

Journal of Consulting and Clinical Psychology

A Multimethod Screening Approach for Pediatric Depression Onset: An Incremental Validity Study

Joseph R. Cohen, Hena Thakur, Katie L. Burkhouse, and Brandon E. Gibb

Online First Publication, December 20, 2018. <http://dx.doi.org/10.1037/ccp0000364>

CITATION

Cohen, J. R., Thakur, H., Burkhouse, K. L., & Gibb, B. E. (2018, December 20). A Multimethod Screening Approach for Pediatric Depression Onset: An Incremental Validity Study. *Journal of Consulting and Clinical Psychology*. Advance online publication.

<http://dx.doi.org/10.1037/ccp0000364>

A Multimethod Screening Approach for Pediatric Depression Onset: An Incremental Validity Study

Joseph R. Cohen and Hena Thakur
University of Illinois at Urbana—Champaign

Katie L. Burkhouse
University of Illinois at Chicago

Brandon E. Gibb
Binghamton University

Objective: Screening protocols that rely on a single informant are inadequate in predicting pediatric depression. Multi-informant and risk factor screening approaches are potentially more sensitive methods for identifying depression risk, but the incremental validity of these protocols has not been adequately tested. Using a translational analytic approach and multimethod, longitudinal study design, we simultaneously tested several multi-indicator approaches to depression screening to identify an optimal algorithm for predicting depression onset in youth. **Method:** Participants were 222 never-depressed children and adolescents ($M_{age} = 10.75$ years old, $SD_{age} = 1.85$; female = 50.45%; 82.88% White), who completed baseline questionnaires for depressive symptoms and cognitive vulnerabilities, in addition to a morphed face task to assess pupil dilation. Mothers, meanwhile, completed baseline questionnaires and a semistructured interview to assess maternal and pediatric depression. Follow-up depression diagnostic assessments with both the mother and youth occurred every 6 months for 2 years. Receiver operating characteristics and reclassification analyses were used to test our aims. **Results:** Overall, we found moderate support for a multi-informant approach, and convincing evidence that individual differences in pupil dilation uniquely predicted depression onset. Youth with subthreshold depressive symptoms and elevated pupil dilation were over twice as likely to develop a first lifetime episode of depression compared to one's risk rate based on sex and age. **Conclusions:** Our study provides one of the first screening batteries for detecting first lifetime episodes of depression in youth. The unique and incremental validity provided by pupil dilation suggests feasible biological indicators of depression risk can improve primary prevention efforts that target depression, such as universal pediatric depression screening.

What is the public health significance of this article?

A multimethod assessment approach, paired with a developmental psychopathology perspective, can strengthen pediatric depression screening initiatives. Compared to existing and recommended screening protocols, we found the use of a brief multi-informant diagnostic interview and pupil dilation assessment dramatically improved our ability to identify those youth at-risk for depression prior to an initial episode.

Keywords: pediatric depression, multimethod screening, developmental psychopathology, pupil dilation, incremental validity

Supplemental materials: <http://dx.doi.org/10.1037/ccp0000364.sup>

Depression remains a significant pediatric public health concern within the United States. Approximately 3.1 million adolescents (12.8%) experience a depressive episode annually with over 70%

of these episodes causing severe functional impairment (Center for Behavioral Health Statistics and Quality, 2016). In addition, pre-adolescent samples are experiencing depressive symptoms at an

Joseph R. Cohen and Hena Thakur, Department of Psychology, University of Illinois at Urbana—Champaign; Katie L. Burkhouse, Department of Psychiatry, University of Illinois at Chicago; Brandon E. Gibb, Department of Psychology, Binghamton University.

This project was supported by National Institute of Child Health and Human Development grant HD057066 and National Institute of Mental

Health grant MH098060 awarded to Brandon E. Gibb. We thank Ashley Johnson, Lindsey Stone, Andrea Hanley, Sydney Meadows, Michael Van Wie, Devra Alper, Cope Feurer, Erik Funk, and Effua Sosoo for their help in conducting assessments for this project.

Correspondence concerning this article should be addressed to Joseph R. Cohen, Department of Psychology, University of Illinois at Urbana—Champaign, 603 East Daniel Street, Champaign, IL 61820. E-mail: cohenj@illinois.edu

increasing rate (Luntamo, Sourander, Santalahti, Aromaa, & Heinenius, 2012). Given these alarming trends and the significant burden linked with depression (Avenevoli, Swendsen, He, Burstein, & Merikangas, 2015; Garber & Rao, 2014), it is important to evaluate protocols aimed at identifying children and adolescents at risk for the onset of depression.

Universal mental health screening protocols can play a role in reducing depression rates by identifying vulnerable youth in pediatric or school settings. Depression, along with substance use, is the only psychiatric condition that the U.S. Preventative Services Task Force (USPSTF) recommends routine screening for in youth. Yet, there is underwhelming evidence that these screening initiatives are effective (Wissow et al., 2013). To improve upon depression screening protocols, the USPSTF recently put forth two research agendas (Siu et al., 2016). First, the USPSTF found that, especially among children younger than 12, current tools have low and variable positive predictive values (i.e., the ability to identify positive cases correctly). Second, the USPSTF emphasized the importance of assessing risk factors to prevent the onset of depression in youth. By addressing these research limitations, the USPSTF suggests that universal screening initiatives for depression can reach their considerable public health potential.

Multi-informant protocols, in which parent–youth dyads complete pediatric depression screens, are the gold standard for depression assessments (Klein, Dougherty, & Olino, 2005) and represent an improvement over traditional single-informant approaches. Only recently have theoretical, methodological, and analytical advancements paved the way for integration of these multiple (and often discrepant) sources of data into clinical decision making processes (Martel, Markon, & Smith, 2017). Multi-informant screening is now used in applied settings and is more sensitively able to identify concurrent depression than single-informant approaches (e.g., Johnson, Hollis, Marlow, Simms, & Wolke, 2014). To date, however, there is little systematic research examining the incremental validity of a multi-informant approach among children and adolescents (Johnston & Murray, 2003), especially for prospective outcomes such as the onset of depression. As multi-informant methods largely focus on current depressive symptoms, they may be limited in predicting the emergence of symptoms.

Another possible method for improving current screening algorithms for depression onset is through the inclusion of known risk factors. The developmental psychopathology literature is rich with examples of vulnerabilities that can be targeted by screening protocols (Garber, Korelitz, & Samanez-Larkin, 2012; Rice & Rawal, 2011). Yet, it is unclear which risk factors should be prioritized in translational efforts. Importantly, investigators differentiate between vulnerabilities for first and recurrent depressive episodes (Monroe & Harkness, 2011), suggesting specific risk factors may be useful for detecting episode onset. Focusing on these risk factors may provide the best opportunity for targeting risk before a potentially chronic and severe depression course emerges. One of the more well-documented risk factors for depression onset is exposure to maternal depression. Across a variety of adolescent populations (e.g., Murray et al., 2011; Pearson et al., 2013), maternal depression uniquely forecasted depression onset, specifically early onset (i.e., prior to age 15; Hammen, Brennan, & Keenan-Miller, 2008). These findings have led some to call for a greater focus on maternal depression in screening initiatives across the developmental spectrum (Halligan, Murray, Martins, & Coo-

per, 2007). However, despite maternal depression's significant effect on early adolescent depression onset (Hammen & Brennan, 2003), most screening protocols using maternal depression inventories focus on young children (e.g., Earls, 2010). It is therefore unclear whether screening for maternal depression during late childhood and early adolescence is an incrementally valid approach compared to other strategies for operationalizing depression risk.

An alternative risk factor screening approach targets cognitive vulnerabilities for depression in youth. Cognitive vulnerability for depression is conceptualized as a linking mechanism between genetic risk and adolescent depression (Gibb, Beevers, & McGeary, 2013). Therefore, cognitive vulnerabilities may confer a stronger signal in screening protocols compared to maternal depression due to its developmental proximity to depression onset. To date, translational efforts have predominately used self-reported cognitive vulnerabilities. For instance, self-report forms of depressogenic inferential styles (e.g., Garber et al., 2012) and rumination (e.g., Young & Dietrich, 2015) have been included in depression screening protocols. However, self-report is just one unit of analysis that can be used to assess cognitive vulnerability (LeMoult, Yoon, & Joormann, 2016; Nejad, Fossati, & Lemogne, 2013). Overall, there is increasing awareness of the ability of biological methods to confer unique insight into prospective risk at the screening stage (Bylsma, Mauss, & Rottenberg, 2016), particularly efficient, cost-effective, psychophysiological methods (De Los Reyes & Aldao, 2015). Using psychophysiological indicators is consistent with recommendations to use multimethod mental health assessments (Hunsley & Mash, 2007), and the National Institute of Mental Health Research Domain Criteria's (RDoC's) emphasis on objective units of analysis in clinical protocols (Insel et al., 2010). Simultaneously using psychophysiological and self-reported measures can clarify whether using one, or both, methods of assessment provides the best approach for predicting depression onset. As youth may not become aware of certain depressive cognitive styles until after a depressive episode (Sheppard & Teasdale, 2004), objective assessments of cognitive vulnerability may be especially useful in identifying risk for depression onset.

A psychophysiological marker of cognitive vulnerability previously recommended for translational protocols is pupil dilation (Burkhouse, Siegle, Woody, Kudinova, & Gibb, 2015; Silk et al., 2009). Greater pupil dilation in response to emotionally salient stimuli is a marker of cognitive–affective processing (for reviews, see Bradley, Miccoli, Escrig, & Lang, 2008; Laeng, Sirois, & Gredebäck, 2012) and self-reported cognitive vulnerabilities (e.g., rumination; Siegle, Steinhauer, Carter, Ramel, & Thase, 2003). Recently, Burkhouse and colleagues (2015) found that elevated pupil dilation to sad (but not happy/angry) faces predicted depressive episodes in offspring of depressed mothers, demonstrating its external validity as a predictor of prospective depression status in at-risk youth. These findings are similar to other studies that demonstrate an association between depression symptoms/status and adolescent pupil dilation (e.g., Price et al., 2016; Silk et al., 2007). We sought to extend these collective findings by examining if peak pupil dilation to sad faces (a) predicted depression onset in an unselected youth sample and (b) provided incremental validity over existing (e.g., multi-informant and maternal depression) and more affordable (e.g., self-reported risk factors) methods for depression screening.

