Clinical Implications of a Dimensional Approach: The Normal:Abnormal Spectrum of Early Irritability

Lauren S. Wakschlag, PhD, Ryne Estabrook, PhD, Amelie Petitclerc, PhD, David Henry, PhD, James L. Burns, MS, Susan B. Perlman, PhD, Joel L. Voss, PhD, Daniel S. Pine, MD, Ellen Leibenluft, MD, Margaret L. Briggs-Gowan, PhD

Objective: The importance of dimensional approaches is widely recognized, but an empirical base for clinical application is lacking. This is particularly true for irritability, a dimensional phenotype that cuts across many areas of psychopathology and manifests early in life. We examine longitudinal, dimensional patterns of irritability and their clinical import in early childhood.

Method: Irritability was assessed longitudinally over an average of 16 months in a clinically enriched, diverse community sample of preschoolers (N = 497; mean = 4.2 years; SD = 0.8). Using the Temper Loss scale of the Multidimensional Assessment Profile of Disruptive Behavior (MAP-DB) as a developmentally sensitive indicator of early childhood irritability, we examined its convergent/divergent, clinical, and incremental predictive validity, and modeled its linear and nonlinear associations with clinical risk.

Results: The Temper Loss scale demonstrated convergent and divergent validity to child and maternal factors. In multivariate analyses, Temper Loss predicted mood (separation anxiety disorder [SAD], generalized anxiety disorder [GAD], and depression/dysthymia), disruptive (oppositional defiant disorder [ODD], attention-deficit/hyperactivity disorder [ADHD], and conduct disorder [CD]) symptoms. Preschoolers with even mildly elevated Temper Loss scale scores showed substantially increased risk of symptoms and disorders. For ODD, GAD, SAD, and depression, increases in Temper Loss scale scores at the higher end of the dimension had a greater impact on symptoms relative to increases at the lower end. Temper Loss scale scores also showed incremental validity over *DSM-IV* disorders in predicting subsequent impairment. Finally, accounting for the substantial heterogeneity in longitudinal patterns of Temper Loss significantly improved prediction of mood and disruptive symptoms.

Conclusion: Dimensional, longitudinal characterization of irritability informs clinical prediction. A vital next step will be empirically generating parameters for the incorporation of dimensional information into clinical decision-making with reasonable certainty.

Key Words: irritability, dimensional, developmental psychopathology, normal:abnormal spectrum, longitudinal modeling

J Am Acad Child Adolesc Psychiatry 2015;54(8):626–634.

Irritability is present in diverse forms of mental illness.[1-3] Prior research has generally focused on extreme irritability. However, because irritability falls along a spectrum and is an early-life precursor to psychopathology,[4,5] dimensional, developmentally specified approaches are needed. Here, we characterize the normal:abnormal spectrum of irritability in early childhood using developmentally informed quantitative methods. Specifically, we model how progression along the dimensional spectrum of irritability relates to subsequent clinical risk and impairment, characterize the variability of irritability over time, and test the value of this longitudinal variation for prediction.

Early identification of abnormal irritability would be of great value for the prevention of mental health disorders. However, irritable behavior is normative in early childhood, and its clinical significance varies based on its context, modulation, and pervasiveness.[5-7] Recent work lays the foundation for making such normal to abnormal differentiations in early childhood.[8,9] For example, we have defined a developmentally based irritability spectrum using the Temper Loss scale of the Multidimensional Assessment Profile of Disruptive Behavior (MAP-DB) questionnaire (this scale was originally titled the "Multidimensional Assessment of Preschool Disruptive Behavior" but has since been renamed to reflect its use and validation across a broader age range) in a prior unselected sample.[10,11] This psychometric work lays the foundation for the present clinical validation study.

