Psychological Assessment
2012, Vol. 24, No. 1, 141–155

1040-3590/12/$12.00 DOI: 10.1037/a0025176

A Transportable Assessment Protocol for Prescribing Youth Psychosocial Treatments in Real-World Settings: Reducing Assessment Burden via Self-Report Scales

Chad Ebesutani, Adam Bernstein, and Bruce F. Chorpita
University of California at Los Angeles

John R. Weisz
Harvard University and Harvard Medical School

Current evidence-based assessment methods, such as structured interviews and lengthy assessment batteries, often require hours to administer, score, and interpret and thus are infrequently used in real-world practice. As evidence-based assessment tools are developed for implementation in real-world youth mental health settings, the transportability properties of assessment procedures (including administration and interpretation burden) need to be considered and improved. In the present study, we thus conducted an initial feasibility study using a clinical sample of community-based youths ($N = 306$) to develop an assessment protocol based on 2 child and 2 parent self-report questionnaires (thus low on administration burden). Using decision-tree analysis, we identified a series of cutoff scores across these scales that may be used to inform treatment need related to anxiety, depression, attention-deficit/hyperactivity disorder (ADHD), and disruptive behavior problems. This algorithm-based approach to interpreting assessment information provided clear and simple guidelines (thus low on interpretation burden) that matched the best estimate treatment determinations derived by trained assessors, supervisors, and expert consultants who integrated information provided by child and parent structured interviews and self-report scales. The present study demonstrated the feasibility of developing an assessment protocol to inform various treatment allocation decisions in a way that imposes little assessment administration and interpretation burden yet maintains adequate classification accuracy. These characteristics make the proposed protocol promising with regard to its transportability and suitability for adoption and implementation in real-world mental health settings.

Keywords: decision tree, youth assessment, classification tools, self-report scales

Community mental health settings are in need of evidence-based assessment tools to assist in informing the provision of youth mental health services in ways that both are feasible and meet best practice standards. Only a small fraction of clinicians report using results from structured assessments in their clinical practice, even among those mandated by state regulations (Garland, Kruse, & Aarons, 2003) and those trained as psychologists (Hatfield & Ogles, 2004). Relatedly, the majority of empirically supported assessment instruments require too much time to be feasibly implemented by "real-world" community practitioners (Garland et al., 2003). For instance, although structured interviews yield more reliable and quantifiable outcomes than do unstructured interviews (which are often used in clinical practice), structured interviews tend to be too time-intensive, resource-heavy and, in many cases, too costly for "real-world" implementation. Managed care has also put pressure on practitioners in recent decades to increase the efficiency and cost-effectiveness of their time spent with clients (Richardson & Austad, 1991), making it increasingly difficult for community practitioners to adopt lengthy evidence-based assessment procedures.

Given these barriers to disseminating (lengthy) empirically supported assessment tools to "real-world" settings, researchers have begun developing briefer assessment instruments for youth problems related to both internalizing and externalizing behaviors (e.g., Chorpita et al., 2010; Webster-Stratton & Spitzer, 1991). Such efforts have used item response theory (e.g., Chorpita et al., 2010) and computerized assessment formats (e.g., Reich, 2000), among other methods, to shorten tests and reduce assessment burden. Although progress is being made in developing briefer instruments to aid in the dissemination of empirically supported assessment, significant barriers remain, and the implementation of evidence-based assessment in "real-world" settings remains limited (Garland et al., 2003).

Figure 1 provides a visual framework for conceptualizing various forms of *burden* associated with common evidence-based assessment tools (i.e., structured interviews, questionnaires). The figure depicts three domains likely contributing to the challenges in disseminating evidence-based assessment procedures to "real-world" settings: (a) assessment *administration* burden imposed on *clinicians* (e.g., due to clinicians administering lengthy assessments), (b) assessment *administration* burden imposed on *clients* (e.g., due to youths/parents participating in lengthy assessments),

This article was published Online First August 22, 2011.
Chad Ebesutani, Adam Bernstein, and Bruce F. Chorpita, Psychology Department, University of California at Los Angeles; John R. Weisz, Department of Psychology, Harvard University, and Harvard Medical School.

Correspondence concerning this article should be addressed to Chad Ebesutani, Department of Psychology, University of California at Los Angeles, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563. E-mail: ebesutani@ucla.edu

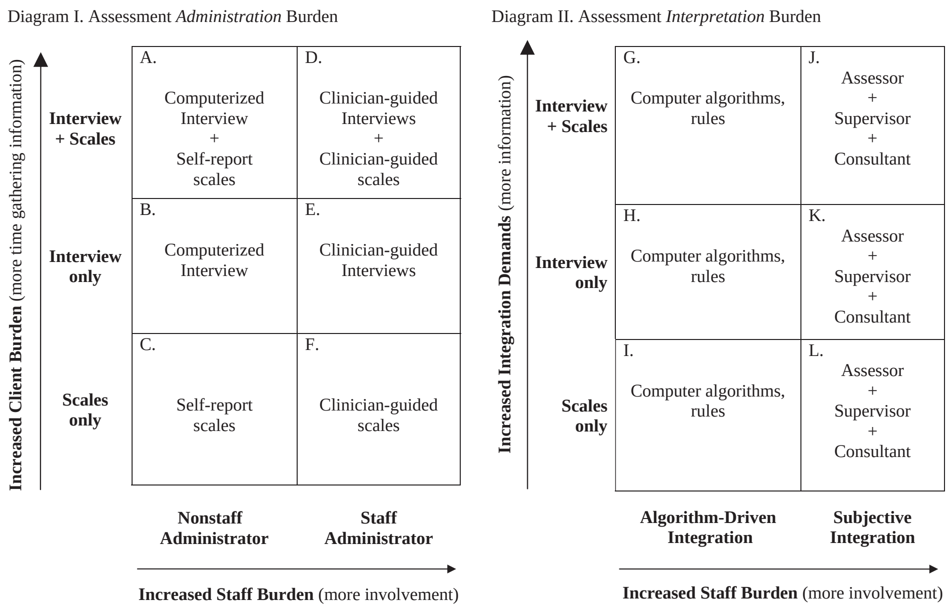

Figure 1. Components of assessment burden on clients and staff. Diagram I: Assessment administration burden. Diagram II: Assessment interpretation burden.

and (c) assessment *interpretation* burden imposed on *clinic staff* (e.g., due to information integration demands). This framework is not exhaustive. For example, it does not incorporate financial and organizational aspects of the burden that affects clinics/agencies (e.g., monetary cost of assessment tools; administrative challenges of incorporating diverse measures with information systems and reimbursement processes) or other factors that impose stress on clients and families (e.g., culture, literacy, and comprehension barriers, Garland et al., 2003; fracturing and redundancy of assessment across multiple sectors of care, Hermann & Palmer, 2002). Rather, the domains included in the framework reflect broad categories of burden commonly perceived to be obstacles to transportability of evidence-based assessment procedures and most applicable to this article's focus on assessment for the purpose of treatment prescription.

Assessment Administration Burden on the Clinic

As seen in Figure 1, Diagram I, assessments associated with greater amounts of clinic burden are procedures that require clinicians to administer assessments (right column: "Staff Administrator"), such as clinician-administered structured interviews (cell E) and, to a lesser extent, clinician-administered checklists and questionnaires (cell F). On the other hand, assessment procedures that remove clinicians from administration procedures (see Figure 1, Diagram I, left column: "Nonstaff Administrator"), such as through the use of *computerized* assessments (cell B; e.g., Reich, 2000) and client *self-report* questionnaires (cell C; e.g., Achenbach & Rescorla, 2001), reduce assessment administration burden on clinics/agencies and may be more likely to be adopted in community settings.

Assessment Administration Burden on the Client

Figure 1, Diagram I, also depicts assessment burden imposed on *clients* as a function of the type of assessment administered. At the low end of the spectrum, assessments based solely on brief questionnaires impose minimal assessment burden on clients. On the other hand, subjecting clients to the combination of longer questionnaires and lengthy structured interviews imposes significantly more assessment administration burden on clients. Indeed, a tension exists between gathering sufficient amounts of information from clients to be able to make reliable and valid inferences regarding patient functioning and impairment and overburdening clients with lengthy assessment procedures.

Assessment Interpretation Burden on the Clinic

Although research efforts have made substantial progress in developing briefer assessment instruments to reduce administration burden on clinics and clients (as noted above), there remains an additional source of burden that has received less attention. Specifically, given that youths seen in community mental health clinics typically present with multiple co-occurring clinical and subclinical problems and recommended assessment approaches involve obtaining information from various sources and informants, multiple sources of information from multiple sources often need to be *integrated* and *interpreted* to identify problem areas and treatment need. Figure 1, Diagram II, depicts two classes of methods for *integrating* and *interpreting* assessment information: (a) manual, subjective integration (right column) and (b) algorithm-driven integration (left column). Currently, there exist very few empirically supported assessment procedures and guidelines for either type of integration procedure. The "best estimate" approach (Klein, Ouimette, Kelly, Ferro, & Riso, 1994) is currently the best available approach by which assessment information is integrated via manual, subjective integration. This approach involves having clinical assessors and supervisors collectively and carefully consider all available information from various sources, informants, and problem areas to make clinical determinations and recommendations regarding treatment need. Specific guidelines that govern this integration process however are not yet well delineated (see De Los Reyes & Kazdin, 2005), although some recent research has reported methods to aid in integrating information (e.g., Kraemer et al., 2003) and identifying top problems in need of psychosocial intervention (e.g., Weisz et al., 2011). In addition to manual methods of integrating information, which can be time-intensive and costly, there has been some (albeit limited) work in this area to develop *algorithm-driven* integration procedures that can be automated (see Figure 1, Diagram II, left column). For example, with respect to integrating *interview-based* information, the "or-rule" and the "and-rule" (cf. Silverman & Albano, 1996) have been used as algorithms for integrating information derived from child- and parent-based structured interviews (see Figure 1, Diagram II, cell H). On the basis of the "or-rule," for example, a youth is considered to have a diagnosis if the child's *or* parent's reports suggest the presence of that diagnosis.

