Predictive Utility of Irritability "In Context": Proof-of-Principle for an Early Childhood Mental Health Risk Calculator

Lauren S. Wakschlag, Leigha A. MacNeill, Lindsay R. Pool, Justin D. Smith, Hubert Adam, Deanna M. Barch, Elizabeth S. Norton, Cynthia E. Rogers, Isaac Ahuvia, Christopher D. Smyser, Joan L. Luby & Norrina B. Allen

Routledge
Taylor & Francis Group

Check for updates

Predictive Utility of Irritability "In Context": Proof-of-Principle for an Early Childhood Mental Health Risk Calculator

Lauren S. Wakschlag[a,b], Leigha A. MacNeill[a,b], Lindsay R. Pool[c], Justin D. Smith[d], Hubert Adam[a], Deanna M. Barch[e,f,g], Elizabeth S. Norton[a,b,h], Cynthia E. Rogers[f,i], Isaac Ahuvia[j], Christopher D. Smyser[g,i,k], Joan L. Luby[f], and Norrina B. Allen[b,c]

[a]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine; [b]Institute for Innovations in Developmental Sciences, Northwestern University; [c]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine; [d]Department of Population Health Sciences, Spencer Fox Eccles School of Medicine at University of Utah; [e]Department of Psychological and Brain Sciences, Washington University in St. Louis; [f]Department of Psychiatry, Washington University School of Medicine; [g]Mallinckrodt Institute of Radiology, Washington University School of Medicine; [h]Department of Communication Sciences and Disorders, Northwestern University; [i]Department of Pediatrics, Washington University School of Medicine; [j]Department of Clinical Psychology, Stony Brook University; [k]Department of Neurology, Washington University School of Medicine

ABSTRACT

Objective: We provide proof-of-principle for a mental health risk calculator advancing clinical utility of the irritability construct for identification of young children at high risk for common, early onsetting syndromes.

Method: Data were harmonized from two longitudinal early childhood subsamples (total $N = 403$; 50.1% Male; 66.7% Nonwhite; $M_{age} = 4.3$ years). The independent subsamples were clinically enriched via disruptive behavior and violence (Subsample 1) and depression (Subsample 2). In longitudinal models, epidemiologic risk prediction methods for risk calculators were applied to test the utility of the transdiagnostic indicator, early childhood irritability, in the context of other developmental and social-ecological indicators to predict risk of internalizing/externalizing disorders at preadolescence ($M_{age} = 9.9$ years). Predictors were retained when they improved model discrimination (area under the receiver operating characteristic curve [AUC] and integrated discrimination index [IDI]) beyond the base demographic model.

Results: Compared to the base model, the addition of early childhood irritability and adverse childhood experiences significantly improved the AUC (0.765) and IDI slope (0.192). Overall, 23% of preschoolers went on to develop a preadolescent internalizing/externalizing disorder. For preschoolers with both elevated irritability and adverse childhood experiences, the likelihood of an internalizing/externalizing disorder was 39–66%.

Conclusions: Predictive analytic tools enable personalized prediction of psychopathological risk for irritable young children, holding transformative potential for clinical translation.

When mental disorders become entrenched, they are "unusually stubborn beasts" (Sonuga-Barke, 2014), making prevention at the vulnerability, rather than frank disorder, phase most impactful. Most youth who develop persistent internalizing or externalizing (INT/EXT) syndromes exhibit dysregulation in early childhood (Shaw, 2013). The severity of early-onset INT/EXT problems is partially driven by negative developmental cascades engendered by early dysregulation and its intersection with environmental exposures (McLaughlin, 2016). As such, identifying clinical risk as early as possible has the greatest impact on preventing or forestalling psychopathology (Wakschlag et al., 2019; Luby et al., 2019).

Despite the identification of markers of behavioral vulnerability to mental health in very young children, translation to routine clinical practice has been slow (Evans et al., 2021; Wakschlag et al., 2022). A barrier to clinical uptake is the lack of clinically integrated tools that provide action-oriented guidance that clearly defines risk thresholds and points to services and resources if prevention is warranted. Decisional uncertainty about "when to worry and when to act" creates high cognitive burden contributing to a benevolent stance to "watch and wait," which can avert the child from support to build self-regulation capacities (Wakschlag et al., 2022). Youngstrom et al. (2018) have highlighted the imperative for evidence-based decision-making tools bridging the gap between the rigors of

psychometrically validated assessments and utility in the real world. For clinical utility, an evidence-based assessment approach must advance at least one of the "three Ps": *prediction, prescription, or process* (Youngstrom & Van Meter, 2016). As our objective is to advance broad-based prevention-oriented identification, we employ a risk calculator model focused on prediction.

In this proof-of-principle paper, we first propose that early childhood mental health risk calculators are uniquely poised to serve as an innovative engine of this sorely needed clinical translation (Hahn et al., 2017; Luby et al., 2019; MacNeill et al., 2021). We then model an early childhood risk calculator, applying validated risk prediction methods (Pool et al., 2018) to a pooled cohort as an example of utility for predicting INT/EXT disorders. Finally, we lay out action steps for future research and implementation.

Irritability as a Developmental Marker of Risk for Psychopathology

Current psychopathology frameworks are neurodevelopmental, unfolding from vulnerability to frank disorder with developmental heterogeneity in expression (Mittal & Wakschlag, 2017). This neurodevelopmental paradigm shifts from reified categorical disorders to broad neurodevelopmental vulnerability to impairing psychopathology (Finlay-Jones et al., 2019). Irritability is part of normal human experience in response to frustration, but is also a core diagnostic feature when pervasively dysregulated (Evans et al., 2017). Dysregulated irritability in early childhood is conceptualized as low frustration tolerance and proneness to anger that is disproportionate to context and relatively unresponsive to support (Wakschlag et al., 2018). It is a robust transdiagnostic indicator of young children's neurodevelopmental vulnerability to INT/EXT syndromes (Beauchaine & Tackett, 2019; Wakschlag et al., 2018). In contrast, research in older youth is more consistently linked to INT syndromes (Evans et al., 2017; Vidal-Ribas et al., 2016). This developmental difference may reflect method variance across age periods, as there are validated early childhood measures specifically designed to capture the dimensional spectrum of irritability. In older youth, however, most studies have employed narrower DSM symptom-based characterization (e.g., oppositional defiant disorder [ODD] irritability symptoms). As the present risk modeling incorporates early childhood irritability, we theorized transdiagnostic predictive utility to preadolescent psychopathology. In the DSM-5 (American Psychiatric Association, 2013), irritability is a diagnostic feature of more than a dozen disorders (e.g., depression, ODD, anxiety), a corollary feature of many others (e.g., autism and attention-deficit/hyperactivity disorder), and is a frequent intervention target (Evans et al., 2017). The clinical salience of irritability was further operationalized in the DSM-5 by adding Disruptive Mood Dysregulation Disorder (DMDD) and an ODD subtype with chronic irritability/anger in ICD-11.

Irritability is also a salient developmental indicator of the broader emotion dysregulation substrates of common psychopathologies, has pragmatic measurement methods for screening, and is identifiable in the first years (Beauchaine & Tackett, 2019; Krogh-Jespersen et al., 2021; Morris et al., 2020; Wakschlag et al., 2019). Tantrums are a key feature of dysregulated irritability in young children, yet are also a normative misbehavior of early childhood (Wakschlag et al., 2019). To address this distinction, the Multidimensional Assessment Profiles (MAPS) Temper Loss scale (Wakschlag et al., 2014) was developed as a parent-report irritability spectrum assessment based on novel developmental specification theory. [Note: Originally called the Multidimensional Assessment for Preschool Disruptive Behavior, we have revised this to be the Multidimensional Assessment Profiles (MAPS) Scales to reflect its extension to older ages, internalizing dimensions, and transdiagnostic nature)]. Findings using the MAPS scale indicate that early childhood atypical irritability is distinguishable from normative misbehavior via frequency, dysregulation, and occurrence in developmentally unexpected contexts (Wakschlag et al., 2018). The MAPS scale has predictive clinical and mechanistic utility for impairment (Damme et al., 2022; Wiggins et al., 2018), holding promise for early identification of INT/EXT risk.

The Promise of Early Childhood Mental Health Risk Calculators

Clinical risk calculators use epidemiologic prognostic risk prediction models to generate probabilistic risk of developing a health condition (Pencina & D'Agostino, 2012). This healthcare application of predictive analytics moves from traditional *post hoc* prediction (hindsight) and investigation of statistical group differences (insight) to predicting likelihood of developing a future condition (foresight) (Hahn et al., 2017). Risk calculators often pool multiple prospective cohorts to capture disease occurrence and achieve power (Pencina & D'Agostino, 2012; Silva Ribeiro et al., 2020). Modeling identifies most parsimonious indicators for *clinical discrimination*. Thus, the key difference between standard regression and risk calculators is that the latter estimates risk in a manner designed for clinical translation.