Currently, empirical approaches for extracting clinically useful information from a dimensionally defined irritability spectrum are underdeveloped. Dimensions are based on the assumption that risk cannot be defined by a single, extreme threshold but instead manifests probabilistically.[4] Thus, a dimensional approach may enhance developmental sensitivity to prodromal phases of risk. Dimensional risk may increase linearly or nonlinearly, with different implications for clinical decision making. Little is also known about the clinical informativeness of longitudinal variation in dimensional patterns. This is of particular importance in early


Supplemental material cited in this article is available online.

TABLE 1 Differences in Temper Loss Score[a] by *DSM-IV* Disorder

Disorder	Prevalence (%)[b]	Mean (SE)		Significance
		Meets Criteria	Does Not Meet Criteria	
Disruptive Disorders				
ODD	14.24	1.00 (0.11)	−0.07 (0.05)	$t(394) = 8.32^{***}$
CD	5.15	0.89 (0.28)	0.04 (0.05)	$t(393) = 2.91^{**}$
ADHD	6.18	0.90 (0.22)	0.02 (0.05)	$t(390) = 3.80^{***}$
Any disruptive disorder	17.94	0.87 (0.11)	−0.10 (0.05)	$t(389) = 7.81^{***}$
Mood Disorders				
GAD	21.22	0.63 (0.10)	−0.07 (0.06)	$t(378) = 5.85^{***}$
SAD	10.55	0.64 (0.20)	0.01 (0.05)	$t(384) = 2.96^{**}$
Depressive disorders[c]	2.41	1.23 (0.38)	0.06 (0.05)	$t(399) = 3.06^{**}$
Any mood disorder	26.59	0.53 (0.10)	−0.09 (0.06)	$t(377) = 5.09^{***}$
Both disruptive and mood disorders	33.88	0.56 (0.08)	−0.17 (0.06)	$t(377) = 6.66^{***}$

Note: ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; GAD = generalized anxiety disorder; ODD = oppositional defiant disorder; SAD = separation anxiety disorder; SE = standard error.
[a]*Calibrated item response theory scores (mean = 0, SD = 1).*
[b]*Adjusted for clinical enrichment via sampling and response weights.*
[c]*Combines depression and dysthymia.*
$^{**}p < .01$; $^{***}p < .001$.

childhood, when the capacity for self-control improves dramatically across relatively short time intervals.[12,13] Dimensional, longitudinal approaches also hold promise for elucidating the substantial heterogeneity in outcomes among young children exhibiting early high irritability (i.e., which young children who are irritable will go on to develop clinical problems and which will not). For example, recent trajectory modeling suggests that more than 25% of young children with high early irritability develop normally when followed longitudinally.[14]

Here we draw on a large, clinically enriched sample of preschoolers to establish the validity of the dimensional Temper Loss scale for clinical prediction and explicate the shape of its relation to clinical outcomes. The goals of this article are as follows: to establish the validity of the Temper Loss scale, including convergent/divergent, clinical, and incremental validity; to characterize the short-term longitudinal variation in Temper Loss scale score; and to test the incremental validity of this variation for clinical prediction.

METHOD

Participants

The Multidimensional Assessment of Preschoolers (MAPS) Study includes a large, diverse sample of preschoolers recruited from the waiting rooms of multiple pediatric clinics in a large urban area of the United States. This unselected sample (N = 1,857) was seen only at baseline and is the sample on which the psychometric modeling of the Temper Loss scale is based.[15] The primary analytic sample for the present study is an intensive subsample of this MAPS pediatric cohort (n = 497), which was clinically enriched by oversampling for child disruptive behavior and parental intimate partner violence. The mean age of the sample at baseline was 4.2 years (T0: mean = 4.2 years, range = 2.9–6.0 years; T1: mean = 4.8 years, range = 3.1–7.7 years; T2: mean = 5.54 years, range = 3.8–8.5 years). Approximately half of the sample were boys and were living in poverty. Participants were predominantly African American, Hispanic, and non-Hispanic white. (For additional sample details, see Supplement 1 and Table S1, available online, and prior published descriptions.[10,11]) All clinical validity analyses used sampling weights that accounted for both unequal probabilities of selection and differential nonresponse rates in this subsample.

