Journal of Consulting and Clinical Psychology
2008, Vol. 76, No. 5, 711–722

0022-006X/08/$12.00 DOI: 10.1037/0022-006X.76.5.711

Diagnostic Agreement Predicts Treatment Process and Outcomes in Youth Mental Health Clinics

Amanda Jensen-Doss
Texas A&M University

John R. Weisz
Judge Baker Children's Center, Harvard University

Several studies have documented low rates of agreement between clinician- and researcher-generated diagnoses. However, little is known about whether this lack of agreement has implications for the processes and outcomes of subsequent treatment. To study this possibility, the authors used diagnostic agreement to predict therapy engagement and outcomes for 197 youths treated in 5 community mental health clinics. Diagnostic agreement predicted better therapy engagement, with the agree group having fewer therapy no-shows and cancellations and a decreased likelihood of therapy dropout. Additionally, support for a link between agreement and treatment outcomes was found, as the agree group obtained larger reductions in parent-reported internalizing problems during treatment. These findings suggest that diagnostic accuracy may be an important precursor to successful treatment and highlight the importance of future research to find ways to incorporate standardized diagnostic procedures into clinical care settings.

Keywords: diagnostic agreement, youth mental health services, treatment outcomes, therapy engagement

Diagnosis plays a central role in the treatment of psychological distress. The assignment of a diagnosis from the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) is often the first step in treatment planning, because many clinics and third-party payers require a *DSM* diagnosis for authorization of treatment. In addition, diagnoses play a role in determining eligibility for particular services (e.g., psychotherapy or wraparound services) within many mental health systems. Diagnoses also can aid clinicians in treatment planning, as many psychosocial and pharmacological interventions have been designed for and tested with particular diagnostic groups.

In clinical practice settings, diagnoses have traditionally been assigned through unstructured interviews, in which clinicians interview clients and/or family members to gather diagnostic information. Surveys indicate that the unstructured clinical interview is the assessment method used most often by clinicians (e.g., Cashel, 2002) and that it is often the only method used to generate diagnoses (Anderson & Paulosky, 2004).

Despite the frequent use of unstructured diagnostic interviews, questions have been raised about whether they generate valid diagnoses. Research has shown that clinicians are susceptible to several information-gathering biases that could influence the diagnostic process, such as deciding on a diagnosis before collecting all relevant data, seeking information to confirm that diagnosis while ignoring conflicting information, combining information in ways that do not match *DSM* criteria, and making decisions based on assumptions about gender, ethnicity, and/or age (Angold & Fisher, 1999; Garb, 1998, 2005). Many researchers have compared diagnoses assigned by clinicians with those generated through use of research diagnostic interviews conducted in accordance with standard rules for information gathering. These studies were combined in a recent meta-analysis (Rettew, Doyle, Achenbach, Dumenci, & Ivanova, 2006). The overall agreement between clinician- and researcher-generated diagnoses for children in Rettew et al. was $\kappa = .15$, which is considered "poor" agreement (Landis & Koch, 1977).

This lack of agreement does not necessarily mean in and of itself that the research instruments are correct and the clinicians are incorrect. However, when indicators of validity are used for the comparison, research instruments have been found to be more valid than are clinician-generated diagnoses. For example, Basco et al. (2000) found that research instruments had higher concordance than did clinician diagnoses with "gold standard diagnoses," which were generated by experts who reviewed all available information—including medical records, the clinician diagnoses, and the results of the research interview—and had the opportunity to ask clients additional follow-up questions before rendering a final diagnosis. Other studies that compared both sets of diagnoses to external indicators of validity, such as independent reports of daily behavior (Jewell, Handwerk, Almquist, & Lucas, 2004) and measures of constructs expected to be associated with the diagnosis assessed (e.g., impaired functioning; Tenney, Schotte, Denys, van Megen, & Westen-

Amanda Jensen-Doss, Departments of Educational Psychology and Psychology, Texas A&M University; John R. Weisz, Judge Baker Children's Center, Harvard University.

The study was supported by National Institutes of Mental Health (NIMH) Research Grant R01 MH 49522, NIMH Senior Research Scientist Award K05 MH01161, and the Research Network on Youth Mental Health, sponsored by the John D. and Catherine T. MacArthur Foundation.

Correspondence concerning this article should be addressed to Amanda Jensen-Doss, Department of Educational Psychology, 4225 TAMU, College Station, TX 77845. E-mail: ajensendoss@tamu.edu

berg, 2003), have suggested higher validity of the research interviews.[1]

Given the important roles diagnoses play in clinical settings, these findings are cause for concern. They suggest that many clients who are being seen in clinical practice settings may be receiving the wrong diagnoses. Theoretically, there are many mechanisms by which this misdiagnosis could impact a client's course of treatment. First, poor diagnostic accuracy could negatively impact the process of treatment. For example, many authors have argued that agreement on therapy goals is an essential component of a strong working alliance during therapy (e.g., Bordin, 1979; Horvath & Luborsky, 1993); this argument suggests that misdiagnosis could negatively impact the therapeutic alliance. Given the significant association between therapy alliance and treatment engagement (e.g., Hawley & Weisz, 2005), one might hypothesize that misdiagnosis would also be associated with decreased therapy engagement (e.g., increased numbers of session no-shows and cancellations) and increased likelihood of therapy dropout. Second, misdiagnosis could negatively impact the outcomes of treatment. Treatment that is focused on the wrong diagnosis might well fail to address the correct diagnosis and might thus lead to worse treatment outcomes. Moreover, therapy outcomes could be adversely affected if misdiagnosis undermines therapy alliance and engagement, given the importance of alliance and engagement to treatment success (e.g., McKay & Bannon, 2004; Shirk & Karver, 2003). In addition, many psychotherapies and medications are developed and tested in research studies in which structured diagnostic interviews are used to determine study eligibility. It is therefore possible that clients diagnosed through unstructured interviews could be assigned to an inappropriate course of treatment that would result in worse treatment outcomes.

Surprisingly, however, little is known about whether an accurate diagnosis is actually important for successful treatment. Many efforts to improve the quality of psychological services, such as the empirically supported treatments movement (e.g., Chambless et al., 1996, 1998), are based on the assumption that treatment outcomes can be improved by matching evidence-based practices to clients with a particular *DSM* disorder. However, the assumption that treatment should be matched to diagnoses is largely untested, and the treatment utility of diagnoses (i.e., whether having accurate diagnostic assessment data actually improves treatment outcomes) is not well established (Nelson-Gray, 2003).

The link between accurate *DSM* diagnosis and treatment outcomes has received more attention in the primary care literature than in the mental health literature. In primary care settings, studies have focused on accurate detection of adult depression, with "accuracy" conceptualized as agreement between the provider-assigned diagnoses and diagnoses generated by a standardized research instrument. Not surprisingly, accurate detection of depression has been found to impact the process of treatment by increasing the likelihood of antidepressant prescription (e.g., Berardi et al., 2005); however, findings are mixed as to whether accurate detection improves treatment outcomes (Pini, Perkonnig, Tansella, & Wittchen, 1999; Tiemens, Ormel, & Simon, 1996). Although interesting, these findings from primary care settings, in which undetected cases likely receive no mental health treatment, do not directly address the question of accurate diagnosis in mental health settings, in which all clients receive some form of mental health treatment.

To our knowledge, only two studies have tested the impact of diagnostic accuracy within mental health settings, with "accuracy" operationalized as agreement between clinician- and researcher-generated diagnoses. Utilizing a sample of adolescent inpatients, Pogge et al. (2001) examined the implication of agreement on diagnoses of manic episodes. When they compared treatment process and outcomes for the youths with mania for whom chart and research diagnoses agreed and for the youths diagnosed with mania in the charts only, they found no impact of agreement on length of stay, readmissions, or post-treatment functioning measures. However, they did find that the agree group had greater reductions in depression symptoms after treatment. Kramer, Robbins, Phillips, Miller, and Burns (2003) examined the impact of accurate detection of substance use disorders among adolescents who were seeking mental health services. They found that adolescents with undetected substance use were less likely to receive substance use services and were more likely to experience legal problems at a 6-month follow-up than were adolescents with detected substance use, although the two groups did not differ on other indices of symptom and functioning outcomes.