With respect to algorithm-driven integration methods for synthesizing information from *scale information only* (see Figure 1, Diagram II, cell I) or from the combination of *clinical interviews and scale information* (see Figure 1, Diagram II, cell G), integration algorithms and guidelines are lacking. Given that the resource-heavy and costly "best estimate" procedure remains the primary method for integrating assessment information, additional empirically supported assessment procedures and guidelines based on algorithm-driven methods are needed to aid in integrating (often discrepant) information (De Los Reyes et al., 2005). Through developing algorithm-driven interpretation methods, evidence-based assessments may more easily inform and become integrated into the selection and implementation of evidence-based treatment protocols. Such methods may be used to strengthen the connection between evidence based assessment and treatment practices, which many investigators have cited as a critical future direction for clinical science and practice (Hunsley & Mash, 2007; Weisz, Chu, & Polo, 2004).

The Present Study

In the present study, we therefore investigated whether it was feasible to develop an assessment protocol that (a) is based on assessment procedures that reduce assessment burden along the three dimensions outlined above and (b) includes an algorithm-driven method for integrating information obtained from multiple sources (i.e., child and parent report scales) to determine which treatment protocol (if any) is most indicated for a given youth. To reduce assessment burden, we attempted to develop a protocol that (a) is based only on relatively *brief* questionnaires (see Figure 1, bottom row), (b) utilizes *self-report* child and parent questionnaires (see Figure 1, Diagram I, cell C), and (c) utilizes *algorithm-driven* data integration (i.e., a series of optimal cutoff points across various scales; see Figure 1, Diagram II, cell I) to minimize assessment administration and interpretation burden on practitioners and clients to be more suitable for "real-world" implementation.

We hypothesized that by using decision-tree methodology, we would be able to develop scale-based algorithms using optimal cutpoints to inform treatment allocation decisions at three different levels of increasing specificity that may be applicable in "real-world" settings.

1. Is the referred youth in need of *any* form of treatment (from among the areas of anxiety, depression, disruptive behavior, or ADHD)? This clinically relevant question is applicable to the intake screening process whereby a parent or school staff may contact a clinic due to some concerns of his or her child or student. At this level, the scale-based guidelines do not specify which problem area may require treatment but rather provide a simple indication of whether any significant problems are present that warrant treatment or whether clinic-related services do not appear needed at the present time for these problem areas.

2. Beyond the indication of need for any treatment provided by Question 1, a clinic may want to determine whether a youth's primary treatment needs fall broadly into the internalizing or externalizing areas. This level of specificity that assesses only for internalizing versus externalizing problem focus may be sufficient for certain treatment allocation purposes given that, for example, Chorpita, Daleiden, et al. (2011) review of the evidence-base of youth psychosocial treatments revealed that a single treatment (i.e., parent management training) is among the most effective psychosocial treatments for *both* ADHD and disruptive behavior problems, the two most common externalizing problems. Relatedly, Barlow, Allen, and Choate (2004) recently developed a unified treatment for emotional disorders that is designed to be applicable to the broad area of internalizing problem (i.e., anxiety and/or depressive problems). Clinicians who take this (transdiagnostic) approach to treatment delivery may therefore not need to make differential diagnoses beyond this level of specificity (i.e., internalizing versus externalizing treatment need) for the purpose of treatment allocation in many cases.

3. Given that many evidence-based treatments have been developed and shown to be efficacious for the more specific problem areas of anxiety (e.g., Coping Cat; Kendall et al., 1997), depression (e.g., PASCET; Weisz, Gray, Bearman, Stark, 2008), disruptive behavior (e.g., Incredible Years; Webster-Stratton & Reid, 2003), and ADHD (stimulant medication; Abikoff et al., 2004), a clinic may decide to prescribe treatment at this level of specificity.[1] We

[1] Importantly, we only sought to classify youths in the present study at the level of general problem area (i.e., anxiety, depression, ADHD and disruptive behavior) as opposed to the more specific diagnostic subtypes related to these problem areas (e.g., separation anxiety disorder) because making diagnostic subtype determinations will likely increase assessment burden and, more important, because (a) broader problem categories are more concordant with the way treatment target decisions are made currently in clinical practice, given the rare usage of structured diagnostic measures noted above (Garland et al., 2003) and (b) evidence-based treatments can often be prescribed without precise diagnostic subtype determinations. Indeed, some evidence-based treatments are in fact specific to certain diagnostic subtypes (e.g., Coping Cat for separation anxiety disorder, generalized anxiety disorder, and social phobia child, Kendall et al., 1997). However, much of the evidence base on treatment protocols has demonstrated effectiveness with respect to youths characterized not by diagnoses but rather by syndrome elevations determined by cutoff scores. This is reflected in the randomized clinical trials (RCTs) testing the efficacy of treatment protocols; only 41% of RCTs even reported participant diagnoses among 435 randomized clinical trials reviewed in a recent study (Chorpita, Bernstein, & Daleiden, 2011). Therefore, from a treatment prescription perspective (as taken in the present study), it was sufficient to only classify youths into the general problem areas of anxiety, depression, ADHD, or disruptive behavior.

thus developed algorithm-driven guidelines to inform treatment allocation decisions at each of these levels of specificity, with the goal of generating more efficient and simple assessment protocols that provide instrumental guidance for each of these three important clinical decisions. Given the complex nature of integrating data from various sources and informants (De Los Reyes & Kazdin, 2005), we did not expect the present algorithms to adequately classify 100% of youths at all levels; rather, we sought to develop algorithms that classified a large portion of youths at each level so that the algorithms might be used, for example, as an initial screening tool before the intake assessment, so as to reduce subsequent assessment burden. For instance, if the algorithm classification tool (generated in the present study) indicates a particular treatment protocol with an acceptable level of accuracy, subsequent assessment may be brief; on the other hand, if the present classification tool is unable to accurately classify a youth, more thorough assessment may be needed, much like a hybrid assessment approach recently suggested by Chorpita and Nakamura (2008).

Method

Participants

Youths in the present study were 306 of 330 consecutively referred children and adolescents who were seeking mental health services in community clinic settings in urban settings in Hawaii and Massachusetts. Inclusion in the study required youths having data available on parent and youth versions of each of the measures described below (see Measures). Nine youths (2.7%) were excluded because no data were available on one or more measures. To ensure validity of the data entered into the analysis, we also excluded 10 youths (3.0%) due to their forms not having at least 90% completed data. Last, five youths (1.5%) withdrew from the study, leaving a final sample size of 306. Information on the total number of youth diagnoses appears in Table 1. Youth and primary caregiver demographic information appears in Table 2.

Measures

All means, standard deviations, and internal consistency reliability estimates for each of the measure's scale scores appear in Table 3. Correlations between corresponding child and parent scale scores appear in Table 4. Further detailed information on each measure used in the present study appears below.

Child Behavior Checklist/Youth Self-Report (CBCL/YSR; Achenbach & Rescorla, 2001). The CBCL and YSR are parent and youth self-report questionnaires that assess youth emotional and behavioral problems. They are rated on a 3-point Likert scale: not *true* (0), *sometimes true* (1), and *often true* (2). Summed items yield eight syndrome scale scores, six *Diagnostic and Statistical*

Table 1
Number of Principal and Anywhere Diagnoses Among Study Participants

	Child-based Dx		Parent-based Dx		Consensus Dx	
Diagnosis	Any	Principal	Any	Principal	Any	Principal
Major depressive Dx	24	10	53	17	68	29
Dysthymic Dx	7	4	25	8	27	7
Depressive Dx NOS	5	2	5	2	16	7
Panic Dx	1	0	1	0	1	0
Specific phobia	37	20	60	7	64	2
Generalized anxiety Dx	15	6	58	20	53	20
Separation anxiety Dx	42	22	64	21	74	27
Social phobia	16	10	28	5	32	6
OCD	9	4	13	3	12	8
PTSD	4	2	13	5	15	5
Anxiety NOS	3	2	1	1	4	3
ADHD-C	22	4	71	22	78	15
ADHD-PI	32	13	53	24	47	13
ADHD-PH	2	0	4	3	5	1
ADHD-NOS	22	11	41	16	43	14
Oppositional defiant Dx	76	50	127	84	140	71
Conduct disorder	32	17	50	17	50	38
DBD NOS	1	0	2	2	2	1
Bipolar	1	1	3	0	6	2
Schizophrenia	3	2	3	0	8	4
PDD	1	0	0	0	1	1
Other	6	3	13	7	20	9
No diagnosis	123	123	42	42	25	25

Note. $N = 306$. Other includes substance abuse/dependence, enuresis, and trichotillomania. Child-based Dx = disorders based solely on the ChIPS interview; Parent-based Dx = disorders based solely on the P–ChIPS interview; Consensus Dx = final disorders based on all available assessment information; Any = a diagnosis that appears anywhere in a child's diagnostic profile (i.e., principal, secondary, tertiary); Principal = a child's primary diagnosis; Dx = disorder; NOS = not otherwise specified; ADHD = attention-deficit/hyperactivity disorder; PDD = pervasive developmental disorder; PTSD = posttraumatic stress disorder; OCD = obsessive-compulsive disorder; DBD = disruptive behavior disorder; C = combined type; PI = primarily inattentive type; PH = primarily hyperactive type; ChIPS = Children's Interview for Psychiatric Syndromes; P–ChIPS = Children's Interview for Psychiatric Syndromes—Parent Version.

Table 2
Youth and Caregiver Demographic Information

Demographic information	*n*	%
Youth gender		
Male	205	67
Female	101	33
Youth ethnicity		
Multiethnic	93	30
White	142	46
African American	28	9
Asian American	10	3
Other	9	3
Missing	2	1
Caregiver type		
Biological mother	173	56.5
Biological father	60	19.6
Adoptive mother	8	2.6
Adoptive father	6	2.0
Grandmother	16	5.2
Grandfather	9	2.9
Other	21	6.9
Missing	13	4.2
Caregiver marital status		
Married	124	40.5
Divorced	69	22.5
Never married	52	17.0
Separated	20	6.5
Living with partner	19	6.2
Widowed	13	4.2
Missing	9	2.9
Caregiver highest level of education		
College degree	152	49.7
High school diploma/GED	97	31.7
No high school diploma	29	9.4
Graduate school	21	6.9
Missing	7	2.3
Family income		
\$0–\$19,000	66	21.6
\$20,000–\$39,000	99	32.4
\$40,000–\$59,000	44	14.4
\$60,000–\$79,000	35	11.4
\$80,000–\$99,000	14	4.6
\$100,000 or more	34	11.1
Missing	14	4.6

Note. GED = general equivalency diploma.