Procedures

Procedures were approved by institutional review boards, and parental informed consent was obtained. The clinical subsample participated in 3 longitudinal assessments over an average period of 15.8 months (SD = 5.7 months; for overview, see Figure S1, available online). At baseline (T0), mothers completed the Temper Loss scale. At T1 (~6 months later), they took part in an intensive clinical and neurocognitive assessment. At T1, 80% also completed the Temper Loss scale again (20% were missing because the MAP-DB was added to the T1 assessment after this phase was underway). At T2 (~9 months later), participants completed the Temper Loss scale and survey measures of clinical symptoms and impairment (94% response rate).

Measures

Irritability. Irritability was assessed via the MAP-DB Temper Loss scale at T0, T1, and T2. The Temper Loss scale measures key features of irritability including mood and tantrums. The 22 Temper Loss scale items capture variations in quality, intensity, and context along an objective frequency scale (ranging from never during the past month to many times each day). There were no significant differences in the structure of the Temper Loss scale from the prior independent sample[11] based on differential item function (DIF) estimations using a weighted least squares approach ($\chi^2[109] = 128.95$, $p = .09$). Confirmatory factor analyses also indicated a unidimensional factor (Comparative Fit Index [CFI] = 0.96,[16] Tucker Lewis Index [TLI] = 0.95,[17] and root mean square error of approximation [RMSEA] = 0.09).[18] Unidimensionality was evident across

TABLE 2 Clinical and Incremental Validity of Temper Loss: Predicting *DSM-IV* Symptoms (T1–T2) and Disorders (T1) and Impairment (T1–T2) From T0 Temper Loss[a]

Clinical Validity Models	T1 Prediction Estimate (SE)[β]	T2 Prediction Estimate (SE)[β]
ODD Symptoms		
T0 Temper Loss (linear)	0.56 (0.15) [0.29]***	0.14 (0.07) [0.12]*
T0 Temper Loss (quadratic)	0.19 (0.07) [0.14]**	0.14 (0.07) [0.17]*
CD Symptoms		
T0 Temper Loss (linear)	−0.03 (0.09) [−0.03]	0.03 (0.04) [0.04]
T0 Temper Loss (quadratic)	0.09 (0.06) [0.12]	0.11 (0.06) [0.21]
ADHD Symptoms		
T0 Temper Loss (linear)	0.73 (0.34) [0.17]*	0.52 (0.23) [0.15]*
T0 Temper Loss (quadratic)	0.23 (0.18) [0.08]	0.29 (0.16) [0.12]
SAD Symptoms		
T0 Temper Loss (linear)	0.31 (0.16) [0.21]	0.12 (0.14) [0.14]*
T0 Temper Loss (quadratic)	0.17 (0.07) [0.16]*	0.14 (0.23) [0.23]
GAD Symptoms		
T0 Temper Loss (linear)		0.06 (0.06) [0.06]
T0 Temper Loss (quadratic)	0.13 (0.07) [0.13]	0.16 (0.08) [0.22]
Depression/Dysthymia Symptoms		
T0 Temper Loss (linear)	0.27 (0.11) [0.19]*	0.13 (0.08) [0.11]
T0 Temper Loss (quadratic)	0.19 (0.06) [0.19]***	0.14 (0.23) [0.17]
Any Mood/Disruptive Disorder		
T0 Temper Loss (linear)	1.20 (0.31) [1.05]***	N/A
Incremental Validity Models		
Functional Impairment		
T0 Temper Loss (linear)	0.05 (0.04) [0.13]	−0.01 (0.02) [−0.03]
T0 Temper Loss (quadratic)	0.07 (0.03) [0.27]**	0.02 (0.01) [0.16]**
Impairment in Preschool/Day Care		
T0 Temper Loss (linear)	0.00 (0.02) [0.02]	−0.05 (0.02) [−0.22]*
T0 Temper Loss (quadratic)	0.01 (0.01) [0.09]	0.02 (0.01) [0.08]
Impairment in Family Context		
T0 Temper Loss (linear)	0.00 (0.04) [0.00]	0.00 (0.02) [0.00]
T0 Temper Loss (quadratic)	0.06 (0.03) [0.23]*	0.01 (0.01) [0.08]

Note: ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; GAD = general anxiety disorder; N/A = not available; ODD = oppositional defiant disorder; SAD = separation anxiety disorder.