These two studies suggest that accurate diagnosis may play a small but substantive role in treatment process and outcome. However, there is a need for additional studies that focus on a broader range of diagnoses. It is possible that agreement is more important for diagnoses with a high level of impact on functioning, such as mania and substance use, but that agreement on other diagnoses might not be as important. Given the high prevalence of other diagnoses, such as depression, disruptive behavior disorders, attention-deficit/hyperactivity disorder, and anxiety disorders in youth mental health settings (e.g., Garland et al., 2001), research should ascertain whether accuracy of these diagnoses impacts treatment process and outcomes.

The present study was designed to address this gap through examination of the implications of diagnostic agreement—the assignment of chart diagnoses that match the results of a structured diagnostic interview—on treatment process and outcomes among children seeking mental health services in community mental health clinics. Using a sample in which low levels of diagnostic agreement had previously been documented (Jensen & Weisz, 2002), we tested whether diagnostic agreement predicted treatment processes and outcomes. To better understand the extent to which diagnostic agreement might be important, we examined agreement on both the client's primary diagnosis, as identified by the clinician, and all diagnoses assigned by either source. We hypothesized that diagnostic disagreement would be associated with changes in therapy process, as evidenced by decreased engagement (shorter time in treatment, increased session cancellations and no-shows, and increased likelihood of therapy dropout) and by increased likelihood of ending treatment due to problems in the therapeutic

[1] A recent meta-analysis (Aegisdottir et al., 2006) that compared diagnoses based on clinical judgment with those generated through statistical methods found mathematical methods to be more efficient, but not more accurate, than were clinicians. However, because this meta-analysis included only studies in which the two forms of decision making were based on the same information, it did not take into account threats to reliability and/or validity that might result from biases during the information-gathering stage of diagnosis.

alliance. We also hypothesized that diagnostic disagreement would be associated with worse treatment outcomes, as indicated by smaller improvements in parent- and child-reported symptoms during treatment and worse symptom outcomes after the end of treatment. Finally, we tested whether any associations between diagnostic disagreement and outcomes were mediated by the therapy process variables and conducted secondary analyses to clarify the role of diagnostic agreement in therapy processes and outcomes.

Method

Participants

Participants were families of children 7–17 years of age who were seeking treatment in five outpatient community mental health clinics in Southern California. All families were participants in a previous study on agreement between researcher- and clinician-generated diagnoses (Jensen & Weisz, 2002), which was part of a larger study on the processes and outcomes of clinical care in community mental health clinics. Participants were recruited at the time of their clinic intake. To be included in the present analyses, children had to have received a diagnosis based on the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev.; *DSM–III–R;* American Psychiatric Association, 1987) at their clinic intake, and their primary intake diagnosis had to fall into one of five categories assessed by the project's research diagnostic instrument (see *Measures,* below). Additionally, they had to have participated in at least one of the project's follow-up assessments (see *Procedures,* below).

The sample from our original study on diagnostic agreement consisted of 245 children (Jensen & Weisz, 2002). Of these, 34 were excluded from the current analyses due to lack of follow-up data, 6 were excluded because their treatment ended after their study participation, which precluded estimations of their change during treatment (see *Data Analysis,* below), and 8 were excluded because their primary intake diagnosis did not fall into one of the five categories. These exclusions resulted in a sample size of 197 children. The sample included 132 boys (67.0%) and 65 girls (32.8%) who ranged from 7 to 17 years of age ($M = 11.10$ years, $SD = 2.58$). Of those sampled, 55.8% were Caucasian, 12.1% ($n = 24$) were African American, 14.6% ($n = 29$) were Latino, 2.5% ($n = 5$) were Asian/Pacific Islander, and 12.1% ($n = 24$) were mixed or other ethnicity (5 participants did not provide information on ethnicity). Fifty-five percent ($n = 108$) of the parents had completed at least 1 year of college. Two thirds of the sample ($n = 132$) had an annual family income of less than $30,000 per year. The children in the sample differed from the 48 excluded children on parental education and ethnicity, in that the included children had more highly educated parents and were more likely to be Caucasian. There were no differences between the two groups on other demographic characteristics, initial parent- or child-rated symptom levels, or therapist-rated functioning.

Procedures

During clinic intake, parents received a description of the study and provided consent to be contacted for an interview. Study recruitment took place from October 1991 through September 1995. At the pretreatment (Time 1) interview, parents and children provided written consent and demographic data; they also separately completed a battery of measures (the measures used in the present analyses are described below). The Time 1 assessment took place as soon as it could be scheduled after the clinic intake appointment ($M = 23.4$ days after intake).[2] During the interviews conducted 6 months (Time 2), 1 year (Time 3), and 2 years (Time 4) after intake, parents and children again provided written consent and completed measures of child symptoms. Following the Time 4 interview, clinic records were reviewed. The study was conducted in compliance with the institutional review boards of the University of California, Los Angeles, and the participating clinics.

Participants received the routine treatment procedures delivered through their respective clinics. Participants were treated by a total of 72 therapists. Degree data were available for 45 of the treating therapists, who were primarily master's level (48.9%, $n = 22$), followed by doctoral level (33.3%, $n = 15$) and bachelor's level (e.g., social work interns; 17.8%, $n = 8$).

Measures

Diagnostic Interview Schedule for Children (DISC). In the research assessment, child diagnoses were obtained with the DISC 2.3 (Shaffer et al., 1993). A highly structured interview consisting of a series of yes/no questions, it was administered by clinical psychology graduate students who had received the standard DISC interviewer training program designed by the DISC development team at Columbia University. They entered the answers into a computer algorithm to generate diagnoses based on *DSM–III–R* criteria (American Psychiatric Association, 1987) for the five primary diagnostic categories relevant to children: anxiety disorders (panic disorder, obsessive-compulsive disorder, separation anxiety disorder, overanxious disorder, avoidant disorder of childhood or adolescence, elective mutism, simple phobia, generalized anxiety disorder, social phobia, and agoraphobia); depressive disorders (major depressive disorder and dysthymia); disruptive behavior disorders (conduct disorder and oppositional defiant disorder); attention-deficit/hyperactivity disorders; and elimination disorders (enuresis and encopresis).[3] The numbers of children who received diagnoses falling into each of these categories are presented in Table 1. Although the clinicians recorded their diagnoses as being either primary or secondary, the DISC does not provide decision rules for ranking diagnoses in this manner.

Test–retest reliability for the parent version of the DISC has been shown to be good to excellent, with retest correlations ranging from .77 to .87 (Schwab-Stone et al., 1993). Interrater reliability has been shown to be high, with intraclass correlations of 1.00 for all of the major diagnoses except major depressive episode, which showed intraclass correlations of .66 (Shaffer et al., 1993). Investigations have established criterion validity of the DISC when

[2] Our original analyses (Jensen & Weisz, 2002) indicated that the amount of time elapsed between the clinic intake and the research assessment was not related to diagnostic agreement.

[3] The DISC also assessed tic disorders (Tourette's disorder, chronic motor or vocal tic disorder, and transient tic disorder). However, our original analyses indicated that the clinicians did not assign tic disorder diagnoses to any clinic clients (Jensen & Weisz, 2002), which precluded their use in the present analyses.

