



Youth Depression Screening with Parent and Self-Reports: Assessing Current and Prospective Depression Risk

Joseph R. Cohen¹ · Felix K. So^{1,3} · Jami F. Young² · Benjamin L. Hankin¹ · Brenda A. Lee¹

Published online: 8 February 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Few studies have examined the incremental validity of multi-informant depression screening approaches. In response, we examined how recommendations for using a multi-informant approach may vary for identifying concurrent or prospective depressive episodes. Participants included 663 youth ($Age_M = 11.83$; $Age_{SD} = 2.40$) and their caregiver who independently completed youth depression questionnaires, and clinical diagnostic interviews, every 6 months for 3 years. Receiver operating characteristic (ROC) analyses showed that youth-report best predicted concurrent episodes, and that both youth and parent-report were necessary to adequately forecast prospective episodes. More specifically, youth-reported negative mood symptoms and parent-reported anhedonic symptoms incrementally predicted future depressive episodes. Findings were invariant to youth's sex and age, and results from person and variable-centered analyses suggested that discrepancies between informants were not clinically meaningful. Implications for future research and evidence-based decision making for depression screening initiatives are discussed.

Keywords Depression · Multi-informant screening · Receiver operating characteristics · Translational research

Introduction

Adolescence is a critical period for the development of depression and its onset confers significant risk for functional impairment, comorbid forms of psychopathology, and suicidal behavior [1]. While less prevalent, childhood depression is also associated with significant impairment and is a predictor of future mental health problems [2]. Given the prevalence, chronicity, and consequences associated with youth depression, there is an urgent need for early intervention and improved depression screening protocols [1].

Increasingly, providers are encouraged to screen early and often for depression [3, 4], but detailed recommendations for how to accomplish this aim are largely missing. Most agree

that a multi-informant approach, in which multiple perspectives are solicited about a youth's symptoms, is necessary [5], but a paucity of incremental validity studies test this claim [6]. In addition, studies that examine multi-informant protocols for depression (a) conduct them in treatment-seeking populations [7, 8], (b) do not include self-reports (i.e., only parent and teacher reports; e.g., [9]), (c) only assess current/short-term prospective (e.g., 6 months) outcomes [10] or (d) rely on questionnaires for diagnoses [11]. Collectively, these limitations inhibit research from informing universal child and adolescent depression screening, at a time when its use is encouraged. The present study sought to address this gap in the literature by integrating current trends in the youth assessment research [12] to provide recommendations for child and adolescent depression screening. Particular attention was paid to how the validity of parent and youth reports may vary for concurrent versus prospective diagnoses, whether informants differ in their ability to report on specific symptoms (e.g., negative mood, anhedonia), discrepant informant reports (e.g., parent reports high symptoms while youth reports low symptoms), and the moderating impact of age and sex. Findings from our study can provide an empirical foundation for feasible, multi-informant depression screening initiatives.

✉ Joseph R. Cohen
cohenj@illinois.edu

¹ Department of Psychology, University of Illinois Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820, USA

² Department of Child and Adolescent Psychiatry and Behavioral Sciences, Children's Hospital of Philadelphia, Philadelphia, PA, USA

³ Department of Psychology, University of California Los Angeles, Los Angeles, CA, USA

Trends in Child and Adolescent Depression Screening

The majority of depression screening utilizes a single informant. In a recent review, 85% of pediatric primary care mental health screening protocols relied on the parent or child report [13]. This trend is in stark contrast to the assessment setting where a multi-informant approach is the most common method [14]. Reliance on single-informant protocols may reflect the challenges of integrating multiple sources of data into clinical decision-making at the screening setting. However, examining the incremental validity of different informants can help reduce the burden of screening protocols by only retaining the relevant information [6]. By prioritizing certain index tests within the screening setting and leveraging technological advancements (e.g., computerized adaptive testing; [15]), multi-informant screening can become a feasible and more targeted step in larger depression prevention initiatives.

The vast majority of research concerning “best practices” for identifying youth mental health diagnoses, including the use of multi-informant approaches, stems from the assessment literature. Collectively, these studies suggest an informant gradient in which parent-report is preferred to youth report, but parent and youth report is preferred to parent-report [16]. However, the majority of these studies have not adequately examined the incremental validity of multi-informant approaches [6], nor distinguished between different mental health diagnoses. For example, De Los Reyes and colleagues did not identify a single study that explicitly examined the incremental validity of multi-informant approaches for depression in their comprehensive review [17]. As parent–child disagreement for depressive symptoms is uniquely common [18, 19], creating decision algorithms specific to depression is necessary.

To date, only a few studies provide insight into how to interpret depression questionnaire data from multiple informants. Recently, Salcedo and colleagues [20] and Johnson and colleagues [10] found that parent-report better predicted mood disorder status compared to teacher-report; however, these studies both noted as limitations the exclusion of youth reports. As self-report may be necessary to capture less observable phenomena (e.g., cognitive and emotional states; [17, 21]), it is important to compare self and parent reports of depressive symptoms. When comparing youth and parent-report, Fristad and colleagues [22] and Lewis and colleagues [7] found that only the youth self-report, and not parent report, discriminated between depressed and non-depressed youth in clinical samples. Yet, both of these studies were focused on clinical samples, and neither assessed prospective outcomes.

As the primary aims of universal depression screening initiatives are to (a) identify current distress/impairment and (b) estimate prospective depression risk in an unselected sample [3], it is important to develop decision rules for both current and future depressive episodes in a general community sample.

Individual Differences in Reporting on Youth Depression

For over 25 years, one of the more robust effects within the child mental health assessment literature is the modest agreement between the self and others when reporting on internalizing distress [17, 23, 24]. Discrepant reports impact depression screening algorithms by forcing the administrator to determine the veracity of each informant. To date, a variety of perspectives provide guidance for interpreting and responding to discrepant reports [25, 26]. Collectively, these can be grouped into *person-centered* explanations, in which discrepant scores reflect a subpopulation, and *variable-centered* explanations, in which discrepancies result from normative individual differences (e.g., demographics, symptom presentations). While there is no consensus for which model best explains informant discrepancies (and a combination of reasons is likely), there is agreement that discrepancies can be meaningful [14] and need to be investigated when developing decision rules for multi-informant protocols.

Person-centered hypotheses are important to consider within a screening framework because they may suggest the need for different decision algorithms for different subpopulations. For instance, the *depression–distortion hypothesis* suggests that negativity biases stemming from the caregiver’s depressive diagnostic status leads to elevated reports of the offspring’s depression [27]. Within this perspective, parental reports are overly biased, and youth reports should be prioritized. To date, support for this particular hypothesis is mixed [17]; however, emerging research does suggest that discrepant depression reports may reflect subpopulations of youth. Specifically, Makol and Polo identified a profile of youth with high self-reported depressive symptoms and low parent-reported symptoms [28]. The authors speculated that this class of individuals represented a subpopulation of youth with parents who were less attuned to their youth’s emotional functioning. Based on these findings, youth reports would be more valid for this profile, but for other youth with converging inventories, both parent and youth-reports may provide incremental validity for determining depression diagnostic status and risk.

More commonly, informant discrepancies are explained via *variable-centered* models. For instance, studies have examined how the validity of self- and parent-reported depressive symptoms vary as a linear function of the

youth's age. In sum, these findings are largely inconclusive, potentially due to issues related to sample size for detecting what may be small but significant effects [29]. A more consistent finding, however, is that the validity of youth and parent reports may vary as a function of symptom quality. As previously stated, parents tend to be better reporters of observable, behavioral symptoms, while self-reports are more sensitive for internal cognitive and emotional states [17]. Traditionally, these discrepancies are studied between diagnoses; however, these findings could have important implications within disorders as well. For example, parents may be better equipped to identify behavioral symptoms of anhedonic depression (e.g., apathy, impaired sleeping and eating behavior) compared to the more internal processes related to negative mood (e.g., depressed mood, feelings of worthlessness). To our knowledge, while studies have examined descriptive differences between informants based on anhedonic and negative mood symptoms (e.g., [28]) they have not examined if these reports differentially predict concurrent and prospective depression diagnostic status.