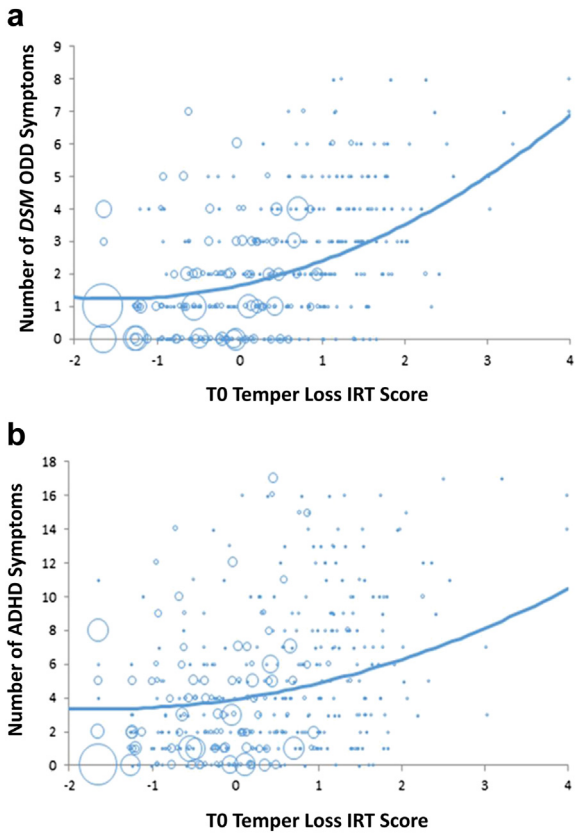
[a]*Symptom and impairment models derived from hierarchical linear regression and DSM disorder model derived from logistic regression model. All models controlled for child age, sex, race/ethnicity, poverty status, harsh physical discipline, child aggression, and time lag between visits. Incremental models predicting impairment also controlled for any mood or disruptive disorder (for full models, see Table S3, available online). T1–T2 symptom model numbers ranged from 370 to 446, based on missing data. Impairment in preschool/day care numbers were smaller due to some children not being in an out-of-home setting (320, 297 respectively).*

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

the 3 time points (α = 0.96–0.98). Scores were derived using item response theory (IRT).[19] IRT is useful for dimensional modeling because it maps the locations of both items and respondents along an underlying latent continuum, scaled from mild, commonly occurring behaviors to severe, rarely occurring behaviors. Baseline (T0) Temper Loss was used as the primary predictor of all T1 and T2 outcomes.

Correlates. Convergent and divergent validity measures were derived from T1. Convergent validity was assessed in relation to 2 surveys and 1 neurocognitive measures. For survey measures, we used the following: a composite child Irritability Symptom Index (α = 0.73; index is detailed by Dougherty *et al.*[9]), derived from the 7 irritability symptoms of the conduct and depression sections of the Preschool Age Psychiatric Assessment (PAPA)[20] (e.g., "is easily frustrated," "has tantrums"); and maternal irritability was assessed with the Patient-Reported Outcomes Measurement Information System (PROMIS) Anger scale (α = 0.93).[21] For a neurocognitive measure, we used a developmentally sensitive test of response reversal, the "Candy Game" task.[22] Response reversal deficits have been demonstrated in older youth with clinical levels of irritability,[23] and we have shown that performance on the Candy Game is associated with prefrontal cortex function at preschool age in a small subsample of the MAPS cohort (n = 28) (Demir OE, Voss J, O'Neill J, Briggs-Gowan M, Wakschlag L, Booth J, unpublished data). The Candy Game is a computerized task in which 2 sets of boxes are presented. One box is designated as "winning," and the other box is designated as "losing" in each set. This designation is reversed in the second half of the trials. Learning occurs via trial and error. Of the 397 children who participated in the neurocognitive assessment, 76% (n = 302)

