Toward Guidelines for Evidence-Based Assessment of Depression in Children and Adolescents

Daniel N. Klein, Lea R. Dougherty, and Thomas M. Olino
Department of Psychology, University at Stony Brook.

We aim to provide a starting point toward the development of an evidence-based assessment of depression in children and adolescents. We begin by discussing issues relevant to the diagnosis and classification of child and adolescent depression. Next, we review the prevalence, selected clinical correlates, course, and treatment of juvenile depression. Along with some general considerations in assessment, we discuss specific approaches to assessing depression in youth (i.e., interviews, rating scales) and briefly summarize evidence on the reliability and validity of a few selected instruments. In addition, we touch on the assessment of several other constructs that are important in a comprehensive evaluation of depression (i.e., social functioning, life stress, and family history of psychopathology). Last, we highlight areas in which further research is necessary and conclude with some broad recommendations for clinical practice given the current state of the knowledge.

In recent years, there has been a growing emphasis on evidence-based medicine. In the field of mental health, this has been reflected by the interest in empirically supported treatments and the development of evidence-based practice guidelines (Chambless & Ollendick, 2000; Drake, Rosenberg, Teague, Bartels, & Torrey, 2003). Although the field of assessment has lagged behind, interest in developing empirically supported approaches to clinical assessment has also begun to emerge (Meyer et al., 2001).

A number of fundamental questions must be resolved to develop evidence-based approaches to clinical assessment. First, what criteria should be considered in evaluating the evidence for various assessment procedures and approaches? Reasonable criteria might include several areas that have traditionally been grouped under the rubric of reliability, such as interrater reliability, scale homogeneity, and test–retest stability, and several areas that have traditionally been considered aspects of validity, such as content validity, concurrent and discriminant validity, and construct validity. Other more specific considerations that are important in the assessment of psychopathology include diagnostic accuracy (e.g., sensitivity and specificity, which are aspects of convergent and discriminant validity, respectively), sensitivity to change, incremental validity, and treatment and prognostic utility.

Second, what constitutes an acceptable level of reliability or validity in each of these areas? Although rules of thumb exist for some of these criteria (e.g., kappa for levels of interrater reliability; Shrout, 2002), there are few guidelines for interpreting the levels of many forms of reliability and validity, and, in most instances, the interpretation depends on the context. For example, the meaning of an internal consistency coefficient depends on the breadth of the construct being assessed, and interpretations of interrater reliability and test–retest reliability coefficients depend on the study design (e.g., paired rater vs. independent interview) and the test–retest interval, respectively. Finally, some forms of validity (e.g., content and construct validity) are complex judgments that cannot be expressed quantitatively using a standard metric.

Third, measures may have acceptable levels of some forms of reliability and validity, unacceptable levels of other forms, and no information bearing on yet other forms. For example, many measures of depressive symptomatology exhibit high concurrent validity but poor discriminant validity and lack data addressing treatment utility and incremental validity. How should these disparate pieces of evidence be weighed and combined into an overall evaluation of the evidence?

Fourth, how many studies are necessary to conclude that there is support for a given form of reliability and validity? What if the results of different studies conflict? Should we insist that positive results be replicated before accepting them, and, if so, do the studies have to be conducted by independent groups?

Fifth, should an evidence-based approach to assessment focus on the level of specific instruments or on more general assessment approaches or strategies? It is more difficult to evaluate the reliability and validity of an approach, or group of measures, than a single instru-
ment. However, an approach- or strategy-based focus may be more clinically useful and could lead to the development of evidence-based assessment algorithms, detailing sequences of assessments that might vary depending on the results of the prior step.

Sixth, the empirical support for various assessment instruments or approaches may vary depending on their purpose. Although there are many reasons to conduct psychological assessments, some of the major purposes are diagnosis, prognosis, treatment planning, and treatment monitoring and evaluation. Some measures or approaches may have empirical support for some purposes but not others (e.g., rating scales may be useful for treatment monitoring and evaluation but not for diagnosis). In addition, the utility of assessment procedures may vary depending on the target population (e.g., developmental level, ethnicity), context (clinics vs. schools), and prevalence (or base rate) of the disorder.

Finally, evidence-based assessment presupposes a clear definition of the constructs to be assessed. For many forms of psychopathology, the diagnostic constructs are controversial and are, at best, approximations of reality (Kendell & Jablensky, 2003). In addition, the targets of treatment planning are often unclear and have limited empirical support, and key constructs in assessing course and outcome (e.g., response, remission, recovery, relapse, recurrence) may be poorly defined and based more on convention than data (E. Frank et al., 1991).

These are general issues that cut across most specific disorders and clinical problems and must be addressed before evidence-based assessments for specific forms of child and adolescent psychopathology can be developed. However, it is also important to consider these problems at the level of the individual disorder, as they can have disorder-specific manifestations. Moreover, there are additional, disorder-specific issues that require consideration. In this article, we consider some of the central issues that must be addressed in developing an evidence-based assessment of depression in children and adolescents. We begin by discussing the classification of child and adolescent depression; review the prevalence, selected clinical correlates, course, and treatment of juvenile depression; discuss specific approaches to assessing depression in children and adolescents; touch on the assessment of several other constructs that are important in a comprehensive evaluation of depression; highlight areas in which further work is necessary; and conclude with some broad recommendations given the current state of knowledge. Although considerable work will be necessary before an evidence-based approach to the assessment of child and adolescent depression can be developed, we believe that this is an important and worthwhile task and hope that this article will help provide a starting point for discussion.

**Diagnosis and Classification of Child and Adolescent Depression**

The recognition of juvenile depression is relatively recent, as prior to the 1970s most investigators and clinicians believed that depression was extremely rare in children. However, by the early 1980s it was apparent that many children and adolescents met adult criteria for major depressive disorder (MDD) and dysthymic disorder (DD; Carlson & Cantwell, 1980).

There are a number of controversies and unresolved issues in the diagnosis and classification of depression in general and in children and adolescents in particular. These issues include the continuity between child, adolescent, and adult depression; whether depression is a discrete entity or a region on a continuum of behavior; whether it should be classified using a categorical or dimensional system; what the boundaries are between depression and normal variations in mood and other forms of psychopathology; the identification of more homogeneous subtypes; whether there are age-specific expressions of depression; and the existence and manifestations of depression in very young children.

**Continuity**

There is considerable evidence for continuity between adolescent and adult depression; however, the evidence for continuity between prepubertal and adult depression is less consistent. The three major lines of evidence that have attempted to address the issue of continuity include studies of the clinical presentation, longitudinal course, and familial aggregation of depression. Depressed children, adolescents, and adults tend to exhibit similar symptoms, although, as discussed later, there may be some developmental variations (Carlson & Kashani, 1988). Follow-up studies indicate that depressed adolescents are at high risk for developing episodes of MDD as adults (Lewinsohn, Rohde, Klein, & Seeley, 1999; Weissman, Wolk, Goldstein, et al., 1999). However, follow-up studies of prepubertal children have yielded mixed results, with some studies indicating that depressed children are at increased risk for depression in adulthood and other studies failing to find evidence of increased risk, except perhaps in particular subgroups (Harrington, Fudge, Rutter, Pickles, & Hill, 1990; Weissman, Wolk, Wickramaratne, et al., 1999). Finally, family studies have found significantly higher rates of MDD in the first-degree relatives of depressed adolescents than adolescents with other forms of psychopathology and adolescents with no history of psychiatric disorder (Klein, Lewinsohn, Seeley, & Rohde, 2001). Family studies have also found higher rates of MDD in the relatives of depressed children than in the relatives of healthy children; however, comparisons of rates of depression in relatives of depressed children and children
with other psychiatric disorders have yielded inconsistent results (Kovacs, Devlin, Pollock, Richards, & Mukerji, 1997; Puig-Antich et al., 1989).

The possibility of discontinuity between prepubertal and postpubertal depression is further reinforced by sex differences in prevalence (discussed more later). Rates of depression are similar among boys and girls in childhood, but the rate of depression is approximately two times higher in female adolescents and adults. If there is truly a discontinuity between child and adolescent and adult depression, it would have important implications for assessment, as what we presently regard as a unitary construct would actually consist of two different disorders that are distinguished by the developmental period in which they are observed.

**Discrete Versus Continuous**

The question of whether mood disorders are discrete entities or regions on a continuum has been debated for much of the past century (Parker, 2000). This topic is not merely of academic interest, as the answer may provide clues regarding the nature of etiological factors, help refine the definition of disorders and subtypes, and indicate which assessment approaches and statistical models provide optimal power (Beauchaine, 2003; Ruscio & Ruscio, 2002).

Currently, the suite of taxometric procedures developed by Meehl (1995) is regarded as the best means of testing whether disorders are discrete or continuous. However, we are aware of only two studies that have applied these techniques to child and adolescent depression. Using a large community sample, Hankin, Fraley, Lahey, and Waldman (2005) did not find evidence that MDD is a discrete entity in either children or adolescents. However, in a clinical sample, Ambrosini, Bennett, Cleland, and Haslam (2002) reported that the melancholic subtype appeared to be discrete.