Risk calculators can accelerate clinical translation, transforming standard of care in preventable physical diseases, most notably cardiovascular disease (CVD) (Pencina & D'Agostino, 2012; Pool et al., 2018). The mental health domain is well suited to risk calculators given the importance of neurodevelopmental indicators for early psychopathology detection (Caye et al., 2020; Oliver et al., 2021). Nascent research on mental health risk calculators in youth is promising, particularly for psychosis (Osborne & Mittal, 2019) and mood/bipolar disorder prediction (Birmaher et al., 2018; Silva Ribeiro et al., 2020). Two recent investigations have used risk calculator methods to determine the probability of young adult INT/EXT from childhood victimization and other risk indicators (Caye et al., 2020; Meehan et al., 2020). These efforts have focused on relatively rare, severe mental illnesses at older ages.

Despite these advances, risk calculators have yet to achieve traction in clinical mental health care outside of research environments. This has been more broadly true of clinical uptake of evidence-based assessments in mental health (Youngstrom et al., 2018). The slow pace of mental health translation may flow from historical factors and the nature of the disorders themselves. Standard of care in physical disease includes reliance on discrete risk markers (e.g., CVD "essential eight," the widely used Framingham risk calculator as well as the pooled cohort equation) (Lloyd-Jones et al., 2022). In contrast, the field of mental health has long focused on subjective experience and biopsychosocial features, which don't lend themselves easily to quantitative risk determination. Second, mental disorders have been conceptualized as developmentally unfolding phenomena for only a few decades. This unfolding approach has been slow to translate to clinical practice, because it does not lend itself easily to decision-making (Wakschlag et al., 2018). In young children, concerns about premature labeling and stigma have further contributed to reluctance to identify early vulnerability, with the hope that it will naturally remit over the rapidly changing developmental course in early childhood (Wakschlag et al., 2019). Nonetheless, the robust evidence on entrenched patterns of psychopathology by preschool, combined with fundamental changes in conceptualizations of mental disorders, provide traction for the advancement of early childhood mental health risk calculators.

Although existing risk calculators predicting young adult clinical outcomes have examined childhood risk, it is imperative to predict the emergence of psychopathology earlier in development and across multiple levels (MacNeill et al., 2021). Because of the substantial variability in mental health problems in early development and the importance of the caregiving environment in risk trajectories, risk calculators that empirically test multi-level risk and protective indicators are critical. On the one hand, early behavioral risk substantially increases risk of subsequent psychopathology. For instance, elevated irritability increases risk of developing impairing problems, including (a) at preschool age associated with 4–14 times greater odds of preadolescent INT/EXT disorders (Wiggins, Ureña Rosario, MacNeill, et al., 2023), and (b) at 1 year of age associated with 4× greater odds of impairment at age 2 (Wiggins, Ureña Rosario, Zhang, et al., 2023). On the other hand, many irritable young children do not exhibit psychopathology (Wiggins et al., 2014). Adverse (e.g., early life stress) and resilience-promoting (e.g., responsive parenting) ecological factors are determinants of this multifinality (Taylor & Sonuga-Barke, 2008). Developmental functioning may also determine risk. To our knowledge, the present work is the first application of risk calculator methodology applied to early irritability and ecological and developmental factors toward precision of estimation (Pencina & D'Agostino, 2012; Pool et al., 2018).

In the current study, we developed a risk calculator leveraging data from two early childhood samples with preadolescent follow-up, with alignment across samples sufficient for harmonization. We tested probabilistic risk for any INT/EXT disorders in preadolescence based on available early childhood indicators across multiple levels (behavioral and developmental functioning, ecological indicators). We also examined whether there were differences in predictive utility for INT and EXT disorders separately.

Methods

Participants

The pooled cohort was derived from two Midwestern NIMH-funded studies: The Mapping the Diversity of Young Children Study (MAPS) and the Preschool Depression Study (PDS). Participants were recruited in early childhood reflecting study goals to identify developmental markers of risk for psychopathology. Parents provided written informed consent for participation at both timepoints, and participants provided written assent at preadolescence. Incentives were provided. Protocols were approved by sites' institutional review boards.

The analytic sample for this harmonized cohort was 403, 55% of the total of the two samples (N = 731). Participants were excluded from this cohort if they did not participate at preadolescence and/or if they had missing data on INT/EXT disorders at

either timepoint (see flow chart, Figure S1). Two-hundred thirty-nine (56%) of the original 425 MAPS participants at the preschool assessment participated in the preadolescent follow-up and had preschool exposure data and were included in the pooled cohort (Figure S1). Of the 306 PDS preschoolers participating in an intensive assessment at baseline, 164 (54%) participated in the preadolescent follow-up and had preschool exposure data. MAPS over-sampled for disruptive behavior and violence exposure. PDS oversampled for child depression. The contrasting enrichment strategies of MAPS (EXT) and PDS (INT) are a strength for examination of transdiagnostic predictive models. (Detailed sample information is in Supplemental Materials, Table S1.)

The pooled cohort was 50.1% male, 66.7% Nonwhite, and socio-demographically diverse (37% poor). Children were M_{age} 4.3 ($SD = 0.8$) at preschool time-point and M_{age} 9.9 years ($SD = 1.4$) at preadolescence. Detailed descriptive statistics are in Table 1.

Harmonization Process

Data harmonization is a systematic process whereby study-specific variables are combined into a dataset based on the constructs they measure, resulting in a set of common variables across both samples to be analyzed as equivalent constructs regardless of study (Fortier et al., 2017). Risk indicators were derived from the preschool waves of the feeder studies based on theory, prior research, and available data. Ability to appropriately combine indicators was necessary for har-monization. Demographic variables were harmonized using codification applied in both studies. Other pre-dictor measures were scored within-sample, and stan-dardized when necessary to compare across samples before harmonization. Score distributions were standar-dized as z-scores ($M = 0$, $SD = 1$). The primary outcome was any preadolescent INT/EXT disorder. In *post hoc* analyses, we also examined INT and EXT separately.

Risk Predictors

Demographics

The harmonized dataset included information on child sex, age, and poverty status (yes/no). The sample was racially/ethnically diverse. Race was not included as it is an imperfect proxy for true risk factors like structural racism and discrimination and may be misconstrued as biologically driven (Helms et al., 2005; Obermeyer et al.,

Table 1. Descriptive statistics for the individual samples and the pooled cohort.

	MAPS (n = 239)	MAPS (n = 239) SD	PDS (n = 164)	PDS (n = 164) SD	Pooled Cohort (N = 403)	Pooled Cohort (N = 403) SD	Test for Differences Between Cohorts
Preschool Demographics							
Sex, %							p = .44
Female	51.5%		47.6%		49.9%		
Male	48.5%		52.4%		50.1%		
Age, M	4.2	0.8	4.5	0.8	4.3	0.8	p < .001
Poverty Status, %							p < .001
Poor	46.9%		22.6%		37.0%		
Non-Poor	53.1%		77.4%		63.0%		
Preschool Behavioral Indicators							
High Irritability, %*	15.9%		17.1%		16.4%		p = .75
High Depressive Behaviors %*	16.3%		12.2%		14.6%		p = .25
Preschool Family Ecological Context Indicators							
Childhood ACES, M	1.7	1.4	2.2	1.7	1.9	1.5	p = .003
High ACEs %**	9.6%		19.5%		13.6%		p = .005
High Parental Responsiveness, %*	51.5%		52.4%		51.9%		p = .85
Preschool Cognitive Ability							
Cognitive Ability, M***	49.5	7.9	49.2	10.1	49.4	8.9	p = .70
High Cognitive Ability%	3.8%		6.1%		4.7%		p = .28
Preadolescence							
Age, M	9.1	1.1	11.2	0.9	9.9	1.4	p < .001
Clinical Outcomes							
Internalizing Disorder, %	4.2%		21.3%		11.2%		p < .001
Externalizing Disorder, %	14.6%		20.7%		17.1%		p = .11
Any INT/EXT Disorder, %	17.2%		31.7%		23.1%		p < .001

Abbreviations: ACEs = adverse childhood experiences; MAPS = Mapping the Diversity of Young Children Study; PDS = Preschool Development Study; INT/EXT = internalizing/externalizing.

*Irritability, depressive behaviors, and parental responsiveness were standardized via z-score within each cohort (mean = 0; standard deviation = 1). High irritability was defined as one standard deviation above the z-score mean, high depressive behaviors was defined as one standard deviation above the z-score mean, and high parental responsiveness was defined as above the z-score mean.

**High ACES was defined as four or more adverse experiences reported.

***Cognitive ability was assessed using a metric with a T-score distribution (mean = 50; standard deviation = 10). High cognitive ability was defined as a score 1.5 standard deviations above the mean.

2019). Poverty was determined by family income-to-needs ratios. Study source was an indicator in the models.