The Present Study

The present study examined the incremental validity of multi-informant (youth and parent report) and risk factor (maternal depression, self-reported cognitive vulnerability, pupil dilation) screening approaches for pediatric depression onset. For the present study, depression onset was defined as having a first lifetime major or minor (referred to as other specified depressive disorder, with the specifier depressive episode with insufficient symptoms in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders [DSM]*) depressive episode. Our decision to focus on both minor and major depression was informed by the literature. Past research shows that minor and major depressive episodes both lead to functional impairment in youth (Lewinsohn, Rohde, Seeley, Klein, & Gotlib, 2000) and including both forms of depression in our criterion reflects current dimensional models for depression status (Balázs et al., 2013). Focusing of both types of depressive episodes is also consistent with the purpose of screening, which is to identify current depressive impairment and calibrate future depression risk (Siu et al., 2016). Ultimately, while discriminating between mood diagnoses is critical for comprehensive treatment planning, it is a complex process best reserved for the assessment setting (De Los Reyes et al., 2015). Thus, using a dimensional depression criterion at the screening stage can cast a wider net for those at risk for depressive impairment.

To develop our screening algorithms, we used an empirically based medicine (EBM) approach (Youngstrom et al., 2017) via receiver operating characteristics (ROC). ROC approaches are ideally suited for translating basic research into screening protocols by generating empirically informed cutoffs that can facilitate clinical decision making (Youngstrom, 2014; Youngstrom et al., 2017). Pairing data-driven cutoffs with diagnostic likelihood ratios (DLRs; Straus, Richardson, Glasziou, & Haynes, 2011) allows protocols to estimate prospective risk and tailor their screening response based on the objectives and resources of the screening initiative. For the present study, we use a “stoplight model” (Youngstrom, Choukas-Bradley, Calhoun, & Jensen-Doss, 2015) to illustrate how to make feasible EBM-informed decisions across multiple, dimensional predictors from the clinical setting.

Although ROC approaches are “best practice” for validating pediatric mental health index tests (Youngstrom et al., 2017), biostatisticians have warned that ROC approaches may be overly conservative with regard to incremental validity (Wang et al., 2006). Thus, we used reclassification analyses as a complementary approach to assess incremental validity (Pencina, D’Agostino, & Steyerberg, 2011). Across medicine, reclassification analyses are used to quantify the costs and benefits of including new prognostic markers in screening protocols (Leening, Vedder, Witteman, Pencina, & Steyerberg, 2014); however, it has sparingly been used in psychosocial research. Pairing ROC with reclassification analyses provides a unique opportunity to optimize a screening protocol for depression onset, and introduces an analytic framework for evaluating multi-indicator screening protocols for other pediatric mental health outcomes.

Method

Participants

Data from 222 children–mother dyads drawn from a larger study of intergenerational transmission of depression (Burkhouse, Siegle, & Gibb, 2014) were used in the current study. An unselected child and adolescent sample was used as the major aim of the study was to inform universal depression screening protocols. Exclusion criteria for the larger study included maternal substance abuse within the last six months, history of bipolar disorder, or symptoms of schizophrenia. For the present study, youth were excluded if they had a current or past diagnosis of a major or minor depressive episode as we were focused on predictors for depression onset. Thirty-three youth reported at baseline either a current or past depressive episode and were excluded from analyses for the present study. The racial/ethnic composition of the pediatric sample was 82.88% White (not Hispanic), 10.36% Biracial, 3.6% Black (not Hispanic), and 1.35% Asian/Pacific Islander. The average age of the youth was 10.75 years old ($SD = 1.85$, range = 8–14) at baseline, and 50.45% were girls. The Appendix provides a narrative on other studies drawn from this sample using shared methods. Of note, a subsample of the present study previously examined the relation between elevated pupil dilation and prospective depressive episodes (approximately 16% of the participants in the current study were represented in this prior publication; Burkhouse et al., 2015).

Measures

Youth depression diagnostic status. The Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime Version (K-SADS; Kaufman et al., 1997) was used to assess current and lifetime history of depression at baseline, and the onset of depressive episodes every 6 months for 2 years. The K-SADS is a widely used diagnostic interview with well-established psychometric properties (Klein et al., 2005). Two trained interviewers administered the K-SADS separately to mother–child dyads. Consistent with research diagnostic criteria (Spitzer, Endicott, & Robins, 1978), as well as past research studies on pediatric depression screening (e.g., Cohen, So, Hankin, & Young, 2018), criteria for minor depression included the presence of a criterion A symptom plus at least one symptom from Criterion B, which lasted for at least 2 weeks and resulted in clinically significant impairment. In the current study, 21 youth reported a first lifetime episode over the course of the study (11 major and 10 minor episodes). Diagnostic status for youth was determined using “best estimate” procedures (Klein et al., 2005). During each follow-up assessment, interviewers asked about the onset of a depressive episode in the past 6-months. A subset of 20 K-SADS interviews from this project were coded by a second interviewer via video recording and kappa coefficients for depressive diagnoses were excellent ($\kappa = 1.00$).

Self-reported depressive symptoms. The Children’s Depression Inventory (CDI) was used to assess pediatric depressive symptoms in the current study. The CDI was chosen because it is the most commonly used measure of youth depression (Myers & Winters, 2002), a recommended measure for assessing depression in applied settings (Klein et al., 2005), and has comparable screen-

ing properties to other depression inventories (e.g., the Center for Epidemiologic Studies Depression Scale; Stockings et al., 2015). For the present study, the CDI ranged from 0 to 27 ($M = 5.31$; $SD = 5.03$) and demonstrated good reliability ($\alpha = .82$).

Self-reported cognitive vulnerability. Rumination was assessed via the Children's Response Style Scale (CRSS; Ziegert & Kistner, 2002). In the present study both the total rumination score and the brooding subscale were used to predict depression onset. The CRSS total score and Brooding subscale have exhibited good reliability and validity in youth (Muris, Fokke, & Kwik, 2009; Orue, Calvete, & Padilla, 2014). In the present study, total and brooding scores ranged between 3 and 90 ($M = 51.80$; $SD = 18.71$) and between 0 and 47 ($M = 28.86$; $SD = 18.71$), respectively. Both the overall scale and Brooding subscale were adequately reliable (Rumination: $\alpha = .79$; Brooding: $\alpha = .68$). Depressogenic inferential styles (DISs) were assessed with the Children's Cognitive Style Questionnaire (Abela, 2001). The questionnaire is comprised of hypothetical negative events to assess the tendency to (a) catastrophize the consequences of a negative event (DIS-consequences) and (b) make attributions toward oneself following a negative event (DIS-Self). DIS-consequences and DIS-Self, as opposed to DISs about causes, have been found to be reliable and predictive of prospective depression in both children and adolescents (Cohen, Young, & Abela, 2012). Scores on the DIS-Consequences and DIS-Self ranged between 0 and 30 ($M = 12.52$; $SD = 5.19$) and between 0 and 21 ($M = 9.06$; $SD = 4.21$) respectively, with good reliabilities on both scales in the present study (DIS-Consequences: $\alpha = .78$; DIS-Self: $\alpha = .75$).

Parent-reported youth depression symptoms. We used the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), specifically the Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, and Internalizing subscales, to query parent perspectives on youth depressive symptoms. These four subscales are valid indicators of depression diagnostic status in youth (Kaufman et al., 1997), and are better index tests than the *DSM*-oriented CBCL subscales (Ebetsutani et al., 2011). All four CBCL subscales demonstrated adequate reliability in past research (see Achenbach & Rescorla, 2001). The range (CBCL-Anxious/Depressed: 0–16; CBCL-Withdrawn: 0–14; CBCL-Somatic: 0–12; CBCL-Internalizing: 0–29) and average raw scores (CBCL-Anxious/Depressed: $M = 3.39$; $SD = 3.35$; CBCL-Withdrawn/Depressed: $M = 1.80$; $SD = 2.34$; CBCL-Somatic: $M = 1.93$; $SD = 2.31$; CBCL-Internalizing: $M = 6.81$; $SD = 6.16$) for the subscales in the current study were comparable to those in past research (Achenbach & Rescorla, 2001). In the present study, the Anxious/Depressed ($\alpha = .79$), Withdrawn/Depressed ($\alpha = .79$), Somatic ($\alpha = .71$), and Internalizing ($\alpha = .86$) subscales demonstrated good internal reliability.

Maternal depression. The Structured Clinical Interview for *DSM-IV* Axis I Disorders (SCID-I; First, Spitzer, Gibbon, & Williams, 1995) was used to assess maternal depression. The SCID-I is a widely used diagnostic interview with well-established psychometric properties (Lobbestael, Leurgans, & Arntz, 2011; First et al., 1995). Trained interviewers administered the SCID-I to all participants. In the present study, 45.9% ($n = 102$) of mothers reported either a current or past depressive episode. In addition, the Beck Depressive Inventory-II (BDI-II; Beck, Steer, & Brown, 1996) was used to assess mothers' current symptoms of depression.

The BDI-II has exhibited excellent reliability and validity in previous research (Storch, Roberti, & Roth, 2004). In the current study, the BDI-II ranged from 0 to 49 ($M = 8.07$; $SD = 9.29$) and exhibited good internal consistency ($\alpha = .94$).

