Youth With Anxiety Disorders in Research and Service Clinics: Examining Client Differences and Similarities

Michael A. Southam-Gerow

Department of Psychology, Virginia Commonwealth University

John R. Weisz

Department of Psychology, University of California at Los Angeles

Philip C. Kendall

Department of Psychology, Temple University

Compared 2 groups of children with anxiety disorders: those treated in a university-based research clinic (RC) and those treated in community-based service clinics (SCs). A widely endorsed goal in intervention research is to disseminate evidence-based treatments from RCs to SCs. Attaining this goal requires an understanding of the similarities and differences between clients in these 2 settings. Youth from SCs showed more comorbid externalizing diagnoses and externalizing problems and were more likely to come from low-income and single-parent families. On measures of internalizing symptomatology and diagnoses, youth from RCs were very similar to SC youth. To facilitate development of treatments with real-world applicability, we describe a model involving the testing of treatments in real-world settings. We also discuss limitations to this project.

Given the relatively poor outcome findings in community mental health clinics (Weiss, Catron, & Harris, 2000; Weisz, Weiss, & Donenberg, 1992), the growing evidence base for child mental health treatments in the research literature has created great optimism in the scientific community for the possibility of empirically based practice in community settings (Chambless et al., 1996; Compton, Burns, Egger, & Robertson, 2002; Farmer, Compton, Burns, & Robertson, 2002; Lonigan, Elbert, & Johnson, 1998). However, in many community service settings, the adoption of evidence-based treatments has been slower than hoped (Norquist, Lebowitz, & Hyman, 1999; Schoenwald &

Hoagwood, 2001; Street, Niederehe, & Lebowitz, 2000; Weisz, 2000; Wells, 1999). Scientists have long lamented the gap between science and practice and research that closes that gap is a major priority for the National Institute of Health (e.g., Hoagwood & Olin, 2002; National Advisory Mental Health Council Workgroup on Child and Adolescent Mental Health Intervention, Development, and Deployment, 2001; Olin & Hoagwood, 2002).

One reason for the slow adoption of evidence-based therapies is the perception that the current evidence base is weak relevant to the realities of a community service clinic (SC). For example, there is almost no research testing evidence-based treatments in real-world clinics under real-world conditions (e.g., Fensterheim & Raw, 1996; Persons & Silberschatz, 1998; Weisz, 2000). If differences do exist between SCs and research clinics (RCs), the external validity of the bulk of the treatment outcome studies must be questioned. To date, few studies have empirically tested the supposition that differences exist between RCs and SCs.

Researchers have discussed several categories of possible differences between RCs and SCs, including therapist differences, clinical context differences, and client differences (see Chorpita et al., 2002; Kendall & Southam-Gerow, 1995; Southam-Gerow, Weisz, Connor-Smith, & Gordis, 2002; Weisz, Donenberg, Han, & Weiss, 1995). For example, therapists in RCs often have small and focused caseloads and receive extensive supervision, whereas in SCs therapists often

carry 15 to 25 cases and receive limited supervision. Further, whereas RC therapists espouse one treatment model (e.g., cognitive–behavioral), SC therapists typically represent a range of theoretical perspectives. Differences also appear likely at the level of the service-providing organization: RCs are designed to support the research enterprise (e.g., staff are committed to the importance of research), whereas SCs are designed for service provision (i.e., must provide service, must make money adequate to cover expenses, research not viewed as a central mission).

Differences at the level of the client may represent the most frequently nominated reason for concern about the applicability of evidence-based therapies. A common assertion is that the problems seen in SC clients are more numerous and more severe than those seen in RCs (e.g., Southam-Gerow & Kendall, 2000; Weisz, Donenberg, et al., 1995). Youth in RC and SC samples may also differ on one or more contextual factors, such as ethnicity, parental psychopathology, or significant life events (e.g., Hammen, Rudolph, Weisz, Rao, & Burge, 1999). Furthermore, RC samples often restrict their sample by IQ or medication use, a constraint that SCs are not permitted to use. Related to all of these issues is a primary external validity question: Are the families who seek help at an RC the same as those who seek help at an SC? Because evidence-based treatment programs were designed for and tested with RC—and not SC—clients, any differences could necessitate changes (minor or major) to the treatment programs for use in SCs. Thus, to facilitate effectiveness research, any differences between the two samples need to be identified and addressed by changes to treatment programs.

To address this question, we tested for RC–SC differences at the client level. Because RCs restrict their focus to a small set of primary diagnoses, we chose to compare youth from a RC that focused on several *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev. [*DSM–III–R*],[1] American Psychiatric Association, 1987) anxiety disorders: generalized anxiety disorder, overanxious disorder, social phobia, avoidant disorder, or separation anxiety disorder. Our SC sample came from six community mental health centers. We tested for RC–SC differences in (a) severity of youth symptoms, (b) the degree of client comorbidity, and (c) family characteristics. We hypothesized that the RC sample would show lower levels of comorbidity and symptoms of disorders other than the target anxiety disorders. In addition, we expected that the SC sample would be more ethnically diverse and economically disadvantaged and would include more single-parent families.

Method

Participants and Recruitment Information

The RC sample came from a large university-based RC in Philadelphia, Pennsylvania. Data for this study include participants recruited into treatment outcome studies at the clinic from 1990 to 1997. The clinic is typical of many RCs in that it (a) adheres to one treatment model, the "probably efficacious" (see Ollendick & King, 1998) cognitive–behavioral Coping Cat program (Kendall, 1994); (b) specializes in the treatment of a focal set of problems, in this case childhood anxiety disorders; (c) is staffed by graduate student therapists, trained and supervised by the treatment developer, who carry relatively small caseloads (three to five cases at a time); and (d) operates within an academic setting committed to research and to the advanced training of students.

Youth in the RC sample were parent-referred, often via an outside agency, mental health, or school professional, or media advertisement. Children and their parent(s) completed informed consent and received a full assessment, including diagnostic interviews and self-report measures. Parent-reported demographic information was completed via mail 1 week in advance of the interview. After the pretreatment assessment, children who received a primary diagnosis of a *DSM–III–R* anxiety disorder (i.e., separation anxiety disorder, generalized anxiety disorder, overanxious disorder, social phobia, or avoidant disorder) as determined by the Anxiety Disorders Interview for Children (ADIS–C; Silverman, 1991) were accepted into study. All families participating in this project qualified for and were offered free treatment through a National Institute of Mental Health funded clinical trial.[2] Clinical outcomes for 77% ($n = 178$) of the youth in this study were reported in previous articles (e.g., Howard & Kendall, 1996; Kendall, 1994; Kendall et al., 1997). Although both parent- and child-reported diagnoses were collected, a parent-report diagnosis was used as the primary criterion for inclusion. The RC sample consisted of 231 youth (97 girls, mean age = 11.1 years, range 7.5 to 15.7 years).