Behavioral Risk

Irritability was the central behavioral risk indicator of the study. We included depressive behavior due to its PDS oversampling. As each study used different indicators of these constructs, data were harmonized: MAPS measured irritability via the 22-item MAPS Temper Loss Scale and depressive behavior via 4 items from the Infant-Toddler-Social-Emotional Assessment (ITSEA) Depressed/Withdrawn Scale (Briggs-Gowan & Carter, 1998); PDS measured irritability with 4 items from the Oppositional scale and depression with 4 items from the Depression scale via the MacArthur Health and Behavior Questionnaire (HBQ) (Essex et al., 2002). Z-scores were calculated within-sample and then harmonized across samples (see Supplemental Materials for details and Tables S2–S3).

Cognitive Ability

Cognitive ability (protective) was measured via nonverbal reasoning in both samples using the Differential Ability Scores (DAS) Picture Similarities scale (Elliott, 1983). Standardized T-scores were calculated within-sample, then harmonized cross-sample.

Family Ecological Context

Two measures of family ecological context were included: maternal responsiveness (protective) and child adverse childhood experiences (ACEs; risk). Maternal responsiveness was measured during observed parent–child interactions: MAPS via the Parenting Clinical Observation Schedule (P-COS) (Hill et al., 2008) and PDS via the Maternal Supportiveness Score (Luby et al., 2006). Within-sample scores were standardized via z-score calculations. The responsiveness z-score was entered continuously. ACEs scores were derived from multiple checklists, using a previously validated approach (Heard-Garris et al., 2020). A set of 15 common ACEs were drawn (Table S4) and included as a count variable.

Psychopathology

Clinical Outcome

Preadolescent DSM-based clinical symptoms were derived from semi-structured interviews with caregivers: the K-SADS (Ambrosini, 2000) in MAPS and the Child and Adolescent Psychiatric Assessment (CAPA) (Angold & Costello, 2000) in PDS.

Participants had an INT/EXT disorder if they met the clinical criteria for EXT disorders (i.e., oppositional defiant, attention-deficit/hyperactivity, and/or conduct disorders) and/or INT disorders (i.e., major depressive. separation anxiety, and/or generalized anxiety disorders).

Psychopathology at Preschool

To test whether behavioral and ecological indicators were merely proxies for concurrent psychopathology, preschool psychopathology was included in *post hoc* models. Psychopathology was measured with the Preschool Age Psychiatric Assessment (PAPA) (Egger et al., 2006), scored within-sample and harmonized cross-sample.

Statistical Analyses

Analyses were conducted using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). Logistic regression models predicted the likelihood that a preschooler would have preadolescent INT/EXT disorder: (**1**) a base model with only demographics, and (**2**) models testing the addition of each risk predictor. Predictors were retained when they added to base model prediction with a significant odds ratio or improvement in discrimination and calibration performance measures. Significant predictors were then sequentially added to measure increased model performance.

We calculated multiple performance measures to determine which sets of risk indicators should be retained for maximal parsimony and precision. *Discrimination* is a measure of how well the model differentiates those at high risk from those at low risk for subsequent psychopathology. Two key discrimination statistics were used: the *area under the receiver operating characteristic curve (AUC)* and the *integrated discrimination index* (IDI). The AUC plots the false-positive against the true positive rate. Therefore, the AUC characterizes the risk model's ability to distinguish between children at preschool age who do/do not have preadolescent INT/EXT. An AUC value of 0.5 means the ability to predict the outcome is no better than random chance, an AUC value between 0.5 and 0.7 indicates a poorly fitting model, and greater than 0.7 is a well-fitting model (Damen et al., 2016). For reference, CVD models with median AUC values of 0.7–0.75 are considered robust (Sniderman et al., 2015). The change in AUC value between the base/demographics model and the models containing additional risk predictor(s) was computed for 95% confidence interval and statistical significance. The discrimination slope measures the difference in the mean of the predicted probability of an

outcome for those with/without the disorder. The discrimination slope ranges from 0 to 1, with higher values indicating better discrimination. We computed the IDI, providing the discrimination slope for each model, and the change between the base/demographics model and the models containing additional risk predictor(s).

Further, we assessed *model calibration*, a measure of how similar predicted risk is to observed outcomes. We produced calibration plots comparing the sequential models identified as showing significant discrimination improvements. Predicted and observed risk were plotted across participants' classified deciles of risk. Perfect calibration occurs when predicted and observed risk are equal. Plots include a line of perfect fit to visually determine whether sample plotted points deviate from the line of perfect fit.

Results

Sequential Models for Risk Calculator Determination

Each potential predictor was added to the base/demographics model one at a time and included in final sequential models if it significantly improved discrimination. Discrimination was weak (AUC = 0.665, discrimination slope = 0.083; Figure 1) in the demographics model. In these single predictor models, **irritability** increased the AUC by 0.085 (95% CI: 0.035, 0.134; p < .001) and the discrimination index by 0.094 (95% CI: 0.059, 0.129; p < .001); **ACEs** improved the discrimination index by 0.032 (95% CI: 0.011, 0.053; p = .003), but did not improve the AUC (0.018; 95% CI: −0.018, 0.055; p = .33). Depressive behavior, cognitive ability, and parental responsiveness did not contribute to discrimination improvement (Table S5).

The final sequential models added the predictors and compared their model statistics in the following order: demographics, irritability, and ACEs. First, we examined discrimination to determine whether the predictive
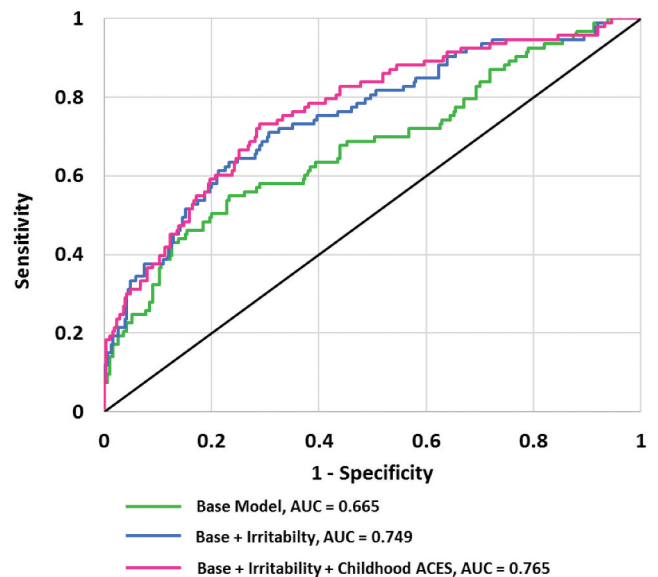

Figure 1. Comparison of receiver operating characteristic (ROC) curves to determine model performance. Three sequential models are compared in this figure. The addition of irritability significantly improved model discrimination, and childhood ACEs further improved discrimination.

ability of the model improved with the addition of predictor variables (Table 2). Adding irritability to the model improved discrimination significantly (AUC = 0.749, discrimination slope = 0.176) as did ACEs based on the IDI (AUC = 0.765, discrimination slope = 0.192). Models with joint consideration of demographics, irritability, and ACEs yielded a strong model fit, whereas models with only one of these did not. Their joint inclusion likely reduces misclassification due to confounding. Model calibration shows how concordant the predicted risk is with the observed outcome across levels of risk, visually assessed by plotting observed versus predicted risk per decile of predicted risk (Figure 2). This shows that the model with demographics, irritability, and ACEs had greatest consistency in observed and predicted risk. Improvements in calibration occurred most dramatically in the middle

Table 2. Sequential model discrimination indices for preschool risk indicators of preadolescent INT/EXT.

Model	Logistic Regression OR (95% CI)	Likelihood Ratio p-value	AUC (95% CI)	AUC Difference (95% CI) p-value	Discrimination Slope	Integrated Discrimination Index (95% CI) p-value
Demographics Only (Base Model)	(Base)	30.46 p < .001	0.665 (0.598, 0.731)	(Base)	0.083	(Base)
Irritability Only	2.25 (1.71–2.96)	38.49 p < .001	0.686 (0.624, 0.747)	0.021 (−0.069,0.111) p = .64	0.100	0.017 (−0.027,0.062) p = .44
Base Model + Irritability	2.29 (1.71–3.07)	66.41 p < .001	0.749 (0.690,0.808)	0.085 (0.035, 0.134) p < .001	0.176	0.094 (0.059,0.129) p < .001
Base Model + Irritability + Childhood ACEs	2.24 (1.65–3.02) 1.26 (1.06–1.49)	73.28 p < .001	0.765 (0.709, 0.822)	0.016 (−0.004,0.037) p = .12	0.192	0.016 (0.000,0.032) p = .05

Abbreviations: ACEs = adverse childhood experiences; AUC = area under the curve; MAPS = Mapping the Diversity of Young Children Study; PDS = Preschool Development Study; INT/EXT = internalizing/externalizing.

deciles, i.e., the final model had improved ability to correctly classify risk in the intermediate risk children (the "gray area" where decision-making is most difficult). This is particularly important for classification of a spectrum of vulnerability to psychopathology.