Multi-informant depression inventory. The present study also used the Children's Depression Rating Scale-Revised (CDRS-R; Poznanski & Mokros, 1996). The CDRS-R is an interviewer administered measure that separately queries the parent and child and then forms a combined depression score. The CDRS-R has demonstrated good reliability and validity in both child and adolescent samples (e.g., Canals, Marti-Henneberg, Fernandez-Ballart, & Domenech, 1995; Poznanski & Mokros, 1996). Recently, an "unfiltered" version of the CDRS-R was shown to adequately distinguish between depression diagnoses in a large at-risk and unselected pediatric sample (Yee et al., 2015). The "unfiltered" version of the CDRS-R differs from standard diagnostic interviews because it does not contextualize the symptoms across other disorders or within a patient's individual history (e.g., concentration symptoms are coded as a symptom of depression regardless of ADHD diagnosis). The use of "unfiltered" administrations significantly reduces the amount of time and training burden for diagnostic interviews (Fristad et al., 2016; Yee et al., 2015), making it more amenable to screening protocols outside the context of outpatient mental health services. Scores on the CDRS-R in the present study ranged between 17 and 42 ($M = 20.40$; $SD = 4.00$). In the present study, the CDRS-R had good reliability ($\alpha = .78$).

Pupil dilation. Pupil dilation in response to sad faces was assessed in a moderately lit room using Tobii T60 and T60XL eye-trackers within the context of a morphed faces computer task (see Burkhouse et al., 2014). The stimulus set consists of full-color pictures of actors taken from a standardized stimulus set (Matsumoto & Ekman, 1988). Sad and neutral photographs from each actor were morphed to form a continuum of 10% increments between the two photographs. During this task, pupil size was recorded using the eye trackers at 60 Hz (every 16.7 ms) for 3 s following the onset of each facial stimulus. Total time for the task was approximately 5 min. Data were cleaned using Siegle, Ichikawa, and Steinhauer's (2008) standard procedures. Linear interpolations replaced blinks throughout the data set and data were smoothed using a 10-point weighted average filter. Data were resampled to 3 Hz (one sample every 333 ms). The average pupil diameter over the 333 ms preceding the onset of the stimulus was subtracted from pupil diameter after stimulus onset to produce stimulus-related pupil dilation waveforms. Consistent with prior studies highlighting the predictive validity of pupil reactivity to emotional faces at the highest, but not medium or low, levels of morphed intensity (Burkhouse et al., 2014, 2015), we focused our analyses on peak pupil dilation to faces during the highest level of sadness intensity (70–90% morph). Peak stimulus-related pupil dilation (i.e., SadPeak) was calculated by taking the maximum pupil response on average across all trials for sad faces. Levels of pupil dilation ranged between 0 and 0.17 ($M = .06$; $SD = .04$). Of note, similar pieces of apparatus to assess pupil dilation are currently being used in applied pediatric settings (Boev et al., 2005; Connelly et al., 2014), albeit without the use of specific stimuli (e.g., sad faces).

Procedure

Participants were recruited from the community through a variety of means (e.g., bus ads, flyers). Mothers responding to these ads were first screened over the phone to determine eligibility. Upon arrival at the laboratory, mothers provided informed consent and children provided assent to be in the study. Next, the SCID-I, CBCL, and CDRS-R were administered to the mother by a research assistant. During this time, the child completed the emotional faces paradigm to assess pupil dilation. An interviewer then administered the K-SADS-PL, self-report measures of depressive symptoms (CDI, CDRS-R), and cognitive vulnerabilities (CRSS, Children's Cognitive Style Questionnaire). Follow-up depression diagnostic assessments with both the mother and youth occurred 6, 12, 18, and 24 months after the first assessment to identify if the youth experienced depression onset. All procedures for this study were approved by the last author's Institutional Review Board.

Data Analytic Strategy

First, logistic regression analyses tested if the relation between predictors and depression onset varied as a function of age or sex. If positive, the area under the curve (AUC) was computed separately for girls and boys and/or children (ages 8–11) and early adolescents (ages 12–14). These AUCs were then compared using the Delong test for paired ROC curves (DeLong, DeLong, & Clarke-Pearson, 1988). If significantly different, incremental validity was examined separately across the sociodemographic variable.

Following recommendations for our sample size (Obuchowski & Lieber, 1998), a bias-corrected accelerated bootstrap method was used to generate the confidence interval for our AUCs. In the present study, an AUC greater than 0.64 conferred a medium, significant effect (Rice & Harris, 2005) while an AUC of 0.70 was prioritized as a *clinically* significant effect (see Swets, 1988). We next examined if index tests achieving an AUC of at least 0.64 differed from baseline CDI scores by using the saved residuals from a linear regression model. The AUC of these residual scores represent the unique variance of the predictor (Edens, Marcus, Lilienfeld, & Poythress, 2006; Hastings, Krishnan, Tangney, & Stuewig, 2011). When testing the residuals, an index test was considered unique if the confidence interval of the AUC did not include 0.50.

To examine the incremental validity of adding index tests to baseline symptoms, a ROC approach was first applied in which predictive probabilities from logistic regression models where predictors were simultaneously entered were plotted within an AUC curve. This multivariate AUC statistic is commonly referred to as the *C* statistic. After identifying the highest *C* statistic between two predictors, we entered a third predictor. This stepwise approach continued until no significant predictors were left or the *C* statistic remained unchanged.

Next, we used reclassification analyses to evaluate the incremental validity of our predictors. Reclassification tables are derived by examining how many additional cases are accurately identified by positive scores on an additional risk indicator (Pencina et al., 2011). The net reclassification improvement (NRI) index, the most common statistic derived from these analyses, is the sum of the false negatives misclassified by the baseline algorithm subtracted by the number of false positives included in the

new algorithm (see Pencina, D'Agostino Sr, D'Agostino Jr, & Vasan, 2008 for further details). We used the continuous model as described by Pencina and colleagues (2011) as it provides less biased results compared to using categorical approaches. At each stage, novel index tests were simultaneously added to a model. If multiple predictors were significant, the indicator with the highest NRI was retained and reentered into a new baseline model and reclassification analyses were conducted with subsequent predictors. This process was repeated until the NRI of any of the predictors was no longer significant (see Cohen, Shorey, Menon, & Temple, 2018 for a demonstration). The NRI, a bootstrapped *C* statistic, the Hosmer–Lemeshow goodness-of-fit statistic, and the integrated discrimination index were calculated to provide complementary perspectives on our final models (Pencina, D'Agostino, Pencina, Janssens, & Greenland, 2012).

Finally, DLRs were calculated to facilitate translation of our findings into applied settings (Straus et al., 2011). DLRs are the proportion of cases with a diagnosis within a certain scoring range on the index test divided by the proportion of cases without the diagnosis in that same range of scores. Ideally, DLRs contain at least three cutoffs so that risk can be operationalized dimensionally across "low," "intermediate," and "high" risk categories. Selection for cutoffs can either be informative (based on established cutoffs) or equal (cutoffs that form three equivalent groups with regard to sample size).

We utilized an informative tertile approach for our baseline predictor, the CDI. Original cutoffs for the CDI were proposed to range between raw scores of 12 and 19 (Kovacs, 1992); however, these cutoffs may not be sensitive within an unselected population (Matthey & Petrovski, 2002). Recently, Cohen, So, et al. (2018) used CDI cutoffs of 7 (subthreshold) and 15 (threshold) based on levels of sensitivity and specificity congruent with current pediatric mental health screening initiatives (70% sensitivity; 90% specificity; Lavigne, Feldman, & Meyers, 2016). For the current study, we formed three groups (i.e., tertiles) based on these cutoffs, but also tested cutoffs that corresponded to 70% sensitivity (subthreshold) and 90% specificity (threshold) for predicting onset in our sample. For all other index tests, equal tertiles were formed. After DLRs for each index test were computed, posterior probabilities were calculated based on the CDI, and then we examined the incremental impact on probability once adding additional risk factors (see Youngstrom, Halverson, Youngstrom, Lindhjem, & Findling, 2018, for an example). DLR scores of 1.0 suggest that one's odds of developing the disorder remain relatively unchanged (i.e., the assessment result was neutral). Higher DLRs suggest that the odds of developing the target diagnosis increase. All analyses were conducted in R (Version 3.4.4), with the exception of DLRs which were calculated in SPSS (Version 24.0). A table in the online supplemental materials available online summarizes each step of our multianalytic plan. Interested readers are encouraged to consult this table to more clearly understand the function of each step.

Results

The number of youth with depression onset exceeded the minimum number of 20 cases needed for ROC analyses ($n = 21$; Kraemer, 1992). Correlations for our baseline predictors are presented in Table 1. Relations between our index tests ranged from

Table 1
Correlations Between Baseline Predictors of Youth Depression Status

Baseline Predictors	Youth-report symptoms and risk (cognitive vulnerability)					Parent-report symptoms and risk (maternal depression)					Multi-informant symptoms	
	CDI	CRRS-T	CRRS-B	CCSQ-C	CCSQ-S	CBCL-A/D	CBCL-W/D	CBCL-SOM	CBCL-INT	SCID	BDI-II	CDRS-R
Youth-report symptoms and risk												
CDI												
CRRS-T	.15*											
CRRS-B	.19**	.88**										
CCSQ-C	.22**	.15*	.21**									
CCSQ-S	.24**	.10	.09		.42**							
Parent-report symptoms and risk (maternal depress)												
CBCL-A/D	.21**	.06	.09	.00	.11							
CBCL-W/D	.28**	.10	.07	-.01	-.02	.55**						
CBCL-SOM	.16*	.08	.07	-.03	.03	.42**	.54**					
CBCL-INT	.24**	.03	.05	-.02	.08	.80**	.80**	.77**				
SCID	.16*	.05	.06	.10	-.03	.24**	.28**	.25**	.30**			
BDI-II	.16*	.00	.05	.03	-.09	.22**	.39**	.31**	.26**	.54**		
Multi-informant symptoms												
CDRS-R	.29**	.15*	.13*	.04	.01	.30**	.36**	.29**	.33**	.38**	.32**	
Pupil dilation												
SadPeak	.08	-.02	.04	.03	-.08	-.08	.13	.04	.02	.22*	.06	.23*

Note. Pearson correlation coefficients for baseline predictors of youth depression status. CDI = Children's Depression Inventory (Kovacs, 1992); CRRS-T = Children's Response Style Scale-total; CRRS-B = CRRS-Brooding subscale (Ziegert & Kistner, 2002); CCSQ-C = Children's Cognitive Style Questionnaire-tendency to catastrophize consequences of a negative event; CCSQ-S = CCSQ-self-attributions following a negative event (Abela, 2001); CBCL-A/D = Child Behavior Checklist-Anxious/Depressed subscale; CBCL-W/D = CBCL-Withdrawn/Depressed subscale; CBCL-SOM = CBCL-Somatic subscale; CBCL-INT = CBCL-Internalizing subscale (Achenbach & Rescorla, 2001); SCID = Structured Clinical Interview for *DSM-IV* Axis I Disorders (no lifetime maternal depression diagnosis = 0; lifetime maternal depression diagnosis = 1; First, Spitzer, Gibbon, & Williams, 1995); BDI-II = Beck's Depressive Inventory-II (Beck, Steer, & Brown, 1996); CDRS-R = Children's Depression Rating Scale-Revised (Poznanski & Mokros, 1996); SadPeak = child's peak pupil dilation to faces during the highest level of sadness intensity.