The SC clients for this study were participants in a longitudinal research project examining treatment outcomes in community mental health centers in Los Angeles, California, from 1991 through 1997. As distinguished from the RC, the SCs (a) used a variety of therapeutic modalities, with therapists free to use the approaches they thought most appropriate (e.g., although this included behavioral and cognitive approaches, psychodynamic approaches were most fre-

---

[1]We focused on *DSM–III–R* diagnoses because collection of these data began before the advent and widespread use of the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., [*DSM–IV*], American Psychiatric Association, 1994).

[2]Although all youth who participated in this study were offered therapy, a small number dropped out of therapy.

quent); (b) saw a wide variety of presenting problems and disorders (e.g., age ranges from 5 to 17; presenting problems ranging from mood or anxiety problems to severe conduct and legal problems); (c) were staffed by therapists from a variety of training backgrounds (e.g., social work, marriage and family therapy, psychology), each carrying about 15 to 20 cases at a time; and (d) were housed in community agencies created primarily for service provision, not research.

We recruited families from six community mental health centers in Los Angeles. These six clinics were selected because they (a) have catchment areas that encompass most of the geographic, ethnic, and socioeconomic diversity of Los Angeles County and (b) account for the majority of the clinic-based youth treatment in the broad zone of the county. At an initial clinic intake interview, families were informed by clinic personnel that the University of California–Los Angeles (UCLA) was conducting a study in which they could choose to participate. Details were provided and families were asked if they would like someone from UCLA to call them. If the family consented, UCLA study staff contacted the family and arranged an assessment. Child participants completed an assent procedure and parents completed a consent procedure. Parent and child received a full assessment, including diagnostic interviews and self-report measures. Parent-reported demographic information was also collected. Fifty dollars and prizes for the children were provided for participation in the project; participation did not influence the clinical services the families were seeking. From a larger SC sample ($n = 478$), we selected only those youth meeting *DSM–III–R* criteria for one of the target anxiety disorders treated in the RC ($n = 108$; 69 boys, mean age = 10.6 years, range 7 to 17 years). The Diagnostic Interview Schedule for Children (DISC; Shaffer, Fisher, Piacentini, Schwab-Stone, & Wicks, 1991) did not permit for the determination of which diagnosis was primary; therefore, it is not known how many of the youth in our SC sample had primary (vs. secondary) anxiety disorder diagnoses. Although we collected parent- and child-report diagnoses, we used parent-report diagnoses as our inclusion criterion to be consistent with the standard used in the RC sample.[3]

Symptom Measures

Child Behavior Checklist (CBCL; Achenbach, 1991a). The CBCL is a widely used 118-item scale that assesses parents' view of an array of behavioral problems and social competencies in their children. Psychometric characteristics of the measure are strong;

for example, stability correlations ranging from .65 to .87 have been reported for periods of as long as 2 years (Achenbach, 1991a). We chose to use the eight narrowband clinical scales (i.e., Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, Aggressive Behavior), all reported in *T* scores with a mean of 50 and a standard deviation of 10. These narrowband scales show moderate correlations with one another (e.g., from .16 to .73 in clinical samples; Achenbach, 1991a). CBCL data were available for both the RC and SC samples. We chose to use the narrowband scales rather than the broadband scales to assess more specific domains of child symptoms.

Children's Depression Inventory (CDI; Kovacs, 1992). The 27-item child self-report CDI is a widely used and researched measure of childhood depressive symptoms. In clinical samples, Cronbach's α for the CDI has ranged from .71 to .89, and retest reliability coefficients have ranged from .50 to .87 (see Kovacs, 1992). CDI data were available for both the RC and SC samples. Differences on depression between the two samples would be of great interest because of the high observed correlation between depression and anxiety.

Teacher Report Form (TRF; Achenbach, 1991b). The TRF parallels the parent version of the CBCL, providing a picture of the youth's classroom functioning. Achenbach (1991b) reported good psychometric characteristics for the TRF, including good retest reliability ($r = .90$ to .92 over an interval of 15 days) and good discriminative validity. As with the CBCL, we included eight narrowband scales. TRF data were available for both the RC and SC samples.

Diagnostic Measures

ADIS–C for *DSM–III–R* (Silverman, 1991). The Anxiety Disorders Interview for Parents (ADIS–P; Silverman, 1991) and ADIS–C are structured diagnostic interviews administered to parents and children, respectively, to assess for *DSM–III–R* diagnoses in children and adolescents. The following diagnostic modules were included from the ADIS–C: major depressive disorder, dysthymic disorder, and conduct disorder. The following diagnostic modules were included from the ADIS–P: separation anxiety disorder, avoidant disorder, overanxious disorder, social phobia, generalized anxiety disorder, simple phobia, agoraphobia, panic disorder, major depressive disorder, dysthymic disorder, attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder. Supportive reliability data have been reported for the ADIS–P (overall interrater kappa .67; Silverman & Eisen, 1992; Silverman & Rabian, 1995) and ADIS–C (overall interrater kappa .76; Silverman & Eisen, 1992;

---

[3]For the SC sample, only a small portion of the DISC was administered to children and none of the anxiety disorder diagnoses was collected by child report.

Silverman & Rabian, 1995). ADIS–C and ADIS–P data were available for the RC sample only. Interrater reliability of project diagnostic interviewers was regularly assessed throughout the project. In this sample, kappas across all diagnoses ranged from .80 to 1.00, with a mean of .87.

DISC 2.3. The DISC (Shaffer et al., 1991) includes child (DISC–C) and parent (DISC–P) forms. The DISC is a highly structured interview, geared to the diagnostic categories of the *DSM–III–R* (APA, 1987). We retained the same diagnostic modules from the DISC–C and DISC–P as we did for the ADIS–C and ADIS–P, respectively. For the diagnoses included in this study, test–retest reliability has been reported in kappa (κ) ranges (for diagnostic reliability) from .55 to .77 for the child form (Schwab-Stone et al., 1993) and from .55 to .88 for the parent form. These data were in the fair to excellent range (Landis & Koch, 1977). DISC–C and DISC–P data were available for the SC sample only.