In sensitivity analyses, we examined model performance measures separately by source sample. Predictive capacity was somewhat stronger in PDS, particularly within the demographics model. However, stratification by sample did not appreciably alter results (Table S6). In post hoc analyses, the addition of preschool INT/EXT status did not appreciably change the model (Table S7). Of note, a model with demographics and preschool INT/EXT was minimally adequate but less discriminating than a model with demographics and irritability.

Final Risk Calculator Model

The following equation was used for the predicted risk of preadolescent INT/EXT:

$$Ln\left(\frac{P}{1-P}\right) = -2.956 - 0.492 * MAPS\ cohort + 0.489$$
$$* male + 0.157 * age + 0.651 * poverty$$
$$+ 0.805 * irritability\ score + 0.228$$
$$* childhood\ ACEs$$

To understand the likelihood of an *individual child's* risk for later INT/EXT, we created personalized risk estimates (Figure 3(a,b)). For context, overall, 23.1% of preschoolers went on to develop a preadolescent INT/EXT disorder. The child most at risk for preadolescent INT/EXT disorder is male, poor, and has high irritability and ACEs (66% likelihood). The child least at risk is female, not poor, and does not have elevated irritability or ACEs (11% likelihood). Poor children were always at greater risk relative to non-poor children. Elevated irritability more than doubled the risk for both girls and boys, regardless of poverty. Irritability was the strongest predictor of INT/EXT (i.e., 29%–56% risk for later INT/EXT), with heightened risk when combined with ACEs (39%–66% risk for later INT/EXT). When we examined model performance measures separately for INT and EXT, irritability had similar discriminative utility. However, strikingly, ACEs added discriminative utility for EXT, but not INT (Table 3).

Discussion

In a first-of-its-kind developmental risk calculator, we demonstrated that preschool irritability adds discriminative value for identification of youth at high probabilistic risk of INT/EXT problems when combined with other contextual factors. Substantial elevated risk is actionable, given that most evidence-based preventions in early childhood are developmentally promotive with no meaningful risk of participation (e.g., increasing self-regulation). For reference, having 7.5% risk of a CVD event within 10 years is elevated, and the basis for clinical decision-making (Goff et al., 2014). Here, we show preschoolers with up to 10× that risk.

Our proof-of-principle analysis highlights three innovations for the early childhood mental health field. (**1**) *It is probabilistic.* The calculator is designed to determine the probability of developing a subsequent mental disorder, to anchor decision-making about "when to worry and when to act." (**2**) *It considers ecological and developmental context.* Clinical risk calculators have typically used individual clinical risk factors, yet the likelihood that neurodevelopmental vulnerability will result in impairing psychopathology is shaped by other factors (Cicchetti & Dawson, 2002). Although our risk model was constrained by available data in our harmonized cohorts, we demonstrated that probabilistic risk increased when ACEs were considered in addition to irritability. (**3**) *It is personalized.* Clinicians do not have a systematic way of integrating multiple sources of information about a child to determine risk of disorder. Our empirical models demonstrate that demographic, behavioral, and ecological risks have value for discrimination of subsequent psychopathology. These innovations represent a foundation for the application of predictive analytics during early childhood, when brain and behavior are most malleable and prevention has maximal long-term benefit.

We have highlighted the promise of early childhood mental health risk calculators as a decision tool to advance precision medicine applications of irritability science. We foresee this based on the impact of risk calculators on routine care in physical disease, their burgeoning use in mental health research, and advances in computational modeling (Luby et al., 2019; MacNeill et al., 2021; Pencina & D'Agostino, 2012). We identified multiple sub-groups of preschoolers with more than a 50% risk of impairing preadolescent INT/EXT. Probabilistic risk identified here is more than *quadruple* the clinically actionable risk level in CVD – preschool-age girls and boys with elevated irritability alone have 29%–56% risk of subsequent INT/EXT, and poor boys with dual behavioral and ecological risk have nearly 70% likelihood of preadolescent INT/EXT.

The transdiagnostic nature of early irritability was evident in its high discriminative value for both broadband syndromes. ACEs also had significant discriminative utility, consistent with prior work showing relations
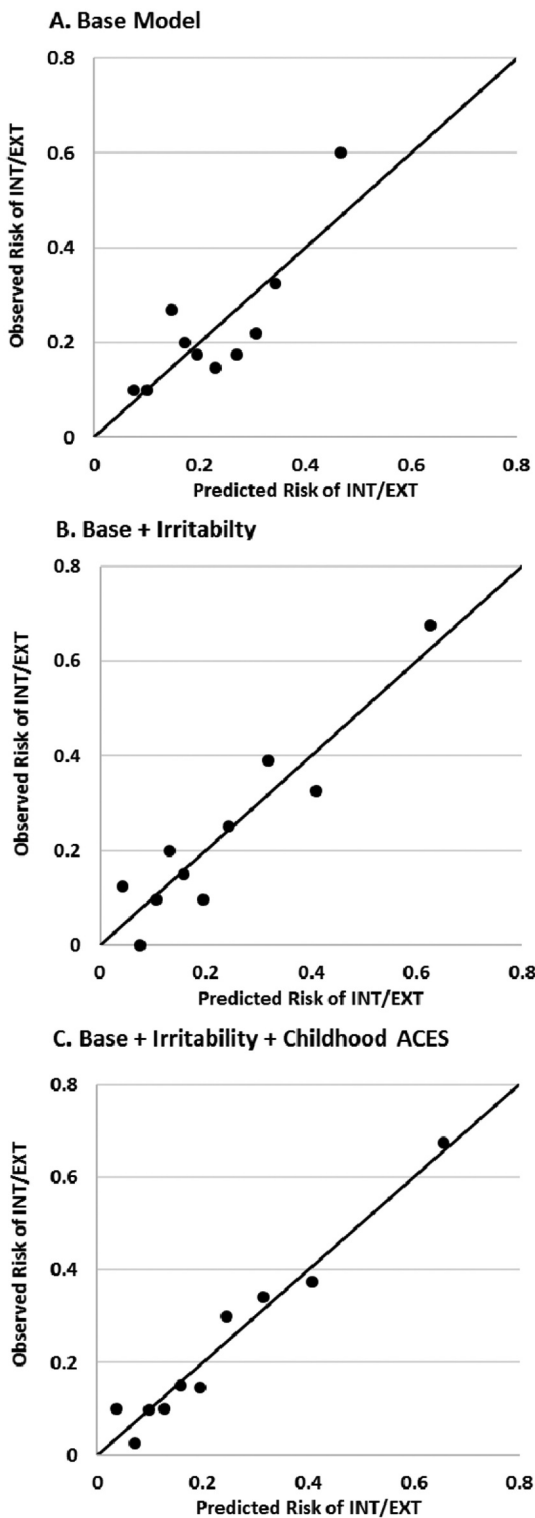
A. Base Model

B. Base + Irritabilty

C. Base + Irritability + Childhood ACES

Figure 2. Calibration plots of predicted vs. observed risk for having a preadolescent INT/EXT disorder. Three sequential models are compared in this figure. Perfect calibration occurs when the predicted vs. observed plots are as close to the line as possible.

between ACEs and earlier-onset disorders (McLaughlin et al., 2012). Although joint INT/EXT models had good performance, there were some model differences when broadband syndromes were considered separately. First, the model predicting INT syndromes had somewhat better discrimination. Depressive and anxious syndromes included in INT outcomes have central features of irritability, whereas EXT syndromes have greater heterogeneity in pathways. In particular, ODD has irritability as core symptom and attention-deficit/hyperactivity disorder is often associated with emotion dysregulation, whereas the conduct disorder-related antisocial patterns have hurtful/callous precursors rather than irritability (Bunford et al., 2015; Frick & White, 2008; Stringaris & Goodman, 2009). Future studies could predict to specific disorders within broadband syndromes, for which the present study is underpowered.

Second, ACEs had added value for prediction of EXT only, when broadband syndromes were separately considered. Findings from multiple samples of youth have demonstrated stronger relations between ACEs and EXT relative to INT syndromes (e.g., Bevilacqua et al., 2021; McLaughlin et al., 2012). Self-control may be a mechanism by which ACEs contribute to psychopathology (McLaughlin et al., 2012), with disruptions of self-control particularly salient to EXT. It is also possible that differential utility of ACEs is a methodologic artifact. ACEs' composition reveals a substantial proportion reflecting dysregulated family patterns of anger (e.g., maltreatment, parental discord). ACEs may be a marker for family violence and disruptive patterns of behavior (and their heritability) rather than a direct effect of adverse exposure. Studies attempting to disaggregate the relation of individual ACEs to these patterns provide some support (Bevilacqua et al., 2021; McLaughlin et al., 2012).

