Chapter Eleven

Instrumentation and Statistical Technologies

OLS vs. HLM: The Significance of Method for Understanding Ethnic Differences in A&D Service Referrals

Carolyn S. Breda

Introduction

This research addresses two questions—one theoretical, one methodological—about juvenile offenders' access to alcohol and drug (A&D) services through the courts. The theoretical question asks whether there is ethnic disparity in courts' decisions to refer offenders to A&D services. Data suggest that a large number of Black and White juvenile offenders have alcohol or substance abuse problems (Dembo, Williams, & Schmeidler, 1994; Lyons, Baerger, Quigley, Erlich, & Griffin, 2001; McManus, Alessi, Grapentine, & Brickman, 1984; Rogers, Powell, & Strock, 1998; Shelton & Phillips, 1999) for which courts could mandate treatment. However, little is known about the impact of offender ethnicity on A&D referral. Research on other types of court decisions (e.g., detention, severity of disposition) has found harsher outcomes for Blacks (Dannefer & Schutt, 1982; Frazier & Bishop, 1985; Marshall & Thomas, 1983; McCarthy & Smith, 1986; Thomas & Cage, 1977; Thomson & Zingraff, 1981; US Dept. of Justice, 1999); harsher outcomes for Whites (Scarpitti & Stephenson, 1971); no ethnic differences (Bell & Lang, 1985; Cohen & Kluegel, 1978; Horwitz & Wasserman, 1980; Minor, Hartmann, & Terry, 1991; Niarhos & Routh, 1992); or differences that did not consistently favor one group over another (Breda, in press). Such mixed and often contradictory findings (e.g., Fagan, Slaughter & Harstone, 1987; Marshall & Thomas, 1983) suggest no clear trend.

The methodological question concerns how best to test for ethnic disparity in courts' A&D referrals. Traditionally, multivariate logistic regression might be used, whereby treatment referral is regressed on a set of theoretically meaningful predictors, including offender ethnicity. However, statisticians have concluded that such approaches, based on Ordinary Least Squares (OLS) models, yield biased estimates. This occurs because, as is often the case with court data, cases (offenders) within a sample are clustered or *nested* within units (courts), therefore violating the OLS assumption of independence among subjects. Relatively new techniques (e.g., Hierarchical Linear Modeling (HLM) are specifically designed for data with nested structures and provide unbiased estimates. This research presents HLM and OLS results to highlight the significance of method for interpreting findings, and relies on HLM findings to inform the theoretical question of ethnic disparity in A&D service delivery.

Method

Sample

The sample includes over 36,000 Black and White youth between the ages of 5 and 18 who were referred to any of 98 courts throughout Tennessee in 1997 for a criminal or status offense (Tennessee Council of Juvenile and Family Court Judges, 1997). The courts preside over urban, suburban, and rural areas, with metropolitan courts likely to be representative of other metropolitan courts in the U.S. (Stapleton, Aday, & Ito, 1982). Sixty-five percent of the offenders in the sample are White, 35% are Black. Offenders are included in the sample regardless of how far they penetrated the system; this was done to avoid bias that can occur when court decisions that restrict the range of the sample or alter its composition are examined (Frazier & Bishop, 1985).

Measures

A&D offense is coded "yes" if any offense included sale or possession of a controlled substance, DUI, public drunkenness, possession of alcohol, or other drug offenses. *Prior record* indicates whether

youth had a previous encounter with the court during the target year. *Police* indicates whether the police took the youth to court or someone else (e.g., school official, family member). *Detention* indicates whether the youth was detained prior to adjudication. *Petition* indicates whether a formal petition was filed against the youth. *Adjudication* indicates whether youth were referred for judicial review and adjudicated delinquent, referred for review but not adjudicated, or not referred for review.

Social factors include youth's *ethnicity, gender,* and *age*. Age categories approximate developmental groupings: less than ten, 10-12, 13-15, and 16-18 (referent group) years old. *Living arrangement* identifies youth who lived with both biological parents (referent group), in a blended household (e.g., with a biological and stepparent), with a single parent, with relatives, or in some other living arrangement (e.g., group home) at the time they were taken to court.

The dependent variable, *treatment referral,* identifies offenders who were or were not referred for community-based A&D services.

Analysis

Cohen and Kluegel (1978) offer three criteria for ethnic disparity in court outcomes, adapted here for treatment referral: (a) ethnicity directly affects the referral decision, (b) ethnicity indirectly affects referrals through a variable stereotypical in nature, or (c) ethnicity moderates the effect of another variable on the treatment decision. To test for disparity, multivariate logistic regression (OLS) and Hierarchical Linear Modeling (HLM) are used, with two equations modeled for each technique. In the first equation, A&D referrals are modeled by youths' social and legal profiles. This allows a direct test of whether offenders' ethnicity affects the chances of referral while simultaneously controlling for other relevant variables. It also identifies whether A&D referral is associated with a stereotypical variable that may disadvantage Black offenders. For example, if courts perceive families headed by a single parent to be less "suitable" for services than others, Black youth could be disadvantaged because they disproportionately live in single-headed households. Cohen and Kluegel's third criterion is tested in the second equations, where interaction terms are included to assess whether ethnicity modifies the effects of other variables on service referral. Significance tests are adjusted for the number of variables in each model to account for the possibility that, with multiple tests, some variables may be significant by chance.

Results

Table 1 suggests that White offenders are twice as likely as Blacks to be referred to A&D services through the courts. However, the data also reveal ethnic differences on other variables that may bear on the referral decision. For example, Black offenders tend to be younger than Whites, and less likely to live with both biological parents. Whites (14%) are more likely than Blacks (9%) to be charged with an A&D-related offense, particularly for alcohol-related offenses. However, Blacks are more likely than Whites to have a prior record (33%), to be taken to court by police (77%), and to be detained (52%).

The first two columns of Table 2 show the direct effects based on the OLS and HLM analyses. The Bonferroni-adjusted alpha level is $p = .003$. The results depend on the technique considered. OLS results suggest that Blacks are significantly less likely than Whites to be referred for A&D services; however, the effect is nonsignificant at even the most liberal levels when HLM is used. Rather, based on HLM, the odds of service referral increase by type of offense (if A&D-related), age (older youth), and prior record. Too, youth living in a single-headed household, a potentially stereotyping variable that could disfavor Blacks, are as likely as others to be referred. (While referral tends to be less likely for youth in other types of arrangements such as group homes; the number of youth in these arrangements is small.)

Table 1
Sample Characteristics (*N* = 36,157)

	Black (35%)	White (65%)
Court Referral to A&D Services	2%	4%
Social Profile		
Gender		
Male	69%	66%
Age		
< 10 yrs.	2%	1%
10-12 yrs.	6%	3%
13-15 yrs.	33%	25%
16-18 yrs.	59%	71%
Living Arrangement		
Both biological parents	12%	36%
Bio + Stepparent	7%	11%
Single-parent	66%	44%
Other relatives	12%	5%
Other	3%	4%
Legal Profile		
Prior record	33%	22%
Taken to court by police	77%	63%
Detained	52%	11%
Petitioned	37%	60%
Adjudicated Status		
Not referred for judicial hearing	4%	11%
Referred, not adjudicated	69%	37%
Referred, adjudicated	27%	52%
Current A&D offense	9%	14%
Alcohol	2%	8%
Drug	7%	7%

Note: All differences are significant $p < .001$.

Results thus far do not support the first two criteria for ethnic disparity. Columns 3-4 show results when moderating effects (interaction terms) related to ethnicity are considered. The Bonferroni-adjusted alpha level is $p = .002$. Again, results depend on technique. Most notably, OLS suggests that detention reduces the odds of A&D referral and that this negative effect is even more negative for Blacks than for Whites. HLM does not support this finding. However, HLM does find that ethnicity moderates the effect of offense type: while having an A&D-related offense increases the chances for treatment referral for Blacks and Whites, it does not increase the chances as much for Blacks as it does for Whites. Thus, the third criterion for ethnic disparity in the court decision to refer offenders to A&D services is met.

Discussion

Very few juvenile justice jurisdictions provide appropriate substance abuse services for youth (McBride, VanderWaal, Terry, & Van Buren, 1999; Schonberg, 1993; Thornberry, Tolnay, Flanagan, & Glynn, 1991), a finding underscored by the low 2-4% overall treatment referral rate found in this research. The underutilization of treatment options may reflect the emphasis in juvenile justice over the past 20 years on accountability and punishment rather than on rehabilitation. At the same time, the study shows that offenders charged with an A&D offense are significantly more likely than others to be referred for care. So, while courts may not be responding in treatment referrals proportionate to

Table 2
OLS and HLM Models of A&D Service Referral
By Social and Legal Characteristics of Offenders
(*N* = 36,157)

	Direct Effects		Direct + Moderating Effects	
	OLS	**HLM**	**OLS**	**HLM**
Social Variables				
Black	-.39 (<.001)	-.21 (.105)	-.81 (.046)	.40 (.054)
Male	.22 (.005)	.17 (.037)	.18 (.029)	.16 (.062)
Age				
LT 10 yrs.	-.95 (.184)	-.66 (.087)	-1.33 (.190)	-1.12 (.052)
10-12 yrs.	-1.38 (.003)	-1.14 (<.001)	-1.27 (.013)	-1.15 (<.001)
13-15 yrs.	.24 (.002)	.14 (.113)	.30 (<.001)	.17 (.120)
Living Arrangement				
Bio/step	.23 (.039)	.14 (.100)	.26 (.028)	.17 (.109)
1-parent	.01 (.935)	-.10 (.253)	.03 (.745)	-.09 (.325)
Relatives	-.22 (.169)	-.26 (.020)	-.20 (.289)	-.23 (.153)
Other living	-.76 (.002)	-.99 (<.001)	-.84 (.004)	-1.00 (<.001)
Legal Variables				
Any A&D Offense	3.22 (<.001)	2.98 (<.001)	3.36 (<.001)	3.18 (<.001)
Prior Record	.94 (<.001)	.76 (<.001)	.91 (<.001)	.74 (<.001)
Police	-.14 (.136)	-.16 (.130)	-.16 (.117)	-.19 (.087)
Detention	-1.44 (<.001)	-.40 (.011)	-1.02 (<.001)	-.39 (.066)
Petition	.82 (<.001)	.17 (.179)	.69 (<.001)	.24 (.089)
Adjudication Status				
No adjudication	.69 (<.001)	-.33 (.329)	.57 (.001)	-.30 (.478)
Adjudicated delinquent	.42 (<.001)	.50 (.058)	.34 (<.001)	.47 (.062)
Interaction Terms				
Black*Male			.36 (.175)	.19 (.116)
Black*LT10			.99 (.490)	1.14 (.045)
Black*Ten12			-.54 (.635)	-.12 (.797)
Black*13-15			-.40 (.054)	-.27 (.043)
Black*Bio/step			-.29 (.472)	-.33 (.123)
Black*1-parent			-.16 (.482)	-.13 (.255)
Black*Relatives			-.15 (.692)	-.18 (.370)
Black*Other living			.07 (.906)	-.09 (.787)
Black*A&D offense			-.85 (<.001)	-1.00 (<.001)
Black*Prior record			.22 (.233)	.09 (.360)
Black*Police			.13 (.580)	.16 (.285)
Black*Detention			-1.1 (<.001)	.07 (.732)
Black*Petition			.58 (.046)	-.48 (.005)
Black*No adjudication			.50 (.259)	-.07 (.910)
Black*Adjudicated delinquent			.45 (.046)	.27 (.160)
Constant	-5.58	-4.98	-5.51	-5.13
Degrees of Freedom	*df*(16)	*df*(16)	*df*(31)	*df*(31)

Note: Significance levels are parenthesized

the need for care, they seem to be targeting appropriate youth for treatment—those with A&D offenses. This finding may offer hope to those who call for a public health approach to drug-related offenses (e.g., Alexander, 1996; Longmire, 1993). Current efforts nationwide to develop juvenile drug court programs for A&D offenders may also represent a shift toward more therapeutic approaches to drug-related crime.

Findings also suggest that the effect of offender ethnicity on the court's decision to refer to A&D counseling is not simple or direct. Rather, it is inextricably linked to other variables—notably, offense type. Black A&D offenders are not given as much access to services as White A&D offenders, all else (considered here) being equal. Some of this disparity may be attributable to the specific type of A&D offense with which youth are charged. Data showed that Whites were more likely than Blacks to have alcohol-related offenses, and were also more likely to be charged with possession rather than sale of controlled substances (data available from the author). National data also show that Blacks are five times more likely than Whites to be placed in custody for drug trafficking (US Dept. of Justice, 1999). Some (Peterson & Hagan, 1984) have suggested that drug laws demonize nonwhite offenders to the extent that they are, or are defined to be, pushers rather than users, villains rather than victims. Similar distinctions can be found in laws regarding crack cocaine (associated with Blacks) versus powder cocaine (associated with Whites). While some (Romer, 1994) advocate a public health response to all A&D offenders, including traffickers, present findings underscore the significance that variable legal definitions of drug-related crime can have for access to care.