Contextual Factor Measure

Demographic information. In addition, demographic information (e.g., family income, family composition) was collected from RC and SC parent(s) at the time of the initial interview using a self-report form.

Results

We compared the RC and SC samples on three domains of variables: (a) child symptoms, as reported by the child (CDI), parent (CBCL), and teacher (TRF); (b) child diagnoses as reported by the child (ADIS–C, DISC–C) and parent (ADIS–P, DISC–P); and (c) contextual factors (i.e., family living arrangement, family income, and child ethnicity). As we were conducting multiple tests, we adjusted our alpha level to minimize Type I errors using the Bonferroni procedure (e.g., Hochberg & Tamhane, 1987) across four broadly defined families of tests: (a) child report (2 tests in the family: CDI, diagnoses); (b) parent report (11 tests in the family: CBCL scales, diagnoses); (c) teacher report (8 tests in the family: TRF scales); and (d) demographic data (3 tests in the family). This procedure resulted in our applying different significance levels for individual tests (i.e., per comparison) but we used the same per-family error rate across the tests, namely $p <$ .05. To determine the per-comparison error rate applied to each test, we divided the per-family error rate by the number of tests in the family. Our per-comparison p values were as follows: (a) for child report: .025 (i.e., .05/2 tests); (b) for parent report: .005 (i.e., .05/11 tests); (c) for teacher report: .006 (i.e., .05/8 tests); (d)

for demographic data: .016 (i.e., .05/3 tests). For the CBCL and TRF, we also conducted clinical significance tests using normative comparisons to see RC and SC groups differed on the proportion of youth in the clinical versus nonclinical range on the scales of those measures. Table 1 contains the results for the child diagnostic and child symptom statistical comparisons.

In addition to these analyses, we also conducted equivalency tests (Jaccard & Guilamo-Ramos, 2002; Rogers, Howard, & Vessey, 1993) for all nonsignificant results. We took this additional step because we were not only interested in differences between the RC and SC samples but similarities and equivalencies as well. Equivalency tests involve calculating z scores based deviations of group differences from a researcher-set equivalency point. In other words, the researcher selects a minimum difference (δ) between the groups that he or she will consider "important enough to make the groups nonequivalent" (Rogers et al., 1993, p. 554). The test involves determining if the two means differ by more or less than δ. The calculated z scores represent a test of the null hypothesis that the means *do* differ by more than δ. Thus, significant results indicate statistical equivalence.

Child Symptom Analyses

On the CBCL, SC youth's scores were significantly higher than those of RC youth for the following subscales: thought problems, attention problems, aggressive behavior, and delinquent behavior. Effect sizes (d) for these subscales ranged from .35 to 1.42, with a mean d of .50. RC youth had significantly higher scores than SC youth on the CBCL Somatic Problems scale; the effect size was .35. On the TRF, SC youth evidenced significantly higher scores on the attention problems, aggressive behavior, and delinquent behavior subscales, with effect sizes ranging from .55 to .96. The two groups did not differ reliably on the CDI, nor did the group differ reliably on the following CBCL or TRF scales: Withdrawn, Social Problems, and Anxiety/Depression.

Using Cohen's (1988) conventional effect size ranges, our results suggest that most differences between SC and RC youth were in the medium (.50 to .79) to large (.80 and up) range, with differences on the CBCL and TRF Delinquent Behavior and Aggressive Behavior scales being quite large (.95 to 1.42). However, some researchers have suggested that use of these conventional effect size ranges have "little empirical justification" (Olejnik & Algina, 2000, p. 277; see also Huberty, 2002; Thompson, 2001); effect sizes could also be evaluated in the context of those achieved in a specific research area (i.e., anxiety treatment research or child treatment research). As such, we used effect size ranges from a recent meta-analysis of the child treatment literature (Weisz, Weiss, Han, & Granger,

Table 1. *Statistics for Comparisons of Research and Service Clinics*

	Service Clinic		Research Clinic					Equivalence z
	M	SD	M	SD	t	df	Cohen's d	Score
Parent report								
Total *DSM* dx	3.62	1.96	2.48	1.06	6.82*	328	0.82	—
DSM extern. dx	1.08	1.01	0.25	0.53	9.80*	328	1.17	—
DSM intern. dx	2.53	1.70	2.23	0.88	2.16	333	0.25	0.58
DSM anxiety dx	1.90	1.26	2.15	0.84	−2.17	335	0.25	−4.02*
CBCL Withdrawn	64.32	9.02	64.54	9.62	−0.19	337	0.02	−6.06*
CBCL Somatic Complaints	63.33	9.65	66.69	9.64	−2.99*	337	0.35	—
CBCL Anxious/Depressed	68.58	8.10	69.99	8.83	−1.40	337	0.16	−8.38*
CBCL Social Problems	64.56	9.27	62.19	10.12	2.07	337	0.24	−3.35*
CBCL Thought Problems	62.62	9.01	59.57	8.45	3.03*	337	0.35	—
CBCL Attention Problems	66.20	9.94	61.74	9.20	4.06*	337	0.47	—
CBCL Delinquent Behaviors	64.27	9.20	54.06	6.03	12.19*	337	1.42	—
CBCL Aggressive Behavior	66.77	10.21	56.28	7.73	10.46*	337	1.22	—
Youth report								
Self *DSM* dx	0.38	0.70	0.03	0.18	6.95*	332	0.82	—
CDI	10.85	8.23	10.76	7.96	0.09	325	0.01	−1.02
Teacher report								
TRF Withdrawn	59.00	8.38	59.93	8.29	−0.83	281	0.11	−6.16*
TRF Somatic Complaints	57.67	9.31	57.98	10.10	−0.24	285	0.03	−4.57*
TRF Anxious/Depressed	61.09	10.83	63.51	9.58	−1.82	287	0.24	−6.59*
TRF Social Problems	61.82	8.40	58.97	8.34	2.54	287	0.34	−2.70*
TRF Thought Problems	57.63	8.77	58.48	9.22	−0.70	288	0.09	−5.48*
TRF Attention Problems	59.93	8.20	55.97	6.78	4.08*	285	0.55	—
TRF Delinquent Behavior	59.57	9.22	53.26	5.37	7.16*	288	0.96	—
TRF Aggressive Behavior	60.97	10.03	54.14	5.93	7.00*	285	0.95	—