Table 1
Diagnostic Data for the Sample

Diagnostic cluster	*N* (%) assigned by DISC	*N* (%) assigned as a primary chart diagnosis	*N* (%) assigned as a primary or secondary chart diagnosis	κ for agreement between primary chart diagnoses and DISC	κ for agreement between all clinician diagnoses and DISC
Anxiety disorders	92 (46.7)	20 (10.2)	23 (11.7)	.078	.070
Depressive disorders	50 (25.4)	62 (31.5)	69 (35.0)	−.045	.009
Disruptive behavior disorders	90 (45.7)	104 (52.8)	122 (61.9)	.186**	.164*
ADHD	68 (34.5)	15 (7.6)	22 (11.2)	.157**	.166**
Elimination disorders	17 (8.6)	2 (1.0)	6 (3.0)	.088*	.408***

Note. DISC = Diagnostic Interview Schedule for Children; ADHD = attention-deficit/hyperactivity disorder.
* $p < .05$. ** $p < .01$. *** $p < .001$.

compared with other structured or semistructured interviews (Schwab-Stone et al., 1996) and with "expert clinician" consensus diagnoses (Fisher et al., 1993).

Of the adults interviewed, 90% were the child's mother figure, 7% were the child's father figure, 2% were another female relative, and 1% were a female nonrelative. Because child-report DISC diagnoses, unlike parent-report diagnoses, have shown only fair-to-poor test–retest reliability for several diagnostic categories (Schwab-Stone et al., 1993) and because poor reliability limits the level of researcher–clinician agreement that can be obtained, we relied on the parent-report DISC.

Clinician-generated diagnoses. Chart diagnoses were the *DSM–III–R* diagnoses generated for each child by the intake clinician assigned to that child, as a part of normal clinic intake, assessment, and treatment-planning procedures. A total of 53 clinicians conducted the clinic intakes. The intake clinicians were primarily master's level (64.2%, $n = 34$), followed by doctoral level (20.8%, $n = 11$) and bachelor's level (15.1%, $n = 8$). Degree information was not available for one intake therapist.

Per clinic policies, all diagnoses assigned by clinicians were classified as either primary or secondary. Clinicians could place multiple diagnoses in each category. Clinician diagnoses were grouped into the same five categories used for the DISC diagnoses. However, because the clinicians used some diagnoses that fell into these categories but were not formally assessed by the DISC, the clinician diagnostic categories included some additional diagnoses. The anxiety disorders category included the diagnosis of adjustment disorder with anxiety; the depressive disorders category included adjustment disorder with depressed mood and depressive disorder, not otherwise specified. Finally, adjustment disorder, disturbance of conduct, pyromania, and intermittent explosive disorder were included in the disruptive behavior disorders category. Rates of clinician diagnoses falling into each category and kappa statistics for agreement between clinician and DISC diagnoses are reported in Table 1. All kappas fell into Fleiss's (1981) "poor" category ($\kappa < .40$), except for agreement on elimination disorders, which fell in the "fair-to-good" range ($\kappa s = .40–.74$).

Symptom outcomes: Child Behavior Checklist (CBCL; Achenbach, 1991a) and Youth Self Report (YSR; Achenbach, 1991b). These widely used, standardized parent-report (CBCL) and child-report (YSR) symptom scales have 118 items, rated on a 3-point scale (0–2). Extensive evidence of the reliability and validity of these measures exists (cf. Achenbach, 1991a, 1991b).[4] Because we were interested in predicting longitudinal change over time and development, we used raw scores on the scales, rather than *t* scores adjusted for age and gender, in the analyses (Achenbach, 1991a, 1991b). The present analyses utilized the Internalizing broadband scores, Externalizing broadband scores, and Total Problems scores from the CBCL and YSR as the parent- and child-report outcome data. Descriptive statistics for the CBCL and YSR can be found in Table 2. Although raw scores were used in the analyses, *t* scores are provided to facilitate comparison to other samples.

Process variables: Clinic records and Reasons for Ending Therapy Questionnaire. Each child's clinic record was examined and several process variables were coded, among them, the number of session cancellations ($M = 2.44$, $SD = 3.82$), the number of session no-shows (i.e., failure to attend a session without notifying the clinician; $M = 1.55$, $SD = 2.58$), the number of weeks the child spent in treatment ($M = 29.31$, $SD = 26.11$), and therapy dropout (whether or not treatment was terminated with concurrence of the clinician; available for 178 children, % dropout = 59.4). One master's level and two bachelor's level research assistants coded the data. Forty records were randomly selected and coded by all coders to establish interrater reliability. Mean Pearson correlations were .83 for cancellations, .91 for no-shows, and .97 for treatment length, and the mean Cohen's kappa (Cohen, 1960) for therapy dropout was .72. All coders were blind to the study hypotheses.

Finally, parent perceptions of therapist behavior that might have been impacted by diagnostic agreement were obtained from the Reasons for Ending Treatment Questionnaire (RETQ; Garcia & Weisz, 2002). The RETQ is a 41-item measure that assesses the degree to which various factors might have led the family to end treatment. The measure asks parents to indicate whether each factor was not important, somewhat important, or very important

[4] Previous analyses published on the present sample have examined the appropriateness of using YSR data for children who were below 11 years of age, the age for which the YSR is normed (Yeh & Weisz, 2001). These analyses suggested comparable coefficient alphas for the YSR scale scores for children below age 11 ($M = 0.76$, $SD = 0.15$) and children age 11 and older ($M = 0.79$, $SD = 0.10$). Six-month test–retest coefficients were comparable for the younger ($M = 0.76$, $SD = 0.15$) and older ($M = 0.76$, $SD = 0.15$) groups. Because internal consistency alphas and test–retest reliability coefficients were very similar for the two groups, it was deemed appropriate to include the YSR data from our younger participants.

Table 2
Descriptive Statistics for CBCL and YSR Raw and T Scores

Measure	Time 1	Time 2	Time 3	Time 4
	M (*SD*)[a]	*M* (*SD*)[b]	*M* (*SD*)[c]	*M* (*SD*)[d]
Raw score				
CBCL Internalizing	16.31 (9.18)	11.67 (8.04)	9.85 (7.63)	9.34 (8.48)
CBCL Externalizing	22.91 (11.36)	18.10 (11.15)	16.31 (11.13)	16.63 (11.00)
CBCL Total Problems	59.77 (27.03)	45.85 (25.60)	39.87 (25.05)	39.51 (27.27)
YSR Internalizing	18.49 (10.15)	14.72 (8.76)	13.42 (8.37)	12.11 (9.27)
YSR Externalizing	17.73 (9.08)	14.85 (9.49)	13.72 (8.47)	13.93 (9.04)
YSR Total Problems	61.82 (27.39)	51.62 (25.48)	47.09 (24.36)	45.01 (26.97)
t score				
CBCL Internalizing	64.41 (10.41)	58.40 (11.08)	55.79 (11.17)	54.01 (12.31)
CBCL Externalizing	65.56 (10.48)	60.88 (11.09)	58.56 (12.01)	59.29 (11.54)
CBCL Total Problems	66.73 (9.33)	61.20 (10.54)	58.40 (11.37)	58.01 (11.71)
YSR Internalizing	58.58 (11.09)	54.21 (10.72)	52.70 (10.83)	50.64 (11.60)
YSR Externalizing	58.09 (11.97)	54.34 (12.55)	52.99 (11.58)	53.30 (12.02)
YSR Total Problems	60.59 (11.22)	56.24 (11.21)	54.26 (11.25)	52.93 (12.05)

Note. CBCL = Child Behavior Checklist; YSR = Youth Self Report.
[a] CBCL, $n = 197$; YSR, $n = 182$. [b] CBCL, $n = 168$; YSR, $n = 136$. [c] CBCL, $n = 156$; YSR, $n = 137$. [d] CBCL, $n = 160$; YSR, $n = 122$.

in their decision to end treatment. The RETQ was completed by 141 parents after their children ended treatment. For the present analyses, the therapeutic relationship factor was used as a measure of whether families ended treatment due to a poor therapeutic alliance. This factor had good internal consistency in the present sample ($\alpha = .92$).