**Categorical Versus Dimensional Classification**

Although classification systems for psychopathology have typically used a categorical format, many investigators believe that dimensional systems are more appropriate (e.g., Widiger & Clark, 2000). This debate is frequently confused with the issue of discreteness, as proponents of categorical models tend to assume that psychiatric disorders are discrete entities, whereas advocates of dimensional models generally believe that they are regions on a continuum. However, it is possible to advocate a categorical approach to classification on the pragmatic grounds that it is an efficient means of summarizing and communicating information about clinical syndromes and guiding decision making for treatment and policy without necessarily believing that the disorders will ultimately prove to be etiologically discrete (Kessler, 2002; Klein & Riso, 1993). Indeed, this is the position taken by the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [DSM-IV]; American Psychiatric Association, 1994), which employs a categorical format but states that it does not assume that each mental disorder is a discrete entity.

Most studies using dimensional approaches to depression have simply summed the number of depressive symptoms. However, it is unlikely that dimensional classification systems that rely entirely on symptoms will be sufficient, as the literature on adults indicates that depressive symptoms wax and wane over time and are not stable across episodes, plus there is poor concordance between patients' reports of symptoms and behavioral ratings by trained observers (Klein, Shankman, & McFarland, in press). Hence, dimensional models must also consider key course variables such as the frequency and duration of episodes (recurrence and chronicity; Shankman & Klein, 2002).

**Boundaries**

Given a categorical classification system, a critical issue concerns the location of the boundaries between depression and normal variations in mood and other forms of psychopathology, such as the anxiety disorders. The distinction between mood disorders and normal variations in mood, nonpathological dysphoria, and responses to stress is complex, as it is influenced by a variety of cultural, social, and economic factors and has changed over time. Some investigators believe that the current boundaries are too broad and include a number of individuals with demoralization and transient responses to stress. The current boundaries may be particularly problematic in adolescence, a period that is characterized by major developmental transitions and high levels of emotional intensity (Arnett, 1999). Consistent with this possibility, Wickramaratne and Weissman (1998) reported that the rates of depression in offspring of depressed and nondepressed parents differed in childhood and young adulthood but that there were no differences in adolescence where there was a marked increase in rates regardless of parental diagnosis. On the other hand, some investigators believe that the current boundaries are too strict, pointing to evidence that subthreshold depressive symptoms are common in juveniles and adults and are frequently associated with significant functional impairment (e.g., Lewinsohn, Solomon, Seeley, & Zeiss, 2000).

There has also been a long-standing debate regarding the boundary between depression and anxiety; for example, are depressive and anxiety disorders distinct conditions, are they variants of a single disorder, or is there a third category of anxious depression that is distinct from both pure depression and pure anxiety? As
discussed later, numerous studies using both clinical and community samples have documented high rates of comorbidity between mood disorders and anxiety disorders (Angold, Costello, & Erkanli, 1999). The relation between depression and anxiety appears to be complex. For example, Eaves and Silburg (2003) recently reported evidence for three distinct pathways in which childhood anxiety influences the development of depression in adolescence: one in which genetic differences in anxiety create later genetic differences in depression; a second in which genes that affect early anxiety increase sensitivity to adverse life events, increasing risk for depression; and a third in which genes that increase risk to early anxiety increase exposure to depressogenic environmental influences. Moreover, the relations between anxiety and depression differ depending on the specific anxiety disorder under consideration (Silberg, Rutter, & Eaves, 2001).

Clark and Watson’s (1991) tripartite model is an important approach to describing the distinctive and overlapping aspects of depression and anxiety. The model parses the domain of anxiety and depression into three dimensions: general distress, anhedonia, and physiological hyperarousal. Depression and anxiety are both characterized by a high level of distress. However, depression is uniquely characterized by a high level of anhedonia, whereas anxiety is uniquely characterized by a high level of physiological hyperarousal. These predictions have been tested in a number of studies of children and adolescents and have generally been supported (e.g., Chorpita & Daleiden, 2002; Lonigan, Phillips, & Hooe, 2003). An important implication of the model for clinical assessment is that the distinction between depression and anxiety can be sharpened by emphasizing the two specific dimensions, anhedonia and physiological hyperarousal, and de-emphasizing symptoms reflecting general distress (e.g., negative affect, insomnia, difficulty concentrating).

Subtypes

Most investigators believe that depression is etiologically heterogeneous. A major focus of the literature on the classification of depression in adults has involved attempting to delineate more homogeneous subtypes (Klein et al., in press). In addition to the unipolar–bipolar distinction (see Youngstrom, Findling, Youngstrom, & Calbrese, 2005), subtypes have been proposed on the basis of symptomatology (e.g., psychotic, melancholic, atypical), course (e.g., age of onset, recurrent, chronic, and seasonal pattern), and the combination of both (e.g., MDD vs. DD vs. double depression). Unfortunately, the issue of heterogeneity has largely been ignored in child and adolescent depression, hence it is unclear whether most of the subtypes of adult depression can be meaningfully applied to children. An important exception, however, is age of onset. Consistent with the previous discussion about the possible discontinuity between child and adolescent depression, there is some evidence that pre- and postpubertal onset depressions may differ with respect to long-term course and familial aggregation (Weissman, Wolk, Wickramaratne, et al., 1999).

Age-Specific Manifestations

Setting aside the issue of continuity at a fundamental, etiological level, the question of whether the clinical presentations of MDD and DD differ as a function of developmental level is complex and, as yet, unresolved. However, it is crucial in determining whether the diagnostic criteria for MDD and DD and the content of interviews and rating scales for depression should be modified for different age groups. As noted previously, the available literature suggests that the symptom picture is fairly similar in school-age children, adolescents, and adults, although there is some evidence that hopelessness and some vegetative and motivational symptoms may be somewhat more frequent in adolescents than children (Carlson & Kashani, 1988; Weiss & Garber, 2003). Nonetheless, the manifestations of particular symptoms do vary as a function of the child’s level of cognitive and social development. For example, younger children may appear sad but have difficulty reporting their mood, and prepubertal children may lose interest in their friends and activities but are unlikely to experience decreased libido. There have also been several studies exploring whether there are developmental differences in the structure of the depressive syndrome, although the findings have been inconsistent (Weiss & Garber, 2003).

Depression in Very Young Children

There has been little systematic research on depression in preschool-age children and infants. Thus, it is unclear whether a syndrome comparable to that identified in school-age children, adolescents, and adults exists in younger children and, if it does, what the defining and associated features are. However, in an important series of articles, Luby, Heffelfinger, Mrakotsky, et al. (2002) recently reported that MDD can be identified in preschool-age children using modified DSM–IV criteria with a shorter duration requirement and provided preliminary evidence of construct validity.

Prevalence

Knowledge of base rates is an important consideration in assessment, as they can have a significant impact on the utility of assessment instruments and the validity of clinical decisions. For example, it is difficult for assessment instruments to improve on chance pre-
Depressive disorders are relatively uncommon in children but are more frequent in adolescents. In community samples, the 6-month prevalence of depressive disorders is 1% to 3% in school-age children and 5% to 6% in adolescents; the lifetime prevalence in adolescents is 15% to 20% (Garber & Horowitz, 2002; Lewinsohn & Essau, 2002). Not surprisingly, the prevalence of depression is much higher in clinical settings, with estimated rates of 8% to 15% in children and more than 50% in adolescents (Schwartz, Gladstone, & Kaslow, 1998). There is no consistent gender difference in the prevalence of depressive disorders in children; however, the rates diverge in early adolescence, and by age 15 the prevalence is approximately two times higher in girls than boys (Hankin et al., 1998).

**Associated Features**

Two associated features that are important to consider in assessing depression are functional impairment and comorbidity, as both may influence course and treatment response, as well as constituting important treatment targets in their own right.

**Functional Impairment**

Depressive disorders are associated with significantly poorer psychosocial functioning in children and adolescents. Depressed children and adolescents often exhibit significant impairment in family, school, and peer functioning, and some degree of impairment may persist after recovery from the depressive episode (Garber & Horowitz, 2002; Lewinsohn & Essau, 2002). Depressed adolescents are also at risk for school dropout and unplanned pregnancy (Waslick, Kandel, & Kakouros, 2002). Depression is the leading risk factor for youth suicide and may be a risk factor for the development of other disorders such as substance abuse (Birmaher, Arbelaez, & Brent, 2002). The causal relation between depression and functional impairment is complex: Depression causes significant impairment, but poor functioning may also be a risk factor for depression. Thus, the association is probably reciprocal and transactional.