Note: All scores reported were from the measures collected *pre-treatment. DSM = Diagnostic and Statistical Manual of Mental Disorders*; CBCL = Child Behavior Checklist; TRF = Teacher Report Form; Total *DSM* dx = parent report total diagnoses on structured diagnostic interviews; *DSM* extern. dx = parent report externalizing diagnoses on structured diagnostic interviews; *DSM* intern. dx = parent report internalizing diagnoses on structured diagnostic interviews; *DSM* anxiety dx = parent report anxiety disorder diagnoses on structured diagnostic interviews; CDI = Children's Depression Inventory; Total *DSM* dx = youth report total diagnoses on structured diagnostic interviews.
T scores reported except as indicated: [a]Raw scores reported. [b]Frequencies reported.
*p < .01.

1995), focusing on two different effect size estimates: clinic-referred children with anxiety disorders or conduct problems. The two effect sizes were as follows: (a) differences between treated versus control group youth in studies of treatment of anxiety and phobias (mean effect size: .60, 16 studies) and (b) differences for treated versus control group youth in studies of treatment of undercontrolled problems such as delinquency and aggression (mean effect size: .52, 59 studies). Considering these data, the effect sizes we achieved in this study were as large and, in most cases, larger than the typical posttreatment differences achieved in treatment outcome research for the populations of interest.

We used an alternative method for assessing the differences between the RC and SC on the CBCL and TRF by comparing the number of youth from each group who were above or below the clinical cutoff *T* score of 70 on each of the 16 scales (8 each for the CBCL and TRF). We used chi-square tests to evaluate the statistical significance of the group differences and used the Bonferroni procedure to set the *p* value for each family of tests at .006 (i.e., .05/8). We found statistically significant differences for the following

scales: CBCL Thought Problems, CBCL Delinquent Behavior, CBCL Aggressive Behavior, TRF Delinquent Behavior, and TRF Aggressive Behavior. In all instances, there were statistically significantly more youth in the clinical range in the SC compared to the RC samples.

We followed up all nonsignificant *t* tests with the equivalency testing procedure described by Rogers et al. (1993; see also Jaccard & Guilamo-Ramos, 2002). We set δ at 10% of the RC mean, following Rogers et al. procedure. Results from these are reported in Table 1. Briefly, the results indicated that statistical equivalence was found for all nonsignificant differences except for the CDI, for which there was not statistical equivalence between the two groups.

Child Diagnoses

According to parent report, SC youth met criteria for significantly more total and externalizing diagnoses than RC youth. Means were 3.62 and 2.48, respectively, for total diagnoses (*d* = .82), and 1.08 and .25 for externalizing diagnoses (*d* = 1.17). Child-report results were similar in that SC youth met criteria for more

overall *DSM* diagnoses than RC youth (*d* = .82). No statistically significant differences were found for number of parent-report internalizing or anxiety disorder diagnoses.

Using the equivalency testing procedure described by Rogers et al. (1993), we followed up the two nonsignificant difference tests, setting δ at 10% of the RC mean. Results from these are reported in Table 1. Briefly, we found statistical equivalence for the number of parent-report *DSM* anxiety diagnoses but not for number of parent-report *DSM* internalizing diagnoses.

Contextual Factors

Table 2 contains the demographic data for the two samples. SC youth were more likely to come from an ethnic minority group. Forty-nine percent of the SC youth were identified as non-Caucasian whereas the proportion of RC youth identified as non-Caucasian was 16%, $\chi^2(1, N = 334) = 39.84, p < .0009$. Because the two samples were drawn from geographically different areas, we considered the possibility that the groups might have differed due to the geographical difference. To investigate this possibility, we examined 1990 Census[4] data (U.S. Department of Commerce, 1992, 1993) to determine if the differences we found would have been expectable based on the geographic difference alone. Using Census data, we determined that the differences were not significantly different (z_{un} = .21)[5] from what would be expected given the ethnic differences in the two metropolitan areas.

Because of the sizable ethnic differences in the two populations (i.e., Philadelphia and Los Angeles) and in our samples, we re-ran the symptom and diagnostic analyses, comparing two subsamples of the SC sample to the overall RC sample. The first set of tests involved comparing all nonminority youth in the SC to RC

Table 2. *Demographic Data for Service and Research Clinic Samples*

	Research Clinic		Service Clinic	
	n	%	*n*	%
Ethnicity[a]				
White	193	84	52	51
Non-White	38	16	51	49
Family income[b]				
More than $50,000	106	50	7	91
Less than $50,000	108	50	67	9
Family living arrangement[c]				
Single parent	40	19	51	52
Two parents	172	81	48	48

Note: Data reported are frequencies.
[a]$\chi^2(1, 334) = 39.84$. [b]$\chi^2(1, 288) = 38.55$. [c]$\chi^2(1, 311) = 34.75$.

youth; the second set involved comparing all minority youth in the SC to RC youth. The pattern of results was identical to the original set of analyses with three exceptions: (a) in the nonminority SC/RC comparison, an additional significant difference emerged on the CBCL Social Problems scale, suggesting that nonminority SC youth exhibited more social problems than RC youth; (b) in the minority SC–RC comparison, the RC and SC samples did not differ on the CBCL Somatic Problems scale, the only difference "favoring" the severity of the RC sample; and (c) in the minority SC–RC comparison, the RC and SC samples did not differ on the CBCL Thought Problems scale—a difference evident in the primary analyses "favoring" the severity of the SC sample. Overall, these analyses do not suggest that ethnicity moderated the findings.

The SC youth were more likely to come from families with low income levels. Fully 91% of the families of the SC youth reported incomes below $50,000, whereas the proportion of RC youth reporting income below that level was 50%, $\chi^2(1, N = 288) = 38.55, p < .0009$. We also evaluated the veracity of our income finding by comparing the two metropolitan areas on household income using 1990 Census data (U.S. Department of Commerce, 1993). We determined that the difference we found, though in the expected direction, was more extreme than would be anticipated (z_{un} = 20.45, $p < .0001$). Specifically, the Census data indicate that in the Philadelphia, Pennsylvania–New Jersey Principal Metropolitan Statistical Area (PMSA), 61% of families report income below $50,000 compared to 63% of families in the Los Angeles–Long Beach, California, PMSA.[6]

Concerning family living arrangement, results indicated that SC youth were more likely to come from a single-parent family or conversely, RC youth were more likely to come from a two-parent family (81% to

---

[4]Census data from 1990 were used for primary analyses instead of 2000 data because the two studies from which the data were drawn occurred in the 1990s. When we reconducted these analyses using 2000 Census data, our conclusions remained the same. Details are available from the first author.