The pattern of findings regarding which indicators did not have discriminative utility requires replication in larger population-based samples. First, we found that irritability had predictive utility, and accounting for the presence or absence of a DSM INT/EXT disorder using symptom-based measures did not have added value. Indeed, *post hoc* models with preschool DSM INT/EXT alone excluding irritability were only weakly discriminative. On its face, the finding that a transdiagnostic behavioral indicator is superior to like disorders predicting like disorders is counterintuitive. This difference may reflect how the irritability indicator covers a broader spectrum of behavior and captures a narrowly defined underlying common mechanism of shared severity. This is relative to the presence/absence categorical indicator of DSM disorders that are heterogeneous and based on extreme symptoms (Hahn et al., 2017). Regardless, a brief survey-based measure which is low burden having strong discriminative utility is of major significance for use in primary care (Wakschlag et al., 2022).
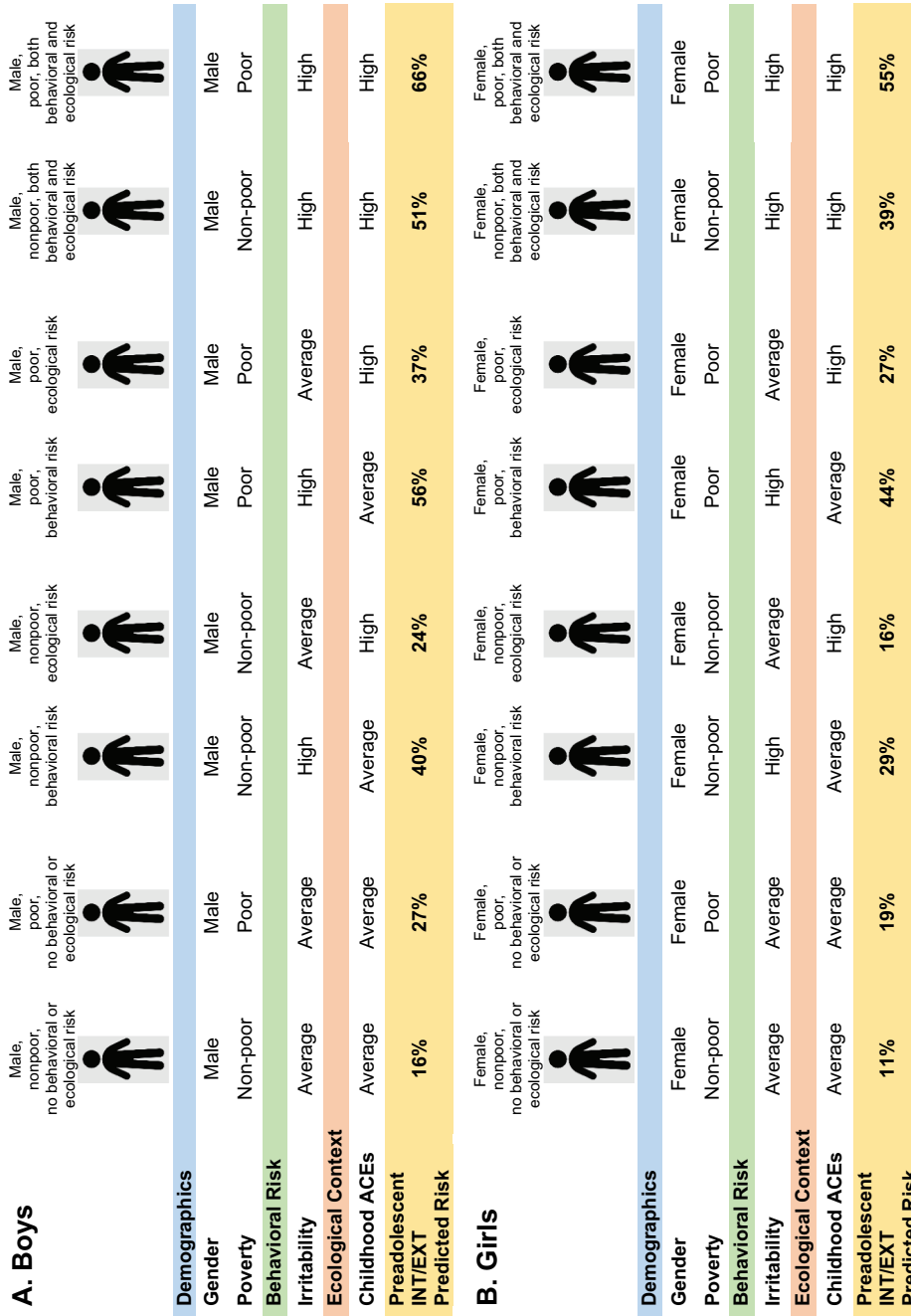
Figure 3. Illustrative personalized clinical risk estimates for having a preadolescent INT/EXT disorder. These predicted risks represent an illustration of the estimates from the best fitting risk equation, base/demographics, irritability, and childhood ACEs, separately for boys (Panel A) and girls (Panel B). Variables other than the ones listed in this figure (i.e., cohort, age) were held at average. Average irritability was defined as z-score value of 0, and high irritability defined as z-score value of 1.5. Average childhood ACEs score defined as z-score value of 1.5. Average childhood ACEs score defined as 2.0, and high ACEs defined as 4.0.

Table 3. Sequential model discrimination indices for preschool risk indicators of preadolescent internalizing and externalizing disorders separately.

Model	Logistic Regression OR (95% CI)	Likelihood Ratio p-value	AUC (95% CI)	AUC Difference (95% CI) p-value	Discrimination Slope	Discrimination Slope Difference (95% CI) p-value
Internalizing Disorders						
Demographics Only (Base Model)	(Base)	48.81 p < .001	0.789 (0.710,0.868)	(Base)	0.155	(Base)
Base Model + Irritability	– 2.61 (1.75–3.89)	75.31 p < .001	0.853 (0.792,0.915)	0.064 (0.014, 0.114) p = .01	0.245	0.090 (0.034,0.145) p = .002
Base Model + Irritability + Childhood ACEs	– 2.58 (1.72–3.86) 1.06 (0.85–1.32)	75.57 p < .001	0.853 (0.792,0.914)	−0.005 (−0.005,0.004) p = .82	0.247	0.002 (−0.003, 0.008) p = .41
Externalizing Disorders						
Demographics Only (Base Model)	(Base)	20.85 p < .001	0.658 (0.586,0.730)	(Base)	0.057	(Base)
Base Model + Irritability	– 2.21 (1.62–3.01)	49.68 p < .001	0.733 (0.665,0.802)	0.075 (0.019, 0.131) p = .008	0.145	0.088 (0.049,0.127) p < .001
Base Model + Irritability + Childhood ACEs	– 2.12 (1.54–2.93) 1.33 (1.10–1.59)	58.87 p < .001	0.761 (0.695,0.827)	0.028 (−0.002,0.058) p = .12	0.171	0.026 (0.005,0.048) p = .02

Abbreviations: ACEs = adverse childhood experiences; AUC = area under the curve; MAPS = Mapping the Diversity of Young Children Study; PDS = Preschool Development Study; INT/EXT = internalizing/externalizing.

Second, despite established relations to psychopathology, observed maternal responsiveness and child cognitive functioning did not have predictive value. This may be due to method variance of the particular measures in this harmonized sample. While it is difficult to pontificate about the absence of discriminative utility of particular indicators in these proof-of-concept analyses, it is important to note that these risk prediction models have a different framework than traditional associative analyses. The latter rely merely on statistical significance of the individual factor, not necessarily its ability to improve the overall discrimination and performance of the model. These patterns are echoed in the cardiovascular field, where established physiologic correlates were not included in standard of care risk calculators because they did not have added value above the available clinical data indicators (Wilson et al., 2005).

Limitations, Real-World Applications, and Future Directions

It would be naïve to suggest that translation to real-world application is a mere step away, as even the most precise methods are ineffective if clinicians do not use them (e.g., Youngstrom et al., 2018). The disconnect between researchers' "best intentions" and systematic uptake of approaches within routine care include clinician mistrust of standardized tools that constrain decision-making autonomy, complexity of presenting predictive analytics in clinically digestible terms, and naiveté about setting specific culture and barriers that require adaptation (Hahn et al., 2017; Keim-Malpass et al., 2018; Oliver et al., 2021; Youngstrom et al., 2018). A "translational mind-set" is

needed so that risk calculator validation is guided by implementation science (Wakschlag et al., 2022). Marrying clinically oriented research with implementation science is essential to moving the dial on this issue (Beidas et al., 2021; Wakschlag et al., 2022; Oliver et al., 2021). Implementation science provides methods to scale scientific discoveries via strategies that rigorously address clinician- and system-level barriers and dynamically adjust based on system response (Beidas et al., 2021). Rigorous investigation of implementation strategies (e.g., alignment with current screening practices, clinical decision supports, clear sight line to action) is crucial (Wakschlag et al., 2022; Oliver et al., 2021). Although the benefits of early identification and prevention are recognized scientifically, community partner engagement must also consider the "gap" between professional and public views of mental health in young children and tackling ethical trade-offs inherent in early risk identification (Shonkoff & Bales, 2011).