**Comorbidity**

The majority of children and adolescents with MDD or DD also meet criteria for other psychiatric disorders. In a meta-analysis of studies using community samples, Angold et al. (1999) reported that the median odds ratios for the associations of depression with anxiety, conduct disorder, and attention deficit disorder were 8.2, 6.6, and 5.5, respectively. Depression is also often comorbid with eating, reading, and developmental disorders and general medical conditions. As noted previously, there are a number of possible reasons for these high comorbidity rates, including inadequacies in the diagnostic system (e.g., splitting what is really a single category, such as internalizing disorders, into multiple subtypes, such as depression and generalized anxiety disorder), one disorder causing the other (e.g., some forms of anxiety leading to depression; e.g., Cole, Peeke, Martin, Truglio, & Seroczyński, 1998), the comorbid condition representing a distinct disorder (e.g., depression and disruptive behavior disorders; e.g., Harrington, Rutter, & Fombonne, 1996), or both disorders having a common cause (Klein & Riso, 1993). Identifying the causes of comorbidity could have important implications for revising the classification system. However, even without understanding the nature of the causal processes, comorbidity is clinically important because it can obscure the existence of a depressive disorder and is associated with greater impairment and a poorer course and treatment response. In addition, comorbid conditions may require attention in their own right.

**Course**

Almost all children and adolescents with an episode of MDD recover, although many continue to experience subsyndromal (or residual) symptomatology. The length of episodes varies. The mean duration of episodes of MDD is approximately 7 to 8 months in clinical samples, and episodes of DD last an average of 48 months (Birmaher et al., 2002; Kovacs, 1996). Rates of relapse and recurrence of MDD are high, with the majority of depressed juveniles experiencing another episode within several years (Birmaher et al., 2002; Kovacs, 1996). Long-term follow-up studies indicate that adolescents with MDD are at high risk for experiencing depressive episodes in adulthood; however, the evidence for children with MDD is less consistent (Birmaher et al., 2002).

The mechanisms and processes that serve to maintain depressive episodes and cause recurrences are poorly understood, in part because most studies have not distinguished between onset, maintenance, and recurrence. However, longitudinal studies of the course of depression in children and adolescents have identified a number of factors that appear to predict the duration of MDD episodes and the probability of recurrence. These data provide important prognostic information, although further work is needed to determine the relative importance of these predictors.
of the prognostic variables may themselves be reasonable targets for intervention, although without a better understanding of their causal role, it is uncertain whether modifying them will hasten recovery or prevent recurrence. Variables that are associated with a longer time to recovery include an early age of onset, greater severity of depression, suicidality, double depression, the presence of comorbid anxiety or disruptive behavior disorders, depressotypic cognitions, and an adverse family environment (Birmaher et al., 2002). Most of these factors also predict recurrence. Variables that have been associated with an increased risk of recurrence include greater severity, psychotic symptoms, suicidality, a prior history of recurrent MDD, double depression, the presence of subthreshold symptons after recovery, a depressotypic cognitive style, recent stressful life events, an adverse family environment, and a family history of MDD (particularly if it is recurrent; Birmaher et al., 2002).

It is difficult to determine the relative importance of these prognostic factors, as there are few instances in which most of these variables have been included in the same study. However, in a study of a community sample of depressed adolescents, Lewinsohn, Rohde, Seeley, Klein, and Gotlib (2000) recently examined most of these variables as predictors of recurrence in young adulthood. They found that a prior history of recurrent MDD, a family history of recurrent MDD, personality disorder traits, and, for girls only, greater conflict with parents were significant and independent predictors of subsequent recurrence.

Children and adolescents with MDD and DD are also at risk for developing manic and hypomanic episodes. The probability of “switching” to bipolar disorder is higher in patients with psychotic symptoms, psychomotor retardation, a family history of bipolar disorder, a high familial loading for mood disorders, or a combination of these (Birmaher et al., 2002; Geller, Fox, & Clark, 1994).

Treatment

There is relatively strong support for the efficacy of cognitive–behavioral therapy (CBT) and interpersonal therapy for depressed adolescents (Asarnow, Jaycox, & Tompson, 2001; Kaslow, McClure, & Connell, 2002). However, the few studies examining the effects of family therapy, either alone or in conjunction with treatment for the adolescent, have generally failed to support its efficacy (Asarnow et al., 2001). Fewer data are available on the efficacy of psychosocial interventions in school-age children. Almost all of the clinical trials in this age group have used variants of CBT, generally administered in a group format to children with elevated levels of depressive symptoms but not necessarily diagnoses. Although the findings have varied, the majority of studies have reported evidence supporting the efficacy of CBT (Kaslow et al., 2002).

Little is known about the mechanisms underlying the effects of CBT and interpersonal therapy on depression in children and adolescents (or adults for that matter; Kazdin, 2003). In addition, despite the efficacy of psychosocial interventions for depressed children and adolescents in clinical trials, there is evidence that this may not translate into effectiveness in community settings (Weisz, Donenberg, Han, & Weiss, 1995). Finally, there are no data on the efficacy of treatments for very young children with depression.

Controlled clinical trials of antidepressant medications in children and adolescents are also limited. The available evidence indicates that the cyclic antidepressants are not efficacious (Emslie & Mayes, 2001). The data on newer medications, such as the selective serotonin uptake inhibitors and atypical antidepressants, is mixed. Four published double-blind placebo-controlled trials have reported benefits in adolescents or mixed samples of children and adolescents, but other published and unpublished studies have failed to find differences (Jureidini et al., 2004; Whittington et al., 2004). In addition, questions have recently been raised about whether several of the newer antidepressants are associated with increased suicidal ideation and behavior in children and adolescents (Jureidini et al., 2004; Whittington et al., 2004).

Similar to adult depression, it appears that there are high rates of relapse and recurrence when psychosocial and pharmacological treatments are terminated. Unfortunately, there are few studies of continuation or maintenance treatments for depressed children and adolescents (Asarnow et al., 2001; Emslie & Mayes, 2001). Extrapolating from the adult literature, it may be prudent to consider continuation and maintenance treatment for patients with partial recovery or characteristics associated with an increased risk of recurrence (e.g., history of recurrent episodes, double depression, family history of MDD, ongoing family conflict, or other stressors).

Data on predictors of treatment response in depressed children and adolescents is limited. However, it appears that many of the same variables that predict a more protracted recovery in naturalistic studies also predict a poorer response to treatment (Emslie, Mayes, Laptook, & Batt, 2003; Kaslow et al., 2002). Unfortunately, there are even fewer data on predictors of differential treatment response, that is, which patients respond better to some treatments than others. An exception is one study that reported that comorbid anxiety disorder predicted a better response to CBT than to systemic-behavioral family therapy or supportive therapy (Brent et al., 1998). An intuitively appealing strategy is to use patients' deficits to guide treatment selection (e.g., CBT for children and adolescents with prominent depressive cognitions, interpersonal ther-
apy for those with interpersonal problems, family therapy for those with family dysfunction; Kaslow et al., 2002). However, the limited evidence available for both depressed youths and adults indicates that deficits in particular areas do not predict a better response to treatments designed to target those deficits (e.g., depressive cognitions do not predict response to cognitive-behavioral therapy; Asarnow et al., 2001; Sotsky et al., 1991). To our knowledge, the alternative strategy of matching treatment to strengths has not been investigated for youth depression, although there have been hints that it may be an effective approach for adults (Sotsky et al., 1991).

**General Considerations in Assessment**

Clinical assessment can be thought of as a sequence including at least three phases: diagnosis and prognosis, treatment planning, and treatment monitoring and evaluation. The major goal of the initial phase is to develop a preliminary diagnosis and case conceptualization. For depression, this includes determining whether criteria are met for MDD or DD and ruling out exclusionary diagnoses such as bipolar disorder and depression due to a general medical condition or substance. As part of the assessment of depression, the clinician must assess key symptoms (e.g., suicidal ideation, psychotic symptoms) that might influence treatment decisions. In addition, it is important to carefully assess the previous course of the depression (e.g., prior episodes, chronicity) due to its prognostic value and possible implications for long-term treatment. It is also important to assess comorbid psychiatric, developmental, and general medical disorders and areas of significant functional impairment (e.g., family, school, peers) both to determine whether depression is the principle diagnosis that should be the primary target of intervention and because of their prognostic implications. Finally, it is important to assess the child's family environment, school functioning, and peer relationships; significant stressors and traumas; and family history of psychopathology, as these factors have considerable prognostic value.