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

nonsignificant to medium effect sizes. We then examined whether any of our indicators varied as a function of sex or age. Logistic regression analyses were nonsignificant, leading us to calculate AUCs across the whole sample. AUCs for all index tests are displayed in Table 2. Overall, CDI scores, CBCL withdrawn-depressed symptoms, CBCL somatic complaints, CBCL internalizing symptoms, and maternal depression diagnostic status, exhibited a significant, medium effect ($AUC \geq .64$). Meanwhile, only CDRS-R scores and peak pupil dilation while viewing sad stimuli (SadPeak) were clinically significant ($AUC \geq .70$). Of note, our self-reported cognitive vulnerabilities did not predict onset ($p > .05$). None of the predictors with AUCs between 0.64 and 0.70 remained significant once covarying out CDI scores ($p > .05$). Alternatively, SadPeak ($AUC = .75, p < .01$) and CDRS-R scores ($AUC = .63, p = .05$) remained significant. The DeLong test suggested that the CDI, CDRS-R, and SadPeak equivalently forecasted depression onset ($p > .10$).

Next, we examined the C statistics for CDRS-R and SadPeak added to CDI scores (Table 3). Overall, the C statistic for CDRS-R together with CDI (0.75) represented a 17% increase from the CDI's AUC, while SadPeak and CDI's C (0.78) demonstrated a 22% increase above the CDI alone. Meanwhile, the C statistic for all three index tests was 0.85, a 33% increase above our baseline CDI score. Although the DeLong test for paired ROC suggested using two index tests did not incrementally predict onset better

than CDI alone ($p > .05$), using three index tests together (CDI, CDRS-R, and SadPeak) better forecasted onset compared to only using CDI scores (DeLong test = $-3.17, p < .01$). As for reclassification analyses, the NRI for CDRS-R scores and SadPeak were both significant compared to our baseline CDI model ($p < .05$). Because SadPeak had the higher NRI it was entered into a new baseline model that included CDI scores. For this model, CDRS-R scores were nonsignificant ($p > .05$). Summary statistics for the ROC algorithm (CDI, CDRS-R, SadPeak) and reclassification algorithm (CDI, SadPeak) are presented in Table 3.

Table 4 displays the DLRs for our algorithms. Within our sample, a score of 4 on the CDI corresponded to 70% sensitivity, and a score of 10 conferred 90% specificity. However, DLRs between these cutoffs, and our a priori cutoffs (7, 15) were similar (e.g., DLRs differed by less than 10% for high scores), prompting us to use the a priori cutoffs to facilitate comparisons across the literature. Subthreshold (i.e., the medium tertile) and threshold (i.e., high tertile) scores were 18 and 20 for CDRS-R, and 0.45 and 0.72 for SadPeak. When examining the CDI alone, individuals with scores above threshold (i.e., ≥ 15) were approximately 2.5 times as likely to experience a first lifetime episode of depression compared to the average prevalence rate for depression onset within the sample (see the first posttest probability in Table 5). Table 5 demonstrates how including SadPeak, our most robust novel indicator, into the decision algorithm influences probabili-

Table 2
Summary of Individual Predictors and Multi-Indicator Algorithms: Area Under the Curve and Effect Sizes for Individual Predictors

Predictor	AUC	BCa CI	Cohen's <i>d</i> effect size
Youth-report symptoms and risk			
CDI	.64*	[.51,.76]	.51 (M)
CRSS-T	.52	[.41,.63]	.07
CRSS-B	.58	[.48,.68]	.29 (S)
CCSQ-C	.59	[.49,.69]	.32 (S)
CCSQ-S	.56	[.45,.67]	.21 (S)
Parent-report symptoms and risk (maternal depress)			
CBCL-A/D	.59	[.44,.74]	.32 (S)
CBCL-W/D	.67*	[.57,.77]	.62 (M)
CBCL-SOM	.68*	[.55,.80]	.66 (M)
CBCL-INT	.65*	[.52,.78]	.55 (M)
SCID-Diagnosis	.64*	[.54,.75]	.51 (M)
BDI-II	.62	[.48,.76]	.43 (S)
Multi-informant symptoms			
CDRS-R	.72**	[.61,.83]	.82 (L)
SadPeak	.76**	[.55,.94]	1.00 (L)

Note. AUC = area under the curve; BCa CI = bias-corrected accelerated confidence intervals; CDI = Children's Depression Inventory (Kovacs, 1992); CRSS-T = Children's Response Style Scale-total; CRSS-B = CRSS-Brooding subscale (Ziegert & Kistner, 2002); CCSQ-C = Children's Cognitive Style Questionnaire-tendency to catastrophize consequences of a negative event; CCSQ-S = CCSQ-self-attributions following a negative event (Abela, 2001); CBCL-A/D = Child Behavior Checklist-Anxious/Depressed subscale; CBCL-W/D = CBCL-Withdrawn/Depressed subscale; CBCL-SOM = CBCL-Somatic subscale; CBCL-INT = CBCL-Internalizing subscale (Achenbach & Rescorla, 2001); SCID = Structured Clinical Interview for *DSM-IV* Axis I Disorders (no lifetime maternal depression diagnosis = 0; lifetime maternal depression diagnosis = 1; First, Spitzer, Gibbon, & Williams, 1995); BDI-II = Beck's Depressive Inventory-II (Beck, Steer, & Brown, 1996); CDRS-R = Children's Depression Rating Scale-Revised (Poznanski & Mokros, 1996); SadPeak = child's peak pupil dilation to faces during the highest level of sadness intensity; S = small effect (*d* > .20); M = medium effect (*d* > .50); L = large effect (*d* > .80). Boldface indicates AUCs $\geq .64$.
* *p* < .05. ** *p* < .01.

ties. These estimates were drawn using a naïve approach (see Youngstrom et al., 2018), in which the posterior probability, as estimated by the CDI, becomes the new pretest probability. The

Table 4
Diagnostic Likelihood Ratios (DLRs) for Significant Predictors

Model	DLR		
	Low	Medium	High
CDI	.77	1.40	2.51
SadPeak	.27	.59	2.39
CDRS-R	.19	.73	1.70

Note. CDI = Children's Depression Inventory (child self-report; Kovacs, 1992); SadPeak = child's peak pupil dilation to faces during the highest level of sadness intensity; CDRS-R = Children's Depression Rating Scale-Revised (combined child self-report and parent-reported scores; Poznanski & Mokros, 1996). Low DLR: CDI < 7, SadPeak < .45, CDRS-R < 18; medium DLR: CDI = 7–14, SadPeak = .45–.71, CDRS-R = 18–19; high DLR: CDI \geq 15, SadPeak \geq .72, CDRS-R \geq 20.

incremental impact of SadPeak is then demonstrated by multiplying the new pretest odds by the DLR derived from SadPeak scores to calculate the new posterior probability. Similar estimates can be generated using the DLRs for CDRS-R both with and without SadPeak. Of note for our reclassification algorithm (CDI, SadPeak), scores above threshold significantly increase risk (see Example Cases 2 and 3), while scores below threshold on the SadPeak attenuate risk (see Example Cases 1 and 4). These findings suggest that our multi-indicator approach can be used to both "rule in" and "rule out" screening cases that are at or near cutoff scores on the CDI.

Discussion

It is now well-established that multiple indicators are needed for an adequate mental health assessment (Hunsley & Mash, 2007). Yet, due to a paucity of incremental validity studies (Johnston & Murray, 2003), it is unclear which indicators should be targeted. Overall, we found moderate support for the incremental validity of multi-informant approaches, and convincing evidence that a specific psychophysiological index, pupil dilation, improves our ability to predict depression onset in youth even when compared to other informant reports or well-documented risk factors. The translational, clinical significance of these findings is discussed below.

As efficiency is prioritized within a screening setting, there is a significant need to demonstrate the incremental validity of a po-

Table 3
Summary of Individual Predictors and Multi-Indicator Algorithms: Discrimination and Calibration Statistics for All Multi-Indicator Algorithms

Baseline model	NRI	Lower bound	Upper bound	IDI	Lower bound	Upper bound	<i>C</i>	Lower bound	Upper bound	H-L χ^2	<i>p</i>
CDI											
CDRS-R	.76*	.09	1.43	.08	−.01	.17	.75	.65	.84	12.69	.12
SadPeak	.96***	.38	1.53	.21**	.03	.38	.77	.60	.94	8.85	.36
CDI and sad peak											
CDRS-R	.63	−.04	1.31	.05	−.06	.16	.84	.71	.96	7.15	.52

Note. NRI = net reclassification improvement index (Pencina, D'Agostino, Pencina, Janssens, & Greenland, 2012); IDI = integrated discrimination index (Pencina et al., 2012); *C* = multivariate area under the curve; H-L χ^2 = Hosmer-Lemeshow test for whether expected rates of the outcome are significantly different than the observed rates; CDI = Children's Depression Inventory (child self-report; Kovacs, 1992); CDRS-R = Children's Depression Rating Scale-Revised (combined child self-report and parent-reported scores; Poznanski & Mokros, 1996); SadPeak = child's peak pupil dilation to faces during the highest level of sadness intensity. Nonsignificant *p* suggests the model is well calibrated.