[5]To test the hypothesis that the difference between the RC and SC samples were (or were not) different from the expected difference based on Census data on the populations, we adapted the z_{un} test for proportion differences for two independent samples. Specifically, we tested whether the difference between the SC and RC sample proportions [$p(1)$ and $p(2)$] was greater than the difference between the population proportions [$pi(1)$ minus $pi(2)$], relative to a standard error term created under the assumption of the difference between population proportions [$pi(1)$ minus $pi(2)$] and using the two total sample sizes [$n(1)$ and $n(2)$]. For the non-White (vs. White) difference: $p(1)$ = .495, $p(2)$ = .165, $pi(1)$ = .592, $pi(2)$ = .247, $n(1)$ = 103, and $n(2)$ = 231. The difference between the sample proportions is .330 and the difference between the population proportions is .345. Thus, we are asking whether the .330 difference could have been sampled from the .345 difference. The calculation is as follows: z_{un} = [.330 – .345] /sqrt [.345 (1/103 + 1/231)] = .014 / .069 = .207.

[6]We found the same pattern when we repeated these analyses for the two SC subsamples (i.e., divided by ethnicity).

47%), $\chi^2(1, N = 311) = 34.75$, $p < .0009$. An examination of 1990 Census data (U.S. Department of Commerce, 1993) suggested that these data were more extreme than would be expected based on the geographical difference ($z_{un} = 60.1$, $p < .0001$). Specifically, Census data indicated that the proportion of "female householder, no husband present" families in the Philadelphia–New Jersey PMSA was 20.7% compared to a rate of 20.5% in the Los Angeles–Long Beach PMSA.[7]

Post Hoc Analyses

Because demographic differences emerged between the two samples, we were concerned that these differences may have accounted for the disparities on the symptom and diagnostic measures. In other words, the demographic disadvantages of the SC sample may have accounted for the symptom differences we were attributing to the RC–SC distinction. Thus, after completing our primary analyses, we conducted hierarchical regression analyses to control for the impact of the demographic factors and see if the RC–SC differences were still significant. For each of the variables for which we obtained significant results in our primary analyses, we conducted a hierarchical regression analysis entering the ethnicity, family income, and family living arrangement variables in Step 1 and RC–SC status in Step 2. Because we were conducting a large number of analyses, we again chose to use a p value of .005 for the CBCL scales and .006 for the TRF scales, consistent with the first set of analyses. We report our results in Table 3. The results suggest that the RC–SC differences on symptom and diagnostic measures, though attenuated, were still evident when we controlled for demographic differences.

Discussion

Although many have contended that RC samples differ in important ways from SC samples, little research to date has examined this concern. This study represents one of the first such efforts. We compared RC clients to SC clients; participants were matched on the presence of at least one of several *DSM–III–R* anxiety disorders. Results showed that differences and similarities existed between the two samples. Specifically, SC youth showed higher levels of externalizing and total problems and disorders and possessed significant demographic disadvantage compared to the RC youth. Differences on the externalizing symptom and diagnosis measures were evident even when demographic dif-

ferences were controlled. Using similarity testing, we also found that the two samples were equivalent on measures of internalizing symptomatology.

Consistent with our hypotheses, SC youth exhibited higher rates of externalizing behavior problems and disorders across parent and teacher report. Effect sizes for the measures of externalizing behavior problems were all above .90, effects considered very large by Cohen's (1988) standards and effects that are comparable to differences seen in treated versus untreated youth in clinical trials for anxiety disorders and conduct disorder (e.g., Weisz, Weiss, et al., 1995). These symptom differences remained robust even when demographic differences were controlled for.

These findings suggest that transporting treatments from research to SCs may require considerable rethinking. As one example, reconsideration of extant treatment programs for youth with anxiety disorders in light of a potentially high level of comorbid externalizing behavior problems and disorders (see also Russo & Beidel, 1994) may prove beneficial. It is possible, for example, that parent training (e.g., Kazdin, 1996; Webster-Stratton, 1994; see Southam-Gerow & Kendall, 1997, for review) and other treatment "curricula" (e.g., child-focused strategies; see Southam-Gerow, in press) that address child externalizing behavior problems may prove important adjuncts. Although our past work has suggested that level of comorbidity was not related to outcome (e.g., Kendall, Brady, & Verduin, 2001), these results suggest RC samples may not possess the high rates of externalizing comorbidity typical in SCs, a difference that could undermine anxiety treatment effectiveness in ways not seen in our previous findings.

In addition to having more severe externalizing symptoms, SC youth were at greater demographic disadvantage than RC youth. Specifically, SC youth were poorer and more likely to have only one caretaking parent. As our analysis of Census data suggested, these demographic differences were not a product of geographical differences. To the extent that family income and household composition moderate youth adjustment and treatment outcome, changes to extant treatment programs are warranted. Research has documented that demographic characteristics like these are associated with premature termination of therapy (e.g., Kazdin & Mazurick, 1994; Wierzbicki & Pekarik, 1993). In fact, in our own RC, single-parent family status predicted dropout (Kendall & Sugarman, 1997). To address this problem, interventions designed to enhance client engagement (e.g., Henggeler, Schoenwald, Rowland, & Cunningham, 2002; McKay, McCadam, & Gonzales, 1996) could be particularly important additions to consider for future adaptations of evidence-based treatments.

In contrast to these sample differences, on most measures of internalizing symptomatology, we found that the two samples were equivalent, using Rogers et

---

[7]We found the same pattern when we repeated these analyses for the two SC subsamples (i.e., divided by ethnicity).