Finally, the research demonstrates that traditional analytical techniques that do not account for the correlation among subjects (e.g., offenders) within units (e.g., courts) can yield biased results. Of particular interest are findings related to ethnicity and detention. OLS methods suggest that ethnicity directly affects courts' referral decision. It is only with the multiplicative model in HLM that the significance of ethnicity emerges. Moreover, prior research on ethnic disparity has focused on the detention decision itself, or on the impact of detention on other, final dispositions of the court. Here, OLS findings lead to results similar to those in prior research—direct and moderating effects of detention (for Blacks) are negative. However, the HLM analyses do not corroborate these findings. Thus, at the same time this research provides empirical support for ethnic disparity in the court decision to refer offenders to A&D services, it also demonstrates that findings based on commonly used, but inappropriate techniques in the past, may be misleading. Future research should consider techniques appropriate for nested designs when addressing critical questions such as ethnic disparity in court outcomes.

References

Alexander, R. (1996). African American youths and drugs: A time to pursue a mental health perspective. *Journal of Black Psychology, 22*, 374-387.

Bell, D., Jr., & Lang, K. (1985). The intake dispositions of juvenile offenders. *Journal of Research in Crime and Delinquency, 22,* 309-328.

Bishop, D. M., & Frazier, C. E. (1988). The influence of race in juvenile justice processing. *Journal of Research in Crime and Delinquency, 25*, 242-263.

Breda, C. (in press). Offender ethnicity and mental health service referrals from juvenile courts. *Criminal Justice and Behavior*.

Cohen, L. E., & Kluegel, J. R. (1978). Determinants of juvenile court dispositions: Ascriptive and achieved factors in two metropolitan courts. *American Sociological Review, 43*, 162-176.

Dannefer, D., & Schutt, R. K. (1982). Race and juvenile justice processing in court and police agencies. *American Journal of Sociology, 87*, 1113-1132.

Dembo, R., Williams, L., & Schmeidler, J. (1994). Psychosocial, alcohol/other drug use, and delinquency differences between urban Black and White male high-risk youth. *International Journal of the Addictions, 29*, 461-483.

Fagan, J., Slaughter, E., & Harstone, E. (1987). Blind justice? The impact of race on the juvenile justice process. *Crime and Delinquency, 33*, 224-258.

Frazier, C. E., & Bishop, D. M. (1985). The pretrial detention of juveniles and its impact on case dispositions. *Journal of Criminal Law and Criminology, 76*, 1132-1152.

Horwitz, A., & Wasserman, M. (1980). Some misleading conceptualizations in sentencing research: An example and a reformulation in the juvenile court. *Criminology, 18*, 411-424.

Tennessee Council of Juvenile and Family Court Judges (1997). *Juvenile Court Data [Data file].* Nashville, TN: Author.

Longmire, D. (1993). Re-medicalizing the connection between drugs and crime: Educating for peace rather than war. In P. B. Kraska (Ed.), *Altered states of mind: Critical observations of the drug war.* (pp. 251-265). New York: Garland.

Lyons, J., Baerger, D., Quigley, P., Erlich, J., & Griffin, E. (2001). Mental health service needs of juvenile offenders: A comparison of detention, incarceration, and treatment settings. *Children's Services: Social Policy, Research, and Practice, 4*(2), 69-85.

Marshall, I. H., & Thomas, C. W. (1983). Discretionary decision-making and the juvenile court. *Juvenile and Family Court Journal, 34*, 47-59.

McBride, D., VanderWaal, C., Terry, Y, & VanBuren, H. (1999). Breaking the cycle of drug use among juvenile offenders. *Final Technical Report*. Washington, DC: National Institute of Justice.

McCarthy, B. R., & Smith, B. L. (1986). The conceptualization of discrimination in the juvenile justice process: The impact of administrative factors and screening decisions on juvenile court dispositions. *Criminology, 24*, 41-64.

McManus, M., Alessi, N. E., Grapentine, W. L, & Brickman, A. (1984). Psychiatric disturbance in serious delinquents. *Journal of the American Academy of Child Psychiatry, 23*, 602-615.

Minor, K. I., Hartmann, D. J., & Terry, S. (1991). Predictors of juvenile court actions and recidivism. *Crime and Delinquency, 43*, 328-344.

Niarhos, F. J., & Routh, D. K. (1992). The role of clinical assessment in the juvenile court: Predictors of juvenile dispositions and recidivism. *Journal of Clinical Child Psychology, 21*, 151-159.

Peterson, R., & Hagan, J. (1984). Changing conceptions of race: Towards an account of anomalous findings of sentencing research. *American Sociological Review, 49*, 56-70.

Rogers, K. M., Powell, E., & Strock, M. (1998). The characteristics of youth referred for mental health evaluation in the juvenile justice system. In J. Willis, C. Liberton, K. Kutash, & R. Friedman (Eds.), *The 10th annual research conference proceedings, a system of care for children's mental health: Expanding the research base* (pp. 329-334). Tampa, FL: University of South Florida.

Romer, D. (1994). Using mass media to reduce adolescent involvement in drug trafficking. *Pediatrics, 93*, 1073-1077.

Scarpitti, F. R., & Stephenson, R. M. (1971). Juvenile court dispositions: Factors in the decision-making process. *Crime and Delinquency, 17*, 142-151.

Schonberg, S. (1993). Guidelines for the treatment of alcohol- and other drug-abusing adolescents. *Treatment Improvement Protocol (TIP) Series #4*. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, DHHS Publication # (SMA) 95-3059.

Shelton, D, & Phillips, R. (1999). Estimates of emotional disorder in detained and committed youth in the Maryland juvenile justice system. In J. Willis, C. Liberton, K. Kutash, & R. Friedman (Eds.), *The 11th annual research conference proceedings, a system of care for children's mental health: Expanding the research base* (pp. 275-278). Tampa, FL: University of South Florida.

Stapleton, V. D., Aday, D., & Ito, J. A. (1982). An empirical typology of American metropolitan juvenile courts. *American Journal of Sociology, 88*, 549-564.

Thomas, C. W., & Cage, R. J. (1977). The effect of social characteristics on juvenile court dispositions. *The Sociological Quarterly, 18*, 237-252.

Thomson, R. J., & Zingraff, M. T. (1981). Detecting sentencing disparity: Some problems and evidence. *American Journal of Sociology, 86*, 869-880.

Thornberry, T., Tolnay, S., Flanagan, T., & Glynn, P. (1991). *Children in custody 1987: A comparison of public and private juvenile custody facilities*. Washington, DC: U.S. Department of Justice. Office of Juvenile Justice and Delinquency Prevention.

U.S. Dept. of Justice, Office of Juvenile Justice and Delinquency Prevention. (1999). Minorities in the juvenile justice system. *National report series: Juvenile Justice Bulletin*. Washington, DC: U.S. Government Printing Office.

CONTRIBUTING AUTHORS

Carolyn S. Breda, Ph.D.
Center for Mental Health Policy, Vanderbilt University, 1207 18th Ave. South, Nashville,
TN 37212; 615-343-8723, fax: 615-322-7049; e-mail: Carolyn.Breda@Vanderbilt.edu

Confirmatory Factor Analysis of the Behavioral and Emotional Rating Scale (BERS)

Qinghong Liao
E. Wayne Holden
Michael Epstein

Introduction

Assessment and provision of services based upon a child and family's strengths is a key component of the system-of-care approach. Reliable and valid measures of these constructs are necessary for comprehensively evaluating outcomes associated with mental health services for children and families provided within community based settings. Several measures for evaluating the strengths of children and adolescents are available (Epstein, 1999; Gilgun, 1999; Lyons, Uziel-Miller, Reyes & Sokol, 2000; Rhee, Furlong, Turner & Harari, 2001). One of these measures, the Behavioral and Emotional Rating Scale (BERS; Epstein & Sharma, 1998) is a standardized, norm referenced scale that has been developed to assess the behavioral and emotional strengths of children and adolescents.

The BERS is designed to be completed by caregivers or professionals (e.g., teachers) to rate the behaviors of children ages 5 to 18. The scale consists of 52 items that can be categorized into the following five domains: Interpersonal strength, Family involvement, Intrapersonal strength, School functioning, and Affective strength. These items are rated on a 4-point scale: 0 = *not all at like the child*, 1 = *not much like the child*, 2 = *like the child*, and 3 = *very much like the child*. Standard scores can be derived based upon expectations for the full population of children within particular age and gender groups (standardization sample N = 2,176) or for specific groups of children with serious emotional disturbance (standardization sample N = 861). An overall strength quotient is also available that provides a single summary score of strengths. A rational/conceptual approach was used for item development that involved multiple stages with input and review from multiple stakeholder groups (Epstein, 1999). The factor structure for the BERS was derived with exploratory factor analysis (EFA) on a relatively small subset of children (N = 258). Evidence for test-retest and interrater reliability (Epstein, Harniss, Pearson & Ryser, 1999; Epstein, Hertzog & Reid, 2001), discriminant validity (Reid, Epstein, Pastor & Ryser, 2000), and convergent validity (Harniss, Epstein, Ryser & Pearson, 1999) is available. Preliminary evidence for significant agreement between parents and teachers ratings on the BERS has also been reported (Friedman, Leone & Friedman, 1999).

The BERS offers a number of important advantages over standard deficit oriented measures used in children's mental health services. The underlying factor structure of the BERS, however, has not been tested outside of the EFA that was conducted as part of the development of the measure. In addition, this EFA analysis was limited to teachers as respondents rather than caregivers. Furthermore, factorial invariance across demographic characteristics within samples of children and adolescents has not been adequately tested. A confirmatory factor analysis (CFA) has not been conducted to evaluate the degree to which the measure conforms to the original factor structure within populations that are being evaluated within community mental health settings with caregivers rather than teachers as respondents. The purpose of the present study was to confirm the factor structure of the BERS among a large population of children and adolescents entering services in community based systems of care where caregivers are the respondents and to evaluate factorial invariance across gender.

Method

Sample

The sample for this study was obtained from Phases II and III of the national evaluation of the Comprehensive Community Mental Health Services for Children and their Families Program (data were obtained from 35 out of the 45 grantee communities initially funded in the 1997, 1998 and 1999 funding cycles). As part of a comprehensive interview battery administered at intake into the

outcomes study for the national evaluation, caregivers were administered the BERS. Demographic information was obtained from responses to the Demographic Information Questionnaire (DIQ) which assesses a wide range of demographic, diagnostic and historical parameters relevant to entering services in systems-of-care.

For the purpose of this analysis, only children ages 5 to 18 with complete BERS data were included, which yielded a total of 1,799 children in the sample. The mean age of the sample was 12.5 (n = 1,799) and 70% of the children were boys (n = 1,798). The racial/ethnic background reported (n = 1,667) was 64% Caucasian, 19% African-American, 10% of Hispanic origin, 10% Native American, and 9% other (percentages do not necessarily add up to 100% because each respondent can identify more than one racial/ethnic background).

Results

Confirmatory factor analysis (CFA) of the five latent factors was conducted using the Amos 3.6 software (Arbuckle, 1997a, 1997b). CFA is a factor analytic technique used to test the validity of factor structures that have been previously established. The procedure optimally matches the observed variables and underlying latent structures for a new sample to determine the goodness-of-fit of the a priori model. The weighted least squares (WLS) method was used to estimate the model due to nonnormality of the data. The overall model fit was assessed with the chi-square test. Various goodness-of-fit indices are provided in Amos, which allow for evaluation of the fit of the factor structure to the current sample in the study. These fit indices include the adjusted goodness of fit index (AGFI), normed fit index (NFI) and relative fit index (RFI). It has been suggested in the literature that a .95 cutoff value is reasonable for both continuous and categorical outcomes with a large enough sample size (n > 250) (Browne & Cudeck, 1993; Hu & Bentler, 1999).

To evaluate the fit of the five-factor model to the data, we allowed correlations among the five factors to be freely estimated. In addition, all items on the BERS were allowed to load freely on their corresponding hypothesized factor, but not across the different factors. Results from the CFA analysis indicated that the five-factor model fit the current data well, with an AGFI value of 0.965, an NFI value of 0.961 and an RFI value of 0.959. For comparison, a model was constructed where all items loaded on one latent factor. The fit of this model was poor (AGFI = .752, NFI = .732, RFI = .695) relative to the five factor solution.

In addition, factor invariance across the subsamples of males and females was examined by conducting CFA for the two groups simultaneously and fitting different nested models. A baseline model was constructed where there was no equality constrained. In Model 1, factor loadings, factor variances and covariances are all constrained to be equal, in Model 2, factor loadings were constrained to be equal, and in Model 3, factor variances were constrained to be equal. Results of the comparison of fit indices across models are presented in Table 1. The results indicated that the five-factor model holds invariant across the two gender groups.