Manual of Mental Disorder (DSM)–oriented scale scores, and total problems scale scores (i.e., total problems, internalizing and externalizing scores). Validity and reliability of all scale scores have been documented (Achenbach & Rescorla, 2001). Ebesutani, Bernstein, Martinez, Chorpita, and Weisz (2011) found the YSR to yield reliable scores across younger (ages 7–10 years) and older (ages 11–14 years) youths, supporting the use of the YSR across both younger and older youths in our sample.

Children's Interview for Psychiatric Syndromes (child and parent versions; ChIPS/P–ChIPS; Weller, Weller, Teare, & Fristad, 1999a, 1999b). The ChIPS and P–ChIPS are semistructured interviews designed to be administered to youths (aged 6–18 years) and their parents by trained assessors[2] to identify Axis I disorders as well as psychosocial stressors based on the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) classification criteria. Concurrent validity, interrater reliability, and test–retest of the ChIPS/P–ChIPS outcomes have been demonstrated in clinical and community samples (e.g., Fristad, Teare, Weller, Weller, & Salmon, 1998).

Revised Child Anxiety and Depression Scales (child and parent versions; RCADS/ RCADS–P; Chorpita, Yim, Moffitt, Umemoto, & Francis, 2000; Ebesutani, Bernstein, Nakamura, Chorpita, & Weisz, 2010). The RCADS and RCADS–P are each a 47-item questionnaire designed to assess *DSM–IV* depression and anxiety disorders in youths. They are each comprised six subscales: Separation Anxiety Disorder, Social Phobia, Generalized Anxiety Disorder, Obsessive-Compulsive Disorder, Panic Disorder, and Major Depressive Disorder. They also yield an anxiety total score and total internalizing problems score. The RCADS and RCADS–P ask youths and their parents to rate items according to how often each applies to the youth. Responses range from 0–3, corresponding to *never*, *sometimes*, *often*, and *always*. The RCADS scores have been shown to have good reliability, high convergent and discriminant validity, and an adequate factor structure in both community and clinical samples of youths ages 7–17 years (Chorpita et al., 2000; Chorpita, Moffitt, & Gray, 2005). The RCADS–P scores also demonstrated good internal consistency, test–retest, and convergent and discriminant validity in a clinical and a school-based sample (Ebesutani, Bernstein, Nakamura, Chorpita, & Weisz, 2010; Ebesutani, Chorpita, et al., 2011).

Top Problems measure (TP; Weisz et al., 2011). The TP is an idiographic, client-guided measure that was administered separately to youths and caregivers following each structured interview. Youths and caregiver were asked to list the problems they were most concerned about. The interviewer wrote these down in the respondent's own words, and then asked which one "*is the biggest problem right now? Which of these is giving you* [or youth's name] the most trouble right now? Which one is the most important to work on?" These procedures led to identification of the most concerning problem area endorsed by the youth and parent (ranked first) as well as other problem areas associated with lower rankings. Weisz et al. (2011) reported on a large subset of the current study's sample and found support for the psychometric strength of the TP measure. The TP scores demonstrated good convergent validity. For example, the youth TP total score correlated significantly with the YSR total score at .25, and the care-

[2] Assessors were PhD-level clinical child psychologists and senior doctoral students in clinical psychology. Although interrater reliability data of these structured interviews were not gathered, assessors were trained to reliability. Becoming trained to reliability involved (a) observation of three *P–ChIPS/*ChIPS interviews conducted by trained assessors, (b) conducting a series of five *P–ChIPS/*ChIPS interviews while being observed by a trained assessor, (c) matching the trained-assessor on all clinical diagnoses in at least three of the five interviews, and (d) matching the experienced interviewer on clinical severity ratings (CSRs) within at least 1 point on all diagnoses given. CSRs ranged from 0–10, and CSRs $\geq$ 5 indicated clinically significant severity for each disorder. The CSR procedures were based on the procedures used with the Anxiety Disorders Interview Schedule for *DSM–IV*, Child and Parent Versions (ADIS–C/P; Silverman & Albano, 1996). In order for a trainee interviewer's diagnostic results to have "matched" the criterion-trained interviewer's diagnostic results, the full diagnostic profiles (consisting of the presence and absence of all diagnostic categories, as well as the designation of which diagnosis or diagnoses were primary) were required to match.

Table 3
Means, Standard Deviations and Internal Consistency Reliability Estimates for the YSR, CBCL, RCADS, and RCADS–P Scales

Scale	*M*	*SD*	Internal consistency
Youth Self-Report			
DSM-oriented scales			
Affective Problems	5.19	3.99	.73
Anxiety Problems	2.97	2.40	.66
Somatic Problems	3.39	3.04	.78
ADH Problems	5.32	3.45	.78
Oppositional Problems	2.87	2.34	.75
Conduct Problems	3.66	3.86	.80
Syndrome Scales			
Anxious/Depressed	5.26	4.31	.79
Withdrawn/Depressed	3.85	2.91	.67
Somatic Complaints	5.10	3.99	.80
Social Problems	4.80	3.84	.75
Thought Problems	5.00	4.07	.75
Attention Problems	5.92	3.92	.79
Rule Breaking Behavior	3.21	3.37	.75
Aggressive Behavior	7.14	5.66	.85
Broad Band Scales			
Internalizing Total	14.21	9.41	.88
Externalizing Total	10.36	8.32	.89
Total Score	24.56	15.72	.92
Child Behavior Checklist			
DSM-oriented scales			
Affective Problems	5.48	3.97	.74
Anxiety Problems	4.11	2.88	.75
Somatic Problems	2.23	2.47	.73
ADH Problems	6.67	3.66	.82
Oppositional Problems	5.07	2.72	.80
Conduct Problems	5.52	5.08	.85
Syndrome Scales			
Anxious/Depressed	7.92	5.05	.83
Withdrawn/Depressed	4.67	3.39	.78
Somatic Complaints	3.49	3.35	.75
Social Problems	5.55	4.06	.77
Thought Problems	4.76	3.97	.73
Attention Problems	8.21	4.70	.85
Rule Breaking Behavior	4.51	3.72	.75
Aggressive Behavior	11.85	7.31	.90
Broad Band Scales			
Internalizing Total	16.08	9.50	.88
Externalizing Total	16.36	10.35	.91
Total Score	32.45	16.39	.92
RCADS			
Generalized Anxiety Disorder	4.75	3.78	.81
Separation Anxiety Disorder	4.24	4.31	.80
Obsessive-Compulsive Disorder	4.20	3.86	.79
Social Phobia	7.72	5.41	.84
Panic Disorder	3.95	4.30	.84
Major Depressive Disorder	6.02	4.82	.80
Anxiety Total	24.87	18.14	.94
Total Score	30.89	21.90	.95
RCADS–P			
Generalized Anxiety Disorder	5.25	4.03	.87
Separation Anxiety Disorder	4.79	4.46	.83
Obsessive-Compulsive Disorder	1.72	2.90	.83
Social Phobia	9.81	5.51	.86
Panic Disorder	2.58	3.09	.80
Major Depressive Disorder	6.68	4.65	.80
Anxiety Total	24.15	15.80	.93
Total Score	30.84	19.04	.94

Note. *DSM* = *Diagnostic and Statistical Manual of Mental Disorders*; YSR = Youth Self-Report; CBCL = Child Behavior Checklist; RCADS = Revised Child Anxiety and Depression Scale; RCADS–P = Revised Child Anxiety and Depression Scale—Parent Version; ADH = attention-deficit/hyperactivity.

giver TP total score correlated significantly with the CBCL total score at .32. Discriminant validity of the TP scores was also supported as the correlations between the TP scores and almost all measures of theoretically distinct constructs were nonsignificant. The TP scores also provided useful and incremental information above and beyond that obtained by the CBCL and YSR scores. For example, the identified top problem involved unique content (i.e., not matching any item on any narrowband scale in the clinical range) for 41% of caregivers and 79% of youths. Additional psychometric analysis focused on caregivers and youths rating the severity of identified problems weekly, during treatment. Evidence on test–retest reliability, sensitivity to change, slope reliability, and the association of TP slopes with scores from standardized measure further supported the TP's psychometric strength. For instance, test–retest reliability estimates of the TP Internalizing, Externalizing, and Total scale scores were uniformly high and significant ($p < .01$), ranging from .69 to .91.

Procedures

Legal guardians of all participating youths underwent standardized institutional review board-approved notice of privacy and consent procedures prior to any data collection. Following consent provided at the initial meeting, youths were asked to fill out the YSR and RCADS, while the youths' caregivers were asked to fill out the CBCL and RCADS–P. Youths and their caregivers also participated in separate structured interviews (i.e., the ChIPS and P–ChIPS). Following administration of the ChIPS and P–ChIPS structured interviews, the assessors had the youths and their parents independently provide clinical severity ratings (CSRs; 0–10) as well as rankings (from *most problematic* [1] to *least problematic* [highest number]) for each module endorsed on the ChIPS and P–ChIPS. Even if only one item was endorsed as problematic in a given section (e.g., the youth only indicated that he or she is "sometimes shy" in the ChIPS Social Anxiety module), that area was considered a potential problem area, and severity ratings and rankings were obtained. Both clinical problems (i.e., meeting diagnostic criteria on the ChIPS/ P–ChIPS) and subclinical problems (meeting some, but not all, diagnostic criteria) were thus included in this severity scoring and ranking process.[3] After the ChIPS (child diagnostic interview), assessors formulated child-based diagnoses (appearing in Table 1) based on information obtained only from the ChIPS interview procedures noted directly above. Similarly, assessors derived parent-based diagnoses following the P–ChIPS (parent-based) diagnostic interview, based on information obtained only from the P–ChIPS interview procedures noted above. These parent-based diagnoses also appear in Table 1.