Table 3. *Summary of Hierarchical Multiple Regression Analyses*

Variable	Total *DSM* Diagnoses			*DSM* Externalizing Diagnoses			Self *DSM* Diagnoses		
	B	SE B	β	B	SE B	β	B	SE B	β
Step 1	$R^2 = .06*$			$R^2 = .10*$			$R^2 = .08*$		
Ethnicity	−.17	.23	−.05	−.04	.13	−.02	−.02	.07	−.02
Family income	−.58	.22	−.19*	−.41	.12	−.24*	−.22	.06	−.25*
Family living arrangement	.30	.22	.10	.22	.12	.13	.06	.06	.07
Step 2	$\Delta R^2 = .08*$			$\Delta R^2 = .15*$			$\Delta R^2 = .08*$		
Research clinic/service clinic	−1.12	.22	−.35*	−.84	.12	−.46*	−.31	.06	−.33*

Variable	CBCL Somatic Complaints			CBCL Thought Problems			CBCL Attention Problems		
	B	SE B	β	B	SE B	β	B	SE B	β
Step 1	$R^2 = .01$			$R^2 = .01$			$R^2 = .03$		
Ethnicity	−.11	1.72	−.01	1.07	1.50	.05	1.80	1.63	.08
Family income	1.48	1.60	.07	−.48	1.40	−.03	−2.72	1.51	−.13
Family living arrangement	−.36	1.61	−.02	.26	1.41	.01	.52	1.53	−.02
Step 2	$\Delta R^2 = .03*$			$\Delta R^2 = .02$			$\Delta R^2 = .01$		
Research clinic/service clinic	4.66	1.73	.20*	−3.72	1.51	−.18	−2.73	1.65	−.12

Variable	CBCL Delinquent Behavior			CBCL Aggressive Behavior			TRF Attention Problems		
	B	SE B	β	B	SE B	β	B	SE B	β
Step 1	$R^2 = .23*$			$R^2 = .16*$			$R^2 = .05*$		
Ethnicity	3.86	1.28	.19*	−.50	1.46	−.02	2.20	1.26	.13
Family income	−5.86	1.19	−.32*	6.62	1.35	−.33*	−1.47	1.15	−.10
Family living arrangement	1.50	1.20	.08	2.4	1.36	.14	1.01	1.15	.07
Step 2	$\Delta R^2 = .10*$			$\Delta R^2 = .08*$			$\Delta R^2 = .02$		
Research clinic/service clinic	−7.68	1.21	−.39*	−7.12	1.41	−.33*	−3.17	1.33	−.19

Variable	TRF Delinquent Behavior			TRF Aggressive Behavior		
	B	SE B	β	B	SE B	β
Step 1	$R^2 = .14*$			$R^2 = .10*$		
Ethnicity	.46	1.16	.03	.43	1.32	.02
Family income	−3.81	1.06	−.26*	−3.31	1.20	−.21*
Family living arrangement	2.25	1.06	.15	2.44	1.21	.15
Step 2	$\Delta R^2 = .08*$			$\Delta R^2 = .11*$		
Research clinic/service clinic	−5.59	1.19	−.34*	−7.48	1.33	−.41*

Note: DSM = Diagnostic and Statistical Manual of Mental Disorders; CBCL = Child Behavior Checklist; TRF = Teacher Report Form; Total DSM diagnoses = total diagnoses on structured diagnostic interviews reported by parent; DSM externalizing diagnoses = externalizing diagnoses on structured diagnostic interviews reported by parent; Self DSM diagnoses = total diagnoses on structured diagnostic interviews reported by youth.
*$p < .01$.

al. (1993) procedures. Because RC youth were selected for internalizing problems (i.e., withdrawn behaviors, anxiety, somatic complaints), the equivalence is not surprising. Differences between research and SCs, at least so far as our data suggest, are primarily an issue of externalizing symptoms and demographics.

Because this is one of the first studies comparing RC and SC samples, an important first future direction should be replication of these findings across multiple samples and including multiple disorders. Further, as discussed in more detail shortly, the use of the same diagnostic interview in these future studies would improve confidence in the findings. If our findings prove robust in future studies, current methods for developing and deploying treatment may benefit from some rethinking. The model most widely employed, the medi-

cal–pharmaceutical model (Friedman, Furberg, & DeMets, 1998), involves a comprehensive set of clinical trials in tightly controlled settings to insure safety and to assess efficacy. However, the medical–pharmaceutical model necessarily sacrifices external validity, and this fact has lead to criticism of sole reliance on the model for mental health treatment development (e.g., Schoenwald & Hoagwood, 2001; Weisz, 2000). Our results suggest the possibility that evidence from clinical trials conducted in RCs will not adequately represent the full range of childhood mental health problems. Thus, alternatives to the MP model may be helpful to improve external validity.

One alternative is the deployment-focused model (Weisz, in press-a, in press-b), which emphasizes the transport and adaptation of treatments to practice set-

tings early in the development of a treatment program. Adherents to the model would conduct much of the development, testing, and refinement of treatment approaches with the clientele and in the contexts for which the treatment is ultimately intended (e.g., referred individuals treated in SCs). The importance of testing treatments in "real-world" settings has also been advocated by researchers and public policy advocates (e.g., Hoagwood, Burns, Kiser, Ringeisen, & Schoenwald, 2001; Schoenwald & Hoagwood, 2001; Street et al., 2000; Strosahl, Hayes, Bergan, & Romano, 1998).

As with all studies, this project has some limitations. Geographic differences between the two samples represent one concern. Although some differences were expected based on geography, the differences we found were more extreme than we expected. In addition, the differences may be related to the methods used to recruit participants. For the RC, fliers and newspaper advertisements describing typical anxiety disorder symptomatology were used to recruit possible participants. The SC sample, on the other hand, was a naturally occurring sample in that no special steps were taken to recruit children with anxiety disorders. Another concern is that we did not have measures for some possibly relevant variables like family relationships, past mental health service use, parental psychopathology,[8] or perceived barriers to treatment. Future studies will be improved by including such comparisons.

Because we used different diagnostic instruments with our two samples, readers should interpret the diagnostic differences cautiously. The two interviews are of different types: the DISC is a highly structured, respondent-based interview that prohibits clinical judgment, whereas the ADIS–C–P is a semistructured, interviewer-based instrument that permits clinical judgment. A related problem is that the DISC does not include a procedure to determine the primary (vs. secondary) diagnosis. Thus, although all youth in the SC sample met criteria for a *DSM–III–R* anxiety disorder, it was not possible to determine if the diagnosis was primary. However, these problems are somewhat mitigated by several factors. First, other measures of child psychopathology used were identical (e.g., CBCL, TRF), and the pattern of differences was robust across measures. Moreover, the two diagnostic instruments were based on the same diagnostic system (i.e., *DSM–III–R*) and used exactly the same symptom criteria to establish diagnoses. In addition, both instruments have been shown to possess strong psychometric characteristics. Finally, though it is not certain that the anxiety disorders of the SC sample were "primary," they

were as severe across measures of internalizing symptoms as those of the RC youth. Still, replication of these results with the same diagnostic instrument applied to both RC and SC samples would bolster these findings.

