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 154–160.

Bickman, L. (1996). A continuum of care: More is not always better. *American Psychologist, 51,* 689–701.

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

*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.

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

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1998). HLM/2L and HLM/3L 4.04 for Windows [Computer software]. Chicago: Scientific Software International.

*Butler, L., Miezitis, S., Friedman, R., & Cole, E. (1980). The effect of two school-based intervention programs on depressive symptoms in preadolescents. *American Educational Research Journal, 17,* 111–119.

Casey, R. J., & Berman, J. S. (1985). The outcome of psychotherapy with children. *Psychological Bulletin, 98,* 388–400.

Cerel, J., & Fristad, M. A. (2001). Scaling structured interview data: A comparison of two methods. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 341–346.

*Clarke, G. N., Hawkins, W., Murphy, M., Sheeber, L. B., Lewinsohn, P. M., & Seeley, J. R. (1995). Targeted prevention of unipolar depressive disorder in an at-risk sample of high school adolescents: A randomized trial of a group cognitive intervention. *Journal of the American Academy of Child and Adolescent Psychiatry, 34,* 312–321.

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., Rohde, P., Lewinsohn, P. M., Hops, H., & Seeley, J. R. (1999). Cognitive–behavioral 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.

Craighead, W. E., Curry, J. F., & Illardi, S. S. (1995). Relationship of Children's Depression Inventory factors to major depression among adolescents. *Psychological Assessment, 7,* 171–176.

Doerfler, L. A., Felner, R. D., Rowlinson, R. T., Raley, P. A., & Evans, E. (1988). Depression in children and adolescents: A comparative analysis of the utility and construct validity of two assessment measures. *Journal of Consulting and Clinical Psychology, 56,* 769–772.

Franklin, M. E., Abramowitz, J. S., Kozak, M. J., Levitt, J. T., & Foa, E. B. (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., et al. (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. D. (1996). Childhood depression. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (pp. 153–195). New York: Guilford Press.

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.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

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

Jayson, D., Wood, A., Kroll, L., Fraser, J., & Harrington, R. (1998). Which depressed patients respond to cognitive-behavioral treatment? *Journal of the American Academy of Child and Adolescent Psychiatry, 37,* 35–39.

*Kahn, J. S., Kehle, T. J., Jenson, W. R., & Clark, E. (1990). Comparison of cognitive-behavioral, relaxation, and self-modeling interventions for depression among middle school students. *School Psychology Review, 19,* 196–211.

Kaslow, N. J., & Thompson, M. P. (1998). Applying the criteria for empirically supported treatments to studies of psychosocial interventions for child and adolescent depression. *Journal of Clinical Child Psychology, 27,* 146–155.

Kazdin, A. E. (2000). Developing a research agenda for child and adolescent psychotherapy. *Archives of General Psychiatry, 57,* 829–836.

Kazdin, A. E., Bass, D., Ayers, W. A., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology, 58,* 729–740.

Kazdin, A. E., & Marciano, P. L. (1998). Childhood and adolescent depression. In E. J. Mash & R. A. Barkley (Eds.), *Treatment of childhood disorders* (2nd ed., pp. 211–248). New York: Guilford Press.

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., Cantwell, D. P., & Kazdin, A. E. (1989). Depression in children and adolescents: Assessment issues and recommendations. *Cognitive Therapy and Research, 13,* 109–146.

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.

**Kent, L., Vostanis, P., & Feehan, C. (1997). Detection of major and minor depression in children and adolescents: Evaluation of the Mood and Feelings Questionnaire. *Journal of Child Psychology and Psychiatry, 38,* 565–573.

Kovacs, M. (1992). *Children's Depression Inventory manual.* North Tonawanda, NY: Multi-Health Systems.

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.

Lahey, B. B., Flagg, E. W., Bird, H. R., & Schwab-Stone, M. E. (1996). The NIMH Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) study: Background and methodology. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 855–864.

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 adolescents. *Behavior Therapy, 21,* 385–401.

Lewinsohn, P. M., & Hops, H. (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.

*Liddle, B., & Spence, S. H. (1990). Cognitive-behavior therapy with depressed primary school children: A cautionary note. *Behavioural Psychology, 18,* 85–102.

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

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.

*Reynolds, W. M., & Coats, K. I. (1986). A comparison of cognitive-behavioral therapy and relaxation training for the treatment of depression in adolescents. *Journal of Consulting and Clinical Psychology, 54,* 653–660.

**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., 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.

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.

*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.

Schwab-Stone, M. E., Shaffer, D., Dulcan, M. K., Jensen, P. S., Fisher, P., Bird, H. B., et al. (1996). Criterion validity of the NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC–2.3). *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 878–888.