After reviewing all information obtained from the assessment (i.e., information from ChIPS/P–ChIPS interviews, child- and parent-based severity and ranking scores for each problem area, raw scores and *T*-scores from the various YSR, CBCL, RCADS and RCADS–P subscales), the assessors formulated Consensus Diagnoses for each youth. Given that Consensus Diagnoses were

[3] Across all youths, 39% and 68% had more than one problem area module endorsed on the ChIPS and P–ChIPS, respectively, thereby requiring youths and parents to provide rankings across multiple problem areas endorsed.

Table 4
Parent–Child Agreement Between Corresponding YSR, CBCL, RCADS, and RCADS–P Scales

Scale	Correlations
ASEBA (CBCL & YSR) Scales	
DSM-oriented scales	
Affective Problems	.21*
Anxiety Problems	.27*
Somatic Problems	.21*
ADH Problems	.19*
Oppositional Problems	.30*
Conduct Problems	.45*
Syndrome Scales	
Anxious/Depressed	.26*
Withdrawn/Depressed	.10
Somatic Complaints	.19*
Social Problems	.26*
Thought Problems	.11
Attention Problems	.19*
Rule Breaking Behavior	.40*
Aggressive Behavior	.34*
Broad Band Scales	
Internalizing Total	.17*
Externalizing Total	.38*
Total Score	.19*
RCADS & RCADS–P Scales	
Generalized Anxiety Disorder	.26*
Separation Anxiety Disorder	.39*
Obsessive-Compulsive Disorder	.26*
Social Phobia	.24*
Panic Disorder	.17*
Major Depressive Disorder	.24*
Anxiety Total	.28*
Total score	.26*

Note. YSR = Youth Self-Report; CBCL = Child Behavior Checklist; RCADS = Revised Child Anxiety and Depression Scale; RCADS–P = Revised Child Anxiety and Depression Scale—Parent Version; ASEBA = Achenbach System of Empirically Based Assessment.
* $p < .01$.

derived from integrating information from different informants (i.e., child, parent) as well as from different assessment modalities (i.e., clinical interviews, self-report scales), it was possible for diagnoses to appear on a youth's Consensus Diagnostic profile that did not appear on his or her child-based or parent-based Diagnostic profile. Assessors also provided their own set of clinical severity ratings (CSRs) and rankings across all problem areas endorsed. The assessors derived these CSRs and rankings with their clinical supervisors, and the results were additionally reviewed along with a report of all scale scores by expert consultants.[4] The process of integrating assessment information thus combined clinical judgment, supervision, and expert consultation. Although no gold standard method yet exists for integrating multi-informant multi-method assessment data (De Los Reyes & Kazdin, 2005), the common "best estimate" approach in which multiple clinicians review all available data and use their clinical judgment to achieve consensus on a diagnosis has been found to produce results with good to excellent reliability (Klein et al., 1994). The method used in this study was similar to the "best estimate" approach but more rigorous, adding expert consultants and rules regarding agreement between multiple sources of information, as explained below.

Determining treatment need. Youths were deemed in need of mental health services if (a) one or more problems were endorsed in a given ChIPS or P–ChIPS problem area (e.g., sleeping problems related to depression), indicative of either clinical or subclinical elevations, (b) severity rankings provided by the youth and/or parent suggested significant impairment, and (c) there was at least one *T*-score in the borderline range from an RCADS, RCADS–P, YSR, or CBCL subscale corresponding to the problem area endorsed on the ChIPS/P–ChIPS. We used these three criteria for determining treatment eligibility for the following reasons: (a) this allowed for subclinical problems to be treated, which is important given that subclinical problems are associated with future emotional/behavioral problems and impairment (e.g., Gotlib, Lewinsohn, & Seeley, 1995); (b) youths with subclinical problems are often treated in community mental health settings (Jensen & Weisz, 2002), thereby increasing the generalizability of the present findings to community settings; and (c) we aimed to balance "real-world" applicability and generalizability (i.e., treating both clinical and subclinical problems) with the best practice approach of considering information from multiple sources and multiple methods—that is, clinician-guided structured interviews and self-report questionnaires. Treatment eligibility thus required problems to be endorsed on both *clinician-guided* interviews (subclinical or clinical levels) and *youth/parent self-reported* questionnaires (in at least the borderline range). Youths who did not meet these requirements were deemed not in need of mental health services at the present time for the four problems areas of anxiety, depression, ADHD, or disruptive behavior.

Determining treatment track. Once a youth was deemed in need of treatment for at least one of the areas of anxiety, depression, ADHD or disruptive behavior, our assessment team used the information on the TP measure and all other available assessment information to generate final *assessor-derived* rankings whereby the first-ranked problem was deemed the primary problem and indicated treatment. Given that youths and parents often provide discrepant reports (e.g., listing different first-ranked problems on the TP measure), the assessment team collectively and carefully considered all available information, using heuristics and clinical judgment to resolve discrepant reports, consistent with the best estimate approach.[5] Among the 252 youths deemed in need of some form of treatment related to anxiety, depression, ADHD, or disruptive behavior, 65 youths were deemed in need of anxiety treatment, 43 were deemed in need of depression treatment, 59 were deemed in need of ADHD treatment, and 85 were deemed in need of disruptive behavior treatment.

[4] The clinical supervisors in the present study were PhD-level postdoctoral fellows, and the expert consultants were members of the Research Network on Youth Mental Health who were each recognized in the field as having extensive backgrounds in the youth problem areas for which they consulted.

[5] Widely used heuristics were utilized to aid in resolving discrepancies between child and parent reports (e.g., giving more weight to parent reports for externalizing problems and more weight to youth reports for internalizing problems). However, guidelines for how to resolve such discrepancies between informants are not available for all situations (see De Los Reyes & Kazdin, 2005). Thus, the trained assessors, supervisors, and consultants had to use some degree of clinical judgment (factoring in informant reports related to severity and functional impairment) to resolve discrepant reports.

Data Analytic Approach

Although missing data levels were low,[6] we used the Missing Value Analysis module of SPSS 15.0 to examine missing data patterns and impute missing values. To help ensure that all subscale scores calculated were valid, each subscale was calculated only if it had at least 80% completed items. We used 80% as the cutoff for subscale inclusion (instead of 90%, as we used for the cutoff for the entire measure), to allow subscales with low item counts to have one item missing and still be calculated.

We employed decision-tree analysis using the Classification Tree module of SPSS 15.0 to identify optimal cutoff scores from the RCADS, RCADS–P, YSR, and CBCL to inform the treatment prescription decisions. Decision-tree analysis is a statistical technique that can be used to identify significant and meaningful patterns and rules hidden in data. Decision-tree analysis employs a hierarchical classification procedure that recursively partitions a data set into smaller subdivisions (or nodes). Beginning from the root node, branching occurs that leads to sets of internal nodes and, finally, to a set of terminal (leaf) nodes. These procedures and the associated growth method (detailed further below) are able to identify optimal cut-points from the most informative predictor variables (e.g., a set of scales) to best classify individuals with respect to the dependent grouping variable (e.g., group status related to treatment allocation).

Growth method. We utilized the chi-squared automatic interaction detection (CHAID) growth method for our decision-tree analysis. The CHAID algorithm examines all independent predictor variables (e.g., scale scores) simultaneously entered into the model and identifies those that have the strongest statistical interaction with the dependent variable (e.g., treatment group status) at each level of the tree until growth stopping criteria are met. The CHAID growth method bands scale scores into discrete groups (e.g., 0–2, 3–4, 5–6, etc.) prior to analysis to identify cut-points that allow for optimal classification. CHAID then merges banded score categories if they are not significantly different with respect to predicting the dependent variable (see Ture, Tokatli, & Kurt, 2009, for a more detailed description on the CHAID algorithm). The significance level for splitting and merging categories using CHAID was set to $p < .01$.

Minimum number of cases. A parameter must be specified to indicate the minimum number of cases needed to fall into a node before CHAID will create an actual split in the tree at that junction. This parameter helps to prevent splitting in the tree due to chance (i.e., too few cases falling into that category in ways that are not meaningful and/or generalizable to other samples). We set the minimum number of cases to five, with the goal of producing only splits with a substantial likelihood of generalizability.

Dependent (criterion) variables. As noted above, the criterion used for this study was the treatment allocation decision made via the "best estimate" method. This criterion indicated one of the following for each youth: (a) treatment not needed, (b) anxiety treatment needed, (c) depression treatment needed, (d) ADHD treatment needed, and (e) disruptive behavior treatment needed. To create the three different decision-tree models depicted in Figure 2 and corresponding to the treatment allocation decisions described above, this five-value criterion variable was used as-is (Decision Tree C; see Figure 2, third row) and also summarized to create (a) a binary criterion variable indicating whether each youth was deemed to be in need of *any* treatment (used to generate Decision Tree A; see Figure 2, first row) and (b) a three-value criterion variable indicating whether each youth was deemed to need internalizing treatment, externalizing treatment, or no treatment (used to generate Decision Tree B; see Figure 2, second row).

Independent (predictor) variables. For all decision-tree analyses, we simultaneously entered all raw scores and *T*-scores from the YSR, CBCL, RCADS, and RCADS–P subscales related to anxiety, depression, ADHD, and disruptive behavior into the model.[7] We also created and entered a composite variable to approximate the "best estimate" assessment integration procedures described above. Specifically, given that one of the criteria for treatment eligibility was having at least one *T*-score in the borderline range on a relevant subscale, we created a binary variable that was scored "1" if the youth had at least one *T*-score in the borderline range or higher on *any* of the subscales related to anxiety, depression, ADHD or disruptive behavior and "0" if *none* of the relevant subscales reached the borderline range. Similarly, we also entered a summary variable indicating the presence of at least one *T*-score in the *clinical* range on a relevant scale.[8]

---

[6] Of the 306 participants, 92.2%, 87.6%, 100%, and 82.4% had no missing RCADS, RCADS–P, YSR, and CBCL items, respectively; 6.2%, 7.8%, and 12.7% had only one missing RCADS, RCADS–P, and CBCL item, respectively.