Most of these variables are also critical for treatment planning. Data on the severity and prior course of depression, key symptoms such as suicidal ideation or behavior and psychotic symptoms, comorbidity, and functional impairment are important for determining the appropriate treatment setting (e.g., inpatient vs. outpatient), the intensity and duration of treatment, and perhaps the treatment modality, although as noted previously, few data are available to guide these decisions. Information on comorbidity is also necessary to determine whether other disorders should be monitored or targeted for treatment. Data on comorbid mental, developmental, and general medical disorders, the family environment, and family history of psychopathology are important because they may suggest, contraindicate, or complicate some treatment options (e.g., parent counseling, family therapy) or suggest the need for additional referrals for the patient or a family member. Finally, it is critical to take a detailed history of previous treatment and assess the goals, attitudes, and motivation of the child and parents with respect to the relevant treatment options. This information is critical both for treatment selection and for engaging the child and family in treatment. As children and parents often disagree on the selection of treatment targets (Hawley & Weisz, 2003), it may take considerable negotiation to develop a treatment plan that is acceptable to all parties.

The third phase of assessment involves treatment monitoring and evaluation. This entails systematically assessing the degree of change in target symptoms and impairments to determine whether treatment should be continued, intensified, augmented, changed, or terminated. As few guidelines are available to help clinicians determine when treatment should be modified, this is an important area for future research.

**Information Source**

It is important to obtain data from multiple informants, including the child, parents, and teachers. Child report is critical, as parents and teachers tend to report lower levels of depressive and other internalizing symptoms in children than youths report themselves (Jensen et al., 1999). However, it is useful to supplement youths' reports with information from collateral sources of psychopathology than adolescents (Edelbrock, Costello, Dulcan, Kalas, & Conover, 1985). In addition, younger children have difficulty reporting on information requiring temporal parameters; therefore, parents must be relied on for information on course such as age of onset, previous episodes, and duration of current episode (Kovacs, 1986). Finally, parents are more involved in the day-to-day lives of children than adolescents and therefore are more knowledgeable about their behavior and activities.

Although obtaining data from multiple informants is optimal, agreement between informants is only fair to moderate (Achenbach, McConaughy, & Howell, 1987). In addition, depressed parents appear to have a lower threshold for detecting depression in their children, hence their reports tend to yield higher rates of both true and false positives (i.e., increased sensitivity but decreased specificity; Najman et al., 2000; Richters, 1992; Youngstrom, Izard, & Ackerman, 1999).
Despite the substantial disagreement between informants, there is evidence for the validity of both parent and child reports (Jensen et al., 1999). In addition, several studies have demonstrated that child, parent, teacher, and clinician ratings all account for significant variance in predicting subsequent outcomes (Ferdinand et al., 2003; Verhulst, Dekker, & van der Ende, 1997).

The low agreement between data sources presents a significant challenge for clinicians who must decide how to interpret and integrate conflicting information. A variety of approaches to integrating data from multiple informants has been discussed in the literature, including assuming that the feature or diagnosis is present if any informant reports it (the “or” rule), requiring several informants to confirm the feature or diagnosis (the “and” rule), or developing various statistical procedures for optimizing prediction (e.g., Baillargeon et al., 2001; Kraemer et al., 2003; Rubio-Stipec, Fitzmaurice, Murphy, & Walker, 2003). The approach that most closely mirrors clinical practice is the “best estimate” procedure, in which the clinician uses his or her best judgment to integrate and resolve conflicting reports. This raises the possibility of bringing through the back door the unreliability and idiosyncrasy that structured interviews and standardized ratings scales were developed to prevent (discussed later). However, there is evidence from the adult literature that when applied, following some general guidelines (e.g., self-report takes precedence for internalizing disorders; informant report is given priority for externalizing disorders), the reliability of best-estimate diagnoses can be very high (Klein, Ouimette, Kelly, Ferro, & Riso, 1994).

Attenuation Effect

Studies of interviews and ratings scales for both juvenile and adult psychopathology have often found that rates of diagnoses and ratings of symptom severity tend to decrease with repeated administrations, a phenomenon referred to as the “attenuation effect” (Egger & Angold, 2004). This has important implications for treatment monitoring and evaluation, as it is difficult to distinguish the attenuation effect from a positive response to treatment for the individual patient. Although there is no solution to this problem at present, it behooves the clinician to be aware of this phenomenon and to consider alternative explanations for what appears to be improvement on rating scales.

Approaches to Assessing Depression in Children and Adolescents

The two major approaches to diagnosing and assessing depression in children and adolescents involve interviews and rating scales. Unfortunately, there are no laboratory measures of psychosocial or biological variables that are useful for clinical assessment at this time (Garber & Kaminski, 2000; Waslick et al., 2002). Several observational measures and coding systems have been employed for assessing depression in children; however, there is insufficient evidence to recommend their use (Garber & Kaminski, 2000; Kendall, Cantwell, & Kazdin, 1989).

Interviews can include unstructured clinical interviews and semistructured or fully structured interviews. Unstructured clinical interviews vary from clinician to clinician with respect to format, duration, focus, and coverage and therefore in the amount and type of information elicited. The literature indicates that clinicians using unstructured interviews often fail to inquire about key aspects of psychopathology, particularly if it is inconsistent with their initial diagnostic impressions (Angold & Fisher, 1999), and end up making fewer diagnoses than clinicians using structured interviews (Zimmerman, 2003). Thus, structured interviews are more comprehensive. In addition, they probably have greater interrater reliability, although we are unaware of studies directly testing this. Although there are few data on the incremental validity of structured diagnostic assessments, Kashner et al. (2003) found that when the results of semistructured diagnostic interviews were provided to clinicians, it resulted in changes in clinicians’ diagnoses and treatment, suggesting that the additional information was perceived as clinically useful and valid. Although most of this work has focused on adults, there is evidence that some of these findings may also apply to children and adolescents (Hughes et al., 2000; Lewczyk, Garland, Hurlburt, Gearly, & Hough, 2003).

Semistructured interviews are often referred to as “interviewer based” because the interviewer is responsible for rating the criteria as accurately as possible, using all information at his or her disposal, and improvising additional questions or confronting the respondent with inconsistencies when necessary. In contrast, fully structured interviews are referred to as “respondent based” because the interviewer’s role is limited to reading the questions as written and recording the respondent’s answers. Both approaches require judgments about the presence and significance of symptoms and other clinical features. In semistructured interviews, these judgments are made by the interviewers, whereas in fully structured interviews they are made by the respondents. As a result, semistructured interviews were designed for use by mental health professionals or well-trained and supervised technicians and seek to capitalize on their clinical training and experience, whereas fully structured interviews were developed for lay interviewers in large-scale epidemiological studies in which the cost of interviewers with clinical training is prohibitive.
There have been few direct comparisons of the validity of semi- versus fully structured interviews. In the absence of evidence to the contrary, we assume that the interviewer-based approach yields higher quality data than the respondent-based approach because the interviewer presumably has a better sense of the constructs being assessed than the respondent. However, there are some data suggesting that fully structured interviews administered by bachelor's degree level technicians may be useful in supplementing clinical evaluations in clinical settings (Hughes et al., 2000).

Rating scales include clinician-administered, self-report, and parent and teacher measures. Clinician-administered rating scales are semistructured interviews that focus on a circumscribed area of symptomatology (e.g., depression). Unlike diagnostic interviews, they do not collect sufficient information to make a diagnosis (e.g., duration and exclusion criteria are generally not assessed). Self-report and parent and teacher rating scales are typically questionnaires that are self-administered by the designated informant (although they can be read to younger children). Similar to clinician-administered rating scales, self-, parent, and teacher rating scales typically focus on current symptoms and behavior and therefore do not provide sufficient information to make diagnoses.

Due to their economy, self-, parent, and teacher rating scales can be used as screening instruments, with elevated scores leading to a more intensive evaluation. Self-rating scales are generally superior to parent and teacher rating scales in screening for internalizing disorders due to their greater sensitivity. However, even the best self-rating scales have only moderate sensitivity and specificity, producing a substantial number of false positives and false negatives (Kendall et al., 1989). The most useful indexes for evaluating the clinical utility of a screening instrument are positive predictive power, the probability that individuals scoring above the cutoff point on the instrument actually have the disorder, and negative predictive power, the probability that individuals scoring below the cutoff point do not have the disorder. Positive predictive power and negative predictive power are strongly influenced by the prevalence, or base rate, of the condition in the population being screened. Positive predictive power tends to be low and negative predictive power tends to be high in low base rate situations, whereas the inverse is characteristic of high base rate situations. As the prevalence of child and adolescent depression tends to be fairly low in most screening contexts, the use of screeners often produces unacceptably low positive predictive power, with the false positives greatly outnumbering the true positives (Matthey & Petrovski, 2002; Roberts, Lewinsohn, & Seeley, 1991). Thus, the potential economy and efficiency of screening must be weighed against the costs of unnecessary extended evaluations for false positive cases and the risks associated with missing false negative cases (Costello & Angold, 1988).

In the next two sections, we briefly describe several of the better researched and more widely used semistructured diagnostic interviews and rating scales. Due to space limitations, we have been highly selective, and there are a number of equally good but less widely used measures that we have not included. For more information, readers are referred to more comprehensive reviews (Angold & Fisher, 1999; Brooks & Kutchner, 2001; Myers & Winters, 2002; Silverman & Rabian, 1999). In addition, we focus on general measures of depression. There are also a number of measures of specific components of the depressive syndrome, such as self-esteem, hopelessness, depressive cognitions, and suicidality (see Winters, Myers, & Proud, 2002, for a review) that may be useful for particular cases, but they are not reviewed here.