Schwartz, J. A., Gladstone, T. R. G., & Kaslow, N. J. (1998). Depressive disorders. In T. H. Ollendick & M. Hersen (Eds.), *Handbook of child psychopathology* (pp. 269–290). New York: Plenum Press.

Shaffer, D., Fisher, P., Dulcan, M. K., & Davies, M. (1996). The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC 2.3): Description, acceptability, prevalence rates, and performance in the MECA study. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 865–877.

Shaffer, D., Fisher, P., Piacentini, J., Schwab-Stone, M., & Wicks, J. (1991). *Diagnostic Interview Schedule for Children (DISC 2.3).* New York: Columbia University.

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.

Silver, L. B., & Silver, B. J. (1983). Clinical practice of child psychiatry: A survey. *Journal of the American Academy of Child Psychiatry, 23,* 552–561.

**Smucker, M. R., Craighead, W. E., Craighead, L. W., & Green, B. J. (1986). Normative and reliability data for the Children's Depression Inventory. *Journal of Abnormal Child Psychology, 14,* 25–39.

*Stark, K. D., Reynolds, W. M., & Kaslow, N. J. (1987). A comparison of the relative efficacy of self-control therapy and a behavioral problem-solving therapy for depression in children. *Journal of Abnormal Child Psychology, 15,* 91–113.

U.S. Census Bureau. (1999). *Population estimates for states by age and Hispanic origin.* Retrieved January 2002 from http://www.census.gov/population/estimates/state/srh/srh99.txt

U.S. Census Bureau. (2000). *Statistical abstract of the United States: The national data book.* Washington, DC: U.S. Government Printing Office.

*Vostanis, P., Feehan, C., Grattan, E., & Bickerton, W. (1996a). A randomised controlled out-patient trial of cognitive-behavioural treatment for children and adolescents with depression: 9-month follow-up. *Journal of Affective Disorders, 40,* 105–116.

*Vostanis, P., Feehan, C., Grattan, E., & Bickerton, W. (1996b). Treatment for children and adolescents with depression: Lessons from a controlled trial. *Clinical Child Psychology and Psychiatry, 1,* 199–212.

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., Brent, D. A., & Gaynor, S. T. (2001). [Effectiveness of CBT for adolescent depression and suicidality.] Research in progress, Western Psychiatric Institute and Clinic.

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

Weiss, B. H., Catron, T., Harris, V., & Phung, T. M. (1999). The effectiveness of traditional child psychotherapy. *Journal of Consulting and Clinical Psychology, 67,* 82–94.

Weiss, B. H., & Weisz, J. R. (1995). Relative effectiveness of behavioral and nonbehavioral child psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 317–320.

Weisz, J. R. (2000). Lab-clinic differences and what we can do about them: The clinic-based treatment development model. *Clinical Child Psychology Newsletter, 15,* 1–2.

Weisz, J. R. (2001). [Community clinic test of youth depression treatment]. Research in progress, University of California, Los Angeles.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 688–701.

Weisz, J. R., Huey, S. M., & Weersing, V. R. (1998). Psychotherapy outcome research with children and adolescents: The state of the art. In T. H. Ollendick & R. J. Prinz (Eds.), *Advances in clinical child psychology* (Vol. 20, pp. 49–92). New York: Plenum.

Weisz, J. R., Southam-Gerow, M. A., & McCarty, C. A. (2001). Control-related beliefs and depressive symptoms in clinic-referred children and adolescents: Developmental differences and model specificity. *Journal of Abnormal Psychology, 110,* 97–109.

*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.

Weisz, J. R., Valeri, S. M., McCarty, C. A., & Moore, P. S. (1999). Interventions for child and adolescent depression: Features, effects, and future directions. In C. A. Essau & F. P. Peterman (Eds.), *Depressive disorders in children and adolescents: Epidemiology, risk factors, and treatment* (pp. 383–435). Northvale, NJ: Jason Aronson.

Weisz, J. R., & Weersing, V. R. (1999). Psychotherapy with children and adolescents: Efficacy, effectiveness, and developmental concerns. In D. Cicchetti & S. L. Toth (Eds.), *Rochester Symposium on Developmental Psychopathology: Vol. 9. Developmental approaches to prevention and intervention* (pp. 341–386). Rochester, NY: University of Rochester Press.

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55,* 542–549.

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., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117,* 450–468.

*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.

Yeh, M., Eastman, K., & Cheung, M. K. (1994). Children and adolescents in community health centers: Does the ethnicity or the language of the therapist matter? *Journal of Community Psychology, 22,* 153–163.

Received January 2, 2001
Revision received May 7, 2001
Accepted May 14, 2001 ■