[7] Predictor variables: anxiety-related: RCADS/RCADS–P Anxiety total raw and *T*-score scales and YSR/CBCL *DSM*-oriented anxiety problems raw and *T*-score scales; depression-related: RCADS/RCADS–P Depression raw and *T*-score scales and YSR/CBCL *DSM*-oriented affective problems raw and *T*-score scales; ADHD-related: YSR/CBCL syndrome attention problems raw and *T*-score scales and YSR/CBCL *DSM*-oriented ADH Problems raw and *T*-score scales; disruptive behavior-related: YSR/CBCL Externalizing total, *DSM*-oriented oppositional and conduct problems scales, and syndrome aggressive behavior and rule-breaking behavior raw and *T*-score scales. We also added the RCADS/RCADS–P total score raw and *T*-score scales and the YSR/CBCL Total Internalizing, Total Externalizing, and Total Problems raw and *T*-score scales into all models as predictors. Both raw scores and *T*-scores were entered as available predictors for all decision tree models because the two types of scores have potentially different predictive power. For example, Achenbach and Rescorla's Achenbach System of Empirically Based Assessment (2001) manual recommends using raw scores when conducting analyses on the narrow-band scales in order to account for the full range of variation in these scales. Importantly, issues of multicollinearity due to the availability of both raw and *T*-scores are not a major concern for the current (CHAID-based decision tree) analyses. This is because predictors are essentially considered one at a time in CHAID when growing the decision tree (Kass, 1980). Multicollinearity thus has the effect of reducing the chances of highly correlated variables jointly appearing in the tree. Resulting trees should not suffer from typical threats of multicollinearity (e.g., instability of coefficient estimates, inflated predictive power), provided that the highly correlated predictors (e.g., the raw and *T*-score version of a given scale) do not both appear in the final tree (cf. Strambi & Van De Bilt, 1998; Horner, Fireman, & Wang, 2010).

[8] Based on published manuals and documentation, borderline ranges were as follows: CBCL/YSR Total Problems, Internalizing, and Externalizing *T*-scores = 60–63, syndrome and *DSM*-oriented *T*-scores = 65–69, all RCADS/RCADS–P *T*-scores = 65–69. Clinical cutoffs were as follows: CBCL/YSR Total Problems, Internalizing, and Externalizing *T*-scores > 63, syndrome and *DSM*-oriented *T*-scores ≥ 70, and all RCADS/RCADS–P *T*-scores ≥ 70 (Achenbach & Rescorla, 2001; Chorpita et al., 2000).

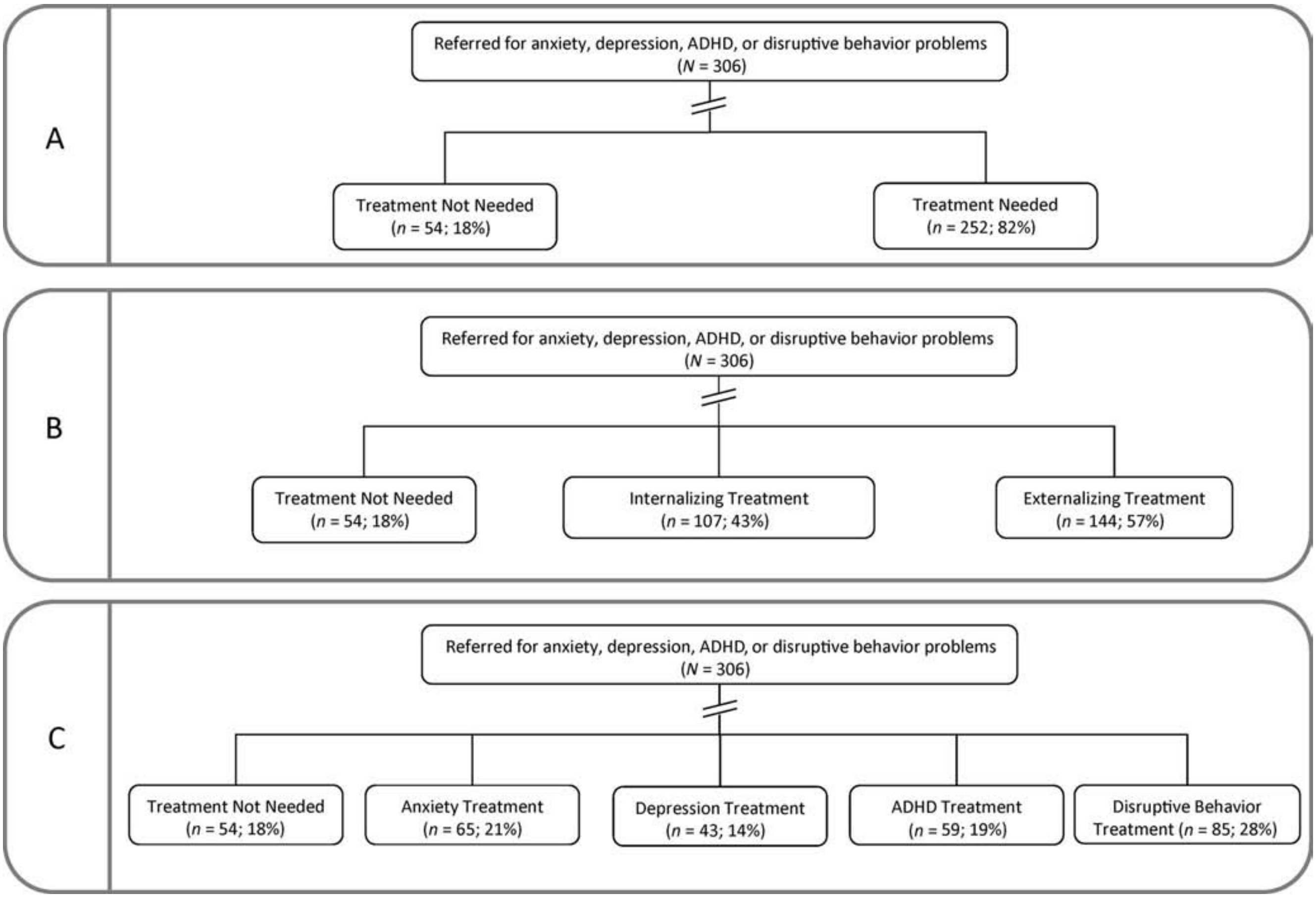

Figure 2. Decision trees for informing treatment prescription. Sample sizes and percentage (relative to the parent nodes) of the youths in each category determined by the gold standard assessment procedures in the present study appear in each box. The demarcations (//) between the root and terminal nodes symbolize where the sequential rule-based guidelines (i.e., a series of cutoff scale scores) apply, which if followed, would lead to group classification with "fair" to "excellent" accuracy. The specific decision rules may be found in the Results section. ADHD = attention-deficit/hyperactivity disorder.

Model validation. It is important that a decision tree be tested and validated for generalizability beyond the sample used to build its decision rules. We thus allowed CHAID to develop each decision tree using a random 70% of the full sample (i.e., the training sample; cf. Friedl & Brodley, 1997). We then used the remaining 30% of the full sample (i.e., the validation sample) to test the classification accuracy of the decision-tree rules generated in the training sample. All classification rates reported in the results section below are those based on this validation hold out sample. Classification accuracy rates of a decision-tree model based on a validation sample are similar to classification accuracy rates obtained via the methodology of receiver operating curve (ROC) analysis (cf. Steadman et al., 2000). Accuracy rates were thus interpreted according to the rubric commonly applied to prediction accuracy: .50–.70, poor; .70–.80, fair; .80–.90, good; and .90–1.00, excellent (cf. Ferdinand, 2008). Importantly, given that the decision-tree-derived guidelines may not adequately classify *all* youths with sufficient precision (as alluded to above), we only reported the algorithm-driven guidelines (i.e., scales and cutoffs) that classified youths in the validation sample with at least "fair" prediction accuracy. The "poor" range and below .50 prediction accuracy are unlikely to be sufficient for any clinical application, and thus, decision tree paths that sorted youths with this low degree of accuracy were not reported. Instead, all youths who fell outside the reported paths were classified as in need of further assessment effort. For this group, additional assessment or integration effort would be required to adequately determine the best treatment option (e.g., via supervisor-based integration of the existing information and/or by obtaining additional information through clinical interviews or more scales). The percentage of youths classified as in need of further assessment is reported in our results. For all three decision trees (A–C), we also reported the optimal cut-points identified for each predictor (to thus be used accordingly in "real world" clinical practice), the classification accuracy rates, and whether the classification performance of each predictor fell in the "poor," "fair," "good," or "excellent" range.

Cross-validation shrinkage. Given that decision tree rules generated in any development sample are expected to perform less well (i.e., with less predictive/classification accuracy) when applied to other samples, we also calculated and reported shrinkage estimates based on Pedhazur's (1982) shrinkage equation relevant for discriminant analyses (such as decision tree analyses): % correctly classified$_{screening}$ − % correctly classified$_{calibration}$ (cf. Collins et al., 1987). For each decision-tree model (A–C), we calculated shrinkage estimates based on (a) the entire development/screening and validation/calibration samples, as well as (b) only the portion of youths to whom the decision tree rules applied in both development/screening and validation/calibration samples.

Results

Decision Tree A

Any treatment needed? The first decision tree (Decision Tree A; see Figure 2) generated for determining whether any form of treatment was needed consisted of three sequential rules: (a) If the CBCL Total Problems raw score > 10, then treatment is needed; (b) if the CBCL Total Problems raw score ≤ 10 AND at least one child or parent *T*-score fell in the clinical range on any related scale, then treatment is needed; and (c) if the CBCL Total Problems raw score ≤ 10 AND no child or parent *T*-score fell in the clinical range on any related scale, then treatment is not needed. This algorithm covered 100% of the validation sample (n = 96) at a classification accuracy rate of 94%, falling in the "excellent" category.

Cross-validation shrinkage. Decision Tree A's classification rules were applicable to 100% of both the development and validation samples at a classification accuracy rate of 94%. These results revealed little to no shrinkage associated with these screening rules.