The Present Study

To test our study's aims, we examined the relation between self- and parent-reports of the Children's Depressive Inventory (CDI; [30]) with concurrent and prospective depressive episodes measured via a semi-structured diagnostic interview [31]. The CDI was chosen as our screening inventory because it is a recommended depression screener [5], is one of the most utilized measure within childhood depression research [32], contains valid subscales for different facets of depression [30] and is one of the few scales previously examined in incremental validity studies [32, 32]. Consistent with past research, we hypothesized that parent reports would not contribute incremental validity to the identification of current episodes [22]. Alternatively, consistent with questionnaire data on internalizing symptoms [24], we predicted that a combination of parent and youth reports would best predict prospective depression. We hypothesized parents' ability to better identify behavioral, anhedonic symptoms of depression, which uniquely contribute towards prospective depressive episodes in adolescence [33, 34], would help explain these findings. Exploratory analyses tested whether these findings would vary across discrepant/convergent profiles and demographic characteristics.

Finally, a theoretically-informed analytic plan was used to test our study's aims. First, we examined how discrepant reports may impact our decision algorithm by analytically testing a person-centered [28] and variable-centered explanation [35]. Second, we utilized a recommended, translational, analytic plan (i.e., receiver operating characteristics paired with multilevel diagnostic likelihood ratios; [12, 36, 37]) to

estimate risk across subthreshold and threshold scores. Using these multiple cutoffs can help balance the tension between capturing the dimensional nature of depression and generating clinically useful cut-off scores [36, 38]. Collectively, our analytic plan can directly inform recommended [3] and emerging [10] youth screening protocols that aim to simultaneously gauge concurrent and prospective depression risk.

Methods

Participants and Procedures

Children and adolescents were recruited at two sites: University of Denver and Rutgers University. Brief information letters were sent home directly to families with a child in the third, sixth, or ninth grades at participating school districts. Of the families to whom letters were sent, 1108 participants responded to the letter and called the laboratory for more information. Over the phone, parents established that the parent and child were fluent in English, the child did not have an autism spectrum or psychotic disorder, and the child had an IQ > 70, making them eligible for the study. At baseline, 663 youth (approximately 60% of the total number of families that initially contacted the laboratory) qualified as participants for the study, as they met criteria and completed self-reports and the diagnostic assessments at baseline. Participants included the youth, who ranged in age from 7 to 16 ($M = 11.83$; $SD = 2.40$), as well as one caregiver. Overall, 91% of caregivers identified as maternal caregivers, 7% identified as paternal caregivers, and 1% identified as other family members (e.g., grandparent).¹ Youth were balanced with regard to sex (Female = 56%) and grade (3rd = 30%; 6th = 37%; 9th = 32%), and reflected the racial/ethnic composition of the United States, with the exception of less Hispanic youth (White = 62.2%; African-American = 11.3%; Hispanic = 7.5%).

Every 6 months, caregiver-youth dyads completed inventories and diagnostic assessments for youth depression for a total of seven assessments over the course of 3 years. At each follow-up visit, we examined whether the youth currently or in the past 6 months experienced depression. At baseline, 18 months, and 36 months, assessments took place in-person as part of a larger laboratory study, while at 6, 12, 24, and 30 months, diagnostic interviews were conducted over the phone, with the CDI completed either over the phone or via mail.² Retention rate from baseline to 36 month follow-up

¹ As the overwhelming number of caregivers were mothers, and past research suggests non-significant differences between informants who are caregivers [23], all caregivers were included and treated equally in the present study.

² Additional analyses showed that all findings presented in this manuscript were invariant to data method collection (e.g., in-person versus phone versus mail).

for the overall study was 93%. Caregivers provided informed written consent for their own and their child's participation; youth provided written assent. Both youth and caregiver were compensated monetarily for participating and institutional review board (IRB) approval was obtained for all study procedures.

Measures

Depression Diagnoses

Trained interviewers administered the Mood disorders section of the Schedule for affective disorders and Schizophrenia for school age children (K-SADS-PL; [30]) to youth and caregiver at baseline and follow-up. Interviewers were trained and supervised by licensed clinical psychologists. Interviewers completed an intensive training program for administering the K-SADS and for making diagnostic decisions. The training program consisted of attending approximately 40 h of didactic instruction, listening to audiotaped interviews, and conducting practice interviews. The PIs also reviewed interviewers' notes and tapes to confirm the presence of a diagnosis. Best estimate procedures were used to determine diagnostic status [5]. Diagnostic interview interrater reliability was good ($K=0.91$) based on approximately 20% of reviewed interviews. Consistent with past research [39], youth were diagnosed with depression if they met DSM-IV criteria for Major Depressive Disorder (MDD) Definite, MDD-Probable (four depressive symptoms lasting at least 2 weeks), or minor Depressive Disorder (mDD) Definite (two or three depressive symptoms lasting at least 2 weeks).

Depression Symptoms

The Children's Depression Inventory (CDI; [30]), a 27-item questionnaire, assessed both self- and parent-reported symptoms. The CDI measures five domains of depression: negative mood (6 items), interpersonal problems (4 items), ineffectiveness (4 items), anhedonia (8 items), and self-esteem (5 items). The youth (CDI-Y) and parent (CDI-P) report on the CDI are identical except parents answer with regard to how they believe their child feels. Scores on the CDI-P have been shown to be effective in discriminating between depressed and non-depressed youth [40]. For the present study, scores on the CDI-Y ranged from 0 to 35 ($M=7.08$; $SD=5.87$ at baseline; $M=4.17$; $SD=4.71$ average across follow-ups) and CDI-P ranged from 0 to 28 ($M=4.73$; $SD=5.13$ at baseline; $M=4.13$; $SD=5.02$ average across follow-ups). Consistent with past research, youth reported more symptoms than parents [28]. Internal reliability on the CDI-Y ($\alpha=0.84$ – 0.89) and CDI-P ($\alpha=0.86$ – 0.90) was excellent. Reliability estimates for the CDI subscales were: Negative Mood (CDI-Y: $\alpha=0.61$;

CDI-P: $\alpha=0.62$), interpersonal problems (CDI-Y: $\alpha=0.43$; CDI-P: $\alpha=0.46$), ineffectiveness (CDI-Y: $\alpha=0.59$; CDI-P: $\alpha=0.65$), anhedonia (CDI-Y: $\alpha=0.59$; CDI-P: $\alpha=0.62$), and self-esteem (CDI-Y: $\alpha=0.62$; CDI-P: $\alpha=0.61$). Overall, reliability was similar to past research [41].

Data Analytic Strategy

Discrepant Reports

We first examined whether discrepant reports represented a meaningful subpopulation of youth (i.e., a person-centered explanation). Latent profile analyses (LPA) following similar steps used in the youth depression literature (e.g., [28, 42]) were initially conducted with ten depression indicators (i.e., five subscales of the CDI-Y and CDI-P respectively) with age and sex entered as covariates. To determine the fewest number of profiles that best characterized distinct profiles of informants, we used the Lo–Mendel–Rubin likelihood ratio test (LMR LRT) and Vuong–LMR LRT significance tests. Once identifying the best fitting solution based on the LMR LRT and Vuong–LMR LMRT, we inspected information criteria based indices (i.e., Akaike information criteria, Bayesian information criteria) and the entropy criterion to confirm model fit. A priori, we hypothesized between a 2- and 8-profile solution. Within our theoretical model, two class solutions represent convergent high and low reports across symptoms subscales, while increasingly more complex models could reflect the classification of profiles comprised of divergent reports. For instance, an 8-profile solution could reflect youth who report elevated internalizing depression subscales (i.e., negative mood and self-esteem), but underreport behavioral symptoms, with parents who report elevated behavioral symptom subscales (i.e., anhedonia, interpersonal, and ineffectiveness), and lower internalizing symptoms. Once establishing the best LPA solution at baseline, we tested whether it replicated across the follow-up. If discrepant subpopulations were identified, separate ROC analyses (described below) were conducted for each profile. Latent profile analyses were conducted using MPlus [43]. All analyses described below were conducted with SPSS (v24.0).