Table 1
Selected Goodness-of-Fit Indices for
Gender Invariance Comparison Model

Model	Chi-square	df	AGFI	NFI	RFI
0. Baseline Model[a]	6242.110	2528	.962	.958	.956
1. Model 1[b]	7057. 596	2410	.959	.952	.951
2. Model 2[c]	6706.245	2575	.960	.955	.953
3. Model 3[d]	6887.194	2579	.959	.953	.952

[a] No equality constrained.
[b] Factor loadings, variances, and covariances constrained as equal.
[c] Factor loadings constrained as equal
[d] Variances constrained as equal

Discussion

The overall fit of the model with the current sample indicated that the original five-factor structure of the BERS was well established. This occurred within a sample of parent/caregiver respondents rather than teachers who were the primary category of respondents for the development of the scale. Use of the BERS factors scores across various categories of respondents may be appropriate given the results of this study. However, factor invariance was examined only for the subsamples of males and females. Further research evaluating other demographic factors, such as age or ethnicity that may influence the factor structure underlying BERS scores, is warranted. In addition, it is important to conduct a CFA of the BERS that will test for second order factors underlying the five factor structure to confirm the utility of the strength quotient. Further exploration of the multidimensionality of behavioral and emotional strengths in children and adolescents is highly warranted.

A number of limitations with the current study should be noted. The most important perhaps is that the strategy employed in this analysis for addressing missing data was to exclude cases with missing information on any of the BERS items. Although a relatively large sample was investigated, it is not known whether the missing data were systematic, missing at random, or missing completely at random; therefore, results of this study may not be generalizable to other samples. Depending upon the nature of the missing data, results of this CFA analysis may differ from future analyses that include missing data. It should be noted that part of the reason that missing data were not handled in the analysis was because maximum likelihood estimation method is the only method that can be used in the CFA with the presence of missing data. However, given that data were non-normal in this case, a different estimation method (WLS) was employed, which did not allow for inclusion of missing data. Analyses using an imputation strategy to handle the missing data prior to their inclusion in the CFA may also be warranted.

References

Arbuckle, J. L. (1997a). Amos 3.6 [Computer software]. Chicago: Smallwaters.

Arbuckle, J. L. (1997b). *Amos user's guide version 3.6.* Chicago: Smallwaters.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230-239.

Epstein, M. H., Harniss, M. K., Pearson, N., & Ryser, G. (1999). The Behavioral and Emotional Rating Scale: Test-retest and inter-rater reliability. *Journal of Child and Family Studies, 8,* 319-327.

Epstein, M. H., Hertzog, M. A. & Reid, R. (2001). The Behavioral and Emotional Rating Scale: Long term test-retest reliability. *Behavioral Disorders, 26,* 314-320.

Epstein, M. H. (1999). The development and validation of a scale to assess the emotional and behavioral strengths of children and adolescents. *Remedial and Special Education, 20,* 258-263.

Epstein, M. H., & Sharma, J. (1998). *Behavioral and Emotional Rating Scale: A strength-based approach to assessment.* Austin, TX: PRO-ED.

Friedman, K. A., Leone, P. E. & Friedman, P. (1999). Strengths-based assessment of children with SED: Consistency of reporting by teachers and parents. *Journal of Child and Family Studies, 8,* 169-180.

Gilgun, J. F. (1999). CASPARS: New tools for assessing client risks and strengths. *Families in Society, 80,* 450-459.

Harniss, M. K., Epstein, M. H., Ryser, G., & Pearson, N. (1999). The Behavioral and Emotional Rating Scale: Convergent validity. *Journal of Psychoeducational Assessment, 17,* 4-14.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Lyons, J. S., Uziel-Miller, N. D., Reyes, F. & Sokol, P. T. (2000). Strengths of children and adolescents in residential settings: Prevalence and associations with psychopathology and discharge placement. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 176-181.

Reid, R., Epstein, M. H., Pastoir, D. A., & Ryser, G. R. (2000). Strengths-based assessment differences across students with LD and EBD. *Remedial and Special Education*, *21,* 346-355.

Rhee, S., Furlong, M. J., Turner, J. A. & Harari, I. (2001). Integrating strength-based assessments in psychoeducational evaluations. *California School Psychologist, 6*, 5-17.

CONTRIBUTING AUTHORS

Qinghong Liao, M.S.
*Senior Scientist, ORC Macro, 3 Corporate Square, Suite 370, Atlanta, GA 30329;
404-321-3211, fax: 404-321-3688; e-mail: liao@macroint.com*

E. Wayne Holden, Ph.D.
*Vice President, ORC Macro, 3 Corporate Square, Suite 370, Atlanta, GA 30329;
404-321-3211, fax: 404-321-3688; e-mail: wholden@macroint.com*

ORC Macro, 3 Corporate Square, Suite 370, Atlanta, GA 30329;

Michael Epstein, Ed.D.
*Professor, Department of Special Education and Communication Disorders,
University of Nebraska at Lincoln, 202 Barkley Memorial Center, Lincoln, NE
68583-0732; e-mail: mepstein@unl.edu*

Case-Mix Adjustment of the Children's Functional Assessment Rating Scale (CFARS)

Michael G. Dow
Timothy L. Boaz
John C. Ward

Introduction

There has been growing pressure to demonstrate the effectiveness of publicly funded human service programs using data on the outcomes of interventions. One of the primary challenges is to develop fair and meaningful indices of outcomes in applied settings in which the treatment results for pre-existing groups of mental health consumers are to be compared and contrasted. Controlled research studies are generally not possible in these settings. Instead, case-mix adjustment, or statistical adjustment of outcomes data from pre-existing groups of consumers, is necessary. Yet, there have been few attempts to implement such strategies and comprehensively compare the relative effectiveness of intact clinical programs, such as community mental health centers. The need for such methods is clear. State governments provide millions of dollars of funding for these programs without clear indications of the relative effects of these investments in public mental health.

Identifying Appropriate Case-Mix Adjustment Variables

The selection of a case-mix variable (CMV) is based on a theoretical assumption that some variable other than the treatment itself has a systematic effect on the dependent variable. The severity of the mental health condition prior to treatment is one such consideration, among others. However, unless the potential CMV correlates with the dependent variable and there are significant provider differences on the CMV, statistical adjustment will have little or no effect. Moreover, if the relationship between the CMV and the dependent variable is different for different providers, no systematic and fair adjustment is possible.

It is also important that the CMV is distinct from the treatment. Level of care (inpatient, day treatment, outpatient), for example, is not an appropriate CMV because this is an aspect of the treatment. When this type of comparison is desired, it makes more sense to "block" the data according to the treatment variable of interest and then compare providers separately for each type of treatment.

Dow, Boaz, and Thornton (2001) proposed four criteria for identifying appropriate case-mix adjustment variables: (a) the variable must be considered an independent variable other than the treatment, (b) the variable accounts for significant variance in the dependent variable, (c) providers differ significantly on the variable, and (d) the variable does not interact meaningfully with provider, i.e., the relationship between the case-mix variable and the dependent variable is relatively consistent across providers.

Conducting Case-Mix Adjustment

It is also important to select appropriate dependent variables of interest when analyzing changes over time. Although raw change scores (posttest score minus pretest score) are commonly used and may appear to make a reasonable adjustment for pretest values, raw change scores are only the best adjustment for initial values when the slope of the regression line (posttest regressed on pretest) is 1. A residual gain score is a more general solution to this problem that offers a better adjustment for initial values (Cronbach & Furby, 1970).

In the paper by Dow et al. (2001), the method of case-mix adjustment was demonstrated by comparing twenty-four providers of adult mental health services on two dependent measures of interest—the Behavioral Healthcare Rating of Satisfaction (BHRS; Dow & Ward, 1996), which is a consumer satisfaction measure, and the Global Assessment of Functioning (GAF; American Psychiatric Association, 1994), which is a clinician rating of functioning. Several demographic and

clinical indices were shown to be appropriate CMVs. However, the case mix adjustment results were somewhat surprising in that, although the rank order of a few providers changed a great deal, the overall unadjusted results correlated very highly with the overall adjusted results, r = .934 (87.2% of the variance). Interestingly, despite all of the concerns about the risk of comparing pre-existing groups of consumers in treatment, the adjustment process did relatively little to change the overall rank order of the providers. Obviously, it is important to replicate this approach with other dependent variables and other populations.

The present study demonstrates the process of case-mix adjustment using outcome data from child mental health programs funded and evaluated by the State of Florida, Department of Children and Families. Several variables are used to evaluate outcomes, including the Children's Global Assessment Scale (CGAS; Shaffer, et al., 1983), which is collected as part of the Children's Functional Assessment Rating Scale (CFARS; Ward, et al., 1998). Given that the CGAS has the greatest national recognition, CGAS results are featured here.

Method

Participant Selection

The State of Florida began collecting CFARS data to be used for performance monitoring in July 1998. The CGAS is one of the items on this scale. For each child receiving services the CGAS rating was to be completed at admission to the agency and quarterly thereafter until a final assessment was completed at discharge from the agency. The data for this study consisted of all valid CFARS records collected prior to October 1, 2001 (n = 115,749). Records were eliminated from the analysis if there was an invalid social security number, a missing CGAS score, or if the age was greater than 18 (resulting n = 92,781). From these records, a "best" follow-up record (posttest) was selected for each child with the priority that a discharge record was selected if available—otherwise, the latest quarterly assessment record was selected. Then, a "best" initial assessment record (pretest) was selected for each child with the priority that an admission record was selected if available—otherwise, the earliest quarterly assessment record was selected. Cases were retained for which the length of time between the initial assessment record and the follow-up assessment record was at least two days and no more than 365 days. This left 11,995 cases. In order to conduct an analysis comparing providers, we restricted the sample to cases from those providers that had complete CGAS data (pretest and posttest) on at least 150 cases. This left a final sample of 9,636 children who were served by 25 provider agencies.

Participants

For the 9,636 children, the mean age was 11.4 years (SD = 3.6); 63.3% were male; 53.1% were white; and 34.4% were black. Primary diagnoses included Adjustment Disorder (28.3%), ADD/ADHD (24.7%), Conduct Disorder (13.2%), Mood Disorder (11.0%), and Other (mostly DSM IV '313.xx', 12.3%). Approximately 66.2% of the children were classified as Severely Emotionally Disturbed (SED), and 33.8% were classified as Emotionally Disturbed (ED). Approximately 11% of the children were involved with the legal system. The mean CGAS at pretest was 49.1 (SD = 8.6) and the mean CGAS at posttest was 54.6 (SD = 9.7).

Procedure

The assigned case manager or primary therapist completed CGAS ratings at intake, three-month intervals, and at discharge. Using the algorithm described above, pairs of pretest and posttest data were selected for each child. The average length of time between the assessments was 152.2 days (SD = 89.5). A CGAS residual gain score was computed by predicting the posttest using the pretest and then using the residual (actual posttest score minus the predicted posttest score) as the dependent variable. The correlation of the pretest scores with the posttest scores was r = .50, p < .0001. The correlation of the pretest scores with the raw difference scores was r = -.41, p < .0001. The correlation of the pretest scores with the residual gain scores was r = 0, as expected.

Results and Discussion

Selecting Case Mix Variables

The correlation of each potential CMV with the CGAS residual gain score is shown in Table 1. Note that race was dummy coded by two binary variables of White/not White and Black/not Black ("other" is coded by 0 on each). Similarly, diagnosis was dummy coded by adjustment disorder/not, ADD/not, and conduct disorder/not. Results showed small, but significant correlations for each variable. The SED variable classification accounted for slightly more than 1% of the variance in the dependent variable; that is, children who were classified as SED improved less than expected.

As a second step in determining the appropriateness of these potential CMVs, analysis of variance (ANOVA) using the Statistical Analysis System (SAS, 2001) Proc GLM, demonstrated that there were significant provider differences on each potential CMV, $p < .05$. Moreover, provider differences on the CGAS residual gain score were significant, $F(24,9611) = 52.99$, $p < .0001$, and accounted for 11.7% of the variance.

Conducting Case Mix Adjustment

Using Proc REG in SAS, the best 1 variable model, 2 variable model, 3 variable model, etc., were considered in the effort to predict the CGAS residual gain score. The decision rule adopted was to keep predictor variables as long as the marginal F for inclusion was significant at $p < .01$ and increased the adjusted percent of variance accounted for. Results showed that the best model had three predictors: SED, adjustment disorder diagnosis, and conduct disorder diagnosis. These variables were entered as covariates (Proc GLM) and provider effects were examined, $F(24, 9608) = 47.82$, $p < .0001$. Each covariate was significant and the total model accounted for 12.9% of the total variance; this was an increase of only 1.2% from the unadjusted model presented above. As a final step, three terms for the interaction of each covariate and provider were added and were shown to be significant, $F(3, 9536) = 5.07$, $p < .01$. That model accounted for 16.1% of the variance. Unfortunately, these interaction effects suggest that the case mix process is somewhat questionable for these data, in that the relationships between the covariates and the dependent variable are not homogeneous across the set of providers. The provider means, adjusted and unadjusted, are shown in Table 2 for illustrative purposes.