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

Table 5
Stoplight model: Examples and Interpretations of Screening Cases

Case	Pretest probability	Scoring profile	Posttest probability	Zone	Interpretation
Female, age 8	10.78%	CDI: 15	CDI alone: 23.27%	Yellow	Despite an above threshold CDI score, a low score for pupil dilation leads to a slightly lower likelihood of experiencing depression onset than someone in the general population. As the CDI score is elevated, continued monitoring is warranted but others with above threshold CDI scores and pupil dilation should be prioritized for services.
Male, age 8	6.25%	SadPeak: .38 CDI: 14	CDI; SadPeak: 7.46% CDI alone: 8.53%	Red	Just relying on CDI scores, one would potentially be missing out on someone who is nearly 3 times more likely to experience depression onset. This child should be immediately referred for a mental health assessment, and any preventative services should be made readily available.
Female, age 14	15.38%	SadPeak: .79 CDI: 14	CDI; SadPeak: 18.03% CDI alone: 20.28%	Red	The subthreshold CDI score confers that this individual is slightly more at risk to develop depression. The threshold pupil dilation score together with the subthreshold CDI score, however, suggest that this person is nearly 3 times as likely to experience depression onset.
Male, age 14	11.77%	SadPeak: .79 CDI: 14	CDI; SadPeak: 37.83% CDI alone: 15.74%	Green	Despite presenting with a CDI score near the threshold, this individual is actually at little risk for experiencing a depressive episode over the next 2 years. Specifically, low scores on pupil dilation together with subthreshold scores on the CDI suggest that this individual is approximately 300% less likely to experience a first lifetime episode of depression compared to his pretest probability.
		SadPeak: .38	CDI; SadPeak: 4.80%		

Note. CDI = Children's Depression Inventory (child self-report; Kovacs, 1992); SadPeak = child's peak pupil dilation to faces during the highest level of sadness intensity; Pretest probability = percentage chance of each depression outcome based on sex and age; Posttest probability = $[\text{prevalence}/(1 - \text{prevalence}) \times \text{DLR}]/([\text{prevalence}/(1 - \text{prevalence})] + 1)$ (Straus, Richardson, Glasziou, & Haynes, 2011). Zones: green = standard screening approach, yellow = increased monitoring; red = implement brief intervention (or prioritize for mental health services in settings now equipped to implement mental health services).

tential index test. Although a multi-informant approach is commonly used for discriminating between concurrent diagnostic statuses within the assessment context (De Los Reyes et al., 2015), research is mixed about whether it may improve forecasts for prospective functioning (Johnson et al., 2014). Overall, our findings suggest that a clinician-administered, multi-informant depression screen (CDRS-R) added incremental validity, while a multi-informant questionnaire for internalizing symptoms (CBCL) did not. Past research suggests that parental perspectives may be especially helpful in identifying behavioral manifestations of psychological distress in youth (De Los Reyes et al., 2015). As the CDRS-R represents a more focused assessment of depressive symptoms and specifically incorporates behavioral observations into its score (Poznanski & Mokros, 1996), it is well-positioned to identify early behavioral manifestations of depression. As depression related behavioral symptoms uniquely predict onset in adolescence (Ong, Bergeman, Bisconti, & Wallace, 2006), the inclusion of multi-informant behavioral screening initiatives may be especially important for detecting first lifetime episodes of depression in youth.

While multi-informant depression assessments are a reasonable approach to improving depression screening initiatives, assessing pupil dilation contributed to optimal models for predicting depression onset. Our study extended past research (e.g., Burkhouse et al., 2015) by (a) showing its association with first lifetime episodes within an unselected sample and (b) its incremental validity com-

pared to other index tests. This finding is also consistent with recent work suggesting that psychophysiological and neurobiological tools can improve prediction models for psychiatric illnesses (De Los Reyes & Aldao, 2015; Gabrieli, Ghosh, & Whitfield-Gabrieli, 2015). It is noteworthy that neither self-reported cognitive vulnerabilities nor indicators of maternal depression were predictive of our outcome. One reason for this may be is that self-reports of cognitive vulnerabilities, and to a lesser extent subjective reports of maternal depression, shared a significant amount of variance with our baseline CDI model (see Table 1). However, within the univariate models, pupil dilation still achieved a higher AUC than these other risk factors, suggesting shared method variance alone did not explain why pupil dilation's incremental validity was unique in this study.

Pupil dilation may have emerged as a superior candidate for multi-indicator screening protocols due to its objective nature. With regard to rumination and depressogenic inferential styles, past research concerning their ability to forecast prospective episodes of depression compared to other indicators is mixed (Cohen, So, et al., 2018; Seeley, Stice, & Rohde, 2009). One possibility is that self-reported cognitive vulnerabilities are more influential in predicting recurrent depressive episodes, as opposed to first lifetime episodes, due to increased metacognitive awareness (Teasdale et al., 2002). Specifically, individuals who have previously experienced a depressive episode may be more likely to be aware of

and report on negative cognitive styles (Sheppard & Teasdale, 2004). In response, autonomic indices of cognitive vulnerability, such as pupil dilation, may be needed for primary depression screening initiatives (Rawal, Collishaw, Thapar, & Rice, 2013).

As for maternal depression, we found a small-medium effect for predicting depression onset in our univariate models. This finding is consistent with past research on community adolescent samples (e.g., Hammen & Brennan, 2003; Hammen et al., 2008) and supports connect adolescent depression screening initiatives to primary care sites that routinely screen for maternal depression (Garber et al., 2009; Halligan et al., 2007). Yet, when determining a youth's need for mental health services, more proximal risk factors should be prioritized. For instance, youth within different age cohorts who are exposed to maternal depression early in life demonstrate dysregulation across a wide array of systems, including the autonomic nervous system (Propper & Holochwost, 2013). Thus, indicators of the pathways stemming from maternal depression to adolescent depression onset (e.g., pupil dilation) may provide better insight into depression risk.

The identification of pupil dilation as a biomarker for depression onset is consistent with RDoC's objective to identify markers of dysfunctional neural circuitry that have translational promise (Insel et al., 2010). More direct assessments of neural activity (e.g., via imaging), may not be as easily integrated into screening protocols from a resource perspective. Therefore, cheaper and briefer psychophysiological assessments linked to neural activity can be a more feasible approach for achieving this goal (De Los Reyes & Aldao, 2015). In the past, pupillary responses to emotionally salient images have been linked with increased dorsolateral prefrontal function (Siegle, Steinhauer, Friedman, Thompson, & Thase, 2011), which is associated with deficits in executive control and emotion regulation in depressed patients (Koenigs & Grafman, 2009). Thus, pupil dilation may not only represent an objective indicator of depression onset, but also a translational window into the neural circuitry underlying depressogenic risk.

A significant challenge to the translation of novel index tests, biological and otherwise, into routine clinical practice is the absence of a standard analytic approach to assess incremental validity. Compared to reliability and other forms of validity (e.g., internal, external), incremental validity has garnered less attention from methodologists (Hunsley & Meyer, 2003), and as a result is rarely adequately tested (Cohen, So, et al., 2018; Garb, 2003). With the introduction of RDoC and the proliferation of developmental psychopathology research (Franklin, Jamieson, Glenn, & Nock, 2015), there is a need to test which risk factors should be prioritized for translational research. By using a ROC analytic plan (Youngstrom et al., 2017), together with reclassification analyses (Pencina et al., 2011), we were able to provide a multimethod perspective on incremental validity. Adopting a similar analytic approach for other potential index tests for pediatric depression, as well as for other pediatric mental health conditions, can help bring the promise of RDoC and developmental psychopathology to the clinical setting.

The present study had several strengths (e.g., multimethod risk assessment, diagnostic depression interview, prospective design); however, there are also some notable limitations. First, we only assessed depression as an outcome. While the USPTF calls for preventative screening research specifically for depression (Siu et al., 2016), others have suggested screening protocols need to capture a broader range of pediatric mental health concerns (Lavi-

gne et al., 2016). Second, we were only able to operationalize pupillary reactivity within the first three seconds of the stimulus presentation. While this brief assessment is ideal for the screening setting, pupillary differences well after the stimulus presentation are also indicative of cognitive vulnerability for emotionally salient information (Siegle et al., 2003; Silk et al., 2009). Third, as is common with multiwave, longitudinal designs, the data in the current study was censored. Specifically, youth who already reported depression were excluded from the study (left-censored) and others will certainly develop a depressive episode following the 2-year follow-up period (right-censored). Thus, our findings are relatively conservative in nature. Fourth, while we did not find any significant demographic differences for age and/or sex, our sample size may have been too limited to test these exploratory hypotheses. Given well-documented sex and age differences in the risk profile for depression (Avenevoli et al., 2015), future studies will want to investigate whether algorithms vary as a function of demographics within pediatric subsamples.

Finally, despite the statistical prowess concerning pupil dilation, and to a lesser extent multi-informant approaches, challenges to integrating these methods into clinical practice remain. Most notably, the indicators that demonstrated incremental validity were also the most expensive methods. Therefore, initial costs of set-up with purchasing apparatus, and the cost of training and employing people to administer a diagnostic interview, may make it challenging to integrate multimethod screening protocols into low-resourced settings. Additionally, the use of these screening measures will likely extend appointments (approximately 5 min for pupil dilation; 10–15 min for the CDRS-R) during an era when pediatricians are increasingly burdened with screening directives and shrinking appointment times (Lavigne et al., 2016). Thus, while we achieved our first aim of introducing a novel indicator into a screening/assessment protocol by demonstrating its incremental validity (Johnston & Murray, 2003), future studies must now replicate these findings in an applied setting (Youngstrom et al., 2017), ideally using publicly available depression measures (e.g., the Patient Health Questionnaire [PHQ]; Richardson et al., 2010) in an effort to help reduce the cost of a multi-indicator approach. These studies are necessary to conduct prior to the implementation of our findings into screening protocols, to ensure that we can replicate our findings, and cross-validate our proposed algorithm.