Next, we used a polynomial regression approach as previously recommended in the multi-informant literature [35]. The full equation for this model is:

$$Y = b_0 + b_1 \text{CDI-Y} + b_2 \text{CDI-P} + b_3 \text{CDI-Y}^2 + b_4 \text{CDI-P}^2 + b_5 \text{CDI-Y} * \text{CDI-P} + e$$

Within this equation, a significant interaction between the youth and parent report ($b_5 \text{CDI-Y} * \text{CDI-P}$) suggests that the validity of youth reports may vary in the presence of certain parental scores (and vice-versa). Inclusion of the quadratic

effects help specify that the interaction is identifying the unique effects of difference scores as opposed to quadratic effects more broadly [44]. If an interaction is significant, post-hoc probes via simple slopes were used to determine if informants disagree regardless of symptom level [35] and whether youth or parent reports are valid within the context of these discrepant profiles [44]. If a significant interaction was identified, ROC analyses for each predictor were conducted with the other informant entered as a covariate. We conducted polynomial regression analyses for both the total scores and symptom subscales (e.g., an interaction between parent and youth reported negative mood).

ROC Analyses

We first tested the validity of the CDI-Y and CDI-P for conferring diagnostic risk. Initially, we examined whether these reports vary as a function of sex and/or age for predicting diagnostic status using logistic regression. For concurrent episodes, CDI-Y and CDI-P scores at each 6 month mark were compared to results from simultaneous K-SADS. These analyses started at the 6 month follow-up to ensure each interview was only covering the past 6 months (i.e., baseline assessments did not specify a 6 month time frame). For prospective episodes, baseline CDI scores predicted episodes over the 3 years. For prospective episodes, a standard significance value of $p < .05$ was utilized, while the significance value for concurrent episodes was conservatively placed a priori at 0.01 due to the serial nature of our analyses.

We next examined if the CDI-Y and CDI-P could adequately discriminate between depressed and non-depressed youth. If findings from the logistic regression were significant, Area under the curve statistics (AUCs) for each subpopulation (e.g., for boys and girls) were calculated separately. We then compared these AUCs to determine if they were statistically different [45]. If AUCs were different, subsequent analyses were conducted separately for these subpopulations; however, if this statistic was non-significant, we calculated an AUC for the whole population. We compared contiguous AUCs to determine whether the association between CDI scores and concurrent episodes varied over the number of assessments [45].

For ROC analyses, the AUC is considered significant if it does not include 0.50 in the asymptotic confidence interval; however, higher cutoffs for clinical utility have been recommended. In the present study, an AUC greater than 0.64 (equivalent to a medium effect size; [46]) was conceptualized as a trending significant predictor, while an AUC of 0.70 was considered a “fair” predictor [47]. If both CDI-Y and CDI-P were above 0.64, we used CDI-Y scores to predict CDI-P scores, and vice versa, and saved the residuals. These residual scores represent the unique variance of each predictor and can be used in formal tests of incremental

validity [36, 48]. If the residuals were significant, both predictors were then entered into binary logistic regression analyses, and AUCs for the saved predictive values were computed. Hanley and McNeil’s method was used to determine whether child, parent, or combined reports differed. Diagnostic likelihood ratios (DLRs) were next created to examine the calibration of each measure [45]. Past research indicates a wide range of cutoffs for the CDI-Y and CDI-P (raw scores between 12 and 19; [30]). Thus, DLRs were based on informative tertiles, with the cut-off for the sub-threshold group placed at 70% sensitivity and the threshold group being formed at 90% specificity for predicting prospective depressive episodes.³ These cutoffs mirror the approximate cutoffs of current screening initiatives for youth mental health conditions [36, 49]. Finally, when both the CDI-Y or CDI-P were incrementally valid, we examined if the validity of symptom clusters (i.e., CDI subscales) varied by informant using the ROC approach described above.

Results

Preliminary Analysis

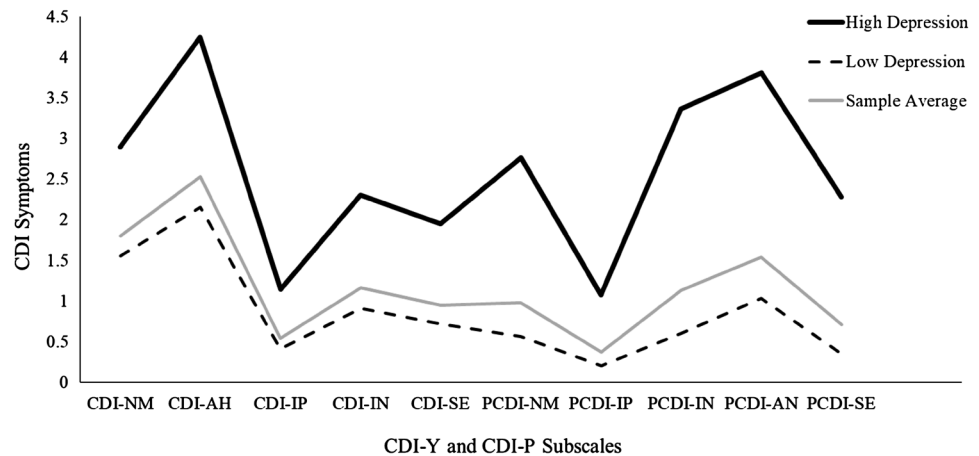
An average of 8.1% of youth were diagnosed with a concurrent depressive episode at each time point ($N_{average} = 45.70$) and 24.3% of the sample met criteria for a new depressive episode during the study ($N = 166$). Chi square analyses showed that females were more likely to have a depressive episode compared to males ($X^2(1) = 8.46, p < .01$) and that 9th graders experienced more episodes compared to 3rd /6th graders ($X^2(2) = 40.46, p < .001$). Bivariate correlations suggested moderate agreement between CDI-Y and CDI-P scores ($r = .34$).

Discrepant Reports

Results from our LPA suggested that a 2-profile solution outperformed a 1-profile solution (LMR LRT = 1445.46, $p < .001$; VLMR LRT = 1428.61, $p < .001$) but none of the higher-ordered solutions were significant. These findings were replicated across follow-ups, suggesting that a 2-profile solution best fit the data (AIC = 25523.03, BIC = 25690.29;

³ Cutoffs for pediatric depression screens ideally have a sensitivity and specificity level of 90% [32]. However, preliminary analyses showed that using a 90% sensitivity cutoff for subthreshold scores was not clinically useful (i.e., over 80% of youth reported scores above the cutoff). Thus, 70% sensitivity was used to determine the subthreshold cutoff as this is the average level of sensitivity for cutoff scores on existing screening measures [36].

Fig. 1 Means for each of the subscales on the CDI-Y and CDI-P for our two profiles identified in our latent profile analyses. The sample average for each subscale is also displayed to provide context for the profiles. For the “High Depression” profile ($N = 129$; 19% of the sample), individuals were classified in the correct profile 96% of the time. For the “Low Depression” ($N = 550$; 81% of the sample) individuals were classified in the correct profile 99% of the time



Entropy = 0.95).⁴ Descriptive statistics for the 2-profile solution can be found in Fig. 1. Subpopulations were defined by “high” (19% of the sample) and “low” (81% of the sample) convergent profiles. Next, polynomial regression models were examined. For concurrent episodes, we did not find significant interactions between CDI-Y and CDI-P ($p = .02$ at 30 month follow-up; p 's range between 0.12 and 0.97 for all other follow-ups). Similarly, for prospective episodes the interaction between CDI-Y and CDI-P was also non-significant ($p = .63$). Findings were replicated for symptom subscales for both prospective (p values ranged between 0.10 and 0.62) and concurrent episodes (average p values ranged between 0.18 and 0.78). Thus, null findings across these analyses suggest that decision rules did not have to vary based on convergent and divergent profiles.