Table 1
Significant Correlations of Possible Case Mix Adjustment Variables with CGAS Residual Gain Score

Case Mix Variable	r
White	.050
Black	-.074
Age	.046
Legal Involvement	.077
SED	-.104
Adjustment Disorder	.080
ADD/ADHD	-.072
Conduct Disorder	.078
Admission Problem Severity Rating	-.085

Note. $p < .0001$ for all correlations listed

Table 2
Provider Means on CGAS Residual Gain Scores, Unadjusted and Adjusted

Provider	CGAS ResGain Mean and Rank	CGAS ResGain Case-Mix Adjusted Mean and Rank
A	5.99 (1)	6.39 (1)
B	5.55 (2)	5.21 (2)
C	5.39 (3)	4.39 (3)
D	4.79 (4)	4.36 (4)
E	3.89 (5)	3.90 (5)
F	3.39 (6)	2.94 (6)
G	2.75 (7)	2.57 (7)
H	2.24 (8)	2.32 (8)
I	0.57 (9)	0.89 (9)
J	0.51 (10)	0.42 (10)
K	0.45 (11)	0.32 (11)
L	0.31 (12)	0.21 (12)
M	-0.64 (13)	-0.43 (13)
N	-1.08 (14)	-1.10 (15)
O	-1.28 (15)	-1.27 (16)
P	-1.28 (16)	-0.96 (14)
Q	-2.17 (17)	-2.15 (17)
R	-2.42 (18)	-2.24 (18)
S	-2.72 (19)	-2.62 (19)
T	-2.78 (20)	-2.98 (21)
U	-3.19 (21)	-3.07 (22)
V	-3.28 (22)	-2.69 (20)
W	-3.79 (23)	-3.67 (23)
X	-4.03 (24)	-4.34 (25)
Y	-4.53 (25)	-4.05 (24)

Conclusions

The importance of using a residual gain score was demonstrated, in that the correlation of the pretest with a raw difference score was almost as large (in absolute value) as the correlation of the pretest and the posttest. The selection process for determining appropriate case mix variables and conducting the analyses was demonstrated. The best case mix variables available were SED/ED categorization, conduct disorder diagnosis, and adjustment disorder diagnosis. However, case mix adjustment made very little difference in the overall results beyond those achieved by using the residual gain scores. Further, the results suggested that case mix adjustment using these variables may have been somewhat inappropriate due to the lack of homogeneous relationships between the CMV and the DV across providers. Providers differed significantly on the average CGAS residual gain score, whether or not case mix adjustment was done. The "best" provider obtained posttest CGAS scores that were an average of 5.99 points (or 6.39 when adjusted) higher than expected. The "worst" provider obtained scores that were 4.53 points below expectation (or -4.05 for CMA). Further research is needed to explore the effects and utility of this method. These results may have been limited by provider differences in the use of the certification variables or diagnosis.

References

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders, 4th edition.* Washington DC: Author.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"—Or should we? *Psychological Bulletin, 74,* 68-80.

Dow, M. G., & Ward, J. C. (1996). *The Behavioral Healthcare Rating of Satisfaction (BHRS).* Odessa, FL: Psychological Assessment Resources.

Dow, M. G., Boaz, T. L., & Thornton, D. (2001). Risk adjustment of Florida mental health outcomes data: Concepts, methods, and results. *Journal of Behavioral Health Services & Research, 28,* 273-286.

Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahila, S. (1983). Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry, 40,* 1128-1231.

Statistical Analysis System (SAS) [Statistical analysis software]. (2001). Cary, NC: SAS Institute, Inc.

Ward, J. C., Dow, M. G., Saunders, T. L., Halls, S. C., Penner, K. F., Musante, K. A., Berry, R. T., & Sachs-Ericcson, N. (1998). *Children's Functional Assessment Rating Scale,* Odessa, FL: Psychological Assessment Resources, Inc.

CONTRIBUTING AUTHORS

Michael G. Dow, Ph.D.
813-974-1928, fax: 813-974-9327; e-mail: dow@fmhi.usf.edu
Timothy L. Boaz, Ph.D.
813-974-1933, fax: 813-974-9327; e-mail: boaz@fmhi.usf.edu
John C. Ward, Ph.D.
813-974-1929, fax: 813-974-9327; e-mail: ward@fmhi.usf.edu

All authors: Department of Mental Health Law & Policy, Louis de la Parte Florida Mental Health Institute, University of South Florida, 13301 Bruce B. Downs Blvd., Tampa, FL, 33612-3807.

Symposium

Toward an Understanding of the Processes and Outcomes of Mental Health Treatment for Adolescents

Chair
James M. Robbins

Discussant
John Landsverk

Authors
James M. Robbins et al.
Susan D. Phillips et al.
Teresa L. Kramer et al.
Terri L. Miller et al.

Symposium Introduction

James M. Robbins

The common theme of the papers presented in this symposium is the measurement of the treatment adolescents routinely receive in specialty mental health settings. Papers presented in this symposium use measures included in the Adolescent Treatment Outcome Module (ATOM) of adolescent emotional and behavioral problems and clinicians' detection of and response to those problems. More specifically, these papers examine alternative ways of conceptualizing psychosocial functioning and methods for adjusting measures of treatment outcomes to allow more equitable between-provider comparisons. Information is also presented on how the treatment provided in routine care compares to treatment suggested by empirical studies and treatment guidelines.

Empirically-Derived Functioning Scales from the Adolescent Treatment Outcomes Module (ATOM)

James M. Robbins, Susan D. Phillips, Teresa L. Kramer, Terri L. Miller, & Barbara J. Burns

Introduction

Adolescents' social functioning can be defined by performance or competency in key life domains: the family, school, community, and relationships with peers. Impairment or disability occurs when an adolescent's performance in one or more of these areas substantially diverges from what is "normal," or expected (Canino, Costello, & Angold, 1999). Whether impaired social functioning occurs in conjunction with or distinct from psychiatric symptoms, difficulty adapting to life demands in these areas is often the cause of referral for mental health care.

The ability to adequately assess functioning is important in establishing diagnoses, determining eligibility for mental health services, and gauging the effectiveness of mental health care. While a number of excellent measures of functioning exist[1], a brief instrument is needed that can be completed by the adolescent and parent, requires no formal training other than reading ability, and is not confounded with measures of psychopathology. In that vein, the functioning scales of the Adolescent Treatment Outcomes Module (ATOM; Robbins et al., 2001) were constructed to assess adolescents' performance in the key domains of home, school, community, and relationship with peers. These measures were derived conceptually from the domains of adolescent functioning identified by Hoagwood, Jensen, Petti, and Burns (1996) and in consultation with an expert panel. Six scales were constructed: (a) Functioning at Home, (b) Functioning at School, (c) Trouble with Rules and Laws, (d) Friendships, (e) Dysfunctional Peers, and (f) Leisure Functioning. These scales were tested in a sample of 67 adolescents in routine care settings and their parents, and subsequently in a sample of 258 adolescent-parent pairs (Robbins, et al., 2001). They are modestly correlated with the more global Columbia Impairment Scale (CIS; (Bird et al., 1993) and the Child Health Questionnaire (CHQ; Landgraf, Abetz, & Ware, 1996) measure of role limitations attributable to problems with emotions or behaviors.

---

[1] *For example, the Child and Adolescent Functional Assessment Scale (CAFAS, Hodges, 1990); the Vanderbilt Functioning Index (VFI; Bickman, Lambert, Karver, & Andrade, 1998); the Child and Adolescent Psychiatric Assessment (CAPA; Angold, et al., 1995); the Child Behavior Checklist (CBCL; Achenbach, 1991); the Children's Global Assessment Scale (C-GAS; Shaffer, et al., 1983); the Columbia Impairment Scale (CIS; Bird, et al., 1993); and the Social Adjustment Inventory for Children and Adolescents (SAICA; John, Gammon, Prusoff, & Warner, 1987).*

Factor analyses of other measures of functioning reported in the literature led us to anticipate that the factors that would be empirically extracted from the ATOM measures of functioning might be less discrete than we had originally conceptualized (Bird et al., 1993; Ezpeleta, Keeler, Erkanli, Costello & Angold, 2001; John, Gammon, Prusoff & Warner, 1987). Consequently, the goal of the current project was to compare the domains of functioning identified through factor analysis of parent and adolescent responses to the initial, conceptually-based domains of functioning.

Method

The ATOM was administered to 258 adolescents receiving specialty mental health care and their parents. Factor models were estimated using Mplus statistical software (Muthén & Muthén, 2000). Mplus approaches the factor analysis of categorical data by assuming items are derived from an underlying normal distribution with a superimposed cut-point. The tetrachoric correlation matrix is analyzed and a weighted least squares approach is used to estimate the factor structure with the asymptotic variance-covariance matrix of the tetrachoric correlations serving as the weight matrix.

Results

Parent Responses

Eight factors displayed Eigenvalues greater than 1. Combined, they accounted for 66.4% of the total variance. A factor loading of .50 or greater was used to interpret the rotated factor loadings. Three of the factors were identical to three functioning scales originally conceptualized: Leisure Activities (Factor 3), Making and Getting Along with Friends (Factor 7), and Friends' Delinquent Activities (Factor 2). The remaining items loaded in an unexpected but intuitively appealing way. Together, these factors can be thought of as representing three areas of functioning: problems with role performance, problems with social relationships, and consequences for poor role performance.

Factor 1 (Role Performance) captured a cross-section of items from prior scales conceptualized as measuring functioning at home, school, and in the community. The items loading on Factor 1 were the items in each scale that measured role performance: (a) trouble following rules at school, (b) not being responsible at school, (c) not doing homework or schoolwork, (d) trouble following rules at home, (e) trouble doing chores, (f) trouble getting along with teachers, and (g) trouble getting along with adults outside the family.

Factor 4 (Legal Consequences) consists of items assessing adolescents' involvement with the criminal justice system that range from breaking the law to being incarcerated. Items from the initial scale assessing functioning at school reflecting consequences of poor school performance (suspension, expulsion) loaded together on Factor 6 (School Consequences).

Factor 5 (Relationship with Family Members) consists of items from the initial scale assessing functioning in the home that measures difficulties in relationships between adolescents and their mothers and fathers, and lack of involvement in family activities. Interestingly, "trouble getting along with siblings," which was part of the initial scale assessing functioning at home, loaded as a separate one-item factor (Factor 8).

Adolescent Responses

Seven factors in this analysis displayed Eigenvalues greater than 1. Combined, these factors accounted for 57.6% of the total variance. Adolescent factors and factor loadings were very similar to factors derived from parental responses.

As in the analysis of parent responses, items assessing leisure activities (Prosocial Activities), making friends and getting along with them (Friendships), and friends' delinquent activities (Delinquent Friends) loaded together. Factors based on the analysis of adolescent responses were almost identical to factors based on parent responses. However, "involvement with family activities" and "trouble with

rules and laws in the community" did not load on any factor in the analysis of adolescent responses. Also, "trouble getting along with siblings," which loaded as a separate one-item factor in the analysis of parent responses, loaded on the same factor with "trouble getting along with mother" and "trouble getting along with father."

Internal Consistency

After analyzing the factor structure of the ATOM functioning items, the ATOM functioning scales were reconceptualized as representing three broad domains of functioning: Role Performance, Relationships, and Consequences. Given the close similarity between parent and adolescent factor-based responses, measures of these three domains were constructed for both informants. Measures, descriptions of content, number of items, and alpha coefficients for parent and adolescent measures are presented in Table 1.

Table 1
Empirically-Derived Measures of Functioning

Factor	*# of Items*	*Domain*	*# of Items*	*Alpha (Parents/Adolescents)*
1. Role performance (trouble following rules at home, school and in the community)	7	**Role performance** (trouble following rules and engaging in positive activities)	12	.83 / .77
3. Prosocial activities (little involvement in positive activities	5			
5. Relations in family (trouble getting along with mother and father)	3	**Relationships** (trouble getting along with others and making positive friendships)	9	.67 / .58
7. Friendships (trouble making and getting along with friends)	3			
2. Delinquent friends (friends regularly get in trouble	3			
4. Legal consequences (contact with police – arrest/ incarceration)	3	**Consequences for misconduct** (disciplinary actions resulting from misbehavior)	7	.71 / .72
6. School consequences (school disciplinary actions)	4			

The alpha coefficient for Role Performance (e.g., trouble with chores; trouble with rules at home, school, and in the community; trouble being responsible at school, trouble getting along with teachers and adults other than parents) was .83 for the parent scale and .80 for the adolescent scale. Role Performance items were combined with the items in the Prosocial Activities scale (e.g., sports, hobbies, extra-curricular activities, leisure interests) to form a broader measure of Role Performance. Resulting alphas were .80 (parent) and .77 (adolescent). The alpha coefficient for items measuring adolescents' relationships with their families (e.g., getting along with mother, father, and siblings, and taking part in family activities) were moderate (.57 parent, .49 adolescent). Items from the Relations in Family scale were combined with items assessing friendships and peers' delinquent activities to create a broader measure of Relationships. Alphas for this broader measure were .67 (parent) and .58 (adolescent). The alpha coefficient for items assessing Consequences for misconduct were .71 (parent) and .72 (adolescent) and remained unchanged when the item assessing out-of-home placement due to emotional and behavioral problems was added (parent scale only).