**Semistructured Diagnostic Interviews**

In this section, we briefly review the three most widely used semistructured diagnostic interviews for child and adolescent psychopathology: the Schedule for Affective Disorders and Schizophrenia in School-Age Children (K–SADS; Puig-Antich & Chambers, 1978), the Diagnostic Interview for Children and Adolescents (Herjanic & Reich, 1982), and the Child and Adolescent Psychiatric Assessment (CAPA; Angold, Prendergast, et al., 1995). We do not discuss fully structured respondent-based interviews, such as the Diagnostic Interview Schedule for Children (Shaffer, Fisher, & Lucas, 1999), as we believe that they are less useful in clinical settings for the reasons described previously. Each interview assesses the criteria for most of the major child and adolescent psychiatric disorders and provides parallel versions for children and parents. Although some of the instruments are used to interview 6- and 7-year-old children, it is questionable whether children younger than 8 or 9 years old can provide valid information in a diagnostic interview (Angold & Fisher, 1999).

Evaluating the validity of semistructured diagnostic interviews is complex, as they are usually used as the "gold standard" against which other measures are compared. Construct validity is probably the best standard (Angold & Costello, 2000), but given the current state of the literature, it is impossible to distinguish the construct validity of semistructured interviews from the diagnoses that they are designed to assess. To try to disentangle the construct validity of interviews from diagnostic constructs, it is necessary to conduct head-to-head comparisons of several interviews using the same sample and the same criteria for construct validation (e.g., family history, course). As such studies have not been conducted, we limit our discussion to the
ASSESSMENT OF CHILD AND ADOLESCENT DEPRESSION

interrater reliability and convergent validity of the interviews. Although the distinction between MDD and DD has important prognostic implications (Kovacs, 1996), the majority of studies combine them in a higher order depressive disorder category or focus solely on MDD. Hence, for the purposes of this article, we focus on depressive disorders as broadly conceived.

The K–SADS (Puig-Antich & Chambers, 1978) is the most widely used semistructured interview for children and adolescents (6 to 18 years). It is also the least structured of the semistructured interviews and therefore requires the greatest amount of clinical training and experience. The K–SADS was modeled after the adult Schedule for Affective Disorders and Schizophrenia (SADS). There are a number of versions of the K–SADS that assess DSM–IV criteria. These versions vary in format, whether they assess lifetime as well as current psychopathology, and whether they also provide dimensional measures of symptom severity (for detailed comparisons see Ambrosini [2000] and Angold & Fisher [1999]). Ratings are based on all sources of information and clinical judgment. Administration time of the parent and child interviews range from 35 min to 2.5 hr each, depending on the severity of the child’s psychopathology. Interrater reliability has been reported to be fair to excellent for depressive disorders in several studies and has been particularly impressive with the more recent versions of the K–SADS (Ambrosini, 2000; Kaufman et al., 1997). Evidence for convergent validity derives from numerous studies reporting correlations between the K–SADS and a variety of clinician, self-, and parent rating measures of depression and internalizing behavior problems (Ambrosini, 2000; Kaufman et al., 1997).

The Diagnostic Interview for Children and Adolescents (Herjanic & Reich, 1982) was originally designed as a fully structured interview. However, more recent versions are semistructured. The most recent version of the Diagnostic Interview for Children and Adolescents (Reich, 2000) assesses both Diagnostic and Statistical Manual of Mental Disorders (3rd ed., rev. [DSM–III–R]; American Psychiatric Association, 1987) and DSM–IV criteria and includes separate interviews for children (6 to 12 years), adolescents (13 to 17 years), and parents. The interview adopts a lifetime time frame and takes approximately 1 to 2 hr to complete. Data on interrater reliability has varied across studies, ranging from poor to good (Boyle et al., 1993; Brooks & Kutcher, 2001; Reich, 2000). The Diagnostic Interview for Children and Adolescents diagnoses are moderately correlated with clinicians’ diagnoses and clinician and self-rated measures of depressive symptoms (Brooks & Kutcher, 2001; Ezpeleta et al., 1997; Reich, 2000), providing some evidence of convergent validity.

The CAPA (Angold, Prendergast, et al., 1995) assesses the criteria for most major diagnoses in children ages 9 to 17 years. The time frame for symptom assessment is the preceding 3 months, and administration time takes from 1 to 2 hr (Angold & Costello, 2000). Although there are few data on the CAPA by investigators who were not involved in its development, the interview has several very attractive features. First, it is unique in that it includes an extensive glossary defining specific symptoms and distress and frequency ratings. As a result, the CAPA can be used by interviewers with minimal clinical experience, as long as they adhere closely to the definitions and conventions in the glossary. Second, it includes a section for assessing impairment in a number of areas, including family, peers, school, and leisure activities, and also includes sections assessing the family environment and life events and traumas. In the one published study of the test–retest reliability of the CAPA, Angold and Costello (1995) reported that the kappas for MDD and DD were .90 and .85, respectively, and that the intraclass correlation for the MDD symptom scale was .88. Data on the association between CAPA diagnoses of depression and other measures of depression have not been published. However, Angold and Costello (2000) reported that depression as diagnosed by the CAPA is associated with significant levels of functional impairment and higher concordance among monozygotic than dizygotic twins, supporting the construct validity of the interview.

Rating Scales

In this section, we briefly review some of the more widely used clinician, self-report, and multi-informant rating scales for depression. Due to space limitations, we do not discuss rating scales designed for adults that are often used with older adolescents (e.g., the Hamilton Rating Scale for Depression [Hamilton, 1967]; the Beck Depression Inventory [Beck, Ward, Mendelsohn, Mock, & Erbaugh, 1961]; and the Center for Epidemiological Studies–Depression Scale [Radloff, 1977]). However, these three measures have similar psychometric properties in adolescent and adult samples (Roberts et al., 1991) and have comparable reliability and validity compared to measures that were specifically designed for juveniles. Hence, these measures appear to be acceptable alternatives for older adolescents.

In discussing reliability and validity, we focus on internal consistency and interrater reliability for clinician-rated scales, internal consistency and test–retest reliability for self- and multi-informant rating scales, and convergent and discriminant validity and sensitivity to change for all rating scales.

The most widely used clinician scale for rating depression is the Children’s Depression Rating Scale (CDRS; Poznanski, Cook, & Carroll, 1979). Based on the Hamilton Rating Scale for Depression, the CDRS
was developed to assess current severity of depression in children ages 6 to 12 years and is often used for adolescents as well. The revised version (Poznanski & Mokros, 1999) contains 17 items assessing cognitive, somatic, affective, and psychomotor symptoms and draws both on the respondent’s report and the interviewer’s behavioral observations. It takes 15 to 20 min to administer. It is designed to be administered separately to the child and an informant, with the clinician subsequently integrating the data using clinical judgment. Cutoff scores are provided to aid in interpreting levels of depression severity. The CDRS has only moderate internal consistency but good interrater reliability (Brooks & Kutcher, 2001; Myers & Winters, 2002). Its convergent validity has been supported by moderate to high correlations with the Hamilton Rating Scale for Depression and several self-rated depression scales (Brooks & Kutcher, 2001; Myers & Winters, 2002; Shain, Naylor, & Alessi, 1990). However, discriminant validity is more problematic, as the CDRS has difficulty distinguishing between depression and anxiety and overestimates depression severity in children with general medical conditions due to its emphasis on somatic symptoms (Brooks & Kutcher, 2001; Myers & Winters, 2002). Finally, the CDRS has detected treatment effects in both psychopharmacology and psychotherapy trials (Brooks & Kutcher, 2001; Myers & Winters, 2002).

There are a number of widely used self-rating scales for child and adolescent depression. We briefly review four: the Children’s Depression Inventory (CDI; Kovacs, 1992), Mood and Feelings Questionnaire (MFQ; Angold, Costello, Messer & Pickles, 1995), Reynolds Child Depression Scale (RCDS; Reynolds, 1989), and Reynolds Adolescent Depression Scale (RADS; Reynolds, 1987).