Decision Tree B

Internalizing or externalizing treatment protocol? The second decision tree (Decision Tree B) generated for determining whether internalizing, externalizing, or no treatment is needed consisted of the following sequential rules: (a) If the CBCL *DSM*-oriented conduct problems raw score > 7, then an externalizing treatment is needed; (b) if the CBCL conduct problems raw score = 0 AND at least one child or parent *T*-score fell in the clinical range on any relevant scale, then an internalizing treatment is needed; and (c) if the CBCL conduct problems raw score = 0 AND no child or parent *T*-score fell in the clinical range on any relevant scale, then no treatment is needed. Although this did not cover all youths, as predicted, this algorithm covered 41% of the validation sample at an accuracy rate of 83%, falling in the "good" category.[9] This corresponded to 38 (of the 93 total youths comprising the validation sample) being classified with "good" accuracy relative to our best practice assessment procedures. Of these 38 youths, 22 were accurately classified as needing an externalizing treatment, and 10 were accurately classified as needing an internalizing treatment. Four of the misclassified youths were inaccurately classified as needing in internalizing treatment, one was incorrectly classified as needing an externalizing treatment, and one was incorrectly classified as needing no treatment. The remaining youths with CBCL conduct problems scores in the range of 1–6 were unable to be classified adequately by this decision-tree model in any of the three groups, thereby informing the need of further assessment efforts for this group.

Cross-validation shrinkage. When applying Decision-Tree Model B to the *entire* development/screening sample, classification accuracy was 71%; when applying the same Decision-Tree Model B to the *entire* cross-validation hold-out (calibration) sample, classification accuracy dropped to 63%. The associated shrinkage estimate was thus .08 (i.e., .71–.63). When calculating shrinkage based on only those youths to whom the generated decision tree rules applied (i.e., 43% and 41% of the development/screening sample and validation/calibration sample, respectively), shrinkage was near zero, as the classification accuracy rates in the screening/development and calibration/validation samples were both 83%.

Decision Tree C

Anxiety, depression, ADHD, disruptive behavior or no treatment? The third decision tree (Decision Tree C) generated for determining whether anxiety, depression, ADHD, disruptive behavior, or no treatment is needed consisted of the following sequential rules: (a) If the CBCL *DSM*-oriented conduct problems raw score > 7, then a *disruptive behavior* treatment is needed; (b) if the conduct problems raw score = 0 AND at least one child or parent *T*-score fell in the clinical range on any relevant scale, then an *anxiety* treatment is needed; and (c) if the conduct problems raw score = 0 AND no child or parent *T*-score fell in the clinical range on any relevant scale, then *no* treatment is needed. This algorithm covered 44% of the validation sample at an accuracy rate of 79%, falling in the "fair" category. This corresponded to 38 (of the 93 total youths comprising the validation sample) being classified with "fair" accuracy relative to our best practice assessment procedures. Of these 38 youths, 19 were accurately classified as needing disruptive behavior treatment, and 11 were accurately classified as needing anxiety treatment. Three of the misclassified youths were inaccurately classified as needing no treatment, two were incorrectly classified as needing anxiety treatment, two were incorrectly classified as needing ADHD treatment, one was incorrectly classified as needing depression treatment, and one was incorrectly classified as needing disruptive behavior treatment. As

[9] Given that the CBCL/YSR Internalizing and Externalizing total scales were developed to aid in this type of discrimination (Achenbach & Rescorla, 2001), we examined a decision tree model in which the Externalizing scale was used instead of the Conduct Problems scale (by eliminating the Conduct Problems scale as a predictor from the model). This tree performed nearly as well. Specifically, the following two sequential rules were generated: (a) If the CBCL externalizing raw score > 20, then a disruptive behavior treatment is needed; (b) if the CBCL externalizing raw score ≤ 7 AND at least one child or parent *T*-score fell in the clinical range on any relevant scale, then an internalizing treatment is needed. All remaining youths, however, were unable to be adequately classified. This rule was associated with a slightly lower classification accuracy rate (80% versus 83%), and it classified 43% of the validation sample. Interestingly, we were unable to grow a decision tree that used the CBCL or YSR *internalizing* scale as a predictor for this classification purpose. This may have been due to externalizing youths (i.e., youths needing an externalizing treatment protocol) also exhibiting internalizing symptoms. To examine this hypothesis, we conducted an analysis of variance (ANOVA) between these groups using the YSR internalizing raw score as the criterion dependent measure for youth internalizing problems. Internalizing and externalizing youths did not significantly differ on this scale (p > .05), suggesting that internalizing problems may also be present among youths needing externalizing treatment.

with Decision Tree B, the remaining youths with CBCL conduct problems scores in the range of 1–6 were unable to be classified adequately in any of the five categories. Interestingly, youths were unable to be classified adequately in the *depression treatment* group with any of the depression scales.[10] Youths could also not be adequately classified in the ADHD treatment group.

Cross-validation shrinkage. When applying Decision Tree Model C to the entire development/screening sample, classification accuracy was 60%; when applying the same Decision Tree Model C to the entire cross-validation hold out (calibration) sample (b), classification accuracy again dropped to 48%. The shrinkage estimate associated with this scenario was .12 (i.e., .60–.48). When calculating shrinkage based on only those youths to whom the generated decision tree rules applied (i.e., 38% and 44% of the development/screening sample and validation/calibration sample, respectively), shrinkage was again near zero, as the classification accuracy rates in the screen/development and calibration/validation samples were 77% and 79%, respectively.

Discussion

The present study demonstrated the feasibility of developing an assessment protocol associated with low administration and interpretation burden to match treatment determinations based on more time-intensive and costly “best estimate” assessment procedures that incorporated clinician-guided structured interviews, child/parent self-report scales, and trained-assessors, supervisors, and expert consultants to integrate and interpret assessment information. Through use of decision-tree statistical procedures, the present study demonstrated the feasibility of identifying a set of predictors and cutoff scores that could serve as general guidelines to aid in the different, yet related, clinical treatment prescription decisions in community settings to adequately classify clinic-referred youths. Although more development is needed to further refine these algorithms before they may be applied with confidence in “real-world” community settings, several key results were demonstrated in the present study that are worth noting.

First, we were able to generate algorithm-driven guidelines that were able to classify 100% of youths with “excellent” accuracy regarding whether some form of treatment is needed (see Results for the three sequential rules). This algorithm relied largely on the CBCL Total Problems raw scale score (cutoff of 10), which is consistent with previous studies showing that this scale can aid in the discrimination between referred and nonreferred youths (Achenbach & Rescorla, 2001). Interestingly, the CBCL (parent report) Total Problems score was selected over the YSR (youth report) Total Problems score with respect to maximally informing this discrimination. Without further research, however, it is unclear whether the CBCL Total Problems score was identified by our decision-tree model over the YSR Total Problems score because (a) parents provided more informative reports relative to youths themselves and/or (b) our criterion, the “best estimate” gold standard assessment procedures (consisting of trained assessors, supervisors and consultants), favored parental reports over youth reports. Nonetheless, the CBCL Total Problems score was found to be particularly useful for this clinical decision and thus is recommended for such interpretation. This discrimination also relied on a composite score based on the “or-rule” (i.e., whether at least one child or parent *T*-score fell in the clinical range on any related scale). This is a simple heuristic that may be easily applied in community settings (i.e., simply reviewing scales for clinical elevations on relevant scales), making this a feasible algorithm for implementation.

For the second decision-tree model (Decision Tree B), we were able to generate three sequential rules (see Results section for specific rules) that classified 41% of the validation sample at 83% accuracy with respect to needing an internalizing treatment, externalizing treatment, or no treatment. Although 59% of youths were classified as needing further assessment, this three-rule algorithm could have been used in our community-based sample to identify the need for internalizing, externalizing, or no treatment for over 40% of the referred youths with “good” accuracy, substantially decreasing the overall assessment burden faced by a clinic and its clients. As noted above, these algorithm-driven guidelines may be particularly useful for clinics who take a “trans-diagnostic” approach to treatment, whereby, for example, related problem areas—ADHD and disruptive behavior (Chorpita, Daleiden, et al., 2011), and anxiety and depression (Barlow et al., 2004)—are treated with the same empirically supported protocol that applies to both problem areas. Regarding the youths who were unable to be classified by the present algorithm for this level of treatment allocation (i.e., Decision Tree B), additional assessment and/or manual integration methods with trained assessors and supervisors (that go beyond the algorithm-driven guidelines generated in the present study) would be needed to classify youths at this level until future studies generate additional guidelines that adequately classify these youths. As noted above, subsequent application of more thorough assessment/integration methods may be viewed as a “hybrid” or dynamic assessment approach, somewhat similar to Chorpita and Nakamura’s (2008) recent demonstration of a dynamic interview algorithm that identified sections on the Anxiety Disorders Interview Schedule for *DSM–IV* that may be skipped based on scores on the RCADS. Through such an approach, the present algorithm-driven guidelines may be used by community clinics to classify as many youths as possible in these three categories, with the remaining unclassified youths administered more thorough assessment procedures as needed.

The third decision-tree model (for classifying youths with respect to more specific treatment need related to anxiety, depression, ADHD, disruptive behavior or no treatment; Decision Tree C) comprised three similar sequential rules (see Results for specific rules). These algorithm-driven guidelines performed at a somewhat lower accuracy level (i.e., “fair”) relative to the guidelines generated in Decision Trees A and B, which was expected, given the increased specificity demands required at this level of classification. Still, however, over 40% of the validation sample

[10] To further examine this, we conducted ANOVAs between the youths who received depression treatment ($n = 43$) and the youths who received anxiety treatment ($n = 65$). Using the RCADS and RCADS–P Depression and Anxiety total scale scores for the criterion measures of depression and anxiety, respectively, results demonstrated that these two groups’ Anxiety and Depression scores did not significantly differ ($p > .01$). These results support the notion that these Anxiety and Depression scales did not aid the clinical decision of whether youths were in need of a depression protocol, as opposed to an anxiety protocol, potentially due to the high comorbidity between anxiety and depression (Brady & Kendall, 1992).

was classified by these guidelines that sorted youths in three of the five targeted categories (i.e., no treatment, disruptive behavior treatment and anxiety treatment). Interestingly, none of the depression scales (the RCADS, RCADS–P, YSR, or CBCL affective-related scales) were identified by the decision tree model to aid in classifying youths in need of depression treatment. Although more research is needed to understand why this was the case, this may be due to the high comorbidity found between anxiety and depression (Brady & Kendall, 1992). In the future, researchers in this area should consider using other approaches to target depressed youths, such as via positive affect (PA) scales, given that PA scales have been shown to be particularly useful for identifying depressed youths apart from youths with other types of psychopathology (Chorpita & Daleiden, 2002). More research is also needed to classify youths needing ADHD treatment.