Concurrent Validity

Scales based on ATOM functioning domains (Role Performance, Relationships, Consequences) were only modestly (.19-.39) correlated with C-GAS ratings of global functioning. The strongest correlation (.39) was with parent ratings of adolescent role performance (including prosocial activities). Parent ratings of adolescent role performance and relationships were also modestly correlated with the CBCL measures of social competency (.35-.50). Correlations with adolescent ratings were lower (.12-.28). Correlations between the ATOM Role Performance, Relationships, and Consequences scales and the Columbia Impairment Scale were in the range of .65-.69 for parent responses and .32 to .41 for adolescent responses.

Within the scope of the project, it was not possible to validate arrests, incarcerations, and school suspensions or expulsions against school or police records. However, the ATOM Consequences measure (illegal activities, questioning by police, arrest, incarceration, school suspension, expulsion) correlated .56 with the CBCL Delinquency scale.

Discussion

In prior studies of measures of functioning, Bird and colleagues (1993) reported that the Columbia Impairment Scale was best represented as a unifactorial measure. John and colleagues (1987) preferred a three-factor solution for the SAICA. Factors were labeled Task Performance, Family Relations, and Spare Time Sociability. Ezpeleta, Keeler, Erkanli, Costello and Angold (2001) also identified a three-factor solution for the CAPA indicated by the domains of School Disability, Family Disability, and Peer Disability.

We have also identified three broad areas underlying both parent and adolescent ratings of adolescent functioning: Role Performance, Relationships, and Consequences. These domains reflect core expectations of adolescent social functioning: (a) behavior according to rules and regulations at home, at school, and in the community; (b) successful relationships with family members, school authorities, and community peers; and (c) ultimately, whether the adolescent remains at home, in school, and out of trouble with the law. These brief scales measuring important areas of functioning were found to be internally consistent and to correlate in the expected direction with longer research-based measures. These scales can be used along with measures of psychiatric symptoms, family impact, and satisfaction with care to monitor changes in social functioning in studies of the effectiveness of mental health care for adolescents.

References

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

Angold, A., Prendergast, M., Cox, A., Harrington, R., Simonoff, E., & Rutter, M. (1995). The Child and Adolescent Psychiatric Assessment (CAPA). *Psychological Medicine, 25*, 739-753.

Bickman, L., Lambert, E. W., Karver, M., & Andrade, A. R. (1998). Two low cost measures of child and adolescent functioning for services research. *Evaluation and Program Planning, 21*, 263-275.

Bird, H. R., Shaffer, D., Fisher, P., Gould, M. S., & Staghezza B (1993). The Columbia Impairment Scale (CIS): Pilot findings on a measure of global impairment for children and adolescents. *International Journal of Methods in Psychiatric Research, 3*, 167-176.

Canino, C., Costello, E. J., & Angold, A. (1999). Assessing functional impairment and social adaptation for child mental health services research: A review of measures. *Mental Health Services Research, 1*, 93-108.

Ezpeleta, L., Keeler, G., Erkanli, A., Costello, E. J., & Angold, A. (2001). Epidemiology of psychiatric disability in childhood and adolescence. *Journal of Child Psychology and Psychiatry, 47*, 901-914.

Hoagwood, K., Jensen, P., Petti, T., & Burns, B. (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.

Hodges, K. (1990, 1994 revision). *Child and Adolescent Functional Assessment Scale*. Ypsilanti, MI: Eastern Michigan University, Department of Psychology.

John, K., Gammon D., Prusoff B. R., & Warner V. (1987). The Social Adjustment Inventory for Children and Adolescents (SAICA): Testing of a new semistructured interview. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 898-911.

Landgraf, J. M., Abetz, L., & Ware, J. E. (1996), *The CHQ User's Manual*. Boston: The Health Institute, New England Medical Center.

Muthén & Muthén. (2000). *Mplus Version 2.12 Statistical Software*. Los Angeles, CA: Author. Retrieved from: http://www.statmodel.com/version2.html

Robbins, J. M., Taylor, J. L., Rost, K. M., Burns, B. J., Phillips, S. D., Burnam, M. A., & Smith, G. R. (2001). Measuring outcomes of care for adolescents with emotional and behavioral problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 315-324.

Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P. J., Fisher, P. W., Bird, H. R., & Aluwahlia, S. (1983). A Children's Global Assessment Scale (C-GAS). *Archives of General Psychiatry, 40*, 1228-1231.

An Exploration of Models for Adjusting Mental Health Treatment Outcomes for Adolescents

Susan D. Phillips, Teresa L. Kramer, Scott N. Compton, Barbara J. Burns, & James M. Robbins

Introduction

If mental health care providers are to be judged fairly using outcome data, it is essential to take into consideration the fact that the outcome of mental health care may not be due solely to the treatment patients receive. Evidence suggests that the extent to which adolescents improve during an episode of care is also influenced by clinical, demographic, and familial characteristics that exist when adolescents present for treatment (Blotcky, Dimperio, Gossett, 1984; Pfeiffer & Strzelecki, 1990; Phillips, et al., 2000). When provider outcomes are compared, those who treat a greater proportion of adolescents with characteristics associated with poorer treatment response may be at a disadvantage. Case-mix adjustment is a way of statistically compensating for between-provider differences in the prevalence of factors that adversely affect treatment response in order to make between-provider comparisons more equitable. The present study had two goals: To compare case-mix adjustment models developed alternately from intake clinical data alone, intake clinical data plus administrative data, and intake clinical data plus parent-reported data describing youths' histories and their family environments, and to explore whether case-mix adjustment makes a difference in interpreting provider performance.

Methods

Participants and Settings

Participants were 258 adolescents and their parents or caregivers (referred to hereafter as parents) who were recruited at intake from inpatient and outpatient treatment sites in Arkansas and Texas as part of a study to validate the Adolescent Treatment Outcomes Module (ATOM; Robbins et al., 2001). Follow-up data were collected on 237 adolescents (92% of the initial sample) approximately six months following intake.

Seven provider agencies participated in this study. Two agencies contributed a minimal number of subjects (5 and 9) to the sample. The site contributing five subjects was a freestanding psychiatric

hospital that closed unexpectedly shortly after the study began. For the purposes of this study, the five subjects from this site were combined with 23 subjects recruited from a like facility in the same community. The site contributing nine subjects was an adolescent medicine program with mental health providers on staff. The nine subjects from this site were combined with 40 subjects from an outpatient mental health program administratively linked to the adolescent medicine program.

After regrouping subjects, there were five provider agencies included in this study. Provider A ($n = 40$) is a mental health program with acute care, residential treatment, and outpatient services operated by a teaching hospital. Providers B ($n = 92$) and D ($n = 50$) are not-for-profit agencies providing a broad spectrum of mental health services for youth. Provider C ($n = 49$) is a hospital-based outpatient mental health program. Subjects from the hospital's adolescent medicine program are included with these subjects. Provider E ($n = 32$) is a psychiatric hospital; its sample is augmented by five subjects from a similar hospital.

Measures

Potential Case-Mix Variables. The case-mix variables included in the present analyses were selected from a pool of variables developed through expert consensus and a review of relevant research (Phillips et al., 2000; Robbins et al., 2001). Of the initial pool of potential variables, those that were significantly ($p < .05$) correlated with any of the outcome measures described above were retained for consideration in regression analyses. These variables were grouped into three categories: (a) factors describing youth clinical status at intake (e.g., symptom severity, role performance, relationships, consequences, family impact); (b) information available from administrative records (e.g., Medicaid, clinician assessment of disruptive disorder, clinician assessment of substance use, age); and (c) parent-reported factors associated with the youth's clinical history and family environment (e.g., history of special education for mental retardation, onset of problems before age six, first treatment before age six, family functioning, prior psychiatric hospitalization, history of abuse/neglect, parental incarceration).

Outcome Measures. The ATOM measures three broad categories of outcomes: symptoms (e.g., caseness, symptom severity), multiple domains of functioning, and the consumer's perspective on care (e.g., satisfaction, impact of adolescent's problems on the caregiver and family). From these three categories, five outcome variables were chosen: (a) symptom severity; (b) impaired role performance (e.g., trouble following rules at home, school, and in the community, getting along with teachers, and doing schoolwork); (c) relationships (trouble getting along with family members, troublemaking and getting along with friends, and peers' involvement in delinquent activities); (d) significant consequences associated with impaired functioning (e.g., school suspension or expulsion, questioning by police, arrest, incarceration); and (e) the impact of the youth's problems on his or her family.

Data Analyses

Adjusted Outcome Scores. Variables with significant bivariate correlations ($p \leq .05$) with a given outcome variable were modeled using ordinary least squares (OLS) regression and predicted outcome scores were calculated based on the resulting models. Only variables representing the clinical status of the adolescent at intake were entered in the first set of models. Administrative data variables were entered into a second set of models with the significant predictors from the first models. In a third series of models, the parent-report variables were entered with the significant predictors from the first series of regression models. A fourth model was intended to be a model based on the significant variables from each of the previous models. In this step, no administrative data variable was significant in a model. These models therefore were essentially the same as the models based on intake clinical status data and parent-report data. Predictive scores were not calculated for these models.

Providers' Performance Based on Adjusted and Unadjusted Outcomes. A 95% confidence interval was calculated for individual providers' mean unadjusted and adjusted (i.e., scores calculated from the predictive equations resulting from the regression models) scores for each outcome. This confidence interval was then compared to the point estimate of the overall mean for all providers.

Results

Regression Models

Variables passing the screen of significant bivariate association with outcomes were considered in regression models. Table 1 shows the unstandardized regression coefficients for each variable that was included in each of these three series of models.

The adjusted R^2 for models based on only the intake measures of youth clinical status ranged from .25 (consequences; i.e., school suspension/expulsions, arrest, incarceration) to .45 (symptom severity). The addition of administrative data variables and parent-report variables did not appreciably increase the value of the adjusted R^2.

Table 1
Coefficients for Variables Entered in Linear Regression Models

	Outcome				
Model	*T_2 symptom severity*	*T_2 role performance*	*T_2 relationship*	*T_2 consequences*	*T_2 Family impact*
Intake measures of clinical status					
Intercept	-3.46	6.42	4.38	.14	10.01
T_1 symptom severity	0.61***	0.01	0.00	0.02	.22**
T_1 role performance	0.04	0.38***	0.03	0.00	0.01
T_1 relationships	0.12	0.12	0.56***	0.01*	0.11
T_1 consequences	0.03	0.01	-0.00	0.39***	0.04
T_1 family impact	0.13	0.01	0.00		0.38***
Adjusted R^2	.45	.33	.40	.25	.30
Significant intake measures of clinical status plus administrative variables					
Intercept	2.84	8.06	5.21	-1.10	
T_1 symptom severity	0.60***				
T_1 role performance		0.42***			
T_1 relationships			0.58***	0.01**	
T_1 consequences				0.37***	
T_1 family impact					
Medicaid	3.34**	0.88			
Disruptive disorder (clinician)	0.98	1.21	0.75*		
Substance use (clinician)				0.24	
Age	1.03	0.00		0.01	
Adjusted R^2	.39	.32	.39	.27	
Significant intake measures of clinical status plus parent-report variables					
Intercept	2.25	6.07	4.93	0.00	10.11
T_1 symptom severity	0.60***				0.22**
T_1 role performance		0.41***			
T_1 relationships			0.62*	0.07*	
T_1 consequences				0.44***	
T_1 family impact					0.36***
Special education (mental retardation)	5.56*		0.53		
Onset ≤ age 6	3.99**	0.42			2.50
First treatment ≤ age 6		2.28**	1.43**		
Family functioning		1.18**	2.61		1.56
Prior hospitalization			1.23		1.40
History abuse/neglect	1.14				
Parent incarceration	3.08*	1.78**		.62**	
Adjusted R^2	.46	.36	.42	.28	.31

T_1 = intake, T_2 = 6 month follow-up
*= $p \le .05$; **= $p \le .01$; *** = $p \le .001$
⁺There were no administrative data variables with significant bivariate correlations to add to the model

Providers' Performance Relative to Unadjusted Outcomes

Table 2 shows the differential effect of the three case-mix adjustment models on individual providers' mean outcome scores relative to unadjusted scores. A provider's mean score (within a 95% confidence interval) that exceeds the point estimate of the mean score for all providers indicates that, on average, youth treated by that provider had significantly more problems at follow-up and are identified in Table 2. A provider's mean score (within a 95% confidence internal) that was lower than the point estimate of the mean score for all providers indicates that, on average, youth treated by that provider had significantly fewer problems at follow-up and are identified in Table 2.

Case-mix adjustment affected the interpretation of some providers' scores more than it did others. For example, adjusting scores by intake clinical status altered the interpretation of four out of five outcome scores for Provider D (i.e., symptom severity, role performance, and consequences). In comparison, only one outcome score (i.e., consequences) changed for Provider E. Sets of scores in which an adjustment score leads to a different interpretation of a provider's performance are identified in Table 2.