ROC Approach

We first examined whether the validity of CDI-Y and/or CDI-P varied as a function of demographics. For concurrent episodes, we did not find that the CDI-Y or CDI-P varied as a function of sex ($p > 0.01$) or grade ($p > 0.01$). For prospective episodes, the CDI-Y-sex ($p = 0.99$) and CDI-Y-grade interactions ($p = 0.99$) were non-significant. As for CDI-P, findings did not vary as a function of grade ($p = 0.11$) but did vary for sex ($p = 0.01$), such that parents more accurately forecasted episodes for boys compared to girls. Separate AUCs for the CDI-P were calculated for boys and girls; however, the difference in the AUCs in forecasting depressive episodes was non-significant ($p = 0.10$). Thus, subsequent analyses were conducted on the whole sample.

AUC statistics are presented in Table 1 along with corresponding Cohen's d scores. For concurrent episodes, CDI-Y and CDI-P averaged large effect sizes and on average exceeded the 0.70 threshold. These AUCs were similar to past screening research with the CDI [32]. AUCs for the residuals of each inventory suggested that the unique variance associated with the CDI-Y was significant, ($p \leq 0.01$ across follow-ups); but not the CDI-P ($p < 0.01$ at 24 months; $p > 0.05$ at every other follow-up). Finally, the difference between the AUCs for the CDI-Y and CDI-P was not statistically different ($p > 0.10$), but the combined model outperformed the CDI-P at each follow-up ($z > 3.00$; $p \leq 0.01$), but never the CDI-Y ($p > 0.20$). As for prospective episodes, CDI-Y and CDI-P exerted a medium effect (AUC > 0.64). Residuals for the CDI-Y ($p = 0.02$) and CDI-P ($p = 0.01$) suggested that both inventories uniquely forecasted future episodes. Overall, findings suggested no difference between the CDI-Y and CDI-P models for prospective episodes ($p > 0.50$), but that the combined model exerted a large effect (AUC = 0.74) and outperformed both inventories (CDI-Y: $z = 4.36$, $p < 0.001$; CDI-P: $z = 3.80$, $p = 0.001$).

DLRs are presented in Table 2. A score of 15 on the CDI-Y and 12 on the CDI-P were cut-off scores for the threshold group, and scores ranging between 8 and 14 (CDI-Y) and 5–11 (CDI-P) constituted the subthreshold group.⁵ These cutoffs for threshold scores fall within the range of cutoffs used in past research [30, 32]. For concurrent episodes, high CDI-Y scores corresponded to an approximately sixfold increase of likelihood for depression. Meanwhile, despite non-significant findings for the CDI-P's incremental influence on concurrent episodes, adolescents with threshold scores on both inventories were 12-times more likely to

⁴ All statistics reported are based on the findings at baseline. As covariates can lead to unstable class solutions [50], analyses were also conducted without age and sex in the model. The pattern of findings was identical. Please contact the first author for statistics for non-significant models or models replicated past baseline.

⁵ Different cut-off scores for boys and girls and youth of different ages were also tested, but did not lead to a significant improvement in sensitivity/specificity, nor alter the pattern of findings.

Table 1 Areas under the curve and effect sizes for concurrent and prospective depressive episodes

Depressive episodes	CDI-Y AUC (SE)	CDI-Y Cohen's <i>d</i>	CDI-P AUC	CDI-P Cohen's <i>d</i>	Combined AUC	Combined Cohen's <i>d</i>
6 months	0.82 (0.03)	1.30 ^{†††}	0.76 (0.04)	1.00 ^{†††}	0.83 (0.03)	1.35 ^{†††}
12 months	0.77 (0.04)	1.04 ^{†††}	0.68 (0.05)	0.66 ^{††}	0.78 (0.04)	1.09 ^{†††}
18 months	0.73 (0.04)	0.87 ^{†††}	0.68 (0.04)	0.66 ^{††}	0.75 (0.04)	0.95 ^{†††}
24 months	0.67 (0.04)	0.62 ^{††}	0.74 (0.04)	0.91 ^{†††}	0.75 (0.04)	0.95 ^{†††}
30 months	0.75 (0.04)	0.95 ^{†††}	0.74 (0.04)	0.91 ^{†††}	0.80 (0.03)	1.19 ^{†††}
36 months	0.77 (0.03)	1.05 ^{†††}	0.70 (0.04)	0.74 ^{††}	0.78 (0.03)	1.09 ^{†††}
Average concurrent episodes	0.75 (0.04)	0.97^{†††}	0.72 (0.04)	0.81^{†††}	0.78 (0.04)	1.10^{†††}
Prospective episodes	0.65 (0.03)	0.55^{††}	0.66 (0.03)	0.58^{††}	0.74 (0.04)	0.91^{†††}

CDI-Y children's depressive inventory (CDI; Kovacs, 1985)-Youth Report; *CDI-P* CDI-parent report; *Combined* predictive probabilities of CDI-Y and CDI-P predicting depressive episodes; *AUC* area under the curve; *Depressive Episodes* depressive episodes as assessed via the Schedule for affective disorders and Schizophrenia for school age children (K-SADS-PL; Kaufman et al., 1997); *Months* concurrent assessment period (non-cumulative); *Prospective Episodes* whether an individual had a depressive episode onset over the 3 years of the study (cumulative); † medium effect; †† large effect; All AUCs significant ($p < .05$)

present with depression than not. For prospective episodes, adolescents with threshold CDI-Y and CDI-P scores were 6-times more likely to have depression in the future than not.

Finally, we examined the incremental validity of CDI subscales for predicting future depression. For the CDI-Y, we found that *negative mood* best forecasted prospective episodes (AUC = 0.64; SE = 0.03; $p < .001$), and was the only CDI-Y symptom cluster that uniquely predicted episodes (AUC = 0.57; SE = 0.03; $p < .01$) after covarying out other CDI-Y symptoms. As for the CDI-P, *anhedonia* best forecasted depressive episodes (AUC = 0.63; SE = 0.03; $p < .001$) and was the only CDI-P subscale that uniquely predicted prospective episodes (AUC = 0.57; SE = 0.03; $p = .006$). Furthermore, the residuals for both the CDI-Y *negative mood* (AUC = 0.62; SE = 0.03; $p < .001$) and CDI-P *anhedonia* (AUC = 0.60; SE = 0.03; $p < .001$) subscales uniquely predicted future episodes after covarying out the other subscale. The combined AUC for negative mood and anhedonia was 0.69 (SE = 0.03; $p < .001$), slightly below the 0.70 benchmark, but only a 7% decrease in the AUC compared to using the full CDI-Y and CDI-P.

Discussion

Recent meta-analyses indicate the importance of using a multi-informant approach to assessing youth mental health [17, 26]. However, few of these studies specifically focus on depression and most have been tested within a clinical setting to examine concurrent diagnostic status. These limitations prevent empirically-based recommendations during a time when governmental and professional organizations are calling for universal depression screening efforts in youth [3, 4]. Below, we contextualize how our findings advance the

existing assessment literature and then conclude by discussing the clinical implications of our study.

Consistent with past research, both youth and parent-reported symptoms conferred current diagnostic status [51]. Furthermore, we found some support for our hypothesis and past research [33], that parent-reported depressive symptoms did not offer incremental validity once accounting for self-reported symptoms as evidenced by the AUC of the CDI-P residuals being non-significant. At the same time, high scores on both inventories, as opposed to only the youth-report, significantly increased one's likelihood for presenting with a depressive episode. Recent research suggests that ROC may underestimate the incremental validity of novel predictors [52] and that for outcomes with lower base rates (i.e., < 10%) additional metrics other than sensitivity and specificity are needed to assess screening protocols [49]. Thus, rather than discard the parent-report, a multi-gated screening method [21], in which youth-report is first examined, followed by the parent report, may be warranted. This approach can help providers make challenging decisions on youth reports that approach, but do not exceed, the clinical cutoff [38].