Data Analysis

Definition of agreement. Before initiating data analyses, we computed two diagnostic agreement variables for each participant. First, we computed a variable that indicated agreement on the child's primary diagnostic category. Using the diagnostic categories described in the *Measures* section above, we coded this variable as 0 (agree) if the child's primary chart diagnosis was in a category that was positive on the DISC ($n = 94$, 47.7%) and as 1 (disagree) if the child did not meet criteria for his or her primary chart diagnostic category according to the DISC ($n = 103$, 52.3%). Second, we computed a variable that indicated agreement on all of the child's diagnostic categories. This variable was coded as 0 (agree) if all diagnostic categories indicated in the charts were positive on the DISC and vice versa ($n = 19$, 9.6%) and as 1 (disagree) if any chart diagnostic categories were negative on the DISC or if the child met criteria for a DISC diagnostic category that was not listed in the charts ($n = 178$, 90.4%). Use of these two variables allowed us to examine the role of agreement on the primary diagnosis or the diagnosis most likely to have impacted the child's treatment. It also allowed us to examine the role of agreement on all diagnoses and to account for problems such as the clinicians failing to detect a comorbid diagnosis.

Statistical analyses. Because the data gathered had a hierarchical structure, with repeated measures nested within clients, who themselves were nested within therapists and within clinics, we used multilevel or hierarchical linear modeling (HLM) with the HLM 6.02a software package (Raudenbush, Bryk, & Congdon, 2005) to conduct the analyses. Because the present sample consisted of only five clinics, we decided to analyze the treatment outcome data using three-level hierarchical models (with repeated measures at Level 1, children at Level 2, and therapists at Level 3), with clinic effects accounted for by inclusion of four dummy-coded variables representing clinic assignment entered at Level 3. Because diagnostic agreement was not associated with intake clinician characteristics, analyses were conducted with children nested within their treating therapists, rather than their intake therapists. In addition, previous studies on the present sample have found that diagnostic agreement levels varied by diagnosis type (Jensen & Weisz, 2002) and that symptom outcomes were significantly predicted by initial symptom severity levels (Jensen Doss & Weisz, 2006). All analyses were therefore conducted controlling for three variables coded to indicate whether or not the child's primary chart diagnoses were internalizing disorders, externalizing disorders, or elimination disorders, and the outcome analyses were conducted also controlling for either the initial CBCL or YSR Total Problems scores (depending on whether the outcomes being analyzed were from the CBCL or the YSR).

Our design consisted of up to 2 years of data for each participant, with variation in treatment end dates across participants. To estimate the amount of change that occurred during treatment, we analyzed the symptom outcome data using piecewise linear growth models (Raudenbush & Bryk, 2002). These models allow one to predict different parts of a growth curve separately by using separate parameters. In the present piecewise models, two time parameters were entered at Level 1. The first time parameter represented the timing of the measures between the Time 1 interview and the treatment end date. A second time parameter was coded to represent only the time after the treatment end date. The parameterization of time was centered around each child's treatment end date. To make the models more easily interpretable, we subtracted each child's initial severity level from the child's outcome data at each time point. When the model was constructed in this fashion, the intercept represented the amount of symptom change that occurred during treatment (e.g., Doss, Thum, Sevier, Atkins, & Christensen, 2005). As all of our therapy process vari-

ables were measured at a single point in time, they were analyzed using two-level HLM analyses (children nested within therapists).

We conducted mediational analyses using the product of coefficients model with asymmetric confidence intervals (PRODCLIN2; MacKinnon, Fritz, Williams, & Lockwood, 2007). This method has more power than do traditional methods for testing mediation (e.g., Baron & Kenny, 1986; cf. Fritz & MacKinnon, 2007; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). Unlike such traditional methods, these analyses are conducted on all possible mediational relations, even in the context of a null direct effect (i.e., the relation between diagnostic agreement and outcomes), due to the possibility that mediators may be functioning as suppressor variables for the direct effect (MacKinnon, Krull, & Lockwood, 2000) or that multiple mediators are operating in opposite directions (e.g., Sheets & Braver, 1999). Finally, we conducted secondary analyses to further explore relationships among agreement, process, and outcomes and to rule out alternative explanations for our findings. These analyses are described in turn below.

Results

Does Diagnostic Agreement Predict Therapy Process Variables?

Because children varied in length of treatment, cancellations and no-shows were analyzed as Poisson variables with variable exposure rates (Raudenbush & Bryk, 2002), with weeks in treatment used as the exposure indicator. Both forms of diagnostic agreement were significantly predictive of session no-shows, such that the disagree group had higher numbers of no-shows (primary diagnoses, $B = 0.34$, $p = .023$; all diagnoses, $B = 0.98$, $p = .001$; due to the nonnormal distribution of these data, effect size estimates are not available for these models). Both forms were also significantly predictive of number of cancellations, with the disagree group again having higher numbers of cancellations (primary diagnoses, $B = 0.34$, $p = .014$; all diagnoses, $B = 0.59$, $p = .011$). Neither form of agreement was significantly predictive of the number of weeks in treatment. All process and outcome results are presented in Table 3.

Because therapy dropout was a dichotomous variable, it was analyzed with Bernouli HLM models (Raudenbush & Bryk, 2002). Agreement on all diagnoses was significantly predictive of therapy dropout, such that the disagree group was more than 5 times as likely to terminate therapy without their clinician's consent ($B = 1.70$, $p = .005$, odds ratio = 5.48). Finally, a relation between diagnostic agreement and the RETQ Alliance scale was tested for with continuous HLM models. Neither form of agreement was significantly predictive of this scale.

Does Diagnostic Agreement Predict Symptom Outcomes?

We ran 12 models to examine the impact of agreement on symptom outcomes, with separate models run for agreement on primary diagnoses and agreement on all diagnoses and for CBCL and YSR Internalizing, Externalizing, and Total Problems scores. In two cases, agreement significantly predicted the amount of symptom change during treatment. First, agreement on primary diagnoses was predictive of outcomes on the CBCL Internalizing score ($B = 1.76$, $p = .047$). Conventions for computing effect sizes in HLM are not well established. As the regression coefficient for diagnostic agreement can be interpreted as the mean difference in symptom improvement between the agree and disagree groups, this coefficient was divided by the standard deviation for that coefficient to approximate a Cohen's *d* effect size (ES; Cohen, 1988) for this effect. The cutoff for a small Cohen's *d* is 0.2; the medium cutoff is 0.5, and the large cutoff is 0.8. The ES for the relation between agreement on primary diagnoses and CBCL Internalizing outcomes was 0.46. The same pattern of results was obtained for the relation between agreement on all diagnoses and improvement in the CBCL Internalizing score ($B = 3.22$, $p = .013$), with an ES that was much larger (0.84). In both

Table 3
Relation Between Diagnostic Agreement and Therapy Process and Outcomes

Measure	Agreement on primary diagnosis			Agreement on all diagnoses		
	B	*SE B*	ES	*B*	*SE B*	ES
Process variable						
Weeks in treatment	−5.42	3.81	−0.54	−6.71	6.28	−0.67
Cancellations	0.34*	0.13		0.59*	0.23	
No-shows	0.37*	0.16		0.98**	0.29	
Dropout	0.74	0.38	2.09	1.70**	0.60	5.48
RETQ Alliance	−0.18	0.65	−0.08	−0.73	1.54	−0.32
Outcome variable						
CBCL Internalizing during treatment	1.76*	0.88	0.46	3.22*	1.29	0.84
CBCL Externalizing during treatment	1.33	1.00	0.33	1.31	1.44	0.32
CBCL Total Problems during treatment	4.77	2.46	0.44	5.71	3.56	0.53
YSR Internalizing during treatment	1.10	0.95	0.26	1.47	1.56	0.34
YSR Externalizing during treatment	1.59	1.02	0.35	1.82	1.68	0.40
YSR Total Problems during treatment	4.48	2.61	0.42	4.81	4.27	0.45

Note. ES = effect sizes (approximation of Cohen's *d* for continuous variables, odds ratio for dichotomous variables); RETQ = Reasons for Ending Treatment Questionnaire; CBCL = Child Behavior Checklist; YSR = Youth Self Report.
* $p < .05$. ** $p < .01$.

cases, disagreement predicted shallower reductions in internalizing problems than agreement. No other analyses predicting symptom outcomes were significant, such that neither form of agreement was significantly predictive of outcomes on the YSR or the CBCL Externalizing or Total Problems scores. However, all outcome analyses trended in the direction of the disagree group having smaller treatment gains, with a mean ES value across all analyses of 0.43.