The CDI (Kovacs, 1992) is the most widely used depression rating scale for children and adolescents. Developed as a modified version of the Beck Depression Inventory, it assesses severity of depression during the previous 2 weeks in children ages 7 to 17 years. It includes 27 items covering a broad range of depressive symptoms and associated features, with a particular emphasis on cognitive symptoms, and takes 10 to 20 min to complete. A number of studies have reported that the CDI has good internal consistency, and many (but not all) studies have also reported good short-term test–retest stability (Brooks & Kutcher, 2001; Kovacs, 1992; Silverman & Rabian, 1999). Studies of the factor structure of the CDI have produced inconsistent findings, with some indication that the factor structure varies by age (Cole, Hoffman, Tram, & Maxwell, 2000; Weiss & Garber, 2003). The CDI is moderately to highly correlated with the CDRS, a number of other self-rated depression scales, and other measures of related constructs (e.g., self-esteem, depressotypic cognitions), supporting its convergent validity (Brooks & Kutcher, 2001; Myers & Winters, 2002; Shain et al., 1990; Silverman & Rabian, 1999). However, the discriminant validity of the CDI is questionable, as it is almost as highly correlated with measures of anxiety as it is with other measures of depression, and studies examining its ability to distinguish depressed from nondepressed patients have yielded conflicting findings (Myers & Winters, 2002; Silverman & Rabian, 1999). Finally, the CDI has been shown to be sensitive to change in several treatment studies (Brooks & Kutcher, 2001; Myers & Winters, 2002; Silverman & Rabian, 1999). One study reported that it was more sensitive in detecting the effects of group CBT in school-age children than the CDRS and the RCDS (Starke, Reynolds, & Kaslow, 1987); however, another study found that it was less sensitive to the effects of medication than the CDRS (Emslie et al., 1997).

The MFQ (Angold, Costello, et al., 1995) was developed to assess depression over the past 2 weeks in youth ages 8 to 18 years. It consists of 32 items covering the DSM-III-R criteria for depression and additional symptoms, such as loneliness and feeling unloved or ugly. Angold, Costello, et al. also developed a shorter, 13-item version by selecting items that yielded optimal discriminating power and internal consistency. The MFQ takes approximately 10 min to complete. It has good internal consistency and test–retest reliability (Angold, Costello, et al., 1995; Wood, Kroll, Moore, & Harrington, 1995). In addition, the MFQ has demonstrated good convergent validity with respect to the CDI, the Diagnostic Interview Schedule for Children, the CAPA, and the K–SADS (Angold, Costello, et al., 1995; Thapar & McGuffin, 1998; Wood et al., 1995). The MFQ was also relatively successful in discriminating youths with diagnoses of depression from those with nonmood disorders (Kent, Vostanis, & Feehan, 1997; Thapar & McGuffin, 1998). Finally, it has demonstrated sensitivity to change in some, but not all, clinical trials (Brooks & Kutcher, 2001).

The RCDS (Reynolds, 1989) and RADS (Reynolds, 1987) are 30-item scales designed to assess depressive symptomatology (as represented in the Diagnostic and Statistical Manual of Mental Disorders [3rd ed.]; American Psychiatric Association, 1980) during the previous 2 weeks in youth ages 8 to 12 and 13 to 18 years of age, respectively. Each scale takes approximately 10 min to complete, although children may require some additional time. The Reynolds scales have been used primarily with school, rather than clinical, samples. Both scales have good internal consistency and test–retest reliability (Brooks & Kutcher, 2001; Myers & Winters, 2002; Reynolds, 1987, 1989). In addition, both are correlated with interview diagnoses and other depression rating scales such as the SADS, CDRS, Hamilton Rating Scale for Depression, CDI, Beck Depression Inventory, and Center for Epidemiological Studies–Depression Scale (Brooks & Kutcher, 2001; Myers & Winters, 2002; Shain et al., 1990; Silverman & Rabian, 1999). However, the discriminant validity of the CDI is questionable, as it is almost as highly correlated with measures of anxiety as it is with other measures of depression, and studies examining its ability to distinguish depressed from nondepressed patients have yielded conflicting findings (Myers & Winters, 2002; Silverman & Rabian, 1999). Finally, the CDI has been shown to be sensitive to change in several treatment studies (Brooks & Kutcher, 2001; Myers & Winters, 2002; Silverman & Rabian, 1999). One study reported that it was more sensitive in detecting the effects of group CBT in school-age children than the CDRS and the RCDS (Starke, Reynolds, & Kaslow, 1987); however, another study found that it was less sensitive to the effects of medication than the CDRS (Emslie et al., 1997).
Discriminant validity has not been well-studied, although like most depression rating scales, the RADS is moderately correlated with measures of anxiety (Myers & Winters, 2002). Both the RCDS and RADS are able to detect treatment effects in controlled clinical trials, although they may be somewhat less sensitive than some other rating scales (Stark et al., 1987).

As noted earlier, it is important to obtain information about child and adolescent depression from informants other than the youths themselves. Several of the self-rating scales, such as the CDI, have been reworded for use by parents and, in some cases, by teachers and peers. Few psychometric data have been reported on these adaptations. However, Cole et al. (2000) recently compared child and parent report versions of the CDI. They reported that the two versions had similar internal consistencies and test–retest reliabilities and that the factor structure of the CDI was relatively similar, although not identical, across informants.

In addition to these modified measures, there are a number of multi-informant rating scales that are designed to assess a broad range of child and adolescent psychopathology using instruments that are comparable across informants (Hart & Lahey, 1999). The most widely used is the parent-report Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), and its accompanying teacher report (Teacher Report Form [TRF]) and child-report (Youth Self-Report [YSR]) versions. The CBCL and TRF are appropriate for youth ages 5 to 18, whereas the YSR is designed for youth ages 11 to 18. The CBCL and YSR assess the child's behavior over the past 6 months, whereas the TRF uses a 2-month time frame. All three measures take approximately 10 to 15 minutes to complete.

The CBCL includes 118 items assessing two broadband and eight narrowband scales identified using factor analysis, as well as a social competence scale. Extensive norms for the CBCL, TRF, and YSR are available for both clinical and community samples, and favorable psychometric properties of the instruments have been documented in hundreds of studies. Unfortunately, the CBCL's utility in assessing depression, at least as conceptualized in the DSM-IV, is limited. The scale that is most relevant to depression is the narrowband Anxious/Depressed scale, which combines symptoms of anxiety and depression. In addition, some other depressive symptoms are included on other narrowband scales. Indeed, a latent class analysis of the Anxiety/Depression scale was unable to distinguish distinct classes for depression and anxiety (Wadsworth, Hudziak, Heath, & Achenbach, 2001).

Several investigators have attempted to develop alternative approaches to scoring the CBCL to assess depression. For example, Lengua, Sadowski, Friedrich, and Fisher (2001) developed a 12-item depression scale from the CBCL that performed better than the original Anxious/Depressed subscale in identifying cases with a depressive disorder diagnosis. The CBCL recently added a series of diagnostic scales that are more closely geared to DSM-IV diagnoses; however, studies of their association with semistructured, interview-derived diagnoses have not yet been reported.

Assessment of Younger Children

As noted previously, the diagnosis and assessment of depression in infants and preschool-age children is still largely uncharted territory (Garber & Horowitz, 2002). Moreover, the administration of semistructured interviews and self-rating scales to children under the age of 8 or 9 years is questionable (Angold & Fisher, 1999), although there is some evidence for the validity of the CDI in children as young as 5 to 6 years old (Ialongo, Edelsohn, & Kellam, 2001).

The most widely used rating scales for psychopathology in preschoolers are the downward extensions of several of Achenbach's instruments: the CBCL 1–5, a parent rating scale, and the TRF 1–5, a teacher rating scale (Achenbach & Rescorla, 2004). To our knowledge, the only rating scale specifically designed to assess depression in preschoolers is the Preschool Feelings Checklist, a brief parent-rated screening measure developed by Luby, Heffelfinger, Koenig-McNaught, Brown, and Spitznagel (2004). Preliminary data on the sensitivity and specificity of the Preschool Feelings Checklist using a structured diagnostic interview with a parent as the criterion was promising. However, cross-validation of these findings is needed.

Egger and Angold (2004) recently developed a downward extension of the CAPA for 2- to 5-year-old children titled the Preschool Age Psychiatric Assessment. The interrater reliability for this measure is comparable to other semistructured interviews with older children and adults. A downward extension of the Diagnostic Interview Schedule for Children for preschoolers is also in development (Shaffer et al., 1999).

Finally, a group of investigators sponsored by the McArthur Foundation recently developed a battery of assessment instruments for children in the early school-age period (ages 4 to 7 years). It includes a parent and teacher rating scale, the McArthur Health and Behavior Questionnaire (Essex et al., 2002), and a semistructured interview, the Berkeley Puppet Interview (Ablow et al., 1999) that makes use of puppets to provide a more developmentally sensitive assessment. Both measures include scales tapping various domains of symptomatology (including a subscale for depressive symptoms), physical health, and peer and school functioning. In the initial reports from this group, the depression scale from the McArthur Health and Behavior Questionnaire parent and teacher forms had adequate internal consistency and good test–retest reli-
ability and discriminated clinic from community participants (Ablow et al., 1999). Although a categorical measure of depression from the McArthur Health and Behavior Questionnaire parent form was not correlated with diagnoses of MDD derived from the parent version of the Diagnostic Interview Schedule for Children, it was associated with a number of teacher-rated indexes of impairment (Luby, Heffelfinger, Measelle, et al., 2002). The Berkeley Puppet Interview depression scale had adequate internal consistency in a clinic sample but poor internal consistency in a community sample, had moderate test–retest reliability in both samples, and discriminated the clinic from community participants (Ablow et al., 1999).