Because our calculator predicts likelihood of psychopathology 6–8 years later, one might ask what is actionable about this long interval. We contend that young children with ~6× greater odds for developing mental health problems would benefit from no/minimal risk preventive interventions that promote developmentally essential self-regulatory skills that are foundational for lifelong mental health and functioning (Wakschlag et al., 2022). The Family Check-Up prevention program, for example, has demonstrated effects on preventing conduct problems in preadolescence by intervening in toddlerhood, with irritability as a mediating mechanism (e.g., Smith et al., 2019). This is aligned with cardiovascular risk calculator best practices, which routinely predict risk up to 10 years later (Goff et al., 2014). However,

utility for mental health prevention might be greater for shorter term prediction (e.g., from infancy risk to preschool psychopathology) prior to school entry, as self-regulatory problems impede school readiness and thus foster a negative cascade that exacerbates problems.

Challenges related to implementation – acceptability to parents and clinicians, uptake and integration within healthcare systems, and sustainability – are critical considerations in the next stage of research. There are a number of considerations and barriers to translation to routine care. First, mixed methods inquiry with multilevel community partners is essential to determine alignment of this approach to their priorities and perspectives on relative tradeoffs for a prevention compared to a treatment orientation, depending on the extent to which symptoms are impairing. This should be done within a framework of ongoing engagement of community partners (e.g., families, health providers, community-based organizations, policymakers) with diverse lived experiences and from diverse sectors, fostering a bi-directional meaning-making process that supports uptake approaches that are co-developed and tailored to settings (McNulty et al., 2019). Relatedly, rigorous implementation trials must be conducted in varied types of clinical settings (e.g., primary care vs. specialty mental health clinics) to determine acceptability and incremental utility relative to standard practice. Third, the notorious lack of prevention programs is a further deterrent to clinicians' willingness to engage in systematic screening of early mental health concerns (Merle et al., 2023). This, however, becomes a chicken or the egg question. That is, should availability of services drive the need to identify children at risk, or should more reliably identifying children who would benefit from such services drive growth in availability and increased access to such services? Optimally, the availability of tools for reliable, developmentally based identification will serve as an impetus to further motivate and strengthen current systems of care, which emphasize early childhood developmental health.

Translation within pediatric primary care provides an avenue for population impact (Wakschlag et al., 2019). This is a trusted, non-stigmatized context with broad reach, such that in the US over 95% of the children have a pediatrician and mental health surveillance is a practice parameter (Boat, 2015). Evidence also suggests they improve clinical accuracy in specialty clinics (Youngstrom et al., 2018). However, tools are only as useful as their uptake. A barrier to routine screening has been uncertainty about when to act regarding young children's social-emotional concerns, given the prevalence of normative misbehavior and extensive variation in early childhood (Wakschlag et al., 2019). However, we

have demonstrated early irritability is equally stable to that at older ages (Wiggins, Ureña Rosario, MacNeill, et al., 2023). The current study highlights the promise of risk calculators for informing policy and program on screening at younger ages.

We may also be at a scientific and societal inflection point to advance risk calculators in real-world mental health care. Recent emphasis on health inequities in clinical decision-making highlights how cognitive burden (a hallmark of differentiating normative variation from clinical risk markers in young children) contributes to biased decisions and perpetuates health disparities via increased reliance on racial/ethnic/social stereotypes (Burgess et al., 2004; Van Ryn & Fu, 2003). The implementation of standardized, quantitative tools for clinical decision support is now widely recognized of high value for bias reduction and equitable decision-making.

The diversity and longitudinal characterization of the harmonized cohort are strengths of the present sample. However, the findings are derived from *post hoc* analyses in studies constrained by available data, including starting relatively "late" in early childhood and not being population-based. Rather than be definitive, they are proof-of-principle for the multi-step research requisite for clinical translation of the "healthier, earlier" approach (Wakschlag et al., 2019). Next steps are to generate an early childhood risk calculator and validate it internally as well as in large, diverse developmentally characterized cohorts, externally, and, using state-of-the-art statistical methods to assess the risk of model bias with steps to mitigate it. Generating a synthetic cohort with pre-alignment of domains and methods may achieve necessary power for empirically testing risk calculator algorithms, a process we have underway (MacNeill et al., 2021). Most developmental psychopathology studies examine the same domains but vary in measures. As such, the risk calculator should determine the domains and methods, rather than specific measures, that add predictive precision. Further, our outcome measurement was based on available data; thus, it did not cover the full INT/EXT spectrum. For example, our version of the K-SADS was based on DSM-IV and did not include DMDD, we were not able to include the full range of anxiety syndromes, and we did not include dysthymia as rates were low. Future predictive models should broaden INT/EXT outcomes across timepoints.

Conclusions

Early childhood mental health risk calculators that include irritability and ecological factors are

a promising, practical tool for altering trajectories via earlier identification of probabilistic risk in routine care. Programmatic work toward their generation, validation, and implementation can benefit from the achievements of the cardiovascular field in integrating risk calculators as standard of care, as well as from the burgeoning research on the use of risk calculators for psychosis and mood disorder. Actualizing the promise of this approach has potential for closing the research: practice gap in early life prevention of psychopathology.

Acknowledgments

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Editor's note

This article is part of a special issue, "The Affective Side of Disruptive Behavior: Toward Better Understanding, Assessment, and Treatment," published in the *Journal of Clinical Child and Adolescent Psychology* in 2024. Spencer C. Evans and Jeffrey D. Burke served as editors of the special issue; Andres De Los Reyes served as conflict-of-interest editor.

Funding

References

Ambrosini, P. J. (2000). Historical development and present status of the schedule for affective disorders and schizophrenia for school-age children (K-SADS). *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(1), 49–58. https://doi.org/10.1097/00004583-200001000-00016

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)* (5th ed.). American Psychiatric Association.

Angold, A., & Costello, E. J. (2000). The child and adolescent psychiatric assessment (CAPA). *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(1), 39–48. https://doi.org/10.1097/00004583-200001000-00015

Beauchaine, T. P., & Tackett, J. L. (2019). Irritability as a transdiagnostic vulnerability trait: Current issues and future directions. *Behavior Therapy*, *51*(2), 350–364. https://doi.org/10.1016/j.beth.2019.10.009

Beidas, R. S., Buttenheim, A. M., & Mandell, D. S. (2021). Transforming mental health care delivery through implementation science and behavioral economics. *JAMA Psychiatry*, *78*(9), 941–942. https://doi.org/10.1001/jamapsychiatry.2021.1120

Bevilacqua, L., Kelly, Y., Heilmann, A., Priest, N., & Lacey, R. E. (2021). Adverse childhood experiences and trajectories of internalizing, externalizing, and prosocial behaviors from childhood to adolescence. *Child Abuse & Neglect*, *112*, 104890. https://doi.org/10.1016/j.chiabu.2020.104890

Birmaher, B., Merranko, J. A., Goldstein, T. R., Gill, M. K., Goldstein, B. I., Hower, H., Yen, S., Hafeman, D., Strober, M., Diler, R. S., Axelson, D., Ryan, N. D., & Keller, M. B. (2018). A risk calculator to predict the individual risk of conversion from subthreshold bipolar symptoms to bipolar disorder i or ii in youth. *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(10), 755–763. e754. https://doi.org/10.1016/j.jaac.2018.05.023

Boat, T. F. (2015). Improving lifetime health by promoting behavioral health in children. *JAMA*, *313*(15), 1509–1510. https://doi.org/10.1001/jama.2015.2977

Briggs-Gowan, M., & Carter, A. (1998). Preliminary acceptability and psychometrics of the infant-toddler social and emotional assessment (ITSEA): A new adult report questionnaire. *Infant Mental Health Journal*, *19*, 422–445. https://doi.org/10.1002/SICI_1097-0355_199824_19:4<422:AID-IMHJ5>3.0.CO;2-U

Bunford, N., Evans, S. W., & Wymbs, F. (2015). ADHD and emotion dysregulation among children and adolescents. *Clinical Child and Family Psychology Review*, *18*(3), 185–217. https://doi.org/10.1007/s10567-015-0187-5

Burgess, D. J., Fu, S. S., & Van Ryn, M. (2004). Why do providers contribute to disparities and what can be done about it? *Journal of General Internal Medicine*, *19*(11), 1154–1159. https://doi.org/10.1111/j.1525-1497.2004.30227.x

Caye, A., Agnew-Blais, J., Arseneault, L., Gonçalves, H., Kieling, C., Langley, K., Menezes, A., Moffitt, T., Passos, I., & Rocha, T. (2020). A risk calculator to predict adult attention-deficit/hyperactivity disorder: Generation and external validation in three birth cohorts and one clinical sample. *Epidemiology and Psychiatric Sciences*, *29*, e37. https://doi.org/10.1017/S2045796019000283