Clinical Implications

Despite some translational barriers, there are trends in the literature that make the integration of pupillometry and diagnostic interviews into universal mental health screening promising. Similar to other psychophysiological assessments (see De Los Reyes & Aldao, 2015), innovations have lowered the financial and temporal burden associated with pupillometry (Nowak, Źarowska, Szul-Pietrzak, & Misiuk-Hojlo, 2014). Across pediatric contexts (e.g., pain clinics) pupil dilation is now regularly measured as an indicator of autonomic nervous system functioning (Boev et al., 2005; Connelly et al., 2014). Advantages of using pupil dilation compared to other biological measures is that it is noninvasive, can be collected across settings via mobile technology, and is easily administered and interpreted by practitioners without significant expertise. The only meaningful difference in how pupil dilation is measured in clinical versus research settings is the use of stimuli.

As the most robust findings for pupil dilation is within the context of viewing depressogenic stimuli (e.g., Silk et al., 2007, 2009), it is important for future studies to test whether integrating this component into clinical settings is problematic in any way.

Similar to pupillometry, advancements in the assessment literature make the use of diagnostic interviews more feasible within a screening setting. The current study used an “unfiltered” version of the CDRS-R as our multi-informant indicator. The focused nature of these interviews, in which comorbid diagnoses and chronicity of symptoms are not assessed, reduces the administration time and the training burden (Fristad et al., 2016; Yee et al., 2015). Thus, while future studies should replicate these findings in applied, clinical settings (Youngstrom et al., 2017), there is reason to believe that pupillometry and diagnostic interviews can be used in cost-effective manners.

To illustrate how our study’s translational approach may be useful with regard to clinical decision making, Table 5 includes four example screening profiles from our study. Below, we briefly guide the reader on the stepped approach to creating and interpreting Table 5 (see Cohen, So, et al., 2018; Youngstrom et al., 2015 for further guidance). The first step toward making evidence-based referrals is calculating the probability of developing a first lifetime episode of depression in the overall sample. Pretest probabilities (Table 5, column 2) in the present study were calculated separately based on age and sex. Example profiles of scores at or approaching the cut-off (15) for the CDI are presented in the next column, as these scores are especially challenging from a referral perspective (Sheldrick et al., 2015). Corresponding DLRs (Table 4) for both a traditional approach (CDI alone) and our reclassification algorithm (CDI and SadPeak) were then used to calculate the posterior prevalence for onset. Referral decisions are ultimately based on the posttest prevalence for developing the target disorder within an EBM approach (Youngstrom, 2014).

Youngstrom and colleagues’ (2015) stoplight model was used to provide a framework to interpret the posterior probabilities (Table 5).¹ The value of multiple indicators is best exemplified when comparing two adolescents in our study (Examples 3 and 4). Despite having the same subthreshold CDI score (14), implementation of a brief intervention (red) is warranted for the girl as opposed to the boy where no behavioral response from the provider is required (green). The girl’s profile, subthreshold CDI and threshold pupil dilation, corresponds to more than a twofold increase in likelihood for experiencing a first lifetime episode. Meanwhile, the boy’s pupil dilation score was below subthreshold, suggesting minimal risk for depression onset in the context of a subthreshold CDI score. By merely examining CDI scores, differentiating between these two cases and tailoring the clinical response would not be possible. Thus, consistent with recent calls for more research on pediatric depression onset (Siu et al., 2016), our findings represent a step toward screening protocols that can target vulnerability for depression onset in youth via a risk factor approach. Continued focus on the incremental validity of novel index tests for primary prevention mental health screening carries the promise of bringing personalized prevention solutions to primary care and school settings.

Therefore, “red” may confer prioritizing for referrals to specialty care, in settings that do not have the capacity for intervention.

References

Abela, J. R. Z. (2001). The hopelessness theory of depression: A test of the diathesis-stress and causal mediation components in third and seventh grade children. *Journal of Abnormal Child Psychology*, 29, 241–254. <http://dx.doi.org/10.1023/A:101033815728>

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the Achenbach system of empirically based assessment school-age forms profiles*. Burlington, VT: Aseba.

Avenevoli, S., Swendsen, J., He, J. P., Burstein, M., & Merikangas, K. R. (2015). Major depression in the National Comorbidity Survey—Adolescent Supplement: Prevalence, correlates, and treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, 54, 37–44. e2. <http://dx.doi.org/10.1016/j.jaac.2014.10.010>

Balázs, J., Miklósi, M., Keresztyén, A., Hoven, C. W., Carli, V., Wasserman, C., . . . Wasserman, D. (2013). Adolescent subthreshold-depression and anxiety: Psychopathology, functional impairment and increased suicide risk. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 54, 670–677. <http://dx.doi.org/10.1111/jcpp.12016>

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Boev, A. N., Fountas, K. N., Karampelas, I., Boev, C., Machinis, T. G., Feltes, C., . . . Troup, C. (2005). Quantitative pupillometry: Normative data in healthy pediatric volunteers. *Journal of Neurosurgery: Pediatrics*, 103, 496–500. <http://dx.doi.org/10.3171/ped.2005.103.6.0496>

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45, 602–607. <http://dx.doi.org/10.1111/j.1469-8986.2008.00654.x>

Burkhouse, K. L., Siegle, G. J., & Gibb, B. E. (2014). Pupillary reactivity to emotional stimuli in children of depressed and anxious mothers. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 55, 1009–1016. <http://dx.doi.org/10.1111/jcpp.12225>

Burkhouse, K. L., Siegle, G. J., Woody, M. L., Kudinova, A. Y., & Gibb, B. E. (2015). Pupillary reactivity to sad stimuli as a biomarker of depression risk: Evidence from a prospective study of children. *Journal of Abnormal Psychology*, 124, 498–506. <http://dx.doi.org/10.1037/abn0000072>

Bylsma, L. M., Mauss, I. B., & Rottenberg, J. (2016). Is the divide a chasm? Bridging affective science with clinical practice. *Journal of Psychopathology and Behavioral Assessment*, 38, 42–47. <http://dx.doi.org/10.1007/s10862-015-9525-7>

Canals, J., Martí-Henneberg, C., Fernández-Ballart, J., & Domènech, E. (1995). A longitudinal study of depression in an urban Spanish pubertal population. *European Child & Adolescent Psychiatry*, 4, 102–111. <http://dx.doi.org/10.1007/BF01977738>

Center for Behavioral Health Statistics and Quality. (2016). *Key substance use and mental health indicators in the United States: Results from the 2015 National Survey on Drug Use and Health*. (HHS Publication No. SMA 16-4984, NSDUH Series H-51). Retrieved from <https://www.samhsa.gov/data/sites/default/files/NSDUH-FFR1-2015/NSDUH-FFR1-2015/NSDUH-FFR1-2015.pdf>

Cohen, J. R., Shorey, R. C., Menon, S. V., & Temple, J. R. (2018). Predicting teen dating violence perpetration. *Pediatrics*, e20172790.

Cohen, J. R., So, F. K., Hankin, B. L., & Young, J. F. (2018). Translating cognitive vulnerability into improved adolescent depression screening: A receiver operating characteristic approach. *Journal of Clinical Child and Adolescent Psychology*. Advance online publication. <http://dx.doi.org/10.1080/15374416.2017.1416617>

Cohen, J. R., Young, J. F., & Abela, J. R. (2012). Cognitive vulnerability to depression in children: An idiographic, longitudinal examination of inferential styles. *Cognitive Therapy and Research*, 36, 643–654. <http://dx.doi.org/10.1007/s10608-011-9431-6>

¹ The original “stoplight” model was designed for the assessment setting with “red” conferring immediate intervention. However, it may not be possible for some settings to implement an intervention immediately.

Connelly, M. A., Brown, J. T., Kearns, G. L., Anderson, R. A., St. Peter, S. D., & Neville, K. A. (2014). Pupillometry: A non-invasive technique for pain assessment in pediatric patients. *Archives of Disease in Childhood*, 99, 1125–1131.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44, 837–845. <http://dx.doi.org/10.2307/2531595>

De Los Reyes, A., & Aldao, A. (2015). Introduction to the special issue: Toward implementing physiological measures in clinical child and adolescent assessments. *Journal of Clinical Child and Adolescent Psychology*, 44, 221–237. <http://dx.doi.org/10.1080/15374416.2014.891227>

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141, 858–900. <http://dx.doi.org/10.1037/a0038498>

Earls, M. F., & the Committee on Psychosocial Aspects of Child and Family Health American Academy of Pediatrics. (2010). Incorporating recognition and management of perinatal and postpartum depression into pediatric practice. *Pediatrics*, 126, 1032–1039. <http://dx.doi.org/10.1542/peds.2010-2348>

Ebesutani, C., Chorpita, B. F., Higa-McMillan, C. K., Nakamura, B. J., Regan, J., & Lynch, R. E. (2011). A psychometric analysis of the Revised Child Anxiety and Depression Scales—Parent Version in a school sample. *Journal of Abnormal Child Psychology*, 39, 173–185. <http://dx.doi.org/10.1007/s10802-010-9460-8>

Edens, J. F., Marcus, D. K., Lilienfeld, S. O., & Poythress, N. G., Jr. (2006). Psychopathic, not psychopath: Taxometric evidence for the dimensional structure of psychopathy. *Journal of Abnormal Psychology*, 115, 131–144. <http://dx.doi.org/10.1037/0021-843X.115.1.131>

First, M. B., Spitzer, R. L., Gibbon, M. W. J. B., & Williams, J. B. (1995). *Structured clinical interview for DSM-IV axis I disorders*. New York: New York State Psychiatric Institute.