FIGURE 1 Relation of Temper Loss dimensional location to clinical symptoms. Part 1a shows oppositional defiant disorder (ODD) symptoms. Part 1b shows attention-deficit/hyperactivity disorder (ADHD) symptoms. Note: IRT = item response theory.



successfully completed the task. (Those children who completed fewer than 50 trials or had less than 50% accuracy were excluded; see Supplement 1 and Table S2, available online, for breakdown of task completion.) Preschoolers with and without complete Candy Game data did not differ on level of Temper Loss ($t = -0.15$, not significant). For the present analyses, reversal learning was computed as post-switch minus pre-switch percent accuracy, indicating the decrement in children's performance after the rule switch (mean pre-switch accuracy = 85%, SD = 22.2%; mean post-switch accuracy = 79%, SD = 25.8).

Divergent validity was assessed in relation to maternal report of children's self-regulation, including the Self Control scale of the Social Skills Improvement System (SSIS),[24] a 10-item scale measuring the child's capacity to regulate emotions and behavior ($\alpha = 0.84$), and the Initiative Scale of the Devereux Early Childhood Assessment (DECA),[25] an 11-item scale measuring problem-solving and persistence ($\alpha = 0.88$).

Covariates. Models controlled for the following: child age and sex; poverty as derived from an income-to-needs ratio; use of harsh physical discipline via the 8-item Parent–Child Physical Assault Scale of the Conflict Tactics Scale (CTS) ($\alpha = 0.65$);[25] and Aggressive Behavior assessed with the MAP-DB Aggression Scale. This 25-item scale encompasses mild to severe aggressive behaviors ($\alpha = 0.95$).[10] In addition, the model predicting Candy Game performance controlled for child nonverbal reasoning, assessed via the Picture Similarities Scale of the Differential Abilities Scales (DAS).[26]

Clinical Outcomes. Clinical symptoms of *DSM-IV* disruptive (ODD, ADHD, and CD) and mood (SAD, GAD, and depression/dysthymia) disorders were assessed at T1 and T2. At T1, these symptom counts were derived from the PAPA. We also created a composite categorical indicator across these disorders, that is, one that meets full criteria for any of these mood or disruptive disorders from T1 PAPA data. PAPA interrater reliability was monitored for 20% of all interviews (with 83%–100% agreement on symptom scores). At T2, symptom counts were derived from the Stonybrook Early Childhood Inventory (ECI).[27] There were high rates of psychopathology in this enriched sample, and all disruptive and mood disorders were associated with higher Temper Loss scores in bivariate analyses (Table 1).

Impairment Outcomes. The Family Life Impairment Scale (FLIS; Briggs-Gowan M, Horowitz S, Carter A. The Family Life Impairment Scale [unpublished rating scale], New Haven CT: Yale University; 1997) was administered at T1 and T2 to assess the extent to which children's emotions and behavior interfered with daily functioning, family functioning, and functioning at preschool/daycare on a 3-point scale (mean $\alpha = 0.75$, range = 0.63–0.81).