Table 2
Provider Performance Based on Unadjusted and Adjusted Mean Outcome Scores

Provider												
	A: Mixed resident & Outpatient		B: Mixed Services		C: Outpatient		D: Mixed Services		E: Hospital		Total	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Symptom Severity												
U	18.23[b]	1.80	18.24[b]	1.22	23.26[a]	1.65	18.64	1.58	25.41[a]	2.28	20.10	.73
I	11.23	1.04	11.07[b]	.70	13.40[a]	1.03	9.46[b]	.86	16.17[a]	1.10	11.78	.42
A	17.04[b]	.98	18.73[b]	.75	21.70[a]	1.17	17.83[b]	.91	24.76[a]	1.28	19.53	.47
P	18.12[b]	1.14	19.42	.85	22.92[a]	1.25	17.67[b]	.95	24.63[a]	1.33	20.07	.50
Role Performance												
U	15.69[b]	.65	17.39	.53	18.55	.81	17.91	.75	20.07[a]	1.09	17.76	.33
I	12.99[b]	.34	13.75[b]	.23	14.24	.28	14.35	.34	15.39[a]	.37	14.02	.14
A												
P	15.32[b]	.44	17.05	.33	17.60	.39	18.12[a]	.49	19.07[a]	.42	17.32	.20
Relationships												
U	14.50	.47	14.69	.33	15.25	.57	13.79[b]	.58	17.47[a]	.78	14.91	.23
I	12.71[b]	.27	13.26	.20	13.02	.28	12.87[b]	.29	14.56[a]	.39	13.20	.12
A	13.96[b]	.31	14.85	.22	14.61	.31	14.54	.34	16.17[a]	.40	14.75	.14
P	14.41[b]	.34	14.98	.23	14.80	.32	14.47[b]	.33	16.45[a]	.46	14.93	.14
Consequences												
U	1.49[b]	.15	2.69	.21	2.64	.28	3.02[a]	.31	2.88	.42	2.59	.12
I	7.03[b]	.35	7.82	.23	8.32	.29	8.43	.35	9.53[a]	.38	8.09	.14
A												
P	1.92[b]	.10	2.76	.12	2.82	.16	2.69	.13	3.50[a]	.25	2.71	.07
Family Impact												
U	35.97	2.04	32.31[b]	1.09	39.07[a]	1.61	32.06[b]	1.48	41.48[a]	3.28	35.12	.77
I	32.17	.96	30.60[b]	.70	33.50[a]	1.04	30.36[b]	.86	35.86[a]	1.03	31.91	.42
A												
P												

U = unadjusted scores; I = adjusted by intake clinical data only; A = adjusted by intake clinical data & administrative data; P = adjusted by intake clinical data & parent-report of youth history and family environment

'Boxes' indicate instances in which the relative performance of a provider is different for unadjusted and adjusted scores.

[a] On average, youth at follow-up had significantly more problems compared to the total sample

[b] On average, youth at follow-up had significantly fewer problems compared to the total sample

Discussion

This study demonstrates that the conclusions that might be drawn about provider performance differ depending on whether unadjusted or adjusted scores are used. As has been observed in studies of adult mental health outcomes, the effect of case-mix adjustment on the interpretation of providers' scores is more pronounced for some providers than for others (Dow, Boaz, & Thornton, 2001). However, after the Time 1 measure of an outcome variable is accounted for, the specific model by

which scores are adjusted appears to make little difference in the bottom line interpretation of providers' performance. That is, when providers' scores adjusted by models based only on intake data are compared with scores adjusted by models that also include administrative or parent-report data, a different conclusion is reached about provider performance in relatively few instances.

References

Blotcky, M. J., Dimperio, T. L., Gossett, J. T. (1984). Follow-up of children treated in psychiatric hospitals: A review of studies. *American Journal of Psychiatry, 141*(12),1499-1507.

Dow, M., Boaz, T., & Thornton, D. (2001). Risk adjustment of Florida mental health outcomes data: Concepts, methods, and results. *The Journal of Behavioral Health Services & Research, 28*(1), 258-272.

Pfeiffer, S. I., Strzelecki, S. C. (1990). Inpatient psychiatric treatment of children and adolescents: A review of outcome studies. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*(6), 847-853.

Phillips, S. D., Hargis, M. B., Kramer, T. L., Lensing, S.Y., Taylor, J. L., Burns, B. J., & Robbins, J. M. (2000) Toward a level playing field: Predictive factors for the outcomes of mental health treatment for adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(12), 1485-1495.

Robbins, J. M., Taylor, J. L., Rost, K. M., Burns, B. J., Phillips, S. D., Burnam, M. A., & Smith, G. R. (2001). Measuring outcomes of care for adolescents with emotional and behavioral problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(3), 315-324.

Quality of Mental Health Care for Depressed Adolescents

Teresa. L. Kramer, Terri. L. Miller, Susan. D. Phillips, James. M. Robbins, & Barbara. J. Burns

Introduction

Although investigators have consistently found deficiencies in the quality of mental health care for depressed adults (Goethe, Szarek, & Cook, 1988; Keller, 1988; Kramer, Daniels, Zieman, Williams, & Dewan (2000), similar studies of adolescent depression treatment have not been conducted. This is particularly discouraging considering the psychosocial morbidity, recurrent episodes and later onset of psychiatric problems associated with depression in youth (c.f., Cicchetti & Toth, 1998). Moreover, considerable progress has been made in identifying efficacious treatments for adolescent depression. For example, cognitive behavioral therapies (CBT) six to eight weeks in duration have had success in ameliorating symptoms in adolescents ages 13 through 18 in specialty care and school settings. Additionally, psychopharmacological interventions, particularly the selective serotonin reuptake inhibitors (SSRIs), have also been successful at reducing symptoms in moderately to severely depressed adults and adolescents when compared with placebo or other antidepressant medications.

Professional agreement and a growing evidence base on depression treatment for adolescents underscore the need to examine variations in the quality of care within the assessment, acute, and maintenance phases of care (American Academy of Child & Adolescent Psychiatry, 1998). It is anticipated that such findings will define the next phase of research: the development and implementation of interventions in child and adolescent treatment settings that will improve the quality of care. Although research on adult depression has progressed to this level, considerable work remains in the child and adolescent field (Hoagwood, 2001). The purpose of this study was to assess the quality of depression treatment received by adolescents across five mental health specialty care settings as documented in medical records. The review addressed several critical areas of treatment, including: (a) assessment of diagnostic criteria, comorbidity, suicide risk, and substance abuse; (b) inclusion of the family in treatment planning; and (c) provision of antidepressant medication, cognitive behavioral therapy and family interventions.

Methods

Participants

Medical record reviews were conducted for 208 patients enrolled in the validation study of the Adolescent Treatment Outcomes Module (ATOM; Robbins et al., 2001). Inclusion criteria included 11-18 years old, who were not psychotic or mentally retarded (IQ ≤ 70), and for whom an adult respondent was available who had at least weekly contact with the adolescent during the previous six months.

Instruments

Quality indicators were developed based on a thorough review of the literature on treatment guidelines, practice parameters, randomized clinical trials and other published studies on adolescent depression. Data pertaining to each indicator were abstracted from the medical record. The ATOM assessed outcomes in three domains: Symptom Severity, Functional Impairment (home, school, community and peers), and Consumer Perspectives (family burden and satisfaction with care) (Robbins et al., 2001). The diagnoses of depression and dysthymia were obtained using either parent or adolescent report on the Diagnostic Interview Schedule for Children–Version 4 (DISC; Fisher et al., 1997).

Procedure

Upon entering treatment, adolescents and their parents were recruited into the ATOM validation study by clinicians. Six months following baseline data collection, three research assistants and two doctorate level clinicians trained in medical record abstraction reviewed the charts and collected data on processes of care for adolescents with depression.

Results

Of the 208 participants, the majority were male (57%) and Caucasian (64%). Forty-nine (19.1%) adolescents were diagnosed with either major depressive disorder (MDD) or dysthymia by parent or adolescent DISC. By comparison, 25 (12%) of the adolescents were identified by the clinician in the medical record as having either disorder with an additional 22 (10.6%) diagnosed by the clinician as having a major depressive disorder not otherwise specified (MDD NOS). Agreement between the DISC and medical record for MDD or dysthymia diagnosis was poor (Kappa = .05 for adolescent DISC and .15 for parent DISC).

Adolescents with a positive MDD or dysthymia DISC diagnosis were categorized as "detected" if the clinician diagnosed MDD or dysthymia or "undetected" if the clinician did not diagnose MDD or dysthymia. (Adolescents diagnosed as MDD NOS by the clinician were excluded from this analysis.) Treatment variables were compared among these two groups and a group of adolescents who did not receive a positive DISC diagnosis of MDD or dysthymia. There were no significant differences among the three groups on documentation of suicide risk, substance abuse, and parent involvement in treatment planning or family or school involvement in treatment. Adolescents whose depression was detected were more likely to have been prescribed an SSRI (68.3%) or any antidepressant (71.4%) when compared to undetected depressed adolescents (13.3% SSRI and 20.0% any antidepressant) and non-depressed adolescents (16.2% SSRI and 22.2% any antidepressant), χ^2 (2, 208) = 17.7, $p < .01$, and χ^2 (2, 208) = 14.6, $p < .001$, respectively. There was also a significant difference among the groups on receipt of cognitive behavioral therapy; those with detected depression had the highest rates (27% non-depressed, 13.3% undetected depressed, 44.4% detected depressed). Despite treatment differences, outcomes were similar across the four groups.

Discussion

The findings suggest that if depression is detected in adolescents seeking mental health services, they may be more likely to receive evidence-based treatments such as CBT or SSRIs. Therefore, it may be advisable to implement routine depression screenings in specialty care settings to assist providers in treatment planning and follow-up. Because of the low percentage of adolescents receiving CBT, findings may also suggest that future research examine outcomes of care for a larger sample of

adolescents in which treatment fidelity and adherence is more closely monitored. Future research should examine patient and provider factors influencing the provision of evidence-based care and implement and evaluate interventions that address these barriers.

References

American Academy of Child and Adolescent Psychiatry. (1998). Practice parameters for the assessment and treatment of children and adolescents with depressive disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *37*(10) suppl: 63S-83S.

Cicchetti, D., & Toth, S. L. (1998). The development of depression in children and adolescents. *American Psychologist*, *53*, 221-241.

Fisher, P., Lucas, C., Shaffer, D., & The NIMH DISC Editorial Board. (1997). *The Diagnostic Interview Schedule for Children (PC-DISC),* 4th ed. New York: Joy & William Ruan Center to Identify and Treat Mood Disorders, Division of Child Psychiatry, Columbia University.

Goethe, J. W., Szarek, B. L., & Cook, W. L. (1988). A comparison of adequately vs. inadequately treated depressed patients. *Journal of Nervous & Mental Disorders*, *176*(8), 465-470.

Hoagwood, K. L. (2001). Evidence-based practice in children's mental health services: What do we know? Why aren't we putting it to use? *Emotional & Behavioral Disorders in Youth, 1*(4), 84-87.

Keller, M. B. (1988). Undertreatment of major depression. *Psychopharmacology Bulletin*, *24*: 75-80.

Kramer, T. L., Daniels, A. S., Zieman, G. L., Williams, C., & Dewan, N. (2000). Psychiatric practice variations in the diagnosis and treatment of major depression. *Psychiatric Services, 51*(3), 336-340.

Robbins, J., Taylor, J., Rost, K., Burns, B., Phillips, S., Burnam, M., & Smith, G. (2001). Measuring outcomes of care for adolescents with emotional and behavioral problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(3), 315-324.

Processes of Adolescent Mental Health Care: Demographic Variations and Relation to Outcomes

Terri L. Miller, Teresa L. Kramer, James M. Robbins, Susan D. Phillips, & Barbara J. Burns

Introduction

Quality of health care services has been conceptualized as the extent to which services "increase the likelihood of desired health outcomes and are consistent with current professional knowledge" (Lohr, 1990, p. 21). Quality improvement requires in part that processes of care associated with desired treatment outcomes be identified and implemented in routine care. Little progress toward establishing measures of treatment processes for children and adolescents has been made to date, and few published studies have focused on specific processes of care (e.g., types of pharmacological and psychosocial interventions) that have demonstrated efficacy and that might be most directly linked to outcomes. Consequently, little is known of the extent to which evidence-based treatment is delivered in routine care, is influenced by important client characteristics, or influences outcomes of care. The authors developed and tested a comprehensive set of quality indicators intended to assess concordance of routine care with standards derived from empirical evidence and expert consensus. We developed these indicators for use in conjunction with outcomes measures to facilitate quality improvement efforts within service delivery systems. This report examines processes of care and their relationship to outcomes in a sample of adolescents.

Method

Indicators were derived from a systematic review of the literature on practice guidelines for emotional and behavioral disorders in youth, randomized clinical trials of psychosocial and pharmacological interventions, relevant observational studies, and existing and proposed performance

measures for children and adolescents. Fifty-eight indicators covering assessment, acute, and continuation phases of treatment assessing global and disorder-specific aspects of care were developed (Kramer et al., 2001). Each indicator represents a specific activity that may occur with varying probabilities. Following performance measurement convention, rates of adherence to indicators can be calculated. Criteria algorithms designed to guide treatment record abstraction were developed, as well as a user's manual detailing the procedure for record review and the evidence/consensus base for all indicators (Miller, Kramer, Robbins, Taylor, & Burns, 2001). Materials were reviewed by an expert panel of clinicians and researchers, and informal feedback was obtained from consumers, providers, and policy makers as well.