The value of a multi-informant screening approach was best exemplified with predicting prospective episodes. Only the combined model was a "fair predictor" that exceeded the AUC cutoff of 0.70, suggesting that utilizing only one inventory is insufficient for predicting future depression. Further, neither inventory was superior in forecasting prospective episodes suggesting that both the CDI-Y and CDI-P should be assessed simultaneously (as opposed to the decision rules for current depression in which the CDI-Y is prioritized). In recent years, different algorithms have been proposed for multi-informant protocols [26]. Some of the most common algorithms are based on "or" or "and" logic for interpreting multiple index tests. For predicting future depression, our

Table 2 Diagnostic likelihood ratios and traditional screening metrics for independent and combined CDI models

		Concurrent episodes = 8.13%						Prospective episodes = 26.40%					
		Traditional screening metrics (subthreshold)						Traditional screening metrics (threshold)					
CDI-Y	Multi-level diagnostic likelihood ratios	Minimal	Subthreshold	Threshold	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV	
Concurrent		0.59	2.88	5.68	0.49	0.88	22.38%	94.93%	0.16	0.97	31.96%	92.84%	
Prospective		0.66	1.40	3.27	0.51	0.76	39.11%	81.05%	0.18	0.95	53.85%	77.00%	
CDI-P		Minimal	Subthreshold	Threshold	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV	
Concurrent		0.60	1.49	4.25	0.57	0.77	15.83%	95.17%	0.25	0.94	27.41%	93.42%	
Prospective		0.64	1.45	3.00	0.50	0.76	39.29%	81.63%	0.15	0.94	50.00%	76.91%	
Multilevel diagnostic likelihood ratios (DLRs) for combined models													
		Minimal, Minimal, Subthreshold	Minimal, Subthreshold, old	Minimal, Subthreshold, old, Threshold	Subthreshold, Subthreshold, old	Subthreshold, old, Threshold	Threshold, Threshold, old	Threshold, old	Subthreshold, Subthreshold, old (PPV/NPV)	Subthreshold, Subthreshold, old (PPV/NPV)	Threshold, Threshold, old (PPV/NPV)	Threshold, Threshold, old (PPV/NPV)	
Concurrent		0.42	1.16	2.03	3.72	5.07	12.04	29.11%	94.10%	34.20%	93.36%	48.26%	
Prospective		0.53	0.88	1.40	1.61	5.14	6.00	52.73%	80.12%	63.93%	78.44%	66.67%	

CDI-Y children's depressive inventory (CDI; [30])—youth report; *CDI-P* CDI-Parent Report; *Combined* predictive probabilities of CDI-Y and CDI-P predicting depressive episodes; *AUC* area under the curve; *DLRs* multilevel diagnostic likelihood ratios [37] also known as positive likelihood ratios; *Minimal* range of lowest risk scores (CDI-Y: 0–7; CDI-P: 0–4); *Subthreshold* range of moderate risk scores (CDI-Y: 8–14; CDI-P: 6–11); *Threshold* range of highest risk scores (CDI-Y: 15 and up; CDI-P: 12 and up); *Sensitivity* likelihood of correctly identifying a depressive episode; *Specificity* likelihood of correctly identifying a non-depressive episode; *PPV* positive predictive value; The probability that subjects with a positive screening test truly have depression; *NPV* negative predictive value; The probability that subjects with a negative screen truly do not have depression; *Combined Models* different combinations of the CDI-Y and CDI-P and the corresponding DLRs for all combinations and PPV/NPV for at-risk combinations on the CDI-Y and CDI-P; Outcomes for concurrent models reflect the average scores across the 6 waves of data collection. Outcomes for concurrent episodes at each wave are available by contacting the first author

findings suggest that “and” rules should be used, as the for the combination of self- and parent-report was superior to the use of either inventory independently.

Low to moderate levels of agreement between informants are problematic for “and” algorithms as there is no clear method for integrating multiple informants that confer opposing information. In the present study, we found low to moderate agreement ($r=0.34$) between youth and parent-reports, which is consistent with past research on internalizing symptoms in general ($r=0.25$; [17]; $r=0.45$; [53]) and for depressive symptoms measured by the CDI specifically ($r=0.23$; [54]; $r=0.37$; [55]). Yet, null findings for our latent profile and polynomial regression analyses suggest that screening protocols would not have to further probe discrepant reports. Instead, the self and parent-reported form should be interpreted independently (e.g., a “15” on the CDI-Y confers the same current or future depression diagnostic status regardless of the CDI-P score). This marks a stark contrast to the assessment context, in which “best practices” suggests one should use a decision tree to understand the nature of the discrepancy [25]. Not only might this not be practical within a screening setting, but based on our findings, there is no incremental validity gained by further understanding discrepant reports.

Analyses concerning the types of depressive symptoms may provide insight into why low to moderate agreement exists between youth and parent reports. In the present study, we found that youth report of negative mood items and parent-reported anhedonia uniquely and incrementally forecasted future symptoms. These results support meta-analytic findings that show parents are better equipped to identify behavioral symptoms, while youth are better reporters on internalizing distress [17]. Further, these findings support past research that suggests parental reports of anhedonia are valid [56], and extend these findings by showing they are incrementally valid compared to youth self-reports of anhedonia. A tension inherent to mental health screening is developing protocols that are sensitive enough to detect specific syndromes, but that can also be administered and scored quickly [3, 49]. Querying negative mood symptoms in youth self-reports and youth anhedonic symptoms in parent-reports may be a fruitful pathway towards reducing the overall burden of a targeted, multi-informant screening protocol.

To date, few studies have examined the screening properties of the CDI, or other depression inventories, within a non-clinical youth sample (see [32] for review). However, within pediatric, non-psychiatric populations with similar base rates for current depression (e.g., 8.13% in the current study versus 7.4%; [57]), comparisons can be made and our study’s findings can be better contextualized. Overall, the positive (31.96%) and negative (92.84%) predictive values for threshold scores on the CDI-Y in the current study are

similar to past research on the CDI-Y (PPV: 21–38%; NPV: 94–100%). While these comparison studies did not include the CDI-P, these studies suggest that the incremental validity of the CDI-P quantified in the current study may generalize above and beyond an established baseline performance for the CDI-Y. As shown in Table 2, the predictive value for current episodes is over 50% higher when considering the CDI-P in addition to CDI-Y scores when predicting concurrent episodes.

As for prospective outcomes, it is more challenging to compare our findings to past research. Cohen and colleagues, in one of the few studies to use an evidence-based approach for future depressive episodes, examined the CDI-Y for first lifetime episodes of depression in youth [58]. Between the two studies a similar estimate for the AUC (0.65 in the current study compared to 0.64) and slightly elevated estimate for the DLR (3.27 in the current study compared to 2.51) was observed.⁶ Interestingly, Cohen and colleagues used a risk factor approach (e.g., assessing pupil dilation) to supplement CDI scores. The inclusion of these risk factors led to an AUC above 0.70 and similar composite DLRs for multiple above threshold scores. Taken together with our current findings, this suggests that reliance on multiple indices of depression is necessary to have a reasonable approach for screening for prospective depression. Based on comparable statistical accuracy between the two algorithms, whether one uses the CDI-P or psychophysiological assessments may depend on the setting’s resources and access to caregivers.

We offer our findings in light of certain limitations. First, baseline data collection for the present study began in 2009, 1 year before the CDI-II self and parent-report were published. Relatedly, despite the CDI-P’s common use in research (e.g., [51, 54]), *t*-scores are not available for this inventory, limiting our ability to use standardized cutoff scores. Second, future research is needed with more parsimonious and ideally publicly available measures for youth depression (e.g., The patient health questionnaire-2; [59]) to confirm that our findings extend beyond the CDI. Third, future studies need to be conducted within applied settings to ensure generalizability beyond research contexts [12]. Fourth, negative mood and anhedonia are multi-faceted constructs and we could not determine which aspects of negative mood (e.g., cognitions versus emotions) or anhedonia (e.g., social versus physical symptoms) parents and youth differed.