Controversy exists about the relative validity of child versus parent report of child anxiety (and other internalizing) symptomatology. Some studies have suggested that children may be the best reporters on their own distress (e.g., Edelbrock, Costello, Dulcan, Conover, & Kala, 1986; Jensen, Traylor, Xenakis, & Davis, 1988), whereas others found that parents might more reliably report their children's anxious distress (e.g., DiBartolo, Albano, Barlow, & Heimberg, 1998; Rapee, Barrett, Dadds, & Evans, 1994; Schniering, Hudson, & Rapee, 2000). Thus, our choice to use parent report of anxiety as the primary entry criterion may be viewed critically. We made the choice because parents invariably were the referral source. Further, we were limited by the data we chose to collect for each project. For example, we did not have composite diagnoses for the RC sample, and we did not have child-report diagnostic data for all of the anxiety disorder diagnoses for the SC sample. Future work should consider the use of both parent and child reports or of composite report.

Overall, our results indicate that where differences exist between research and SCs samples of children with anxiety disorders, they are not in the nature or severity of the target problem but in comorbid externalizing difficulties and demographic characteristics. The pattern and magnitude of differences suggests the need to adapt and test treatments in SC settings to examine their applicability and effectiveness there. Effectiveness research represents an emerging field for clinical researchers and one for which federal funding opportunities are increasing (e.g., Hoagwood & Olin, 2002; Street et al., 2000). Our findings illustrate the need for continued work on this important topic.

References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklists/4–18 and 1991 profile.* Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Teacher Report Form and 1991 Profile.* Burlington: University of Vermont, Department of Psychiatry.

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

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

Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *The Clinical Psychologist, 49,* 5–18.

---

[8]Although both studies included measures of parental symptomatology, the measures were not the same and thus not comparable.

Chorpita, B. F., Yim, L. M., Donkervoet, J. C., Arensdorf, A., Amundsen, M. J., McGee, C., et al. (2002). Toward large-scale implementation of empirically supported treatments for children: A review and observations by the Hawaii Empirical Basis to Services Task Force. *Clinical Psychology—Science & Practice, 9,* 165–190.

Compton, S. N., Burns, B. J., Egger, H. L., & Robertson, E. (2002). Review of the evidence base for treatment of childhood psychopathology: Internalizing disorders. *Journal of Consulting and Clinical Psychology, 70,* 1240–1266.

DiBartolo, P. M., Albano, A. M., Barlow, D. H., & Heimberg, R. G. (1998). Cross-informant agreement in the assessment of social phobia in youth. *Journal of Abnormal Child Psychology, 26,* 213–220.

Edelbrock, C., Costello, A. J., Dulcan, M. K., Conover, N. C., & Kala, R. (1986). Parent–child agreement on child psychiatric symptoms assessed via structured interview. *Journal of Child Psychology and Psychiatry, 27,* 181–190.

Farmer, E., Compton, S. N., Burns, J. B., & Robertson, E. (2002). Review of the evidence base for treatment of childhood psychopathology: Externalizing disorders. *Journal of Consulting and Clinical Psychology, 70,* 1267–1302.

Fensterheim, H., & Raw, S. D. (1996). Psychotherapy research is not psychotherapy practice. *Clinical Psychology: Science and Practice, 3,* 168–171.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998). *Fundamentals of clinical trials* (3rd ed.). New York: Springer-Verlag.

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.

Henggeler, S. W., Schoenwald, S. K., Rowland, M. D., & Cunningham, P. B. (2002). *Serious emotional disturbance in children and adolescents: Multisystemic therapy.* New York: Guilford.

Hoagwood, K., Burns, B. J., Kiser, L., Ringeisen, H., & Schoenwald, S. J. (2001). Evidence-based practice in child and adolescent mental health services. *Psychiatric Services, 52,* 1179–1189.

Hoagwood, K., & Olin, S. S. (2002). The NIMH blueprint for change report: Research priorities in child and adolescent mental health. *Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 760–767.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures.* New York: Wiley.

Howard, B. L., & Kendall, P. C. (1996). Cognitive–behavioral family therapy for anxiety-disordered children: A multiple-baseline evaluation. *Cognitive Therapy and Research, 20,* 423–443.

Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement, 62,* 227–240.

Jaccard, J., & Guilamo-Ramos, V. (2002). Analysis of variance frameworks in clinical child and adolescent psychology: Advanced issues and recommendations. *Journal of Clinical Child and Adolescent Psychology, 31,* 278–294.

Jensen, P. S., Traylor, J., Xenakis, S. N., & Davis, H. (1988). Child psychopathology rating scales and interrater agreement: I. Parents' gender and psychiatric symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry, 27,* 442–450.

Kazdin, A. E. (1996). Problem-solving and parent management in treating aggressive and antisocial behavior. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescents disorders: Empirically based strategies for clinical practice* (pp. 377–408). Washington, DC: American Psychological Association.

Kazdin, A. E., & Mazurick, J. L. (1994). Dropping out of child psychotherapy: Distinguishing early and late dropouts over the course of treatment. *Journal of Consulting and Clinical Psychology, 62,* 1069–1074.

Kendall, P. C. (1994). Treating anxiety disorders in youth: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology, 62,* 100–110.

Kendall, P. C., Brady , E. U., & Verduin, T. L. (2001). Comorbidity in childhood anxiety disorders and treatment outcome. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 787–794.

Kendall, P. C., Flannery-Schroeder, E. C., Panichelli-Mindel, S. P., Southam-Gerow, M. A., Henin, A., & Warman, M. J. (1997). Treating anxiety disorders in youth: A second randomized clinical trial. *Journal of Consulting and Clinical Psychology, 65,* 366–380.

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

Kendall, P. C., & Sugarman, A. (1997). Attrition in the treatment of childhood anxiety disorders. *Journal of Consulting and Clinical Psychology, 65,* 883–888.