Franklin, J. C., Jamieson, J. P., Glenn, C. R., & Nock, M. K. (2015). How developmental psychopathology theory and research can inform the research domain criteria (RDoC) project. *Journal of Clinical Child and Adolescent Psychology*, 44, 280–290. <http://dx.doi.org/10.1080/15374416.2013.873981>

Fristad, M. A., Wolfson, H., Algorta, G. P., Youngstrom, E. A., Arnold, L. E., Birmaher, B., . . . the LAMS Group. (2016). Disruptive mood dysregulation disorder and bipolar disorder not otherwise specified: Fraternal or identical twins? *Journal of Child and Adolescent Psychopharmacology*, 26, 138–146. <http://dx.doi.org/10.1089/cap.2015.0062>

Gabrieli, J. D. E., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, 85, 11–26. <http://dx.doi.org/10.1016/j.neuron.2014.10.047>

Garb, H. N. (2003). Incremental validity and the assessment of psychopathology in adults. *Psychological Assessment*, 15, 508–520. <http://dx.doi.org/10.1037/1040-3590.15.4.508>

Garber, J., Clarke, G. N., Weersing, V. R., Beardslee, W. R., Brent, D. A., Gladstone, T. R., . . . Iyengar, S. (2009). Prevention of depression in at-risk adolescents: A randomized controlled trial. *Journal of the American Medical Association*, 301, 2215–2224. <http://dx.doi.org/10.1001/jama.2009.788>

Garber, J., Korelitz, K., & Samanez-Larkin, S. (2012). Translating basic psychopathology research to preventive interventions: A tribute to John R. Z. Abela. *Journal of Clinical Child and Adolescent Psychology*, 41, 666–681. <http://dx.doi.org/10.1080/15374416.2012.710161>

Garber, J., & Rao, U. (2014). Depression in children and adolescents. In M. Lewis & K. Rudolph (Eds.), *Handbook of developmental psychopathology* (pp. 489–520). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-9608-3_25

Gibb, B. E., Beevers, C. G., & McGeary, J. E. (2013). Toward an integration of cognitive and genetic models of risk for depression. *Cognition and Emotion*, 27, 193–216. <http://dx.doi.org/10.1080/02699931.2012.712950>

Halligan, S. L., Murray, L., Martins, C., & Cooper, P. J. (2007). Maternal depression and psychiatric outcomes in adolescent offspring: A 13-year longitudinal study. *Journal of Affective Disorders*, 97, 145–154. <http://dx.doi.org/10.1016/j.jad.2006.06.010>

Hammen, C., & Brennan, P. A. (2003). Severity, chronicity, and timing of maternal depression and risk for adolescent offspring diagnoses in a community sample. *Archives of General Psychiatry*, 60, 253–258. <http://dx.doi.org/10.1001/archpsyc.60.3.253>

Hammen, C., Brennan, P. A., & Keenan-Miller, D. (2008). Patterns of adolescent depression to age 20: The role of maternal depression and youth interpersonal dysfunction. *Journal of Abnormal Child Psychology*, 36, 1189–1198. <http://dx.doi.org/10.1007/s10802-008-9241-9>

Hastings, M. E., Krishnan, S., Tangney, J. P., & Stuewig, J. (2011). Predictive and incremental validity of the Violence Risk Appraisal Guide scores with male and female jail inmates. *Psychological Assessment*, 23, 174–183. <http://dx.doi.org/10.1037/a0021290>

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

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment*, 15, 446–455. <http://dx.doi.org/10.1037/1040-3590.15.4.446>

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, 167, 748–751. <http://dx.doi.org/10.1176/appi.ajp.2010.09091379>

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. *Developmental Medicine and Child Neurology*, 56, 453–459. <http://dx.doi.org/10.1111/dmcn.12360>

Johnston, C., & Murray, C. (2003). Incremental validity in the psychological assessment of children and adolescents. *Psychological Assessment*, 15, 496–507. <http://dx.doi.org/10.1037/1040-3590.15.4.496>

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., . . . Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36, 980–988. <http://dx.doi.org/10.1097/00004583-199707000-00021>

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

Koenigs, M., & Grafman, J. (2009). The functional neuroanatomy of depression: Distinct roles for ventromedial and dorsolateral prefrontal cortex. *Behavioural Brain Research*, 201, 239–243. <http://dx.doi.org/10.1016/j.bbr.2009.03.004>

Kovacs, M. (1992). *Manual for the Children's Depression Inventory*. North Tonawanda, NJ: Multi-Health Systems.

Kraemer, H. C. (1992). *Evaluating medical tests: Objective and quantitative guidelines*. Newbury Park, CA: Sage.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, 7, 18–27. <http://dx.doi.org/10.1177/1745691611427305>

Lavigne, J. V., Feldman, M., & Meyers, K. M. (2016). Screening for mental health problems: Addressing the base rate fallacy for a sustainable screening program in integrated primary care. *Journal of Pediatric Psychology*, 41, 1081–1090. <http://dx.doi.org/10.1093/jpepsy/jsw048>

Leening, M. J., Vedder, M. M., Witteman, J. C., Pencina, M. J., & Steyerberg, E. W. (2014). Net reclassification improvement: Computation, interpretation, and controversies: A literature review and clinician's guide. *Annals of Internal Medicine*, 160, 122–131. <http://dx.doi.org/10.7326/M13-1522>

LeMoult, J., Yoon, K. L., & Joormann, J. (2016). Rumination and cognitive distraction in major depressive disorder: An examination of respiratory sinus arrhythmia. *Journal of Psychopathology and Behavioral Assessment*, 38, 20–29. <http://dx.doi.org/10.1007/s10862-015-9510-1>

Lewinsohn, P. M., Rohde, P., Seeley, J. R., Klein, D. N., & Gotlib, I. H. (2000). Natural course of adolescent major depressive disorder in a community sample: Predictors of recurrence in young adults. *The American Journal of Psychiatry*, 157, 1584–1591. <http://dx.doi.org/10.1176/appi.ajp.157.10.1584>

Lobbestael, J., Leurgans, M., & Arntz, A. (2011). Inter-rater reliability of the Structured Clinical Interview for *DSM-IV* Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clinical Psychology & Psychotherapy*, 18, 75–79. <http://dx.doi.org/10.1002/cpp.693>

Luntamo, T., Sourander, A., Santalahti, P., Aromaa, M., & Helenius, H. (2012). Prevalence changes of pain, sleep problems and fatigue among 8-year-old children: Years 1989, 1999, and 2005. *Journal of Pediatric Psychology*, 37, 307–318. <http://dx.doi.org/10.1093/jpepsy/jsr091>

Martel, M. M., Markon, K., & Smith, G. T. (2017). Research review: Multi-informant integration in child and adolescent psychopathology diagnosis. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 58, 116–128. <http://dx.doi.org/10.1111/jcpp.12611>

Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian facial expressions of emotion (JACFEE)* [Slides]. California: Intercultural and Emotion Research Laboratory, Department of Psychology, San Francisco State University.

Matthey, S., & Petrovski, P. (2002). The Children's Depression Inventory: Error in cutoff scores for screening purposes. *Psychological Assessment*, 14, 146–149. <http://dx.doi.org/10.1037/1040-3590.14.2.146>

Monroe, S. M., & Harkness, K. L. (2011). Recurrence in major depression: A conceptual analysis. *Psychological Review*, 118, 655–674. <http://dx.doi.org/10.1037/a0025190>

Muris, P., Fokke, M., & Kwik, D. (2009). The ruminative response style in adolescents: An examination of its specific link to symptoms of depression. *Cognitive Therapy and Research*, 33, 21–32. <http://dx.doi.org/10.1007/s10608-007-9120-7>

Murray, L., Arteche, A., Fearon, P., Halligan, S., Goodyer, I., & Cooper, P. (2011). Maternal postnatal depression and the development of depression in offspring up to 16 years of age. *Journal of the American Academy of Child & Adolescent Psychiatry*, 50, 460–470. <http://dx.doi.org/10.1016/j.jaac.2011.02.001>

Myers, K., & Winters, N. C. (2002). Ten-year review of rating scales. II: Scales for internalizing disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 41, 634–659. <http://dx.doi.org/10.1097/00004583-200206000-00004>

Nejad, A. B., Fossati, P., & Lemogne, C. (2013). Self-referential processing, rumination, and cortical midline structures in major depression. *Frontiers in Human Neuroscience*, 7, 666. <http://dx.doi.org/10.3389/fnhum.2013.00666>

Nowak, W., Żarowska, A., Szul-Pietrzak, E., & Misiuk-Hojło, M. (2014). System and measurement method for binocular pupillometry to study pupil size variability. *Biomedical Engineering Online*, 13, 69. <http://dx.doi.org/10.1186/1475-925X-13-69>

Obuchowski, N. A., & Lieber, M. L. (1998). Confidence intervals for the receiver operating characteristic area in studies with small samples. *Academic Radiology*, 5, 561–571. [http://dx.doi.org/10.1016/S1076-6332\(98\)80208-0](http://dx.doi.org/10.1016/S1076-6332(98)80208-0)

Ong, A. D., Bergeman, C. S., Bisconti, T. L., & Wallace, K. A. (2006). Psychological resilience, positive emotions, and successful adaptation to stress in later life. *Journal of Personality and Social Psychology*, 91, 730–749. <http://dx.doi.org/10.1037/0022-3514.91.4.730>

Orue, I., Calvete, E., & Padilla, P. (2014). Brooding rumination as a mediator in the relation between early maladaptive schemas and symptoms of depression and social anxiety in adolescents. *Journal of Adolescence*, 37, 1281–1291. <http://dx.doi.org/10.1016/j.adolescence.2014.09.004>

Pearson, R. M., Evans, J., Kounali, D., Lewis, G., Heron, J., Ramchandani, P. G., . . . Stein, A. (2013). Maternal depression during pregnancy and the postnatal period: Risks and possible mechanisms for offspring depression at age 18 years. *Journal of the American Medical Association Psychiatry*, 70, 1312–1319. <http://dx.doi.org/10.1001/jamapsychiatry.2013.2163>

Pencina, M. J., D'Agostino, R. B., Sr., D'Agostino, R. B., Jr., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, 27, 157–172. <http://dx.doi.org/10.1002/sim.2929>

Pencina, M. J., D'Agostino, R. B., Pencina, K. M., Janssens, A. C., & Greenland, P. (2012). Interpreting incremental value of markers added to risk prediction models. *American Journal of Epidemiology*, 176, 473–481. <http://dx.doi.org/10.1093/aje/kws207>

Pencina, M. J., D'Agostino, R. B., Sr., & Steyerberg, E. W. (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine*, 30, 11–21. <http://dx.doi.org/10.1002/sim.4085>

Poznanski, E. O., & Mokros, H. B. (1996). *Children's Depression Rating Scale, Revised (CDRS-R)*. Los Angeles, CA: Western Psychological Services.