Do Process Variables Mediate the Relation Between Agreement and Outcomes?

To test for possible mediational relationships, we fit models in which each of our six outcome variables was predicted from both diagnostic agreement and each of our process variables. We then utilized the results of these models in combination with those of our analyses predicting the process variables from agreement to test whether each process variable mediated the relationship between diagnostic agreement and each outcome variable by computing a 95% confidence interval for the mediated effect with the PRODCLIN2 program (MacKinnon et al., 2007).

For the relationship between agreement and child-reported outcomes, we found evidence for two mediators working in opposite directions. First, for the YSR Externalizing scale, dropout was a significant mediator of the relation between both forms of agreement and outcomes ($ps < .05$), such that the disagree group had an increased likelihood of dropout (primary diagnoses, $B = 0.74$, $p = .052$; all diagnoses, $B = 1.70$, $p = .005$) and dropouts had smaller symptom decreases (primary diagnoses, $B = 2.74$, $p = .013$; all diagnoses, $B = 2.78$, $p = .013$). The same pattern was found for the relation between agreement on all diagnoses and YSR Total Problems and dropout (disagreement predicting dropout, $B = 1.70$, $p = .005$; dropout predicting smaller symptom decreases, $B = 6.24$, $p = .038$). Number of cancellations was a significant mediator of the relation between agreement on all diagnoses and both child-reported Externalizing and Total Problems ($ps < .05$); the disagree group had a higher number of cancellations ($B = 0.59$, $p = .011$), but higher numbers of cancellations were associated with greater symptom decreases (YSR Externalizing, $B = -0.41$, $p = .009$; YSR Total Problems, $B = -0.87$, $p = .033$). Thus, the overall nonsignificant positive direct relationship between diagnostic disagreement and smaller decreases in YSR Externalizing and Total Problems was mediated through increased likelihood of dropout (associated with worse outcomes), increased numbers of cancellations (associated with better outcomes), and, possibly, other unmeasured mediators.

For parent-reported outcomes, one significant mediator was found. Dropout was a significant mediator of the relation between agreement on all diagnoses and outcomes on the CBCL Internalizing scale ($p < .05$). Again, disagreement was associated with an increased likelihood of dropout ($B = 1.70$, $p = .005$), but dropout was associated with greater symptom decreases ($B = -1.95$, $p = .042$). Thus, the overall significant positive direct relationship between diagnostic disagreement and smaller decreases in CBCL Internalizing symptoms was likely mediated through increased likelihood of dropout (associated with better outcomes) and through unmeasured mediators more strongly associated with worse outcomes.

Secondary Analyses to Further Explore the Relations Between Agreement, Process, and Outcomes

Our mediational analyses suggested some ways in which changes in therapy process help explain the relation between diagnostic agreement and treatment outcomes. We next examined other methodological and clinical considerations that might further clarify our findings.

Did our definition of "agreement" impact our findings? Because we were interested in examining the impact of agreement on therapy process and outcomes, we focused on agreement that might have been expected to actually impact treatment planning. As such, we considered agreement at the level of categories. Included in these categories were diagnoses assigned by clinicians that were not formally diagnosed by the DISC (e.g., depressive disorder, not otherwise specified). To assess whether this decision impacted our findings, we reran our outcome and process analyses excluding the 25 children who had clinician diagnoses that fell into one of our categories but that were not formally assessed by the DISC. The findings obtained for this subsample were highly similar to our original findings. For the process analyses, the only change in results was that the relation between disagreement on the primary diagnosis and higher numbers of no-shows was no longer significant ($B = 0.26$, $p = .14$); all other significant findings held. For the outcome analyses, the relation between disagreement on the primary diagnosis and worse outcomes on the CBCL Internalizing scale was reduced to marginal significance ($B = 1.80$, $p = .058$, ES = 0.57), apparently primarily because of reduced sample size, because the overall effect size for the outcome analyses was essentially unchanged (mean ES = 0.45).

Could our findings be explained by the "third variables" of impairment and severity? One alternative explanation for our findings (rather than agreement directly and or indirectly impacting process and outcomes) is that children who are harder to diagnose are also harder to treat. It is therefore possible that our agreement, process, and outcome variables were related to each other only because all were associated with factors such as higher symptom severity or impaired functioning at the beginning of treatment. Our original outcome analyses controlled for initial severity, and we controlled for diagnosis type in all analyses; as an added check, we reran all of our process and outcome analyses controlling for chart diagnosis type, clinician-rated global assessment of functioning, and either the initial CBCL or the initial YSR Total Problems score (i.e., we ran two sets of analyses, one including parent-rated severity and one including child-rated severity). Controlling for these variables had little impact on our findings. For the process variables, when we controlled for CBCL Total Problems and the other variables, the association between disagreement on the primary diagnosis and higher numbers of cancellations was only marginally significant ($B = 0.28$, $p = .074$) and the association between disagreement on all diagnoses and higher numbers of no-shows was no longer significant ($B = 0.66$, $p = .203$). However, controlling for these variables led to significant prediction of increased likelihood of therapy dropout from disagreement on the primary diagnosis ($B = 0.91$, $p = .053$, odds ratio = 2.49), and all other significant relationships held. For the outcome analyses, all original significant findings held, and the relationship between disagreement on the primary diagnosis and smaller decreases in the CBCL Total Problems became significant

($B = 5.09$, $p = .043$, ES = 0.51). The overall effect size for the outcome analyses was essentially unchanged (mean ES = 0.48).

What was the role of clinic policies requiring a diagnosis to qualify for services? One factor that might have impacted agreement was the fact that, in order to qualify for services, children had to have a diagnosis; thus, clinicians may have felt pressured to assign diagnoses they might not otherwise have assigned. To explore what impact this policy might have had on our findings, we reran our process and outcome analyses, excluding the 42 children who did not meet criteria for a diagnosis according to the DISC. This exclusion resulted in a 21% reduction in sample size and a reduced number of significant findings; the only significant agreement-process relationships that held were between agreement on all diagnoses and numbers of no-shows and likelihood of dropout. On the other hand, exclusion of these children led to an additional agreement–outcome relationship, with agreement on the primary diagnosis and change in the CBCL Total Problems scale becoming significant ($B = 6.14$, $p = .033$, ES = 0.57). The overall mean ES for the outcome analyses was essentially unchanged (mean ES = 0.38).

Does the impact of agreement vary developmentally? Given the broad age range in our sample, we explored whether the impact of agreement differed by age. For each of our process and outcome analyses, we first tested whether there was an interaction between age and diagnostic agreement. In cases of significant interactions, we then ran our analyses separately for children (ages 7–10 years, $n = 90$) and adolescents (ages 11–17 years, $n = 107$; mean age = 11 years). In two cases, there was a significant Age × Agreement interaction. When predicting change in YSR Total Problems, agreement on the primary diagnosis had a significant interaction with age. Among children, agreement on the primary diagnosis was a significant predictor of change in the YSR Total Problems score, such that the disagree group had smaller treatment gains ($B = 11.07$, $p = .009$, $d = 1.07$). Agreement on the primary diagnosis was not a significant predictor of change in the YSR Total Problems score for adolescents ($B = -0.73$, $p = .82$, ES = −0.062). When predicting number of no-shows, agreement on the primary diagnosis also interacted with age. Disagreement on the primary diagnosis was significantly associated with a higher number of no-shows for children ($B = 1.32$, $p < .001$, ES = 1.30) but not for adolescents ($B = 0.036$, $p = .89$, ES = 0.030).