Finally, even for our highest risk youth, the positive predictive value (PPV) is only moderate (approximately 40%

⁶ As this study specifically focused on first lifetime episodes as opposed to prospective episodes more broadly, base rates were dissimilar between the two studies. We therefore used the AUC and DLRs, which are unaffected by base rate, to compare the two studies.

for current depression 65% for prospective depression). While this is partially tied to the base rate for depression [12], it also suggests that over half of the youth that would be referred would not be currently depressed and approximately one-third will never go on to develop depression. Thus, although these PPVs are higher than current depression screening protocols [32], it is important that future research aim to increase the predictive value of depression screening initiatives. At the same time, it may be reasonable for depression screens to have a high NPV, but only a moderate PPV like in the current study [60]. A moderate PPV suggests that several youth may be exposed to further assessment or even preventative interventions that are not warranted. Yet, in the case of depression screening, these services may not be too burdensome or invasive and could even be helpful. For instance, a more extensive mental health assessment could identify other patterns of psychological distress distinct from depression. Meanwhile, cognitive behavioral and socio-emotional depression preventative interventions can be effective even for those at lower levels of risk (albeit to a lesser extent to those at high-risk; [61]). Thus, we recommend that a multi-informant screening approach can be clinically useful, especially for identifying prospective depression risk in youth.

Clinical Implications

Translational studies that leverage the strengths of basic research to inform clinical decision-making is necessary in child and adolescent mental health [12, 36]. Using a multi-wave, longitudinal study and multi-faceted analytic plan, we were able to provide concrete recommendations to the clinical setting. First, self-reports should be prioritized for identifying current depression diagnostic status. We recommend only using parent-report for when self-reported scores are at or near the cutoff. Second, reliable clinical estimates of prospective depression can *only* be made by using both parent and youth reports. This finding is critical, as a primary aim of universal depression screening is to identify prospective depression risk [3]. Finally, our study highlights how clinical decision making should differentially consider assessment approaches for negative mood and anhedonia when predicting future depression risk.

Table 3 provides a summary of the study's findings, and an example of how our results can be used to inform clinical decision making from the screening setting. Using the DLRs from Table 2, we calculated the probability of concurrent and prospective depression for five scoring profiles based on their pre-test probability (i.e., the likelihood of having depression based on your age and gender). We next used an evidence-based medicine, "stoplight" approach [62], which categorizes patients based on risk: "Green" (i.e., minimal/no

risk), "yellow" (i.e., continued monitoring) and "red" (i.e., refer to mental health providers)⁷ based on their probability of presenting with depression in light of their CDI-Y and CDI-P scores. Posttest probabilities for both the CDI-Y and CDI-P, as well as the combination of scores, are presented as a way to quantify the value gained by using a multi-informant approach. We note the "stoplight" column is just an example for how to interpret this table and that ultimate inclusion/referral decisions rely on cost-benefit analyses associated with different screening settings and goals (see [62] for additional guidance on how to interpret Table 3). Ultimately, use of a translational analytic plan [12] paired with continuing education in applied settings on evidence-based medicine, can serve as a bridge for the notorious translational gap and ultimately facilitate better depression recognition in vulnerable children and adolescents.

Summary

To date, few studies have adequately examined the incremental validity of multi-informant assessments for the screening setting. In response, we examined how clinical decision making within a multi-informant approach may vary for predicting concurrent or prospective depressive episodes. To accomplish this aim we tested whether the external and incremental validity of parent and youth reports varied within the context of convergent/divergent profiles, as a function of symptom presentation (e.g., negative mood and anhedonia), or child characteristics (i.e., sex and age) for predicting depression outcomes. Participants included 663 youth ($Age_M = 11.83$; $Age_{SD} = 2.40$) and their caregiver who independently completed youth depression questionnaires, and clinical diagnostic interviews, every 6 months for 3 years. Receiver operating characteristic (ROC) analyses showed that youth self-report best predicted concurrent episodes, and that both youth and parent-report were needed to predict prospective episodes. More specifically, youth-reported negative mood symptoms and parent-reported anhedonic symptoms provided incrementally valid forecasts for prospective episodes. Latent profile and polynomial regression analyses suggested that different decision rules were not necessary for profiles of discrepant reports. Furthermore, these findings were invariant to youth's sex and age. Results were presented and discussed in a manner to facilitate evidence-based decision making for depression screening initiatives.

⁷ We note that in Youngstrom's original model, red references "acute treatment." As our findings are based on non-clinician administered inventories, we recalibrated the recommendations within the model.

Table 3 “Stoplight” recommendations based on different screening scenarios

Sample	Concurrent DLR	Prospective DLR	Post-test probability	Interpretation
3rd Grade girls and boys: Concurrent: 4.02%; over 3 years: 13.5%				
Subthreshold CDI-Y (8–14); Subthreshold CDI-P (5–11)	CDI-Y: 2.88 CDI-P: 1.49 Combined: 3.72	CDI-Y: 1.40 CDI-P: 1.45 Combined: 1.61	CDI-Y Concurrent: 10.32% Prospective: 18.30% CDI-P Concurrent: 5.62% Prospective: 18.83% Combined Concurrent: 12.95% Prospective: 20.40%	Green: Despite DLRs above 1.00 for both parents and youth report, low base rates for girls and boys of this age suggest that it is still rather unlikely they are currently experiencing depression or will go on to experience depression. Specifically, 4 out of 5 youth with this profile will not go on to develop depression in the next 3 years
6th Grade boys: Concurrent: 4.13%; over three years: 15.4%				
Threshold CDI-Y (15+) Minimal CDI-P (0–4)	CDI-Y: 5.68 CDI-P: 0.60 Combined: 2.03	CDI-Y: 3.27 CDI-P: 0.64 Combined: 1.40	CDI-Y Concurrent: 18.51% Prospective: 37.05% CDI-P Concurrent: 2.34% Prospective: 10.33% Combined Concurrent: 7.51% Prospective: 20.12%	Yellow: It is unlikely that boys with these scoring profiles are experiencing depression currently-despite the elevated DLR for the CDI-Y. However, the elevated CDI-Y should give one pause for prospective depressive episodes. Probably best to monitor symptoms in the immediate future, though a referral to outpatient mental health services may be premature absent any critical symptoms (i.e., suicidal thoughts)
6th Grade girls: Concurrent: 9.03%; over 3 years: 26.7%				
Subthreshold CDI-Y (8–14) Threshold P-CDI-P (12+)	CDI-Y: 2.88 CDI-P: 4.25 Combined: 5.07	CDI-Y: 1.40 CDI-P: 3.00 Combined: 5.14	CDI-Y Concurrent: 22.36% Prospective: 34.23% CDI-P Concurrent: 29.82% Prospective: 52.60% Combined Concurrent: 33.64% Prospective: 65.53%	Red: Across parent and youth report, findings suggest an approximate threefold increase in the risk for currently having a depressive episode when compared to the base rate for this subsample. Even more concerning, approximately 2/3 of these girls will go on to develop a depressive episode in the next 3 years with this demographic and scoring profile. A referral for an assessment should be made
9th Grade boys: Concurrent: 8.63%; over 3 years: 31.4%				
Threshold CDI-Y (15+) Subthreshold CDI-P (5–11)	CDI-Y: 5.68 CDI-P: 1.49 Combined: 5.07	CDI-Y: 3.27 CDI-P: 1.45 Combined: 5.14	CDI-Y Concurrent: 36.22% Prospective: 59.53% CDI-P Concurrent: 12.97% Prospective: 39.48% Combined Concurrent: 33.64% Prospective: 69.81%	Red: Similar to the profile described prior a referral should probably be made for this youth. The CDI-Y, the superior indicator for concurrent episodes, suggests a fourfold increase in risk compared to the pre-test base rate. Furthermore, nearly 70% of youth with this demographic and scoring profile will experience an episode in the next 3 years
9th Grade girls: Concurrent: 17.16%; over 3 years: 47.2%				
Minimal CDI-Y (0–7); Threshold P-CDI-P (12+)	CDI-Y: 0.59 CDI-P: 4.25 Combined: 2.03	CDI-Y: 0.64 CDI-P: 3.00 Combined: 1.40	CDI-Y Concurrent: 10.55% Prospective: 36.29% CDI-P Concurrent: 45.94% Prospective: 72.75% Combined Concurrent: 28.87% Prospective: 55.47%	Yellow: Minimal CDI-Y suggest that a current episode is unlikely. Although the combined report suggests that over half of these youth will experience an episode this is only a slight increase from the pre-test base rate for this subsample. Elevated parent report suggests a considerable risk for a prospective episode warranting continued monitoring but not yet a referral