Assessment of Psychosocial Functioning, Life Stress, and Family History of Psychopathology

In this section we briefly discuss measures of psychosocial functioning, life stress, and family history of psychopathology due to the importance of these variables for prognosis, treatment planning, and treatment monitoring and evaluation.

Psychosocial Functioning

Depression in children and adolescents is associated with significant impairment in family and peer relationships and academic performance. The families of depressed youths are often characterized by a lack of cohesion and high levels of disengagement and criticism (Sheeber, Hops, & Davis, 2001). The parents of depressed children and adolescents exhibit less warmth and support and greater control and rejection than parents of controls (Garber & Horowitz, 2002; Stein et al., 2000). Depressed youths have significant peer difficulties and social skills deficits (Rudolph, Hammen, & Burge, 1994) and often exhibit academic underachievement, school attendance problems, and school failure (Hammen, Rudolph, Weisz, Rao, & Burge, 1999).

There are a variety of approaches and instruments for assessing impairments and competencies in psychosocial functioning (for reviews see Bird, 1999; Canino, Costello, & Angold, 1999; John, 2001). Some of the interviews and rating scales discussed previously include measures of functional impairments and competencies. For example, the CAPA includes a comprehensive assessment of the major areas of child psychosocial functioning (Angold & Costello, 2000); the CBCL (Achenbach & Rescorla, 2001) has a 16-item social competence scale; and the Berkeley Puppet Interview includes scales tapping academic and social competence and peer acceptance (Ablow et al., 1999). In this section, we briefly discuss several additional measures of psychosocial functioning in children and adolescents. Although we do not comment on their psychometric properties due to space limitations, these measures generally have acceptable interrater reliability and convergent validity.

One group of measures consists of global or dimensional scales, such as the Child Global Assessment Scale (C–GAS; Shaffer et al., 1983) and the Columbia Impairment Scale (Bird et al., 1993). The C–GAS, adapted from the Global Assessment Scale for adults (which is also the basis for DSM–IV Axis V), is a single 100-point scale designed for clinicians to rate the severity of symptomatology and functional impairment. The rating is based on information collected through other means (i.e., a diagnostic interview), as the C–GAS does not provide questions. The Columbia Impairment Scale is a respondent-based interview that consists of 13 items tapping a variety of domains of social functioning and symptomatology that are combined to form a single score. The C–GAS and Columbia Impairment Scale are both very economical. However, they each yield only one score, hence they do not provide information on the nature of impairment in specific areas of functioning. In addition, both measures combine symptoms and psychosocial functioning so that a child’s score could reflect problems in either or both domains.

Several more extensive instruments are also available. The Child and Adolescent Functional Assessment Scale (Hodges, 1999) is a widely used clinician-rated instrument that takes approximately 30 min to administer (see Bates, 2001, for a review). It assesses five domains: Role Performance (including School/Work, Home, and Community), Behavior Toward Others, Moods/Self-Harm, Substance Use, and Thinking. The Functional Impairment Scale for Children and Adolescents (S. J. Frank, Paul, Marks, & Van Egeren, 2000) was developed to cover the same domains as the Child and Adolescent Functional Assessment Scale using a parent-report format and is briefer, requiring 15 to 20 min to complete. Both the Child and Adolescent Functional Assessment Scale and the Functional Impairment Scale for Children and Adolescents assess symptomatology as well as psychosocial functioning. As these domains are assessed in separate sections, it is possible to obtain information on functioning alone. However, there is considerable overlap between the symptom-focused sections and the interviews and rating scales discussed previously, and the symptom sections are not comprehensive enough to substitute for a diagnostic interview or depression rating scale.

There are several semistructured interviews for psychosocial functioning that are designed to accompany semistructured diagnostic interviews without overlapping with more symptom-focused assessments. The Social Adjustment Inventory for Children and Adoles-
cents (SAICA; John, Gammon, Prusoff, & Warner, 1987) assesses school functioning, peer relations, home life, and spare time activities in youth ages 6 to 18 years. It takes about 30 min to complete and is administered separately to the parent and child. One study has reported that the SAICA performed better than a global functioning scale in predicting the course of depression in adolescents (Sanford et al., 1995). The Psychosocial Schedule for School Age Children-Revised (Lukens et al., 1983; Puig-Antich, Lukens, & Brent, 1986) was designed to assess school functioning; relationships with parents, siblings, and peers; and the parents' marital relationship in children ages 6 to 16 years. Like the SAICA, it is administered separately to the parent and child. Finally, an 11-item interview based on the adult Longitudinal Interval Follow-up Evaluation-Range of Impaired Functioning Tool (Leon et al., 1999) is currently being developed for children and adolescents (Fisher, Leon, & Coles, 2002). Like the SAICA and Psychosocial Schedule for School Age Children-Revised, it assesses school, family, peer, and recreational functioning but is briefer and can be easily incorporated into a more comprehensive evaluation.

We are aware of only one self-report measure of social adjustment for children. Weissman, Orvaschel, and Padian (1980) modified the Social Adjustment Scale—Self-Report, which was originally developed for adults, for use with children and adolescents. This measure for children includes 23 items assessing social adjustment during the past 2 weeks in the areas of school behavior, friends and spare time, family behavior, and, for adolescents only, dating. Rohde, Clarke, Mace, Jorgensen, and Seeley (2004) recently reported that the Social Adjustment Scale—Self-Report was sensitive to treatment effects in a sample of adolescents with comorbid MDD and conduct disorder, whereas the C-GAS was not. Although Weissman et al. included children as young as 6 years in their original validation sample, they did not examine age differences in the psychometric properties of the instrument. Hence, it is unclear whether the Social Adjustment Scale—Self-Report is reliable and valid in preadolescents.

In addition to the instruments noted here, there are numerous more focused measures designed to assess specific areas of functioning. For example, there are a number of widely used inventories assessing key dimensions of family functioning (e.g., Epstein, Baldwin, & Bishop, 1983; Moos & Moos, 1981) and parenting behavior (e.g., Schaefer, 1965), laboratory tasks that have been used to examine the interaction patterns of families of depressed children (see Garber & Kaminski, 2000), and interview and laboratory measures of expressed emotion (e.g., Asarnow, Goldstein, Tompison, & Guthrie, 1993). Finally, there are a variety of peer nomination measures that can be used to assess children's social status in school (e.g., Coie, Dodge, & Coppotelli, 1982). Unfortunately, peer nomination measures are often difficult to use, as they require access to all the students in the target child's class. Therefore it is often necessary to rely on teacher ratings of peer functioning, which are moderately to highly correlated with peer nomination data (Huesmann, Eron, Guerra, & Crawshaw, 1994).

Stressful Life Events

Prospective studies in children have shown that stress, and particularly events related to loss, disappointments, conflict, or rejection, predict the onset and persistence of depressive symptoms (Garber & Horowitz, 2002; Goodyer, 2001). Life stress can be assessed through self-administered questionnaires (e.g., Johnson & McCutcheon, 1980) or interviewer-based interviews (e.g., Goodyer, Kolvin, & Gatzanis, 1987; Hammen & Brennan, 2001; Sandberg et al., 1993; Williamson et al., 2003). Questionnaires are more economical, but semistructured interviews have a number of critical advantages, including the ability to assess the temporal relation between the stressor and the onset of the depressive episode; to distinguish potentially important features of events such as long-term threat and whether the event is independent of, versus dependent on, the child's behavior; and to minimize idiosyncratic interpretations of items (Duggal et al., 2000; Williamson et al., 2003).

Family History of Psychopathology

A number of studies have reported elevated rates of depression, and often other forms of psychopathology, in the relatives of depressed children and adolescents (e.g., Klein et al., 2001). Family history data can be elicited using diagnostic interviews conducted directly with family members (the family interview method) or by interviewing key informants about the other relatives (the family history method). A number of semistructured interviews for eliciting family history information from informants are available, including the Family History Research Diagnostic Criteria (Andreasen, Endicott, Spitzer, & Winokur, 1977), the Family Informant Schedule and Criteria (Mannuzza & Fyer, 1990), and a brief family history screening interview developed by Weissman et al. (2000). Family history data collected from informants tends to have high specificity but only moderate sensitivity. Hence, it is advisable to obtain information from at least two informants to increase the probability of detecting psychopathology in relatives.

Issues for Future Research

Developing guidelines for evidence-based assessments of child and adolescent depression is a challeng-
ing task, and a number of issues must be resolved to accomplish this aim. In this section, we briefly highlight 10 issues that we believe are particularly important.