The Adolescent Treatment Outcomes Module (ATOM; Robbins et al., 2001) is a set of self-report questionnaires designed to assess outcomes of care across multiple domains. The ATOM validation study included 256 adolescents 10 to 18 years old receiving care in specialty mental health settings at six sites in Arkansas and Texas. Each adolescent and a parent/caregiver completed the ATOM at entry into treatment and at six-month follow-up, as well as a set of validation measures including the Child Behavior Checklist (CBCL; Achenbach, 1991a) and the Youth Self Report (YSR; Achenbach, 1991b). Trained reviewers applied quality indicators to treatment records covering this six-month period for 210 participants. Two research team members completed duplicate reviews for 25 participants to permit assessment of interrater reliability of indicators. Ongoing feedback from reviewers resulted in modifications to the indicators, algorithms, and user's manual to address aspects of care not initially covered, to clarify procedures for review, and to specify criteria in further detail.

This report is based on results for 191 participants for whom both adolescent and parent report ATOM measures were completed at baseline and follow-up. This subsample was predominantly made up of younger (68% 10-14 years), male (54%), Caucasian (73%), urban (69%) adolescents. Approximately half (52%) were Medicaid-eligible, and participants were approximately equally distributed across income categories (22% less than $10,000; 26% $10,000-19,999; 26% $20,000-39,000; 25% $40,000 or more). The most common diagnoses included attention deficit hyperactivity disorder (ADHD; 32% of participants), disruptive behavior disorders (DBD; 38%), depressive disorders (46%), and anxiety disorders (8%). Eighty-six percent of participants had at least one of these diagnoses, and 53% had two or more comorbid diagnoses. The majority (65%) were admitted into treatment in outpatient settings. However, emergency, inpatient, and residential treatment (18%); partial hospitalization (8%); and home-based and school-based service settings (8%) were also represented.

Results

Percent agreement was unacceptably low (below 70%) for several indicators, which were dropped from further analyses. Results of interrater reliability analyses for indicators retained are presented in Table 1. Kappas were generally low to moderate, even where percent agreement was relatively good. As kappa is designed to correct for chance levels of agreement, it is influenced by base rates of observed phenomena. Consequently, its value can be misleading regarding actual levels of agreement between raters (Feinstein & Cicchetti, 1990).

Adherence rates are also presented in Table 1. Rates varied widely using fairly lenient scoring criteria, ranging from 6% of records containing evidence of informed consent to treatment (i.e., discussion of treatment options, risks, and benefits) to 82% containing evidence of family interventions (i.e., parent attendance at one or more treatment sessions beyond the initial assessment). Rates were low to moderate in general, being higher than 70% for only three of 14 indicators examined.

Results of chi square analyses of demographic variations in adherence rates are presented in Table 2. Results indicate few significant findings for either age or gender. However, they provide evidence to suggest some variation—although not systematic—by race/ethnicity, socioeconomic status, and urban

Table 1
Interrater Reliability of Indicators and Rates of Adherence to Standards

	Percent agreement	Kappa	Adherence rate
Global assessment indicators			
Participation of multiple informants	75	.36⁺	18
Screening for psychiatric comorbidity	71	.26	48
Screening for medical comorbidity	71	.04	58
Suicide screening	88	.65***	73
Violence screening	92	.78***	62
Substance use screening	78	.48**	73
Child abuse screening	75	.14	61
Global treatment indicators			
Informed consent	96	ª	6
Family intervention	83	.60**	82
Disorder-specific assessment indicators			
ADHD	88	.71*	48
DEP	80	.60*	61
Disorder-specific treatment indicators			
ADHD	86	.70*	67
DBD	100	1.00***	43
DEP	70	.40	66

Note: ADHD = Attention Deficit Hyperactivity Disorder. DBD = Disruptive Behavior Disorders. DEP = Depressive Disorders. Disorder-specific assessment indicators require documentation of at least three symptoms of the indicated disorder. Disorder-specific treatment indicators require behavior therapy or stimulant for ADHD; behavior therapy or appropriately prescribed antipsychotic or mood stabilizer for DBD; and cognitive-behavioral therapy or SSRI for depressive disorders.

***$p < .001$. **$p < .01$. *$p < .05$.
ª Could not be calculated as variable was a constant for rater #2 (i.e., indicator was scored as "absent" in all cases).

versus rural residence. Particularly noteworthy is the finding that documentation of assessments for violence and child abuse was more common for African American, Hispanic, and other non-Caucasian participants and Medicaid-eligible participants than for Caucasians who were not eligible for Medicaid. In addition, documentation of assessment for violence was less common for rural residents than it was for urban residents. Rates of documentation of appropriate disorder-specific interventions differed only in one instance; appropriate treatment of depression (i.e., either SSRI medication or cognitive-behavioral therapy) was less common for rural residents than for urban residents.

Results of correlational analyses of process and outcome variables (not shown) indicate only a few significant findings overall, and no meaningful pattern among them. To investigate the possibility that lack of measured change in clinical and functional status might account for this lack of findings, *t*-test analyses were conducted on baseline and follow-up scores for all ATOM symptom and impairment scales. Results (not shown) indicate significant mean decreases on all scales ($p < .001$). Similar analyses using the CBCL and YSR as outcome measures yielded similar results.

Discussion

Findings suggest significant gaps in global and disorder-specific processes of care for adolescents. Rates of adherence to evidence/consensus-based standards as determined by review of treatment records were lower than 70% for the majority of indicators examined. Although additional research is needed to confirm whether poor performance on these indicators truly represents deficiencies in quality of care or simply reflects inadequacies in documentation, findings strongly suggest the need for interventions to improve the quality of routine mental health care for children and adolescents. The fact that such limited adherence was found in specialty care settings suggests the possibility of even more substantial problems in other service sectors (e.g., general medical, juvenile/criminal justice) in which many youth receive care.

Table 2
Demographic Variations in Processes of Care

	Age	Gender	Race/Ethnicity
Global assessment indicators	χ^2	χ^2	χ^2
Participation of multiple informants	1.519	.125	.150
Screening for psychiatric comorbidity	.134	.752	11.917**[a]
Screening for medical comorbidity	.129	1.378	15.501***[a]
Suicide screening	.028	.398	3.419
Violence screening	.448	.021	12.717**[a]
Substance use screening	7.252**[d]	.086	4.495
Child abuse screening	2.534	.702	6.907*[a]
Global treatment indicators			
Informed consent	.139	.664	4.234
Family intervention	.520	.292	1.959
Disorder-specific assessment indicators			
ADHD	.092	2.064	.063
DBD	1.204	.031	.616
Disorder-specific treatment indicators			
ADHD	1.729	.854	.348
DBD	.234	.436	.872
DEP	.293	1.356	.534

	Income	Medicaid	Urban/Rural
Global assessment indicators			
Participation of multiple informants	4.232	.270	2.346
Screening for psychiatric comorbidity	6.146	1.988	7.635**[b]
Screening for medical comorbidity	1.952	.267	5.588*[b]
Suicide screening	1.623	.128	2.673
Violence screening	.273	7.387**[c]	6.871**[b]
Substance use screening	6.540	.283	.835
Child abuse screening	10.884*[e]	5.237*[c]	1.810
Global treatment indicators			
Informed consent	11.174*[f]	1.602	.084
Family intervention	3.617	2.543	.052
Disorder-specific assessment indicators			
Attention Deficit Hyperactivity Disorder	4.854	.021	.582
Depressive Disorders	4.915	8.724**[g]	.045
Disorder-specific treatment indicators			
Attention Deficit Hyperactivity Disorder	2.953	2.291	.014
Disruptive Behavior Disorders	2.133	.098	.944
Depressive Disorders	6.501	.108	5.799*[b]

Note: ADHD = Attention Deficit Hyperactivity Disorder. DBD = Disruptive Behavior Disorders. DEP = Depressive Disorders. Degrees of freedom = 1 for analyses of age, gender, Medicaid status, and urban vs. rural differences; 2 for analyses of racial/ethnic differences; 3 for analyses of income differences.

[a]More common for African Americans, Hispanics, and others than Caucasians.
[b]More common for urban residents.
[c]More common for those eligible for Medicaid.
[d]More common for 15- to 18-year-olds.
[e]Rates increase as income decreases.
[f]More common for those in income category $40,000 or more.
[g]More common for those not eligible for Medicaid.

***p < .001. **p < .01. *p < .05.

Results suggest that there may be variation in at least some domains of care by client race/ethnicity, socioeconomic status, and urban versus rural residence. Interestingly, certain findings were consistent with widespread beliefs (which may not be an accurate reflection of reality) regarding the prevalence of violence and child abuse across groups defined on the basis of these characteristics. Notably, the results provided little evidence of the demographic differences in disorder-specific treatment that have been found in recent studies of treatment for depressive and anxiety disorders (Wang, Berglund, & Kessler, 2000; Young, Klap, Sherbourne, & Wells, 2001) as well as a range of other health conditions (Fiscella, Franks, Gold, & Clancy, 2000) in adults. Given that these variables are likely confounded with treatment site to some degree in this sample, further analyses are warranted to rule out site variation as an explanation for the differences that were found.

Results suggest a general lack of meaningful relationships between processes and outcomes of routine mental health care for adolescents, consistent with findings from previous investigations demonstrating the difficulty in improving depression care for adults (see Rubenstein et al., 1999). The difficulty in demonstrating such relationships despite the known efficacy of many psychosocial and pharmacological interventions is likely to be overcome only with further research addressing issues such as treatment fidelity, treatment adherence, and the validity of medical record documentation as a measure of treatment processes. The potential criticism that this study assessed only the documentation of care provided at a single site for each child during the review period—and thus may not reflect treatment received concurrently from multiple providers—is a valid one, and may help to explain these results. However, designing interventions that will effect significant change in the quality of routine care remains a substantial challenge given the implication that improvement in processes of care may not necessarily be reflected in improved outcomes as measured by existing instruments.

References

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

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

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543-549.

Fiscella, K., Franks, P., Gold, M. R., & Clancy, C. M. (2000). Inequality in quality: Addressing socioeconomic, racial, and ethnic disparities in health care. *Journal of the American Medical Association, 283*(19), 2579-2584.

Kramer, T. L., Miller, T. L., Robbins, J. M., Taylor, J. L., Feldman, J. H., & Burns, B. J. (2001). *Arkansas quality indicators for child and adolescent mental health care*. Unpublished manuscript.

Lohr, K. N. (Ed.). (1990). *Medicare: A strategy for quality assurance* (Vol. 1). Washington, DC: National Academy Press.

Miller, T. L., Kramer, T. L., Robbins, J. M., Taylor, J. L., & Burns, B. J. (2001). *Arkansas quality indicators for child and adolescent mental health care: User's manual*. Unpublished manuscript.

Robbins, J. M, Taylor, J. L., Rost, K. M., Burns, B. J., Phillips, S. D., Burnam, M. A., & Smith, G. R. (2001). Measuring outcomes of care for adolescents with emotional and behavioral problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(3), 315-324.

Rubenstein, L. V., Jackson-Triche, M., Unutzer, J., Miranda, J., Minnium, K., Pearson, M. L., & Wells, K. B. (1999). Evidence-based care for depression in managed primary care practices. *Health Affairs, 18*(5), 89-105.

Wang, P. S., Berglund, P., & Kessler, R. C. (2000). Recent care of common mental disorders in the United States. *Journal of General Internal Medicine, 15*, 284-292.

Young, A. S., Klap, R., Sherbourne, C. D., & Wells, K. B. (2001). The quality of care for depressive and anxiety disorders in the United States. *Archives of General Psychiatry, 58*, 55-61.

Symposium Discussion

John Landsverk

The papers in this symposium represent an excellent example of cutting edge work in the development of functioning measures that show promise for use in real-world clinical settings. Several noteworthy aspects of this research can be highlighted. First, the scales were conceptually derived, short enough to be used in regular clinical practice, and capture the perspectives of both the adolescent and the adult caregiver. Second, this is an excellent example of work from the Arkansas research program that began development of these types of scales for adult populations and now has extended that work into the youth population. The downward extension from more advanced work in adult systems to child and adolescent service systems is rarely seen and should be applauded. Third, the factor analyses conducted show promising comparability in the scales from the two perspectives of the adolescent and the adult caregiver. Finally, the simplicity of the resulting three broad areas of role performance, relationships, and consequences to behavior suggest that the scales will be easily understood by a range of stakeholders involved in outcome systems within clinical services. The major issue confronting this research program is determining the logical steps and resource base for extending these studies into different regional and ethnic populations in order to produce an outcome package that can be used across the United States.