The following are a presentation of potentially challenging screening cases. Prevalence estimates derived from our sample for concurrent and prospective episodes are provided after describing demographic details (i.e., sex, grade) for each exemplar

Table 3 (continued)

DLR (True positives within a specific scoring range/total number of positive cases)/(The number of false positives within a scoring range/total number of negative cases; [37]). *DLRs* are presented for both concurrent and prospective episodes; *Post-test probability* (pre-test odds x *DLR*)/(pre-test odds x *DLR* + 1); *Interpretation* Decision making based on post-test probability; *Green* No action; *Yellow* Monitor; *Red* Refer for assessment; *CDI-Y* children's depressive inventory (CDI; [30])-Youth Report; *CDI-P* CDI-parent report; *Combined* predictive probabilities of *CDI-Y* and *CDI-P* predicting depressive episodes; Minimal, subthreshold, threshold *CDI* scoring categories derived from Table 2

Funding This research was supported by National Institute of Mental Health Grants 5R01MH077195 and 5R01MH077178 awarded to Benjamin Hankin and Jami Young by the National Institute of Mental Health. The authors have no other conflicts of interest to report.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in our study were in accordance with the ethical standards of the institutional and/or national research committee with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent and parental consent was obtained from all participants.

References

1. Avenevoli S, Swendsen J, He JP, Burstein M, Merikangas KR (2015) Major depression in the national comorbidity survey-adolescent supplement: prevalence, correlates, and treatment. *J Am Acad Child Adolesc Psychiatry* 54:37–44. <https://doi.org/10.1016/j.jaac.2014.10.010>
2. Garber J, Horowitz JL (2002) Depression in children. In: Gotlib I, Hammen's C (eds) *Handbook of depression*. The Guilford Press, New York, pp 510–540
3. Siu AL (2016) Screening for depression in children and adolescents: US preventive services task force recommendation statement. *Pediatrics* 137:3. <https://doi.org/10.1542/peds.2015-4467>
4. Tanski S, Garfunkel LC, Duncan PM, Weitzman M (2010) Performing preventive services: a Bright Futures handbook. American Academy of Pediatrics, Itasca
5. Klein DN, Dougherty LR, Olino TM (2005) Toward guidelines for evidence-based assessment of depression in children and adolescents. *J Am Acad Child Adolesc Psychiatry* 34:412–432. https://doi.org/10.1207/s15374424jccp3403_3
6. Johnston C, Murray C (2003) Incremental validity in the psychological assessment of children and adolescents. *Psychol Assess* 15:496–507. <https://doi.org/10.1037/1040-3590.15.4.496>
7. Lewis AJ, Bertino MD, Bailey CM, Skewes J, Lubman DI, Toumbourou JW (2014) Depression and suicidal behavior in adolescents: a multi-informant and multi-methods approach to diagnostic classification. *Front Psychol* 5:766. <https://doi.org/10.3389/fpsyg.2014.00766>
8. Weisz JR, Sandler IN, Durlak JA, Anton BS (2005) Promoting and protecting youth mental health through evidence-based prevention and treatment. *Am Psychol* 60:628–648. <https://doi.org/10.1037/0003-066X.60.6.628>
9. Aitken M, Martinussen R, Tannock R (2017) Incremental validity of teacher and parent symptom and impairment ratings when screening for mental health difficulties. *J Abnorm Child Psychol* 45:827–837. <https://doi.org/10.1007/s10802-016-0188-y>
10. Johnson S, Hollis C, Marlow N, Simms V, Wolke D (2014) Screening for childhood mental health disorders using the strengths and difficulties questionnaire: the validity of multi-informant reports. *Dev Med Child Neurol* 56:453–459. <https://doi.org/10.1111/dmcn.12360>
11. Fristad MA, Emery BL, Beck SJ (1997) Use and abuse of the children's depression inventory. *J Consult Clin Psychol* 65:699–702
12. Youngstrom EA, Van Meter A, Frzier TW, Hunsley J, Prinstein MJ, Ong M, Youngstrom JK (2017) Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clin Psychol Sci* 24:331–363. <https://doi.org/10.1111/cpsp.12207>
13. Wissow LS, Brown J, Fothergill KE, Gadomski A, Hacker K, Salmon P et al (2013) Universal mental health screening in pediatric primary care: a systematic review. *J Am Acad Child Adolesc Psychiatry* 52:1134–1147. <https://doi.org/10.1016/j.jaac.2013.08.013>
14. Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, Kupfer DJ (2003) A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *Am J Psychiatry* 160:1566–1577. <https://doi.org/10.1176/appi.ajp.160.9.1566>
15. Streiner DL, Norman GR, Cairney J (2015) *Health measurement scales: a practical guide to their development and use*, 5th edn. Oxford University Press, New York
16. Kuhn C, Aebi M, Jakobsen H, Banaschewski T, Poustka L, Grimmer Y et al (2017) Effective mental health screening in adolescents: should we collect data from youth, parents or both? *Child Psychiatry Hum Dev* 48:385–392. <https://doi.org/10.1007/s10578-016-0665-0>
17. De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DAG, Burgers DE et al (2015) The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychol Bull* 141:858–900. <https://doi.org/10.1037/a0038498>
18. Dolle K, Schulte-Körne G, O'Leary AM, von Hofacker N, Izat Y, Allgaier AK (2012) The Beck Depression Inventory-II in adolescent mental health patients: cut-off scores for detecting depression and rating severity. *Psychiatry Res* 200:843–848
19. Lauth B, Arnelsson GB, Magnússon P, Skarphéðinsson G, Ferrari P, Pétursson H (2010) Parent–youth agreement on symptoms and diagnosis: assessment with a diagnostic interview in an adolescent inpatient clinical population. *J Physiol Paris* 104:315–322
20. Salcedo S, Chen Y, Youngstrom EA, Fristad MA, Gadow KD, Horwitz SM et al (2017) Diagnostic efficiency of the child and adolescent symptom inventory (CASI-4R) depression subscale for identifying youth mood disorders. *J Clin Child Adolesc Psychol*. 1–15. <https://doi.org/10.1080/15374416.2017.1280807>
21. Stiffler MC, Dever BV (2015) *Mental health screening at school: instrumentation, implementation, and critical issues*. Springer, New York
22. Fristad MA, Weller RA, Weller EB, Teare M, Preskorn SH (1991) Comparison of the parent and child versions of the children's depression inventory (CDI). *Ann Clin Psychiatry* 3:341–346