1. There are a number of critical questions concerning the target construct of depression that have significant implications for the design and content of assessment instruments. These include the issues of continuity and developmental variations in the expression of depression, whether depression is a discrete entity or a continuous phenomenon, the boundaries between depression and normal variations in mood and other forms of psychopathology, and the identification of more homogeneous subgroups. Thus, if the rates or structure of depressive symptoms vary as a function of developmental period, this should be reflected in the content and scoring of assessment instruments (e.g., it may be necessary to develop multiple forms of the same instrument for different developmental periods). Similarly, whether depression is a categorical or dimensional construct must be taken into account in the design and scoring of the instrument to optimize reliability and validity (Ruscio & Ruscio, 2002). Shifts in the boundaries of depression would have to be reflected in the content of assessment instruments and would have significant implications for convergent and discriminant validity. Finally, if more homogeneous subtypes are identified, the information necessary to identify them should be included in standard assessment instruments.

2. There is a need for longitudinal studies of the course of mood disorders in children and adolescents that focus specifically on the predictors and processes associated with recovery and recurrence. In addition, there is a need for more research on predictors of differential treatment response (i.e., variables that predict response to one treatment but not another). In both cases, valuable information regarding potential targets for assessment and treatment could be provided.

3. Related to the problem of boundaries is the poor discriminant validity of most depression rating scales, particularly vis-à-vis anxiety disorders. This may be due to the high representation of items tapping general distress or negative affectivity that are common to both depression and anxiety. Discriminant validity may be improved by increasing the emphasis on aspects of depression with greater diagnostic specificity, such as anhedonia and low positive affectivity (Clark & Watson, 1991; Chorpita & Daleiden, 2002).

4. Few studies have directly compared the reliability or validity of different diagnostic interviews or rating scales in the same sample. As a result, we are forced to base comparisons on different studies, each using a single instrument. Hence, it is unclear whether any advantages for particular instruments are due to the properties of the instrument or to differences in the methods used in the studies (Angold & Fisher, 1999).

5. Few data on the incremental validity (see Hunsley & Meyer, 2003) of these measures vis-à-vis unstructured assessments or “assessment as usual” are available. Although there is good reason to believe that the semistructured interviews and standardized rating scales reviewed here should provide more reliable and comprehensive assessments, the evidence for this assumption is limited and requires further testing.

6. Several treatments of child and adolescent depression have at least preliminary evidence of efficacy in clinical trials (although this does not assure effectiveness in nonresearch contexts; Weisz et al., 1995). Hence, the most critical criterion, at least from a clinical perspective, is treatment utility (Hayes, Nelson, & Jarrett, 1987): Does using these assessment instruments lead to better treatment outcomes than not using them? Hayes et al. and Nelson-Gray (2003) have described a number of research designs that can be used to test treatment utility that are easily implemented. Conducting these studies should be a high priority.

7. Many of the measures discussed in this article are time consuming and, therefore, costly to use. Even if these measures are shown to have incremental validity and treatment utility, it will be necessary to demonstrate that they are “cost effective” in the sense that the incremental gain associated with their use equals or exceeds their marginal cost.

8. The selection of specific measures is only part of an evidence-based approach to assessment. Perhaps even more important is how the data produced by the instruments is integrated and used for clinical decisions. The problem of integrating data from multiple informants for diagnosis and treatment evaluation was discussed previously. This is clearly an area that requires further work. In addition, the reliability and validity of case formulations and treatment planning is a critical area in which little work has been done for child and adolescent depression. However, if recent research with adults is any indication (Bieling & Kuyken, 2003), this is likely to be highly challenging.

9. Although it is important for clinicians to systematically monitor patient progress, there are few guidelines on the correspondence between specific scores on particular rating scales and clinically meaningful change points in course trajectories (e.g., response, remission, recovery, relapse, recurrence) that guide decisions regarding whether treatment should be intensified, changed, or discontinued. Fortunately, there are some models in the adult literature that may be useful in beginning to consider this issue (e.g., Jacobson, Roberts, Berns, & McGlinchey, 1999; Lueger et al., 2001; Riso et al., 1997).

10. Work on assessing depression in children under the age of 8 years is just getting underway, and only limited data on measures for this age group are available. The development and validation of assessment
instruments for depression and its precursors in young children is an important area for future research.

**Recommendations for Clinical Practice**

Although it will be some time before an evidence-based approach to the assessment of child and adolescent depression is available, we believe that it is possible to offer some tentative recommendations.

One of the major purposes of assessment is for diagnosis, prognosis, and treatment planning. An evaluation of child and adolescent depression that is designed to address these issues should include (a) determining whether criteria are met for MDD or DD; (b) ruling out exclusionary diagnoses such as bipolar disorder and depression due to a general medical condition or substance; (c) assessing key symptoms such as suicidal ideation and psychotic symptoms that might influence treatment decisions; (d) carefully assessing the previous course of the depression (e.g., prior episodes, chronicity); (e) evaluating comorbid psychiatric, developmental, and general medical disorders; (f) assessing family, school, and peer functioning; (g) exploring significant stressors and traumas; and (h) assessing family history of psychopathology.

To assess these areas, we recommend using interviewer-based (or semistructured) interviews covering psychopathology, psychosocial functioning, and family history. This recommendation is not offered lightly, as semistructured interviews require much more time than clinicians typically devote to assessment. However, in light of evidence that key areas of psychopathology are frequently overlooked when less systematic approaches are used, we believe that this is the "best practice" approach to assessment. Given the time requirements and cost of semistructured diagnostic interviews, an important priority for future research is the development of more efficient and streamlined procedures, such as creating screening modules or making limited use of self- or parent-report screening measures, to reduce the time needed to complete the interviews.

Although each of the semistructured diagnostic interviews discussed in this article are reasonable choices, our inclination is to use the K-SADS or the CAPA. The K-SADS is the most widely used interviewer-based diagnostic interview, has accumulated strong support for its psychometric properties, and is available in several versions that vary in terms of the time period assessed and the precision of information about specific symptoms. Although it is newer and less extensively studied, the CAPA is also a good choice, particularly because it also assesses social functioning and stress.

If the K-SADS or a semistructured diagnostic interview other than the CAPA is used, it should be supplemented with a measure of psychosocial functioning. We recommend using a measure that assesses specific areas of functioning (as opposed to a global assessment) and does not overlap with the diagnostic interview. The SAICA is the best-studied measure that meets these requirements, although the Longitudinal Interval Follow-up Evaluation-Range of Impaired Functioning Tool is also attractive due to its brevity and may prove to be a good alternative if future studies support its psychometric properties. Finally, the assessment should include a semistructured family history interview, such as the Family History Research Diagnostic Criteria.

We strongly recommend obtaining data on psychopathology, social functioning, and family history from multiple informants whenever possible. Data on psychopathology and social functioning should be obtained from both the child and the primary caretaker, particularly for school-age children. It may be less useful to conduct direct interviews with younger children, and informant data may not be necessary for older adolescents, particularly if the focus is on internalizing disorders such as depression. It is often difficult to obtain family history data from multiple sources, but whenever possible, assessments should be conducted with both the primary caretaker and another adult (or older adolescent) who knows the family well.

When information from multiple sources is available, it raises the problem of how to integrate the diverse data. We suggest using the best-estimate approach (e.g., Klein et al., 1994), although other approaches such as the "or" rule (which maximizes sensitivity at the cost of reduced specificity) are also reasonable.

After treatment has been initiated, it is important to monitor and evaluate its effects on a regular basis using clinician, self-rating, and parent-rating scales. Although a reduction in scores should be viewed cautiously in light of the possibility of attenuation effects, this provides a standard means of assessing progress and may ultimately lead to the development of evidence-based algorithms indicating when change in treatment is needed to optimize patient outcomes. We recommend using a clinician rating scale, such as the revised version of the CDRS. It may also be useful to supplement the revised CDRS with a self-rating scale. A number of self-rating scales for depression in children and adolescents are available, and the literature does not provide a basis for choosing among them. Finally, it is important to supplement clinician and self-rating scales with information from the primary caretaker, particularly for preadolescents. Unfortunately, few clinician or rating scales for child depression have been developed and validated specifically for use with parents, although measures designed for children are commonly adapted for this purpose. At present, it appears best to use the parent version of the CDI or a broadband parent-rating scale such as the CBCL to obtain information from the primary caretaker.
Finally, self- and parent-rating scales may be used in screening for child and adolescent depression in a variety of settings. Screening measures can be useful in increasing vigilance for patients with high scores. However, their limited sensitivity and specificity indicate that they should be used with caution, particularly in contexts with very low (or very high) base rates. At present, the literature does not provide a basis for choosing among the various self- and parent-rating scales for screening purposes.

References


nostic assessment in child and adolescent psychopathology (pp. 65–87). New York: Guilford.


ASSESSMENT OF CHILD AND ADOLESCENT DEPRESSION