CONTRIBUTING AUTHORS

James M. Robbins, Ph.D.
Professor, Department of Pediatrics and Department of Psychiatry and Behavioral Sciences, University of Arkansas for Medical Sciences, Arkansas Children's Hospital, 800 Marshall St., Slot 512-26, Little Rock, AR 72202; 501-320-3300, fax: 501-320-1552; e-mail: robbinsjamesm@uams.edu

Susan D. Phillips, M.S.W.
Research Associate, Services Effectiveness Research Programs, Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Box 3454, Durham, NC 27710; 919-687-4686 ext. 251, fax: 919-687-4737; e-mail: sphillips@psych.mc.duke.edu

Teresa L. Kramer, Ph.D.
Assistant Professor, Department of Psychiatry and Behavioral Sciences, University of Arkansas for Medical Sciences, Centers for Mental Healthcare Research, Freeway Medical Tower, 5800 West 10th St., Suite 605, Little Rock, AR 72204; 501-660-7550, fax: 501-660-7543; e-mail: kramerteresal@uams.edu

Terri L. Miller, Ph.D.
Assistant Professor, Department of Psychiatry and Behavioral Sciences, University of Arkansas for Medical Sciences, Centers for Mental Healthcare Research, Freeway Medical Tower, 5800 West 10th St., Suite 605, Little Rock, AR 72204; 501-660-7583, fax: 501-660-7542; e-mail: millerterril@uams.edu

Barbara J. Burns, Ph.D.
Professor of Medical Psychology, Director, Services Effectiveness Research Program, Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Box 3454, Durham, NC 27710; 919-687-4676, fax: 919-687-4737; e-mail: bjb@geri.duke.edu

Scott N. Compton, Ph.D.
Assistant Professor, Duke University School of Medicine, Box 3454, Durham, NC 27710; 919-687-4676 ext. 251 ,fax: 919-687-4737 ext. 243; e-mail: sphillips@psych.mc.duke.edu

John Landsverk, Ph.D.
Child & Adolescent Services Research Center (CASRC), 3020 Children's Way MC 5033, San Diego, CA 92123; 858-966-7703 ext. 3508, fax: 858-966-7704; e-mail: jlandsverk@aol.com

The Ohio Scales Youth Form: A Self-Report Outcome Measure for Young Children

Kathy Dowell
Benjamin Ogles

Introduction

Competition for limited resources among publicly supported mental health agencies has created a boom to demonstrate the success of services offered. Consumers, as well as third party payers are requesting verification of the treatment effectiveness and cost efficiency as part of the demand for increased accountability among service providers. Consequently, the assessment of treatment outcomes has become an increasingly common and important part of services conducted by public agencies.

The impact of the outcome assessment movement is varied, however, as the field of children's mental health has lagged behind that of adult mental health (Maruish, 1999; Pratt & Moreland, 1996). Not only have fewer outcome studies been conducted with children, but also fewer psychometrically sound outcome measures, particularly self-report, are available for children. However, researchers now find that young children are able to provide valuable clinical information (e.g., substance abuse unknown to parents) despite the common practice of consulting parents as primary informants when the children are referred for mental health services (Kazdin & Heidish, 1984; Kenny & Faust, 1997; Weissman et al., 1987). Although large discrepancies often exist between the parent's and the child's report, it is not possible to judge at this point who is the more valid reporter; hence, it seems premature to exclude young children as valuable informants. Current research has identified specific biases that may weaken the accuracy of parent and teacher reports (Beitchman & Corradini, 1988; Pratt & Moreland, 1996; Weisz & Weersing, 1999), supporting the call for continued research on children's value as sources of information regarding their mental health status.

The purpose of this study was to expand the utility of an existing self-report outcome measure available for youth ages 12-18 by examining its psychometric properties for children ages 8 to 11. The Ohio Scales Short Form Problem Severity scale is a 20-item outcome measure including three parallel forms for the parent, agency worker, and child. The Problem Severity scale of the Youth Form used in this study assesses internalizing and externalizing clinical symptoms. Raters rank each item on a six-point scale, ranging from 0 (*not at all*) to 5 (*all the time*).

Method

Subjects included 32 children ages 8 through 11 receiving outpatient mental health services in a mid-sized Ohio city. The comparison sample consisted of 97 children ages 8 to 11, with no prior history of receiving mental health services. Chi square analysis indicated that youth in the samples differed on several demographic variables including race, parent's education level, grade retainment, and socioeconomic status. Participation for the children included completion of the Behavioral Assessment Scale for Children (BASC; Reynolds & Kamphaus, 1998), a multi-informant diagnostic measure used as a validation tool, and the Ohio Scales (Ogles, Melendez, Davis, & Lunnen, 2000). Following a one-week period, the entire clinical sample and a randomly selected a subset from the comparison sample were readministered the Ohio Scales to examine test-retest reliability.

Results

Reliability

Across all samples, internal consistency ratings are high ($N = 141$, alpha = .88), indicating items of the Ohio Scales are part of the same underlying construct: problem severity. Internal consistency ratings remained high within samples, suggesting the items of the Ohio Scales remain part of the same construct whether measured within the comparison ($n = 87$, alpha = .84) or clinical sample ($n = 54$,

alpha = .89). Test-retest correlations across (*r* = .719) as well as within each sample (comparison *r* = .681; clinical *r* = .675) indicate good reliability.

Validity

Concurrent validity was demonstrated by the significant correlation between the children's scores on the Emotional Symptoms Index of the BASC and the total score on the Ohio Scales across all groups, *r* (117) = .617, *p* < .000. Concurrent validity was examined separately within the comparison and clinical samples. The BASC was significantly positively related to the Ohio Scales within the comparison sample, *r* (85) = .678, *p* < .000. However there was no relationship between the Ohio Scales and the BASC within the clinical sample, *r* (32) = .18, *p* > .05.

An independent samples *t*-test indicated that the mean clinical sample score on the Ohio Scales (*M* = 29.24, *SD* = 20.05) was significantly higher (i.e. more symptomatic) than the comparison sample (*M* = 16.99, *SD* = 11.89), *t* (75) = 4.09, *p* < .000, providing evidence for the measure's construct validity.

Race Effects

Because the comparison and the clinical samples differed in their racial composition, ethnicity was examined to assess the degree to which it contributed to group differences between comparison and the clinical samples on the Ohio Scales. There were no differences between the Caucasian subjects (*M* = 27.22, *SD* = 19.47) and all other minority subjects (*M* = 33.28, *SD* = 21.12) on the Ohio Scales, *t* (25) = -1.05, *p* > .05, suggesting the differences between the comparison and clinical samples on the Ohio Scales are not likely due to ethnicity, but rather to the child's clinical status.

In order to minimize the effect of the significant demographic differences between the comparison and clinical samples, a matched subsample based on age, gender and level of parent education was included in an independent samples t-test between the comparison and clinical samples on the BASC and the Ohio Scales. Similar to the unmatched samples, the mean Ohio Scales score of the clinical sample (*M* = 29.57, *SD* = 17.26) was significantly higher than the comparison sample (*M* = 13.76, *SD* = 8.11), *t* (40) = -3.80, *p* < .000. It appears that when parent education was controlled for, the BASC was unable to distinguish the comparison sample with the clinical sample, while the Ohio Scales demonstrated significant differences. This suggests that differences in mean Ohio Scales scores were not a function of sample differences.

Age Effects

The entire sample was divided into four age groups (i.e., 8, 9, 10, 11 years of age) to examine the reliability and validity within varying age groups (see Table 1). Internal consistency remains acceptable across all age groups. Test-retest reliability estimates are all significant, with the correlation coefficient increasing fairly consistently from younger to older children. Concurrent validity was demonstrated for all age groups by the correlation between the BASC and the Ohio Scales, although there does not appear to be any pattern related to increasing age as was found with reliability.

Table 1
Reliability and Validity Data per Age Group

	Internal Consistency		Test-Retest		BASC vs. Ohio Scales	
	N	α	*N*	*r*	*N*	*r*
8 year olds						
Comparison	19	.85	8	.212	18	.816**
Clinical	14	.91	12	.596*	7	.078
Combined	3	.89	20	.591**	25	.601**
9 year olds						
Comparison	24	.80	9	.883**	24	.787**
Clinical	15	.81	13	.435	8	-.022
Combined	39	.82	22	.581**	32	.442
10 year olds						
Comparison	28	.88	8	.889**	28	.646**
Clinical	9	.91	9	.802**	3	-.859
Combined	37	.90	17	.786**	31	.555**
11 year olds						
Comparison	17	.82	6	.884**	16	.679**
Clinical	18	.93	16	.845**	7	.888**
Combined	35	.93	22	.867**	23	.516*

* *p*<.05 ** *p*<.01

In order to examine the construct validity within each age group, an independent samples *t*-test was conducted, comparing the mean Ohio Scales score of the clinical sample to that of the comparison sample for 8, 9, 10, and 11 year-old children (see Table 2). It appears that the Ohio Scales is significantly better at discriminating between children of the clinical sample and children of the comparison sample when the child is older than 8 years of age.

Table 2
***T*-test Within Age Groups**

	Comparison			Clinical			
	N	M	SD	N	M	SD	t
8 year olds	19	20.89	14.55	14	26.93	21.37	-.97
9 year olds	24	17.42	11.71	15	27.47	16.19	-2.25[*]
10 year olds	28	15.25	11.96	9	28.22	21.30	-2.31[*]
11 year olds	17	15.29	8.82	18	31.06	22.55	-2.69[*]

[*]denotes *p*<.05

Discussion

Internal consistency analysis indicates that the items of the Ohio Scales all function as part of the same construct: problem severity. The test-retest reliability coefficients of the 8 to 11 year olds, although slightly lower than the adolescent norming samples (*r* = .88 and .72), are good, demonstrating that the Ohio Scales reliably measures the same construct when administered on two separate occasions within a one-week interval (Ogles, et al., 2000).

Construct validity was demonstrated by the significant correlation between the BASC and the Ohio Scales across both samples. Although the BASC was not correlated with the Ohio Scales within the clinical sample, this is most likely due to the small sample size. Evidence supporting the validity of the Ohio Scales within a heterogeneous group (i.e. comparison and clinical samples combined) holds more external validity insofar as children who are referred for mental health services include those that are clinically impaired as well as children exhibiting sub-clinical impairment. The Ohio Scales also demonstrated adequate construct validity by distinguishing children currently receiving mental health services from children who reported no prior mental health referrals.

The age effects analysis supports the use of the Ohio Scales for children 9 and older. Reliability estimates increased gradually from younger to older children. Because validity depends upon the scale's consistency to accurately measure a stable construct, it follows that construct validity is also strongest for children 9 and up. These results support the use of the Ohio Scales Youth form, a brief, practical, psychometrically sound, self-report outcomes measure, for children between the ages of 9 to 11. Based on the results of the age-group analysis, caution should be used when interpreting outcome data gathered from children under the age of 9, as the validity of their self-report has not been determined by this study.

The findings of this study lend support to previous studies that have emphasized the unique information provided by young children and recommended the inclusion of children's self-report data in addition to parent and teacher reports (Kazdin & Heidish, 1984; Kenny & Faust, 1997; Weissman et al., 1987). The ultimate goal of this and other related research is to improve the quality of mental health services available for children by encouraging the regular inclusion of self-report information, as well as to encourage the comparison of child, parent, and teacher/counselor reports in the assessment of treatment outcomes.

References

Beitchman, J. H., & Corradini, A. (1988). Self-report measures for use with children: A review and comment. *Journal of Clinical Psychology, 44*, 477-490.

Kazdin, A. E., & Heidish, I. E. (1984). Convergence of clinically derived diagnoses and parent checklists among inpatient children. *Journal of Abnormal Child Psychiatry, 12*, 421-436.

Kenny, M. C., & Faust, J. (1997). Mother-child agreement on self-report of anxiety in abused children. *Journal of Anxiety Disorders, 11*, 463-472.

Maruish, M. E. (1999). Introduction. In M. E. Maruish (Ed.), *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment, Second Edition,* (pp. 1-39). New Jersey: Lawrence Erlbaum Associates.

Ogles, B. M., Melendez, G., Davis, K. C., & Lunnen, K. M. (2000). *The Ohio Scales: A Manual*. Ohio University.

Pratt, S. I., & Moreland, K. L. (1996). Introduction to treatment outcome: Historical perspectives and current issues. In S. I. Pfeiffer (Ed.), *Outcome Assessment in Residential Treatment* (pp. 1-27). New York: The Hawthorn Press, Inc.

Reynolds, C. R. & Kamphaus, R. W. (1998). *Manual for the Behavior Assessment System for Children*. Circle Pines, MN: American Guidance Services, Inc.

Weisz, J. R., & Weersing, V. R. (1999). Developmental outcome research. In W. K. Siverman & T. H. Ollendick (Eds.), *Developmental Issues in the Clinical Treatment of Children* (pp.457-469). Boston: Allyn & Bacon.

Weissman, M. M., Gammon, G. D., John, K., Merikangas, K. R., Warner, V., Prusoff, B. A., & Sholomskas, D. (1987). Children of depressed parents: Increased psychopathology and early onset of major depression. *Archives of General Psychiatry, 44*, 847-853.

CONTRIBUTING AUTHORS

Kathy Dowell, M.S.
Co-Principal Investigator; 740-593-1707, fax: 740-593-0579;
e-mail: kd289091@ohiou.edu

Benjamin Ogles, Ph.D.
Co-Principal Investigator, 740-593-1077, fax: 740-593-0579; e-mail: ogles@ohiou.edu

All Authors: Department of Psychology, Ohio University, 200 Porter, Hall Athens,
OH 45701