Another interesting finding of Decision Tree C was that a subset of youths was classified reasonably well with respect to needing anxiety treatment if they endorsed no conduct problems per parent report (i.e., the CBCL *DSM*-oriented conduct problems raw score = 0) and if at least one child or parent *T*-score fell in the clinical range on any relevant scale. This finding is particularly notable for the following reasons. First, although very few guidelines exist in general regarding how to integrate data from multiple sources (cf. Klein, Dougherty, & Olino, 2005), one previously existing guideline has been that youth self-reports should be given more weight relative to parent/teacher reports when assessing internalizing problems (based on the assumption that youths have more insight on their internal states). The current findings, however, suggest that parent reports on children's externalizing (conduct) behaviors may also strongly inform need for anxiety treatment. Second, although more research is needed to understand why an externalizing scale (i.e., the CBCL *DSM*-oriented conduct problems scale) performed the best with respect to identifying youths needing anxiety treatment (as opposed to, for example, an anxiety scale), this finding suggests that there may be a subset of youths with heightened and/or particular types of anxiety associated with little to no externalizing symptoms due to their anxiety serving as a preventative buffer against conductlike behaviors. Consistent with this idea, some studies have found that anxiety does appear to serve as a preventative buffer against externalizing behaviors (e.g., Woolston et al., 1989). These findings are also consistent with Gray's (1982) research on the neurobiological systems of the Behavioral Activation System (BAS) and Behavioral Inhibition System (BIS), showing that these two systems work in opposition to each other, as well as Barkley's (1997) ADHD model that links ADHD (externalizing) symptoms with particularly low BIS activity. Consequently, BIS activity—typically associated with anxiety—may thus compete with the presentation of externalizing symptoms. Interestingly, however, Verhulst & van der Ende (1993) and Ollendick, Seligman, and Butcher (1999) did not find support for such palliative effect of BIS activity/anxiety on externalizing behaviors. Although results are mixed on this issue, the present findings adds to the literature on this topic and suggest that there may be a *subset* of anxious youths whose anxiety does have a palliative/buffering effect on externalizing behaviors (who thus may be identified by such characteristics). More research, however, is needed to determine whether this is the case and, if so, what specifically characterizes the type(s) of anxiety that buffer against externalizing behaviors.

Across all models, it is interesting to note that the most informative scales identified by the decision-tree analysis upon which to base predictions of treatment need were CBCL externalizing scales. These findings highlight the importance of obtaining parent reports of externalizing problems, even when assessing youth internalizing problems. Therefore, even if a community clinic specializes in the treatment of anxiety and depression, the present findings suggest that it would be particularly useful to also assess for externalizing problems via parent reports to aid in treatment allocation. It is also notable that the CBCL *DSM*-oriented conduct problems scales were identified as a highly informative scale for this treatment prescription process (i.e., identifying youths in need of externalizing treatment, internalizing treatment or an anxiety treatment in Decision Trees B and C). This finding supports the utility of the recently developed *DSM*-oriented scales, which are comparatively underexamined relative to the syndrome scale. This support for the utility of these scales is particularly important given that a recent study found that the CBCL *DSM*-oriented scales did not evidence incremental clinical utility above the syndrome scales with respect to corresponding with *DSM* diagnoses (with the exception of the *DSM*-oriented anxiety problems scale; Ebesutani, Bernstein, Nakamura, Chorpita, Higa-McMillan, & Weisz, 2010). The present findings thus provide support for the incremental clinical utility of the CBCL *DSM*-oriented conduct problems scales relative to the other scales and measures considered in this study for informing treatment allocation at these levels.

Regarding limitations pertaining to dissemination feasibility, two of the four self-report scales included in the present assessment procedures were proprietary instruments—including the CBCL, which was identified as one of the most informative measures. Since these measures (i.e., the CBCL and YSR) are not available for free, future studies should identify ways to reduce assessment burden still further, such as by also considering including freely available parent and child self-report measures that assess disruptive behavior/conduct problem areas, such as the Strength and Difficulties Questionnaire. Such future work could further reduce the fiscal burden of routine assessment procedures and concurrently increase the transportability and adoption of evidence-based assessment procedures in "real-world" community settings. Another limitation of the present study was that all decision trees (A–C) targeted only the four problem areas of anxiety, depression, ADHD, and disruptive behavior. Although these are arguably the most frequently presenting youth problems in clinic settings, other problems such as eating disorders and pervasive developmental disorders, which are found at lower base rates, are also very important to identify. Future studies should thus also attempt to target other problem areas to be maximally applicable to the diversity of clients seen in community settings. Since more than half of youths were classified as needing additional assessment or integration effort in Decision Trees B and C, additional studies are also needed to improve classification within the problem areas covered by this study. Using larger community-based samples and considering different scale options are both promising avenues for improving the classification rates and increasing the generalizability of integration guidelines developed in this area of research. Relatedly, it is important to note that we used an "internal" cross-validation strategy to obtain less inflated classification accuracy estimates relative to the classification accuracy estimates associated with our development sample. This was one strategy to obtain classification accuracy estimates that would be more reflective of the expected accuracy estimates of the decision tree rules if applied to other similar samples. Although "internal" cross-validation strategies are a commonly used method for reducing the inflation of classification accuracy estimates (Shtatland, Klienman, & Cain, 2004), "external" cross-validation (e.g., applying the decision tree rules to an independent and separate sample from a different referral source, with different base rates of problem behaviors, etc.) is ideal for estimating and reporting the classification accuracy rates of such decision tree rules—as these rates would be more reflective of the classification accuracy rates of decision tree rules applied in other settings (e.g., with different disorder base rates). Further research should be conducted to apply these decision-tree rules to such external samples to obtain more accurate estimates of the classification accuracy rates expected if applied in "real-world" settings.

Despite such limitations, the present study demonstrated the feasibility of utilizing decision-tree analysis to develop relatively less burdensome assessment procedures (in the form of integration guidelines) based solely on self-report scales to classify a large portion of youths referred to community clinic settings. For a considerable portion of the validation sample examined, these automated algorithms were able to match several outcomes of the "best estimate" integration assessment procedure, ranging from "fair" to "excellent" accuracy. The best estimate criterion used in this study is as close to a gold standard as presently exists and had access to rich additional information not considered by the decision tree, including structured interview results and client- and assessor- determined top-problem rankings. Although the algorithms and scale-based classification rules developed in the present study did not have perfect correspondence (i.e., 100% accuracy) relative to the resource-intensive best estimate assessment procedure, this initial feasibility study demonstrated substantial promise with respect to developing assessment procedures that (a) are more likely to be adopted and implemented in community settings (due to being less burdensome) and (b) still provide a degree of predictive power (ranging from "fair" to "excellent") and assistance in guiding the treatment prescription process. Further, although not all youths were classified adequately in Decision Trees B and C—with further refinement and additional cross-validation of these algorithm rules in independent samples in future studies—clinics may be able to use similar assessment protocols and algorithm as reported in the present study to support clinical treatment decisions in their clinic settings pertinent to Decision Trees A, B, and/or C while minimizing overall assessment burden on clinic resources.

Relatedly, when developing assessment instruments and procedures to be adopted by "real-world" community clinicians, a trade-off exists between the burden and the likelihood of adoption. Accordingly, as we move forward, the utility of assessment procedures designed for implementation in "real-world" settings should not be measured solely on the basis of their classification accuracy. Although a considerable degree of accuracy should be a minimum requirement, just as important may be "transportability properties," such as being associated with low assessment burden. This may be seen as a critical distinction between *efficacious* assessment procedures (procedures associated with high precision, even at the cost of high burden and thus low transportability) and *effective* assessment procedures that have increased transportability due to lower assessment burden properties. We hope that this distinction between efficacious and effective assessment procedures provides a guiding framework for researchers when developing additional assessment tools for use in community settings. Low burden assessment procedures may play an important role in increasing the use of evidence-based assessment procedures in community settings and, thus, ultimately better serving youths in the "real-world."

References

Abikoff, H., Hechtman, L., Klein, R. G., Weiss, G., Fleiss, K., Etcovitch, J., . . . Pollack, S. (2004). Symptomatic improvement in children with ADHD treated with long-term methylphenidate and multimodal psychosocial treatment in children with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry, 43,* 802–811. doi:10.1097/01.chi.0000128791.10014.ac

Achenbach, T. M., & Rescorla, L. A. (2001). *The manual for the ASEBA school-age forms & profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

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

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121,* 65–94. doi:10.1037/0033-2909.121.1.65

Barlow, D., Allen, L., & Choate, M. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy, 35,* 205–230. doi:10.1016/S0005-7894(04)80036-4

Brady, E. U., & Kendall, P. C. (1992). Comorbidity of anxiety and depression in children and adolescents. *Psychological Bulletin, 111,* 244–255. doi:10.1037/0033-2909.111.2.244

Chorpita, B. F., Bernstein, A., & Daleiden E. L. (2011). Empirically guided coordination of multiple evidence-based treatments: An illustration of relevance mapping in children's mental health services. *Journal of Consulting and Clinical Psychology, 79,* 470–480. doi:10.1016/S0005-7967(99)00130-8

Chorpita, B. F., & Daleiden, E. L. (2002). Tripartite dimensions of emotion in a child clinical sample: Measurement strategies and implications for clinical utility. *Journal of Consulting and Clinical Psychology, 70,* 1150–1160. doi:10.1037/0022-006X.70.5.1150

Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., . . . Starace, N. (2011). Evidence based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice, 18,* 154–172.