Discussion

Does diagnostic agreement in child psychotherapy predict improved therapy process or better treatment outcomes? We addressed this question by examining agreement on both primary diagnosis (i.e., the diagnosis most likely to impact treatment planning) and all diagnoses (i.e., the full picture, including comorbidities). Results were highly similar for the two forms of agreement. Both forms of agreement (i.e., on primary diagnosis and on all diagnoses) were found to predict lower numbers of session no-shows and cancellations. In addition, agreement on all diagnoses was associated with a decreased likelihood of therapy dropout. Both types of agreement also predicted greater symptom improvement during treatment for parent-reported internalizing symptoms. Although analyses predicting gains in parent-reported Externalizing symptoms, parent-reported Total Problems, and child-reported outcomes were not significant, all outcome findings were in the direction of the disagree group having fewer treatment gains, with the average effect size approximating Cohen's (1988) standard for a medium effect.

Our mediation analyses suggested that the relation between agreement and outcomes is likely complex and is mediated through multiple pathways. For example, the modest but nonsignificant relation between diagnostic disagreement and worse child-reported outcomes was mediated through two variables examined in the present study: (a) increased likelihood of therapy dropout, which in turn was associated with worse outcomes; and (b) increased numbers of cancellations, which were actually associated with better outcomes, perhaps due to families whose children were doing better in therapy being less motivated to attend every week. The significant relation between diagnostic agreement and worse parent-reported outcomes also was mediated through an increased likelihood of dropout; however, in this case, increased likelihood of dropout was associated with better parent-reported outcomes. Given that this mediation pathway works in opposition of the direct effect of agreement on parent-reported outcomes, there are likely stronger unmeasured mediators of this relation. Possible candidates for these mediators could be assignment to an inappropriate treatment or parent or child perceptions that the therapist is focusing on the wrong problem. The fact that dropout was associated with worse child-reported outcomes but better parent-reported outcomes may be due to the possibility that parents might be more motivated to go to the effort of taking their children to treatment if they feel their children still need assistance, whereas children might be less likely to want to continue participating in a treatment they do not perceive to be helpful.

Finally, our secondary analyses provided some useful information about the role of diagnostic agreement in clinical care. First, by dropping children who did not meet criteria for any diagnosis according to the DISC, we were able to explore the potential impact of clinic policies that may pressure clinicians to assign diagnoses in order to help a child qualify for services. Excluding these children from our analyses led to fewer significant agreement–process relationships; this result suggested that the impact of diagnostic agreement on therapy engagement might be particularly strong in families that are told their child has a diagnosis when, in fact, he or she may not. Second, we examined the role of development in the relationship between agreement, process, and outcomes. We found that the relation between agreement and change in child-reported Total Problems was significant for children but not for adolescents; the same was true for the relation between agreement and no-shows. It may be that, as children move into adolescence and begin to play a more active role in their own treatment, agreement between therapist- and parent-reported diagnoses may become less important.

Before the implications of these findings are discussed, several limitations must be considered. First, without additional data to assess the validity of the DISC and chart diagnoses, we cannot say for certain that diagnostic agreement was equivalent to diagnostic accuracy. However, considerable data exist regarding the validity of the DISC (Fisher et al., 1993; Schwab-Stone et al., 1996), and previous studies have concluded that structured diagnostic interviews are more valid than are clinician-generated diagnoses (Jewell et al., 2004; Basco et al., 2000; Tenney et al., 2003). Such studies suggest that it is reasonable to consider the DISC diagnoses as more accurate than are the chart diagnoses. However, given that

some studies have suggested that structured diagnostic interviews, such as the DISC, may have lower validity than do semistructured diagnostic interviews that allow the interviewer to clarify questions for confused participants and to explore participant responses (e.g., Angold & Fisher, 1999), future examination of the implications of agreement with a semistructured interview would be useful.

Second, the treatment outcomes considered were limited to parent- and child-reported symptoms. In future studies, inclusion of other indicators of outcome, such as teacher reports and functioning measures, would be useful. Third, 48 children from our original study sample were excluded due to lack of usable follow-up data or lack of primary diagnoses falling into one of the categories of interest; these excluded children were more likely than were our included children to be members of ethnic minorities and to have parents with lower levels of education. Although this difference potentially places some limits on the generalizability of our findings, we note that diagnostic agreement was not related to child ethnicity in our original analyses (Jensen & Weisz, 2002); in addition, a substantial number of our participants were members of ethnic minorities (41.2%), and nearly half (45%) of the participating parents had received no postsecondary education.

Fourth, although we were able to examine the impact of agreement on several process variables, as mentioned above, the present study lacked measures of other processes that could have been affected. For example, youths with mismatched diagnoses could have been assigned to treatments that were not appropriate for their "true" presenting problems, which would lead to worse treatment outcomes. In addition, although we found that agreement was not predictive of self-reports that parents had ended treatment due to a poor therapeutic relationship, we possibly would have found a relation between agreement and alliance if we had included a measure designed specifically to measure the quality of the therapy alliance. In addition, although we were able to document a link between diagnostic agreement and therapy engagement as measured by session attendance and dropout, we were not able to measure other aspects of therapy engagement, such as the quality of parent and child participation during sessions or adherence to treatment (e.g., completion of assigned therapy homework). Future studies with in-depth measures of within-session therapy process and with parent and child reports of the therapy alliance would help elucidate the relation between agreement and therapy process.

Finally, our analyses of diagnostic agreement were limited to analyses of agreement between a parent-report diagnostic interview and clinician chart diagnoses that were likely generated after interviews of both children and parents. Given the extensive literature documenting poor concordance between various reporters of childhood psychopathology (cf. De Los Reyes & Kazdin, 2005), one challenge faced by the diagnosing clinician likely was combining disparate reports of the child's symptoms and behaviors. It is therefore possible that agreement between clinician-generated and child-report diagnoses might have a different relation to process and outcomes than would parent-report diagnoses. In particular, it is possible that we might have more links between child-reported symptom outcomes and agreement had we utilized the child-report DISC, particularly among adolescents. However, studies involving child-report diagnoses are complicated by the low reliability of such diagnoses (e.g., Schwab-Stone et al., 1993). In addition, it is possible that inclusion of child-report diagnoses would not lead to different results, given the suggestion of previous analyses that combining child- and parent-report diagnoses does not improve agreement with clinician diagnoses (Jensen & Weisz, 2002).

Despite these limitations, this study had several strengths that make it a useful contribution to the literature. First, our methodology focused on the implications of diagnostic agreement in everyday clinical practice. We employed naturally occurring diagnostic procedures rather than contrived experimental conditions. Second, our chart review methodology provided us with several objective indicators of therapy engagement, including session no-shows and cancellations, weeks in treatment, and therapy dropout. Third, unlike previous studies, which focused on diagnostic agreement in only one problem area (Kramer et al., 2003; Pogge et al., 2001), this study investigated agreement on a broad range of diagnoses. Finally, and most important, this study investigated the implications of a well-documented finding in the field, that clinician-generated diagnoses do not generally map onto those diagnoses generated by research instruments. Although this phenomenon has been replicated numerous times, few studies have investigated whether this diagnostic disagreement is associated with any aspect of the processes or outcomes of care, and none has investigated this question in relation to such common youth mental health problems as disruptive behavior disorders, depression, and anxiety. Consistent with the two prior studies on this topic (Kramer et al., 2003; Pogge et al., 2001), our findings suggest that diagnostic disagreement may have significant implications for treatment process and outcomes.