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

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

Lonigan, C. J., Elbert, J. C., & Johnson, S. B. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology, 27,* 138–145.

McKay, M. M., McCadam, K., & Gonzales, J. J. (1996). Addressing the barriers to mental health services for inner city children and their caretakers. *Community Mental Health Journal, 32,* 353–361.

National Advisory Mental Health Council Workgroup on Child and Adolescent Mental Health Intervention Development and Deployment. (2001). *Blueprint for change: Research on child and adolescent mental health.* Washington, DC: Author.

Norquist, G., Lebowitz, B., & Hyman, S. (1999). Expanding the frontier of treatment research. *Prevention and Treatment, 2,* np.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25,* 241–286.

Olin, S. C., & Hoagwood, K. (2002). The Surgeon General's national action agenda on children's mental health. *Current Psychiatry Reports, 4,* 101–107.

Ollendick, T. H., & King, N. J. (1998). Empirically supported treatments for children with phobic and anxiety disorders: Current status. *Journal of Clinical Child Psychology, 27,* 156–167.

Persons, J. B., & Silberschatz, G. (1998). Are results of randomized controlled trials useful to psychotherapists? *Journal of Consulting and Clinical Psychology, 66,* 126–135.

Rapee, R. M., Barrett, P. M., Dadds, M. R., & Evans, L. (1994). Reliability of the *DSM–III–R* childhood anxiety disorders using structured interview: Interrater and parent–child agreement. *Journal of the American Academy of Child and Adolescent Psychiatry, 33,* 984–992.

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

Russo, M. F., & Beidel, D. C. (1994). Comorbidity of childhood anxiety and externalizing disorders: Prevalence, associated characteristics, and validation issues. *Clinical Psychology Review, 14,* 199–211.

Schniering, C. A., Hudson, J. L., & Rapee, R. M. (2000). Issues in the diagnosis and assessment of anxiety disorders in children and adolescents. *Clinical Psychology Review, 20,* 453–478.

Schoenwald, S. K., & Hoagwood, K. (2001). Effectiveness, transportability, and dissemination of interventions: What matters when. *Psychiatric Services, 52,* 1190–1197.

Schwab-Stone, M., Fisher, P., Piacentini, J., Shaffer, D., Davies, M., & Briggs, M. (1993). The Diagnostic Interview Schedule for Children–revised version (DSC–R): II. Test–retest reliability.

Journal of the American Academy of Child and Adolescent Psychiatry, 32, 651–657.

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

Silverman, W. K. (1991). *Anxiety Disorders Interview Schedule for Children.* Albany, NY: Graywind.

Silverman, W. K., & Eisen, A. R. (1992). Age difference in the reliability of parent and child reports of child anxious symptomatology using a structured interview. *Journal of the American Academy of Child and Adolescent Psychiatry, 31,* 117–124.

Silverman, W. K., & Rabian, B. (1995). Test–retest reliability of the *DSM–III–R* childhood anxiety disorders symptoms using the Anxiety Disorders Interview Schedule for Children. *Journal of Anxiety Disorders, 9,* 139–150.

Southam-Gerow, M. A. (in press). Child-focused cognitive–behavioral therapies for antisocial children. In C. A. Essau (Ed.), *Conduct and oppositional defiant disorders in children and adolescents: Epidemiology, risk factors, and treatment.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Southam-Gerow, M. A., & Kendall, P. C. (2000). Cognitive–behavior therapy with youth: Advances, challenges, and future directions. *Clinical Psychology and Psychotherapy, 7,* 343–366.

Southam-Gerow, M. A., & Kendall, P. C. (1997). Parent-focused and cognitive–behavioral treatments of antisocial youth. In D. Stoff, J. Breiling, & J. D. Maser (Eds.), *Handbook of antisocial behavior* (pp. 384–394). New York: Wiley.

Southam-Gerow, M. A., & Kendall, P. C. (2000). Cognitive–behavior therapy with youth: Advances, challenges, and future directions. *Clinical Psychology and Psychotherapy, 7,* 343–366.

Southam-Gerow, M. A., Weisz, J. R., Connor-Smith, J. K., & Gordis, E. B. (2002, August). The UCLA Project: The nuts and bolts of an effectiveness trial. In S. Hinshaw (Chair), *Empirically supported treatments: Transporting effective procedures into practice.* Symposium conducted at the American Psychological Association Convention, Chicago, IL.

Street, L. L., Niederehe, G., & Lebowitz, B. D. (2000). Toward greater public health relevance for psychotherapeutic intervention research: An NIMH workshop report. *Clinical Psychology: Science and Practice, 7,* 127–137.

Strosahl, K. D., Hayes, S. C., Bergan, J., & Romano, P. (1998). Assessing the field effectiveness of Acceptance and Commitment Therapy: An example of the manipulated training research method. *Behavior Therapy, 29,* 35–64.

Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *Journal of Experimental Education, 70,* 80–93.

U.S. Department of Commerce. (1992). *1990 Census of population: General population characteristics: Metropolitan areas.* Washington, DC: Author.

U.S. Department of Commerce. (1993). *1990 Census of population: Social and economic characteristics: Metropolitan areas.* Washington, DC: Author.

Webster-Stratton, C. (1994). Advancing videotape parent training: A comparison study. *Journal of Consulting and Clinical Psychology, 62,* 583–593.

Weiss, B., Catron, T., & Harris, V. (2000). A 2-year follow-up of the effectiveness of traditional child psychotherapy. *Journal of Consulting and Clinical Psychology, 68,* 1094–1101.

Weisz, J. R. (2000). Lab–clinic differences and what we can do about them: I. The Clinic-Based Treatment Development Model. *Clinical Child Psychology Newsletter, 15,* 1–3, 10.

Weisz, J. R. (in press-a). Milestones, methods, and a new deployment-focused model of treatment development and testing. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically-based approaches* (2nd ed.). Washington, DC: American Psychological Association.

Weisz, J. R. (in press-b). *Psychotherapies for children and adolescents: Evidence-based treatments and case samples.* Cambridge, UK: Cambridge University Press.

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., 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. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117,* 450–468.

Wells, K. B. (1999). Treatment research at the crossroads: The scientific interface of clinical trials and effectiveness research. *American Journal of Psychiatry, 156,* 5–10.

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice, 24,* 190–195.