23. Achenbach TM, McConaughy SH, Howell CT (1987) Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychol Bull* 101:213
24. van de Looij-Jansen PM, Jansen W, Wilde EJ, Donker MC, Verhulst FC (2010) Discrepancies between parent-child reports of internalizing problems among preadolescent children: relationships with gender, ethnic background, and future internalizing problems. *J Early Adolesc* 31:443–462
25. De Los Reyes A, Lerner MD, Thomas SA, Daruwala S, Goepel K (2013) Discrepancies between parent and adolescent beliefs about daily life topics and performance on an emotion recognition task. *J Abnorm Child Psychol* 41:971–982. <https://doi.org/10.1007/s10802-013-9733-0>
26. Martel MM, Markon K, Smith GT (2017) Research review: multi-informant integration in child and adolescent psychopathology diagnosis. *J Child Psychol Psychiatry* 58:116–128. <https://doi.org/10.1111/jcpp.12611>
27. Gartstein MA, Bridgett DJ, Dishion TJ, Kaufman NK (2009) Depressed mood and maternal report of child behavior problems: another look at the depression–distortion hypothesis. *J Appl Dev Psychol* 30:149–160. <https://doi.org/10.1016/j.appdev.2008.12.001>
28. Makol BA, Polo AJ (2017) Parent-child endorsement discrepancies among youth at chronic-risk for depression. *J Abnorm Child Psychol* 46:1077–1088. <https://doi.org/10.1007/s10802-017-0360-z>
29. De Los Reyes A, Kazdin AE (2005) Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychol Bull* 131:483–509. <https://doi.org/10.1037/0033-2909.131.3.483>
30. Kovacs M (1992) Children's depression inventory (CDI) manual. Multi-Health Systems, New York
31. Kaufman J, Birmaher B, Brent D, Rao UMA, Flynn C, Moreci P et al (1997) Schedule for affective disorders and Schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 36:980–988. <https://doi.org/10.1097/00004583-199707000-00021>
32. Stockings E, Degenhardt L, Lee YY, Mihalopoulos C, Liu A, Hobbs M, Patton G (2015) Symptom screening scales for detecting major depressive disorder in children and adolescents: a systematic review and meta-analysis of reliability, validity and diagnostic utility. *J Aff Dis* 174:447–463. <https://doi.org/10.1016/j.jad.2014.11.061>
33. Pine DS, Cohen E, Cohen P, Brook J (1999) Adolescent depressive symptoms as predictors of adult depression: moodiness or mood disorder? *Am J Psychiatry* 156:133–135. <https://doi.org/10.1176/ajp.156.1.133>
34. Gabbay V, Johnson AR, Alonso CM, Evans LK, Babb JS, Klein RG (2015) Anhedonia, but not irritability, is associated with illness severity outcomes in adolescent major depression. *J Child Adolesc Psychopharmacol* 25:194–200. <https://doi.org/10.1089/cap.2014.0105>
35. Laird RD, Reyes AD (2013) Testing informant discrepancies as predictors of early adolescent psychopathology: why difference scores cannot tell you what you want to know and how polynomial regression may. *J Abnorm Child Psychol* 41:1–14. <https://doi.org/10.1007/s10802-012-9659-y>
36. Cohen JR, So FK, Hankin BL, Young JF (2018) Translating cognitive vulnerability theory into improved adolescent depression screening: a receiver operating characteristic approach. *J Clin Child Adolesc Psychol*. <https://doi.org/10.1080/15374416.2018.15374416>
37. Straus SE, Glasziou P, Richardson WS, Haynes RB (2011) Evidence-based medicine: how to practice and teach EBM, 4th edn. Churchill Livingstone, New York
38. Sheldrick RC, Benneyan JC, Kiss IG, Briggs-Gowan MJ, Copeland W, Carter AS (2015) Thresholds and accuracy in screening tools for early detection of psychopathology. *J Child Psychol Psychiatry* 56:936–948. <https://doi.org/10.1111/jcpp.12442>
39. Hankin BL, Young JF, Abela JRZ, Smolen A, Jenness JL et al. (2015) Depression from childhood into late adolescence: influence of gender, development, genetic susceptibility, and peer stress. *J Abnorm Psychol* 124:803–816. <https://doi.org/10.1037/abn0000089>
40. Garber J (1984) The developmental progression of depression in female children. *New Dir Child Adolesc Dev* 1984:29–58
41. Joiner TE, Jr Catanzaro SJ, Laurent J (1996) Tripartite structure of positive and negative affect, depression, and anxiety in child and adolescent psychiatric inpatients. *J Abnorm Psychol* 105:401–409. <https://doi.org/10.1037/0021-843X.105.3.401>
42. Cohen JR, Spiro CN, Young JF, Gibb BE, Hankin BL, Abela JRZ (2015) Interpersonal risk profiles for youth depression: a person-centered, multi-wave, longitudinal Study. *J Abnorm Child Psychol* 43:1415–1426. <https://doi.org/10.1007/s10802-015-0023-x>
43. Muthén LK, Muthén BO (1998–2011) Mplus user's guide, 6th edn. Muthén & Muthén, Los Angeles
44. Edwards JR (2002) Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In: Drasgow F, Schmitt N (eds) *Measuring and analyzing behavior in organizations: advances in measurement and data analysis*. Jossey-Bass, San Francisco, pp 350–400
45. Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148:839–843. <https://doi.org/10.1148/radiology.148.3.6878708>
46. Rice ME, Harris GT (2005) Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law Hum Behav* 29:615–620. <https://doi.org/10.1007/s10979>
47. Swets J (1988) Measuring the accuracy of diagnostic systems. *Science*. 240: 1285–1293
48. Hastings ME, Krishnan S, Tangney JP, Stuewig J (2011) Predictive and incremental validity of the violence risk appraisal guide scores with male and female jail inmates. *Psychol Assess* 23:174–183. <https://doi.org/10.1037/a0021290>
49. Lavigne JV, Meyers KM, Feldman M (2016) Systematic review: classification accuracy of behavioral screening measures for use in integrated primary care settings. *J Pediatr Psychol* 41:1091–1109. <https://doi.org/10.1093/jpepsy/jsw049>
50. Tofighi D, Enders CK (2008) Identifying the correct number of classes in growth mixture models. In: Hancock GR, Samuelsen's KM (eds) *Latent variable mixture models*. Library of Congress, Washington, DC, pp 317–342
51. Kim Park II, Garber J, Ciesla JA, Ellis BJ (2008) Convergence among multiple methods of measuring positivity and negativity in the family environment: relation to depression in mothers and their children. *J Fam Psychol* 22:123–134. <https://doi.org/10.1037/0893-3200.22.1.123>
52. Pencina MJ, Steyerberg EW, D'Agostino RB (2011) Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 30:11–21. <https://doi.org/10.1002/sim.4085>
53. Rescorla LA, Ginzburg S, Achenbach TM, Ivanova MY, Almqvist F, Begovac I et al (2013) Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *J Clin Child Adolesc Psychol* 42:262–273. <https://doi.org/10.1080/15374416.2012.717870>
54. Dallaire DH, Pineda AQ, Cole DA, Ciesla JA, Jacquez F, Lagrange B et al (2006) Relation of positive and negative parenting to children's depressive symptoms. *J Clin Child Adolesc Psychol* 35:313–322. https://doi.org/10.1207/s15374424jccp3502_15

55. Wierzbicki M (1987) A parent form of the children's depression inventory: reliability and validity in nonclinical populations. *J Clin Psychol* 43: 390–397. [https://doi.org/10.1002/1097-4679\(198707\)43:4%3C390::AID-JCLP2270430409%3E3.0.CO;2-S](https://doi.org/10.1002/1097-4679(198707)43:4%3C390::AID-JCLP2270430409%3E3.0.CO;2-S)
56. Thapar AK, Hood K, Collishaw S, Hammerton G, Rice F (2016) Identifying key parent-reported symptoms for detecting depression in high risk adolescents. *Psychiatry Res* 242:210–217
57. Allgaier A-K, Fruhe B, Pietsch K, Saravo B, Bathmann M, Korne-Schulte G (2012) Is the children's depression inventory short version a valid screening tool in pediatric care? A comparison to its full-length version. *J Psychosom Res* 73:369–374. <https://doi.org/10.1016/j.jpsychores.2012.08.016>
58. Cohen JR, Thakur H, Burkhouse KL, Gibb BE (2018) A multimethod screening approach for pediatric depression onset: an incremental validity study. *J Con Clin Psychol*. <https://doi.org/10.1037/ccp0000364>
59. Kroenke K, Spitzer RL, Williams JB (2003) The patient health questionnaire-2: validity of a two-item depression screener. *Med Care* 41:1284–1292. <https://doi.org/10.1097/01.MLR.0000093487.78664.3C>
60. Trevethan R (2017) Sensitivity, specificity, and predictive values: foundations, pliabilitys, and pitfalls in research and practice. *Front Public Health*. <https://doi.org/10.3389/fpubh.2017.00307>
61. Stice E, Shaw H, Bohon C, Marti CN, Rohde P (2009) A meta-analytic review of depression programs for children and adolescents: factors that predict magnitude of intervention effects. *J Consult Clin Psychol* 77:486–503. <https://doi.org/10.1037/a0015168>
62. Youngstrom EA (2013) Future directions in psychological assessment: combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *J Clin Child Adolesc Psychol* 42:139–159. <https://doi.org/10.1080/15374416.2012.736358>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.