Cicchetti, D., & Dawson, G. (2002). Editorial: Multiple levels of analysis. *Development and Psychopathology*, *14*(3), 417–420. https://doi.org/10.1017/S0954579402003012

Damen, J. A., Hooft, L., Schuit, E., Debray, T. P., Collins, G. S., Tzoulaki, I., Lassale, C. M., Siontis, G. C., Chiocchia, V., Roberts, C., Schlüssel, M. M., Gerry, S., Black, J. A., Heus, P., van der Schouw, Y. T., Peelen, L. M., & Moons, K. G. (2016). Prediction models for cardiovascular disease risk in the general population: Systematic review. *BMJ*, *353*, i2416. https://doi.org/10.1136/bmj.i2416

Damme, K. S. F., Norton, E. S., Briggs-Gowan, M. J., Wakschlag, L. S., & Mittal, V. A. (2022). Developmental patterning of irritability enhances prediction of psychopathology in preadolescence: Improving RDoC with

developmental science. *Journal of Psychopathology and Clinical Science*, 131(6), 556–566. https://doi.org/10.1037/abn0000655

Egger, H. L., Erkanli, A., Keeler, G., Potts, E., Walter, B., & Angold, A. (2006). The test-retest reliability of the preschool age psychiatric assessment (PAPA). *Journal of the American Academy of Child & Adolescent Psychiatry*, 45(5), 538–549. https://doi.org/10.1097/01.chi.0000205705.71194.b8

Elliott, C. (1983). *Differential abilities scales: Introductory and technical handbook*. Psychological Corporation.

Essex, M., Boyce, W., Goldstein, L., Armstrong, J., Kraemer, H., & Kupfer, D. (2002). The confluence of mental, physical, social, and academic difficulties in middle childhood. II: Developing the MacArthur Health and Behavior Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(5), 588–603. https://doi.org/10.1097/00004583-200205000-00017

Evans, S. C., Burke, J. D., Roberts, M. C., Fite, P. J., Lochman, J. E., Francisco, R., & Reed, G. M. (2017). Irritability in child and adolescent psychopathology: An integrative review for icd-11. *Clinical Psychology Review*, 53, 29–45. https://doi.org/10.1016/j.cpr.2017.01.004

Evans, S. C., Wei, M. A., Harmon, S. L., & Weisz, J. R. (2021). Modular psychotherapy outcomes for youth with different latent profiles of irritability and emotion dysregulation. *Frontiers in Psychiatry*, 12, 446. https://doi.org/10.3389/fpsyt.2021.618455

Finlay-Jones, A., Varcin, K., Leonard, H., Bosco, A., Alvares, G., & Downs, J. (2019). Very early identification and intervention for infants at risk of neurodevelopmental disorders: A transdiagnostic approach. *Child Development Perspectives*, 13(2), 97–103. https://doi.org/10.1111/cdep.12319

Fortier, I., Raina, P., Van den Heuvel, E. R., Griffith, L. E., Craig, C., Saliba, M., Doiron, D., Stolk, R. P., Knoppers, B. M., & Ferretti, V. (2017). Maelstrom research guidelines for rigorous retrospective data harmonization. *International Journal of Epidemiology*, 46, 103–105. https://doi.org/10.1093/ije/dyw075

Frick, P. J., & White, S. (2008). Research review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 49(4), 359–375. https://doi.org/10.1111/j.1469-7610.2007.01862.x

Goff, D. C., Jr., Lloyd-Jones, D. M., Bennett, G., Coady, S., D'Agostino, R. B., Gibbons, R., Greenland, P., Lackland, D. T., Levy, D., O'donnell, C. J., Robinson, J. G., Schwartz, J. S., Shero, S. T., Smith, S. C., Jr., Sorlie, P., Stone, N. J., & Wilson, P. W. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American College of Cardiology/American Heart Association task force on practice guidelines. *Circulation*, 129(25_suppl_2), S49–73. https://doi.org/10.1161/01.cir.0000437741.48606.98

Hahn, T., Nierenberg, A. A., & Whitfield-Gabrieli, S. (2017). Predictive analytics in mental health: Applications, guidelines, challenges and perspectives. *Molecular Psychiatry*, 22(1), 37–43. https://doi.org/10.1038/mp.2016.201

Heard-Garris, N., Davis, M. M., Estabrook, R., Burns, J., Briggs-Gowan, M., Allen, N., Carnethon, M., Aguayo, L., Wakschlag, L., & Penedo, F. (2020). Adverse childhood experiences and biomarkers of inflammation in a diverse cohort of early school-aged children. *Brain, Behavior, & Immunity-Health*, 1, 100006. https://doi.org/10.1016/j.bbih.2019.100006

Helms, J. E., Jernigan, M., & Mascher, J. (2005). The meaning of race in psychology and how to change it: A methodological perspective. *American Psychologist*, 60(1), 27–36. https://doi.org/10.1037/0003-066X.60.1.27

Hill, C., Maskowitz, K., Danis, B., Keenan, K., Burns, J., & Wakschlag, L. (2008). Validation of a clinically sensitive observational coding system for parenting behaviors: The parenting clinical observation schedule. *Parenting: Science & Practice*, 8(2), 53–85. https://doi.org/10.1080/15295190802045469

Keim-Malpass, J., Kitzmiller, R. R., Skeeles-Worley, A., Lindberg, C., Clark, M. T., Tai, R., Calland, J. F., Sullivan, K., Moorman, J. R., & Anderson, R. A. (2018). Advancing continuous predictive analytics monitoring: Moving from implementation to clinical action in a learning health system. *Critical Care Nursing Clinics*, 30(2), 273–287. https://doi.org/10.1016/j.cnc.2018.02.009

Krogh-Jespersen, S., Kaat, A., Gray, L., Petitclerc, A., Burns, J., Briggs-Gowan, H., Adam, M., & Wakschlag, L. (2021). Calibrating temper loss severity in the transition to toddlerhood: Implications for developmental science. *Applied Developmental Science*, 26(4), 1–14. https://doi.org/10.1080/10888691.2021.1995386

Lloyd-Jones, D. M., Allen, N. B., Anderson, C. A., Black, T., Brewer, L. C., Foraker, R. E., Grandner, M. A., Lavretsky, H., Perak, A. M., Sharma, G., Rosamond, W., & American Heart Association. (2022). Life's essential 8: Updating and enhancing the American Heart Association's Construct of Cardiovascular Health: A presidential advisory from the American Heart Association. *Circulation*, 146(5), e18–43. https://doi.org/10.1161/CIR.0000000000001078

Luby, J., Allen, N., Estabrook, R., Pine, D. S., Rogers, C., Krogh-Jespersen, S., Norton, E. S., & Wakschlag, L. (2019). Mapping infant neurodevelopmental precursors of mental disorders: How synthetic cohorts & computational approaches can be used to enhance prediction of early childhood psychopathology. *Behaviour Research and Therapy*, 123, 103–484. https://doi.org/10.1016/j.brat.2019.103484

Luby, J., Sullivan, J., Belden, A., Stalets, M., Blankenship, S., & Spitznagel, E. (2006). An observational analysis of behavior in depressed preschoolers: Further validation of early-onset depression. *Journal of American Academy of Child Adolescent Psychiatry*, 45(2), 203–212. https://doi.org/10.1097/01.chi.0000188894.54713.ee

MacNeill, L. A., Allen, N., Poleon, R., Vargas, T., Osborne, J., Damme, K., Barch, D., Krogh-Jespersen, S., Norton, E., Rogers, C., Smyser, C., Louby, J., Mittal, V., & Wakschlag, L. S. (2021). Translating RDoC to real-world impact in developmental psychopathology: A neurodevelopmental framework for application of mental health risk calculators. *Development and Psychopathology*, 33(5), 1665–1684. https://doi.org/10.1017/S0954579421000651

McLaughlin, K. A. (2016). Future directions in childhood adversity and youth psychopathology. *Journal of Clinical Child & Adolescent Psychology*, 45(3), 361–382. https://doi.org/10.1080/15374416.2015.1110823

McLaughlin, K. A., Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2012).

Childhood adversities and first onset of psychiatric disorders in a national sample of US adolescents. *Archives of General Psychiatry*, *69*(11), 1151–1160. https://doi.org/10.1001/archgenpsychiatry.2011.2277

McNulty, M., Smith, J., Villamar, J., Burnett-Zeigler, I., Vermeer, W., Benbow, N., Gallo, C., Wilensky, U., Hjorth, A., & Mustanski, B. (2019). Implementation research methodologies for achieving scientific equity and health equity. *Ethnicity & Disease*, *29*(Suppl 1), 83–92. https://doi.org/10.18865/ed.29.S1.83

Meehan, A. J., Latham, R. M., Arseneault, L., Stahl, D., Fisher, H. L., & Danese, A. (2020). Developing an individualized risk calculator for psychopathology among young people victimized during childhood: A population-representative cohort study. *Journal of Affective Disorders*, *262*, 90–98. https://doi.org/10.1016/j.jad.2019.10.034

Merle, J., Carroll, A. J., Mohanty, N., Berkel, C., Scherr, C., Davis, M. M., Wakschlag, L. S., & Smith, J. D. (2023). Pediatric clinicians' perspectives on identification and intervention for young children's social-emotional well-being in primary care. *Submitted Manuscript*.