Chorpita, B. F., Moffitt, C., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy, 43,* 309–322. doi:10.1016/j.brat.2004.02.004

Chorpita, B. F., & Nakamura, B. (2008). Dynamic structure in diagnostic structured interviewing: A comparative test of accuracy and efficiency. *Journal of Psychopathology and Behavioral Assessment, 30,* 52–60. doi:10.1007/s10862-007-9075-8

Chorpita, B. F., Reise, S., Weisz, J. R., Grubbs, K., Becker, K. D., Krull, J. L., & The Research Network on Youth Mental Health. (2010). Evaluation of the Brief Problem Checklist: Child and caregiver interviews to measure clinical progress. *Journal of Consulting and Clinical Psychology, 78,* 526–536. doi:10.1037/a0019602

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L., & Francis, S. (2000). Assessment of symptoms of *DSM–IV* anxiety and depression in children: A Revised Child Anxiety and Depression Scale. *Behaviour Research and Therapy, 38,* 835–855.

Collins, L., Sussman, S., Rauch, J., Dent, C., Johnson, A., Hansen, W., & Flay, B. (1987). Psychosocial predictors of young adolescent cigarette smoking: A sixteen-month, three-wave longitudinal study. *Journal of Applied Social Psychology, 17,* 554–573. doi:10.1111/j.1559-1816.1987.tb00330.x

De Los Reyes, A., & Kazdin, A. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131,* 483–509. doi:10.1037/0033-2909.131.4.483

Ebesutani, C., Bernstein, A., Martinez, J. I., Chorpita, B. F., & Weisz, J. R. (2011). The Youth Self-Report: Applicability and validity across younger and older youths. *Journal of Clinical Child and Adolescent Psychology, 40,* 338–346.

Ebesutani, C., Bernstein, A., Nakamura, B., Chorpita, B. F., Higa-McMillan, C., & Weisz, J. (2010). Concurrent validity of the Child Behavior Checklist *DSM*-oriented scales: Correspondence with *DSM* diagnoses and comparison to syndrome scales. *Journal of Psychopathology and Behavioral Assessment, 32,* 373–384. doi:10.1007/s10862-009-9174-9

Ebesutani, C., Bernstein, A., Nakamura, B. J., Chorpita, B. F., & Weisz, J. R. (2010). A psychometric analysis of the Revised Child Anxiety and Depression Scale—Parent Version in a clinical sample. *Journal of Abnormal Child Psychology, 38,* 249–260. doi:10.1007/s10802-009-9363-8

Ebesutani, C., Chorpita, B. F., Higa-McMillan, C., Nakamura, B. J., Regan, J., & Lynch, R. (2011). A psychometric evaluation of the Revised Child Anxiety and Depression Scale–Parent version in a school sample. *Journal of Abnormal Child Psychology, 39,* 173–185.

Ferdinand, R. F. (2008). Validity of the CBCL/YSR *DSM–IV* scales anxiety problems and affective problems. *Journal of Anxiety Disorders, 22,* 126–134. doi:10.1016/j.janxdis.2007.01.008

Friedl, M., & Brodley, C. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment, 61,* 399–409. doi:10.1016/S0034-4257(97)00049-7

Fristad, M. A., Teare, M., Weller, E. B., Weller, R. A., & Salmon, P. (1998). Study III: Development and concurrent validity of the Children's Interview for Psychiatric Syndromes—Parent Version (P–ChIPS). *Journal of Child and Adolescent Psychopharmacology, 8,* 221–226. doi:10.1089/cap.1998.8.221

Garland, A. F., Kruse, M., & Aarons, G. A. (2003). Clinicians and outcome measurement: What's the use? *The Journal of Behavioral Health Services and Research, 30,* 393–405. doi:10.1007/BF02287427

Gotlib, I. H., Lewinsohn, P. M., & Seeley, J. R. (1995). Symptoms versus a diagnosis of depression: Differences in psychosocial functioning. *Journal of Consulting and Clinical Psychology, 63,* 90–100. doi:10.1037/0022-006X.63.1.90

Gray, J. A. (1982). *The neuropsychology of anxiety: An inquiry into the functions of the septo-hippocampal system.* Oxford, England: Oxford University Press.

Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice, 35,* 485–491. doi:10.1037/0735-7028.35.5.485

Hermann, R. C., & Palmer, R. H. (2002). Common ground: A framework for selecting core quality measures for mental health and substance abuse care. *Psychiatric Services, 53,* 281–287. doi:10.1176/appi.ps.53.3.281

Horner, S. B., Fireman, G. D., & Wang, E. W. (2010). The relation of student behavior, peer status, race, and gender to decisions about school discipline using CHAID decision trees and regression modeling. *Journal of School Psychology, 48,* 135–161. doi:10.1016/j.jsp.2009.12.001

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology, 3,* 29–51. doi:10.1146/annurev.clinpsy.3.022806.091419

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. doi:10.1037/0022-006X.70.1.158

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C. Applied Statistics, 29,* 119–127.

Kendall, P. C., Flannery-Schroeder, E., Panichelli-Mindel, S., Southam-Gerow, M., Henin, A., & Warman, M. (1997). Therapy for youths with anxiety disorders: A second randomized clinical trial. *Journal of Consulting and Clinical Psychology, 65,* 366–380. doi:10.1037/0022-006X.65.3.366

Klein, D. N., Dougherty, L., & Olino, T. (2005). Toward guidelines for evidence-based assessment of depression in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34,* 412–432. doi:10.1207/s15374424jccp3403_3

Klein, D. N., Ouimette, P. C., Kelly, H. S., Ferro, T., & Riso, L. P. (1994). Test–retest reliability of team consensus best-estimate diagnoses of axis I and II disorders in a family study. *The American Journal of Psychiatry, 151,* 1043–1047.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *The American Journal of Psychiatry, 160,* 1566–1577. doi:10.1176/appi.ajp.160.9.1566

Ollendick, T. H., Seligman, L. D., & Butcher, A. T. (1999). Does anxiety mitigate the behavioral expression of severe conduct disorder in delinquent youths? *Journal of Anxiety Disorders, 13,* 565–574. doi:10.1016/S0887-6185(99)00023-7

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). New York, NY: Holt, Rinehart, and Winston.

Reich, W. (2000). Diagnostic interview for children and adolescents (DICA). *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 59–66. doi:10.1097/00004583-200001000-00017

Richardson, L., & Austad, C. (1991). Realities of mental health practice in managed-care settings. *Professional Psychology: Research and Practice, 22,* 52–59. doi:10.1037/0735-7028.22.1.52

Shtatland, E. S., Klienman, K., & Cain, E. M. (2004). *A new strategy of model building in PROC LOGISTIC with automatic variable selection, validation, shrinkage and model averaging: SUGI 29 proceedings* (Paper No. 191–29). Cary, NC: SAS Institute. Retrieved from http://www2.sas.com/proceedings/sugi29/191–29.pdf

Silverman, W., & Albano, A. M. (1996). *The anxiety disorders interview schedule for children* (*DSM–IV*). San Antonio, TX: Psychological Corporation.

Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P., Robbins, P. C., Mulvey, E. P., . . . Banks, S. (2000). A classification tree approach to the development of actuarial violent risk assessment tools. *Law and Human Behavior, 24,* 83–100. doi:10.1023/A:1005478820425

Strambi, O., & Van De Bilt, K.-A. (1998). Trip generation modeling using CHAID, a criterion-based segmentation modeling tool. *Transportation Research Record: Journal of the Transportation Research Board, 1645,* 24–31. doi:10.3141/1645-04

Ture, M., Tokatli, F., & Kurt, I. (2009). Using Kaplan-Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5, and ID3) in determining recurrence-free survival of breast cancer patients. *Expert Systems with Applications, 36,* 2017–2026. doi:10.1016/j.eswa.2007.12.002

Verhulst, F. C., & van der Ende, J. (1993). "Comorbidity" in an epidemiological sample: A longitudinal perspective. *Journal of Child Psychology and Psychiatry, 34,* 767–783. doi:10.1111/j.1469-7610.1993.tb01070.x

Webster-Stratton, C., & Reid, M. (2003). The Incredible Years Parent, Teachers, and Children Training Series: A multifaceted treatment approach for young children with conduct problems. In A. Kazdin & J.

Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. 224–240). New York, NY: Guilford Press.

Webster-Stratton, C., & Spitzer, A. (1991). Development, reliability, and validity of the daily telephone discipline interview. *Behavioral Assessment, 13,* 221–239.

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S., . . . Langer, D. (2011). Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and track change during psychotherapy. *Journal of Consulting and Clinical Psychology, 79,* 369–380.

Weisz, J. R., Chu, B. C., & Polo, A. J. (2004). Treatment dissemination and evidence-based practice: Strengthening intervention through practitioner-researcher collaboration. *Clinical Psychology: Science and Practice, 11,* 300–307. doi:10.1093/clipsy.bph085

Weisz, J. R., Gray, J. S., Bearman, S. K., & Stark, K. (2008). *Therapist manual PASCET: Primary and secondary control enhancement training* (3rd ed.). Cambridge, MA: Harvard University.

Weller, E., Weller, R., Teare, M., & Fristad, M. (1999a). *Children's interview for psychiatric syndromes: ChIPS.* Washington, DC: American Psychiatric Association.

Weller, E. Weller, R., Teare, M., & Fristad, M. (1999b). *Children's interview for psychiatric syndromes—Parent version: P–ChIPS.* Washington, DC: American Psychiatric Association.

Woolston, J. L., Rosenthal, S. L., Riddle, M. A., Sparrow, S. S., Cicchetti, D., & Zimmerman, L. D. (1989). Childhood comorbidity of anxiety/affective disorders and behavior disorders. *Journal of the American Academy of Child & Adolescent Psychiatry, 28,* 707–713. doi:10.1097/00004583-198909000-00011

Received June 18, 2010
Revision received January 3, 2011
Accepted February 1, 2011 ■