Price, R. B., Rosen, D., Siegle, G. J., Ladouceur, C. D., Tang, K., Allen, K. B., . . . Silk, J. S. (2016). From anxious youth to depressed adolescents: Prospective prediction of 2-year depression symptoms via attentional bias measures. *Journal of Abnormal Psychology*, 125, 267–278. <http://dx.doi.org/10.1037/abn0000127>

Propper, C. B., & Holochwost, S. J. (2013). The influence of proximal risk on the early development of the autonomic nervous system. *Developmental Review*, 33, 151–167. <http://dx.doi.org/10.1016/j.dr.2013.05.001>

Rawal, A., Collishaw, S., Thapar, A., & Rice, F. (2013). A direct method of assessing underlying cognitive risk for adolescent depression. *Journal of Abnormal Child Psychology*, 41, 1279–1288. <http://dx.doi.org/10.1007/s10802-013-9760-x>

Rice, F., & Rawal, A. (2011). Can basic risk research help in the prevention of childhood and adolescent depression? Examining a cognitive and emotional regulation approach. *Depression Research and Treatment*, 2011, 871245. <http://dx.doi.org/10.1155/2011/871245>

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, 29, 615–620. <http://dx.doi.org/10.1007/s10979-005-6832-7>

Richardson, L. P., McCauley, E., Grossman, D. C., McCarty, C. A., Richards, J., Russo, J. E., . . . Katon, W. (2010). Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents. *Pediatrics*, 126, 1117–1123. <http://dx.doi.org/10.1542/peds.2010-0852>

Seeley, J. R., Stice, E., & Rohde, P. (2009). Screening for depression prevention: Identifying adolescent girls at high risk for future depression. *Journal of Abnormal Psychology*, 118, 161–170. <http://dx.doi.org/10.1037/a0014741>

Sheldrick, R. C., Benneyan, J. C., Kiss, I. G., Briggs-Gowan, M. J., Copeland, W., & Carter, A. S. (2015). Thresholds and accuracy in screening tools for early detection of psychopathology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 56, 936–948. <http://dx.doi.org/10.1111/jcpp.12442>

Sheppard, L. C., & Teasdale, J. D. (2004). How does dysfunctional thinking decrease during recovery from major depression? *Journal of Abnormal Psychology*, 113, 64–71. <http://dx.doi.org/10.1037/0021-843X.113.1.64>

Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by

pupillary responses. *Psychophysiology*, 45, 679–687. <http://dx.doi.org/10.1111/j.1469-8986.2008.00681.x>

Siegle, G. J., Steinhauer, S. R., Carter, C. S., Ramel, W., & Thase, M. E. (2003). Do the seconds turn into hours? Relationships between sustained pupil dilation in response to emotional information and self-reported rumination. *Cognitive Therapy and Research*, 27, 365–382. <http://dx.doi.org/10.1023/A:1023974602357>

Siegle, G. J., Steinhauer, S. R., Friedman, E. S., Thompson, W. S., & Thase, M. E. (2011). Remission prognosis for cognitive therapy for recurrent depression using the pupil: Utility and neural correlates. *Biological Psychiatry*, 69, 726–733. <http://dx.doi.org/10.1016/j.biopsych.2010.12.041>

Silk, J. S., Dahl, R. E., Ryan, N. D., Forbes, E. E., Axelson, D. A., . . . Siegle, G. J. (2007). Pupillary reactivity to emotional information in child and adolescent depression: Links to clinical and ecological measures. *American Journal of Psychiatry*, 164, 1873–1880. <http://dx.doi.org/10.1176/appi.ajp.2007.06111816>

Silk, J. S., Siegle, G. J., Whalen, D. J., Ostapenko, L. J., Ladouceur, C. D., & Dahl, R. E. (2009). Pubertal changes in emotional information processing: Pupillary, behavioral, and subjective evidence during emotional word identification. *Development and Psychopathology*, 21, 7–26. <http://dx.doi.org/10.1017/S0954579409000029>

Siu, A. L., Bibbins-Domingo, K., Grossman, D. C., Baumann, L. C., Davidson, K. W., Ebell, M., . . . the U.S. Preventive Services Task Force (USPSTF). (2016). Screening for depression in adults: U.S. Preventive Services Task Force recommendation statement. *Journal of the American Medical Association*, 315, 380–387. <http://dx.doi.org/10.1001/jama.2015.18392>

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry*, 35, 773–782. <http://dx.doi.org/10.1001/archpsyc.1978.01770300115013>

Stockings, E., Degenhardt, L., Lee, Y. Y., Mihakopoulos, C., Liu, G., . . . 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. *Journal of Affective Disorders*, 174, 447–463. <http://dx.doi.org/10.1016/j.jad.2014.11.061>

Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory-2nd ed. in a sample of college students. *Depression and Anxiety*, 19, 187–189. <http://dx.doi.org/10.1002/da.20002>

Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2011). Acquiring the evidence: How to find current best evidence and have current best evidence find us. *Evidence-based medicine: How to practice and teach EBM* (4th ed., pp. 73–77). London, United Kingdom: Elsevier.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293. <http://dx.doi.org/10.1126/science.3287615>

Teasdale, J. D., Moore, R. G., Hayhurst, H., Pope, M., Williams, S., & Segal, Z. V. (2002). Metacognitive awareness and prevention of relapse in depression: Empirical evidence. *Journal of Consulting and Clinical Psychology*, 70, 275–287. <http://dx.doi.org/10.1037/0022-006X.70.2.275>

Wang, T. J., Gona, P., Larson, M. G., Tofler, G. H., Levy, D., Newton-Cheh, C., . . . Vasan, R. S. (2006). Multiple biomarkers for the prediction of first major cardiovascular events and death. *The New England Journal of Medicine*, 355, 2631–2639. <http://dx.doi.org/10.1056/NEJMoa055373>

Wissow, L. S., Brown, J., Fothergill, K. E., Gadomski, A., Hacker, K., Salmon, P., & Zelkowitz, R. (2013). Universal mental health screening in pediatric primary care: A systematic review. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52, 1134–1147. e23. <http://dx.doi.org/10.1016/j.jaac.2013.08.013>

Yee, A. M., Algorta, G. P., Youngstrom, E. A., Findling, R. L., Birmaher, B., & Fristad, M. A., & the LAMS Group. (2015). Unfiltered Administration of the YMRS and CDRS-R in a Clinical Sample of Children. *Journal of Clinical Child and Adolescent Psychology*, 44, 992–1007. <http://dx.doi.org/10.1080/15374416.2014.915548>

Young, C. C., & Dietrich, M. S. (2015). Stressful life events, worry, and rumination predict depressive and anxiety symptoms in young adolescents. *Journal of Child and Adolescent Psychiatric Nursing*, 28, 35–42. <http://dx.doi.org/10.1111/jcap.12102>

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*, 39, 204–221. <http://dx.doi.org/10.1093/jpepsy/jst062>

Youngstrom, E. A., Choukas-Bradley, S., Calhoun, C. D., & Jensen-Doss, A. (2015). Clinical guide to the evidence-based assessment approach to diagnosis and treatment. *Cognitive and Behavioral Practice*, 22, 20–35. <http://dx.doi.org/10.1016/j.cbpra.2013.12.005>

Youngstrom, E. A., Halverson, T. F., Youngstrom, J. K., Lindhjem, O., & Findling, R. L. (2018). Evidence-based assessment from simple clinical judgments to statistical learning: Evaluating a range of options using pediatric bipolar disorder as a diagnostic challenge. *Clinical Psychological Science*, 6, 243–265. <http://dx.doi.org/10.1177/2167702617741845>

Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M. L., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, 24, 331–363. <http://dx.doi.org/10.1111/cpsp.12207>

Ziegert, D. I., & Kistner, J. A. (2002). Response styles theory: Downward extension to children. *Journal of Clinical Child and Adolescent Psychology*, 31, 325–334. http://dx.doi.org/10.1207/S15374424JCCP3103_04

Appendix

Stepwise Summary of Translational Analytic Approach

Narrative of Related Studies

There have been two previously published reports that involve findings concerning pupil dilation in youth previously published from the larger study from which we drew our sample. The first study (Manuscript 1) examined only baseline data and showed individual differences in pupil dilation of children of depressed and anxious mothers. In the second study (Manuscript 2), the authors found that in a subsample of youth with a depressed mother, greater pupillary reactivity predicted prospective depressive symptoms and episodes. Approximately 16% of the current study was also used in Manuscript 2. The current study is distinct

from these studies by examining pupil dilation (a) within a larger, never-depressed subsample of youth; (b) examining first lifetime episodes of depression as the criterion; and (c) examining the incremental validity and subsequent clinical utility of pupil dilation within the context of other informant and risk factor screening approaches.

Received February 1, 2018

Revision received September 17, 2018

Accepted September 22, 2018 ■