Mittal, V. A., & Wakschlag, L. S. (2017). Research domain criteria (RDoC) grows up: Strengthening neurodevelopmental investigation within the RDoC framework. *Journal of Affective Disorders*, *216*, 30–35. https://doi.org/10.1016/j.jad.2016.12.011

Morris, A., Wakschlag, L., Krogh-Jespersen, S., Fox, N., Planalp, B., Perlman, S., Shuffrey, L., Smith, B., Lorenzo, N., Amso, D., Coles, C., & Johnson, S. (2020). Principles for guiding the selection of early childhood neurodevelopmental risk and resilience measures: Healthy brain and child development study as an exemplar. *Adversity and Resilience Science*, *1*(4), 247–267. https://doi.org/10.1007/s42844-020-00025-3

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Oliver, D., Spada, G., Colling, C., Broadbent, M., Baldwin, H., Patel, R., Stewart, R., Stahl, D., Dobson, R., & McGuire, P. (2021). Real-world implementation of precision psychiatry: Transdiagnostic risk calculator for the automatic detection of individuals at-risk of psychosis. *Schizophrenia Research*, *227*, 52–60. https://doi.org/10.1016/j.schres.2020.05.007

Osborne, K. J., & Mittal, V. A. (2019). External validation and extension of the napls-2 and sips-rc personalized risk calculators in an independent clinical high-risk sample. *Psychiatry Research*, *279*, 9–14. https://doi.org/10.1016/j.psychres.2019.06.034

Pencina, M. J., & D'agostino, R. B., Sr. (2012). Thoroughly modern risk prediction. *Science Translational Medicine*, *4*(131), 131fs110. https://doi.org/10.1126/scitranslmed.3004127

Pool, L. R., Ning, H., Wilkins, J., Lloyd-Jones, D. M., & Allen, N. B. (2018). Use of long-term cumulative blood pressure in cardiovascular risk prediction models. *JAMA Cardiology*, *3*(11), 1096–1100. https://doi.org/10.1001/jamacardio.2018.2763

Shaw, D. (2013). Future directions for research on the development and prevention of early conduct problems. *Journal of Clinical Child & Adolescent Psychology*, *42*(3), 418–428. https://doi.org/10.1080/15374416.2013.777918

Shonkoff, J. P., & Bales, S. N. (2011). Science does not speak for itself: Translating child development research for the public and its policymakers. *Child Development*, *82*(1), 17–32. https://doi.org/10.1111/j.1467-8624.2010.01538.x

Silva Ribeiro, J., Pereira, D., Salagre, E., Coroa, M., Santos Oliveira, P., Santos, V., Madeira, N., Grande, I., & Vieta, E. (2020). Risk calculators in bipolar disorder: A systematic review. *Brain Sciences*, *10*(8), 525. https://doi.org/10.3390/brainsci10080525

Smith, J. D., Wakschlag, L., Krogh-Jespersen, S., Walkup, J. T., Wilson, M. N., Dishion, T. J., & Shaw, D. S. (2019). Dysregulated irritability as a window on young children's psychiatric risk: Transdiagnostic effects via the family check-up. *Development and Psychopathology*, *31*(5), 1887–1899. https://doi.org/10.1017/S0954579419000816

Sniderman, A. D., D'Agostino, R. B., Sr., & Pencina, M. J. (2015). The role of physicians in the era of predictive analytics. *JAMA*, *314*(1), 25–26. https://doi.org/10.1001/jama.2015.6177

Sonuga-Barke, E. J. S. (2014). Editorial: 'What's up, RDoC?' - can identifying core dimensions of early functioning help us understand, and then reduce, developmental risk for mental disorders? *Journal of Child Psychology and Psychiatry*, *55*(8), 849–851. https://doi.org/10.1111/jcpp.12293

Stringaris, A., & Goodman, R. (2009). Longitudinal outcome of youth oppositionality: Irritable, headstrong, and hurtful behaviors have distinctive predictions. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(4), 404–412. https://doi.org/10.1097/CHI.0b013e3181984f30

Taylor, E., & Sonuga-Barke, E. (2008). Disorders of attention and activity. In M. Rutter, D. Bishop, D. Pine, S. Scott, J. Stevenson, E. Taylor, & A. Thapar (Eds.), *Rutter's child and adolescent psychiatry* (pp. 521–542). Blackwell.

Van Ryn, M., & Fu, S. S. (2003). Paved with good intentions: Do public health and human service providers contribute to racial/ethnic disparities in health? *American Journal of Public Health*, *93*(2), 248–255. https://doi.org/10.2105/AJPH.93.2.248

Vidal-Ribas, P., Brotman, M. A., Valdivieso, I., Leibenluft, E., & Stringaris, A. (2016). The status of irritability in Psychiatry: A conceptual and quantitative review. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(7), 556–570. https://doi.org/10.1016/j.jaac.2016.04.014

Wakschlag, L., Briggs-Gowan, M., Choi, S., Nichols, S., Kestler, J., Burns, J., Carter, A., & Henry, D. (2014). Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(1), 82–96. https://doi.org/10.1016/j.jaac.2013.10.011

Wakschlag, L., Perlman, S., Blair, R., Leibenluft, E., Briggs-Gowan, M., & Pine, D. (2018). The neurodevelopmental basis of early childhood disruptive behavior: Irritable and callous phenotypes as exemplars. *American Journal of Psychiatry*, *175*(2), 114–130. https://doi.org/10.1176/appi.ajp.2017.17010045

Wakschlag, L. S., Finlay-Jones, A. L., MacNeill, L. A., Kaat, A. J., Brown, C. H., Davis, M. M., Franklin, P., Berkel, C., Krogh-Jespersen, S., & Smith, J. D. (2022). Don't get lost in translation: Integrating developmental and implementation sciences to accelerate real-world

impact on children's development, health, and wellbeing [Perspective]. *Frontiers in Public Health*, *10*, 827412. https://doi.org/10.3389/fpubh.2022.827412

Wakschlag, L. S., Roberts, M. Y., Flynn, R. M., Smith, J. D., Krogh-Jespersen, S., Kaat, A. J., Gray, L., Walkup, J., Marino, B. S., Norton, E. S., & Davis, M. M. (2019). Future directions for early childhood prevention of mental disorders: A road map to mental health, earlier. *Journal of Clinical Child & Adolescent Psychology*, *48*(3), 539–554. https://doi.org/10.1080/15374416.2018.1561296

Wiggins, J. L., Briggs-Gowan, M. J., Estabrook, R., Brotman, M. A., Pine, D. S., Leibenluft, E., & Wakschlag, L. S. (2018). Identifying clinically significant irritability in early childhood. *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(3), 191–199. https://doi.org/10.1016/j.jaac.2017.12.008

Wiggins, J., Mitchell, C., Stringaris, A., & Leibenluft, E. (2014). Developmental trajectories of irritability and bidirectional associations with maternal depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53* (11), 1191–1205. https://doi.org/10.1016/j.jaac.2014.08.005

Wiggins, J., Ureña Rosario, A. MacNeill, L. A., Smith, J. D., & Wakschlag, L. (2023). Prevalence, stability, and predictive utility of the MAPS clinically optimized irritability score: Pragmatic early assessment of mental disorder risk. *Submitted Manuscript*.

Wiggins, J., Ureña Rosario, A., Zhang, Y., MacNeill, L. A., Yu, Q., Norton, E. S., Smith, J. D., & Wakschlag, L. (2023). Advancing earlier transdiagnostic identification of mental health risk with the MAPS temper loss scale: A pragmatic approach to irritability characterization at the transition to toddlerhood. *Submitted Manuscript*.

Wilson, P. W., Nam, B. H., Pencina, M., D'agostino, R. B., Benjamin, E. J., & O'donnell, C. J. (2005). C-reactive protein and risk of cardiovascular disease in men and women from the Framingham heart study. *Archives of Internal Medicine*, *165* (21), 2473–2478. https://doi.org/10.1001/archinte.165.21.2473

Youngstrom, E. A., Halverson, T. F., Youngstrom, J. K., Lindhiem, 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*(2), 243–265. https://doi.org/10.1177/2167702617741845

Youngstrom, E. A., & Van Meter, A. (2016). Empirically supported assessment of children and adolescents. *Clinical Psychology: Science and Practice*, *23*(4), 327–347. https://doi.org/10.1037/h0101738