Our findings suggest that diagnostic disagreement has a relationship with therapy engagement and treatment outcomes but that the relation between disagreement and outcomes is complex and is likely to be explained in part by mediators not measured in the present study. These findings lend themselves to three possible interpretations. First, it is possible that diagnostically complex youths, for whom it is more difficult to assign diagnoses, are also more difficult to engage in therapy and to treat; such increased complexity might lead to associations between agreement and engagement and outcomes. This interpretation is consistent with prior literature suggesting that low concordance between different informants of childhood psychopathology may be related to characteristics of the child and family, such as problem type and parental psychopathology (cf. De Los Reyes & Kazdin, 2005). However, the fact that we found associations among these variables even after controlling for symptom severity, diagnosis type, and clinician-rated global assessment of functioning suggests that this was likely not the case. In addition, our prior analyses of diagnostic agreement found that agreement between clinician and DISC diagnoses in our sample was not a function of demographic or clinical characteristics of the child or parent (Jensen & Weisz, 2002). However, it is possible that controlling for other indicators of diagnostic complexity, such as other reports of impaired functioning and other contextual variables (e.g., parent–child communication), might have led to different results. Future studies that include more indicators of diagnostic complexity could shed light on this issue.

A second possibility is that, by including clinician diagnosis not formally assessed by the DISC (e.g., adjustment disorders), we artificially inflated the relation between agreement, process, and outcomes. If these diagnoses were associated with worse process

and outcomes and children with these diagnoses were more likely to be placed in the disagree group, inclusion of these clinician diagnoses in our analyses could have artificially created associations between agreement, processes, and outcomes. However, excluding children with these diagnoses from our analyses had minimal impact on our findings, other than what might be expected from somewhat reducing the power of our analyses to detect significant relationships.

A third possibility is that the assignment of an accurate diagnosis is an important precursor to treatment success, such that incorrect diagnoses negatively impact therapy process and outcomes. This interpretation is consistent with previous literature documenting long-term implications of poor concordance between parents and children for child outcomes (e.g., Ferdinand, van der Ende, & Verhulst, 2004, 2006; Pelton & Forehand, 2001) and for parental involvement in treatment (Israel, Thomsen, Langeveld, & Stormark, 2007). If this is the case, these findings suggest that interventions to improve the accuracy of clinician-generated diagnoses have the potential to positively impact the processes and outcomes of clinical care. Studies in the primary care literature suggest that providing medical professionals with feedback from standardized mental health measures can impact the process of care (Espallargues, Valderas, & Alonso, 2000), but little work has been conducted on this topic in mental health settings. Two studies have addressed this topic by randomly assigning children (Hughes et al., 2005) and adults (Kashner et al., 2003) either to diagnostic assessment as usual or to participation in a research diagnostic battery, with feedback provided to the treating psychiatrist. The authors of both studies concluded that providing psychiatrists with this feedback had a positive impact on treatment process and outcomes. However, as neither study conducted a research evaluation on the assessment-as-usual group, it is difficult to know whether these results are attributable to increased diagnostic accuracy or to other effects of participating in a research battery, such as priming clients to consider symptom areas they might not have readily discussed with their providers. In addition, these two studies focused on psychiatrists and the process variables involved in medication practices. Future studies involving therapy clients should show whether these interventions would improve the processes and outcomes of child therapy.

Given the present findings and those of previous studies (Kramer et al., 2003; Pogge et al., 2001) suggesting that diagnostic agreement is associated with improved therapy process and outcomes, further randomized studies of the effects of incorporating standardized assessment tools into clinical practice settings are warranted. This work would build upon findings suggesting that incorporating standardized measures of therapy outcome into practice settings can have a positive impact on therapy outcome (e.g., Lambert, Hansen, & Finch, 2001). To rule out the possibility that the experience of participating in the research interview, not the use of the results of the interview, impacts processes and outcomes, researchers should administer the interview to all participants and should provide the results only to the experimental group's providers. By randomly assigning participants to feedback conditions, such studies would help rule out alternative explanations (e.g., diagnostic complexity) for the link between diagnostic agreement, process, and outcomes. They would also help establish the treatment utility of diagnosis, an assumption made by many in the field that is actually not well tested (Nelson-Gray, 2003). Given the length of many standardized interviews (e.g., 90–120 min per participant for the current version of the DISC; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), these studies could explore the benefit of utilizing less time-consuming or costly methods, such as rating scales, screening tools (e.g., the DISC Predictive Scales; Lucas et al., 2001), or self-administered, computer-guided diagnostic interviews (e.g., the Voice DISC; Shaffer et al., 2000). Finally, such studies would allow the field to explore the exciting possibility that the incorporation of standardized diagnostic instruments into clinical practice settings might have a positive impact on therapy processes and outcomes for youths.

References

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

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

Aegisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist, 34*(3), 341–382.

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

Anderson, D. A., & Paulosky, C. A. (2004). A survey of the use of assessment instruments by eating disorder professionals in clinical practice. *Eating and Weight Disorders, 9*(3), 238–241.

Angold, A., & Fisher, P. W. (1999). Interviewer-based interviews. In D. Shaffer, C. P. Lucas, & J. E. Richters (Eds.), *Diagnostic assessment in child and adolescent psychopathology* (pp. 34–64). New York: Guilford Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Basco, M. R., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., et al. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry, 157*(10), 1599–1605.

Berardi, D., Menchetti, M., Cevenini, N., Scaini, S., Versari, M., & De Ronchi, D. (2005). Increased recognition of depression in primary care: Comparison between primary-care physician and *ICD-10* diagnosis of depression. *Psychotherapy and Psychosomatics, 74*(4), 225–230.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*(3), 252–260.

Cashel, M. L. (2002). Child and adolescent psychological assessment: Current clinical practices and the impact of managed care. *Professional Psychology: Research and Practice, 33*(5), 446–453.

Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L., Calhoun, K. S., Crits-Christoph, P., et al. (1998). Update on empirically validated therapies, II. *Clinical Psychologist, 51*(1), 3–16.

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

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational & Psychological Measurement, 20,* 37–46.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical

framework, and recommendations for further study. *Psychological Bulletin, 131,* 483–509.

Doss, B. D., Thum, Y. M., Sevier, M., Atkins, D. C., & Christensen, A. (2005). Improving relationships: Mechanisms of change in couple therapy. *Journal of Consulting and Clinical Psychology, 73,* 624–633.

Espallargues, M., Valderas, J. M., & Alonso, J. (2000). Provision of feedback on perceived health status to health care professionals: A systematic review of its impact. *Medical Care, 38*(2), 175–186.

Ferdinand, R. F., van der Ende, J., & Verhulst, F. C. (2004). Parent–adolescent disagreement regarding psychopathology in adolescents from the general population as a risk factor for adverse outcome. *Journal of Abnormal Psychology, 113,* 198–206.

Ferdinand, R. F., van der Ende, J., & Verhulst, F. C. (2006). Prognostic value of parent–adolescent disagreement in a referred sample. *European Child & Adolescent Psychiatry, 15*(3), 156–162.

Fisher, P. W., Shaffer, D., Piacentini, J., Lapkin, J., Kafantaris, V., Leonard, H., et al. (1993). Sensitivity of the Diagnostic Interview Schedule for Children, 2nd edition (DISC-2.1) for specific diagnoses of children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*(3), 666–673.

Fleiss, J. R. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18,* 233–239.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment.* Washington, DC: American Psychological Association.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*(1), 67–89.

Garcia, J. A., & Weisz, J. R. (2002). When youth mental health care stops: Therapeutic relationship problems and other reasons for ending youth outpatient treatment. *Journal of Consulting and Clinical Psychology, 70,* 439–443.

Garland, A. F., Hough, R. L., McCabe, K. M., Yeh, M., Wood, P. A., & Aarons, G. A. (2001). Prevalence of psychiatric disorders in youths across five sectors of care. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(4), 409–418.

Hawley, K. M., & Weisz, J. R. (2005). Youth versus parent working alliance in usual clinical care: Distinctive associations with retention, satisfaction, and treatment outcome. *Journal of Clinical Child and Adolescent Psychology, 34*(1), 117–128.

Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology, 61,* 561–573.

Hughes, C. W., Emslie, G. J., Wohlfahrt, H., Winslow, R., Kashner, T. M., & Rush, A. J. (2005). Effect of structured interviews on evaluation time in pediatric community mental health settings. *Psychiatric Services, 56*(9), 1098–1103.

Israel, P., Thomsen, P. H., Langeveld, J. H., & Stormark, K. M. (2007). Parent–youth discrepancy in the assessment and treatment of youth in usual clinical care setting: Consequences to parent involvement. *European Child & Adolescent Psychiatry, 16*(2), 138–148.

Jensen, A. L., & Weisz, J. R. (2002). Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology, 70,* 158–168.

Jensen Doss, A., & Weisz, J. R. (2006). Syndrome co-occurrence and treatment outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology, 74,* 416–425.

Jewell, J., Handwerk, M., Almquist, J., & Lucas, C. (2004). Comparing the validity of clinician-generated diagnosis of conduct disorder to the Diagnostic Interview Schedule for Children. *Journal of Clinical Child and Adolescent Psychology, 33*(3), 536–546.

Kashner, T. M., Rush, A. J., Surís, A., Biggs, M. M., Gajewski, V. L., Hooker, D. J., et al. (2003). Impact of structured clinical interviews on physicians' practices in community mental health settings. *Psychiatric Services, 54*(5), 712–718.

Kramer, T. L., Robbins, J. M., Phillips, S. D., Miller, T. L., & Burns, B. J. (2003). Detection and outcomes of substance use disorders in adolescents seeking mental health treatment. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(11), 1318–1326.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69,* 159–172.

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

Lucas, C. P., Zhang, H., Fisher, P. W., Shaffer, D., Regier, D. A., Narrow, W. E., et al. (2001). The DISC Predictive Scales (DPS): Efficiently screening for diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(4), 443–449.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODLIN. *Behavior Research Methods, 39*(3), 384–389.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1*(4), 173–181.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*(1), 83–104.

McKay, M. M., & Bannon, W. M., Jr. (2004). Engaging families in child mental health services. *Child and Adolescent Psychiatric Clinics of North America Special Issue: Evidence-Based Practice, Part I: Research Update, 13*(4), 905–921.

Nelson-Gray, R. O. (2003). Treatment utility of psychological assessment. *Psychological Assessment, 15,* 521–531.

Pelton, J., & Forehand, R. (2001). Discrepancy between mother and child perceptions of their relationship: I. Consequences for adolescents considered within the context of parental divorce. *Journal of Family Violence, 16*(1), 1–15.

Pini, S., Perkonnig, A., Tansella, M., & Wittchen, H.-U. (1999). Prevalence and 12-month outcome of threshold and subthreshold mental disorders in primary care. *Journal of Affective Disorders, 56*(1), 37–48.

Pogge, D. L., Wayland-Smith, D., Zaccario, M., Borgaro, S., Stokes, J., & Harvey, P. D. (2001). Diagnosis of manic episodes in adolescent inpatients: Structured diagnostic procedures compared to clinical chart diagnoses. *Psychiatry Research, 101*(1), 47–54.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S., Bryk, A., & Congdon, R. (2005). *HLM 6: Hierarchical linear and nonlinear modeling.* Chicago: Scientific Software.

Rettew, D. C., Doyle, A., Achenbach, T. M., Dumenci, L., & Ivanova, M. (2006). *Meta-analyses of diagnostic agreement between clinical evaluations and standardized interviews.* Manuscript submitted for publication.

Schwab-Stone, M., Fisher, P. W., Piacentini, J., Shaffer, D., Davies, M., & Briggs, M. (1993). The Diagnostic Interview Schedule for Children—Revised version (DISC–R): II. Test–retest reliability. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*(3), 651–657.

Schwab-Stone, M., Shaffer, D., Dulcan, M. K., Jensen, P. S., Fisher, P., Bird, H. R., et al. (1996). Criterion validity of the NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3). *Journal of the American Academy of Child & Adolescent Psychiatry, 35*(7), 878–888.

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(1), 28–38.

Shaffer, D., Schwab-Stone, M., Fisher, P. W., Cohen, P., Piacentini, J.,

Davies, M., et al. (1993). The Diagnostic Interview Schedule for Children–Revised version (DISC–R): I. Preparation, field testing, inter-rater reliability, and acceptability. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*(3), 643–650.

Sheets, V. L., & Braver, S. L. (1999). Organizational status and perceived sexual harassment: Detecting the mediators of a null effect. *Personality and Social Psychology Bulletin, 25*(9), 1159–1171.

Shirk, S. R., & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71,* 452–464.

Tenney, N. H., Schotte, C. K. W., Denys, D. A. J. P., van Megen, H. J. G. M., & Westenberg, H. G. M. (2003). Assessment of *DSM–IV* personality disorders in obsessive-compulsive disorder: Comparison of clinical diagnosis, self-report questionnaire, and semi-structured interview. *Journal of Personality Disorders, 17*(6), 550–561.

Tiemens, B. G., Ormel, J., & Simon, G. E. (1996). Occurrence, recognition, and outcome of psychological disorders in primary care. *American Journal of Psychiatry, 153*(5), 636–644.

Yeh, M., & Weisz, J. R. (2001). Why are we here at the clinic? Parent–child (dis)agreement on referral problems at outpatient treatment entry. *Journal of Consulting and Clinical Psychology, 69,* 1018–1025.

Received August 27, 2007
Revision received March 27, 2008
Accepted March 28, 2008 ■

Call for Papers: Special Section on "Posttraumatic Stress Disorder and Trauma in Children and Adolescents"

Papers Due on or Before January 15, 2009

The *Journal of Consulting and Clinical Psychology* (*JCCP*) invites submission of empirical papers and scholarly reviews that focus on research pertaining to posttraumatic stress disorder (PTSD) and trauma reactions in children and adolescents. The inspiration for this section was derived, in part, from the 2008 American Psychological Association (APA) Presidential Task Force on this topic (http://www.apa.org/about/president/initiatives.html) and from APA resolutions on *The Psychological Needs of Children Exposed to Disasters* (http://www.apa.org/pi/cyf/res_needs.html) and on *Children's Mental Health* (http://www.apa.org/pi/resolution/childmentalhlth.html).

Stress reactions resulting from different types of trauma (e.g., natural disasters, terrorism, child sexual abuse, community violence, medical trauma/injury) will be considered. Papers may focus on risk and resilience factors, including potential variations among groups (e.g., sex, ethnicity/culture, socioeconomic status, age/developmental stage); issues of comorbidity and related trauma reactions; impact on adaptive functioning in children, youths, and families; and effective prevention and treatment interventions.

Preference will be given to papers that provide clear articulation of the conceptual or theoretical basis for the variables that are selected for evaluation in the research. It is essential that papers directly discuss the following: (a) areas of research need and important "next steps" that will help guide future research, prevention, and treatment efforts, and (b) recommendations for disseminating information to stakeholders interested in helping children and their families in the aftermath of trauma, such as parents/caregivers, health-care providers, practitioners, policy makers, and government agencies. Findings are intended to help inform the next generation of studies for PTSD and trauma reactions in children and adolescents, as well as the practice of psychologists working with children, adolescents, and families.

Manuscripts must be consistent with the submission guidelines for *JCCP* (see www.apa.org/journals/ccp); papers that do not follow the guidelines may be returned without review. Papers should be submitted electronically through the Manuscript Submission Portal for *JCCP*, and authors must request consideration for the Special Section in the cover letter.

To be eligible for inclusion in the Special Section, papers must be submitted by **January 15, 2009**; early submissions are encouraged. Papers that do not meet the deadline will be considered as "regular" submissions to this journal.

Questions should be addressed to the Journal Office via phone (305-284-8823) or e-mail (jccp.psy@miami.edu).