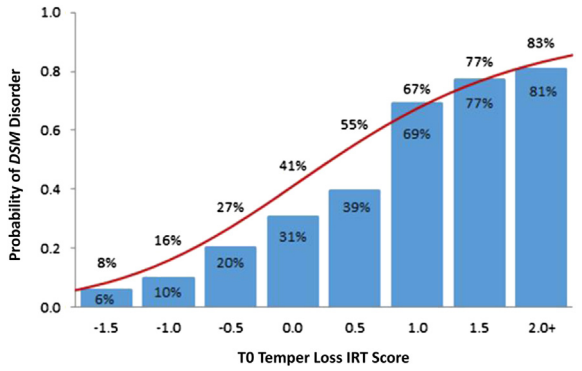
Analytic Plan Overview. Convergent and divergent validity analyses were conducted via multiple regressions controlling for child age, sex, race/ethnicity, and poverty status. The task-based model also controlled for nonverbal reasoning, task version, and pre-switch accuracy. Predictive clinical and incremental validity models added control for harsh physical discipline, child aggression, and the time interval between assessments. Incremental analyses also controlled for having any mood or disruptive disorders at T1. Predictive clinical and incremental validity analyses tested the shape (linear and quadratic) of the relation of T0 Temper Loss in relation to symptoms and impairment at subsequent time points (T1 and T2). A linear association indicates that increases in the clinical outcome occur equally across the full spectrum of Temper Loss scale scores, whereas a quadratic effect indicates that the association of the chemical outcome to Temper Loss depends on the value of Temper Loss. Linear and curvilinear patterns are tested simultaneously. When both are present, the linear pattern must be interpreted as conditional on the curvilinear. In the incremental analyses predicting impairment, T1 mood and disruptive disorders were controlled. We used structural equation models (SEM) to characterize longitudinal variability in Temper Loss across T0 to T2, and to test its incremental clinical validity. These longitudinal incremental analyses predicted T2 symptoms from T0 to T1 Temper Loss. Numbers (n) varied somewhat across analyses because of differential missing data across waves.

RESULTS

Establishment of the Validity of the Temper Loss Scale

Convergent/Divergent Validity. Temper loss was associated in expected directions with convergent and divergent measures. Temper Loss correlated with preschoolers' composite irritability symptoms scores ($\beta = .52$, $p < .0001$) and maternal PROMIS anger scores ($\beta = 0.37$, $p < .0001$). It was also associated with poorer response reversal on the Candy Game ($\beta = -0.11$, $p < .05$). Specifically, higher Temper Loss scores were associated with greater decrements in Candy Game performance after the rule switch, signifying poorer capacity to flexibly shift behavior in response to contextual cues. In contrast, Temper Loss was negatively associated

FIGURE 2 Relation of Temper Loss dimensional location to probability of *DSM* disorders. Note: Numbers below red line indicate percentage of children at each value of Temper Loss predicted to have a *DSM-IV* mood or disruptive disorder, based on Temper Loss score alone. Percentages above red line indicate predicted percentage for every value of T0 Temper Loss, controlling for all covariates. IRT = item response theory.



with indicators of self-regulation and competence, that is, the SSIS Self-Control scale ($\beta = -0.32$, $p < .0001$) and the DECA Initiative Scale ($\beta = -0.20$, $p < .01$).

Clinical Validity. In multivariate models, Temper Loss scale scores added significant variance to prediction of mood and disruptive symptoms. Temper Loss scale scores predicted ODD, ADHD, and SAD symptoms at both T1 and T2, and GAD and depression at T1. As shown in Table 2, the shape of the relationship between Temper Loss and symptoms varied by type of symptoms (for full models, see Table S3, available online). Specifically, for T1 outcomes, there were significant curvilinear patterns for ODD, depression, and SAD symptoms, whereas ADHD and GAD symptoms showed a linear association with Temper Loss (Figure 1). Figure 1a illustrates the quadratic effect for ODD symptoms; increases in ODD symptoms accelerated as scores increased on the Temper Loss scale. In contrast, ADHD symptoms increased relatively evenly regardless of Temper Loss values (Figure 1b). Although Temper Loss scale scores did not predict CD symptoms, aggression was highly predictive ($\beta = 0.34$, $p < .0001$). In addition to Temper Loss, the most consistent predictors of mood and disruptive symptoms were demographic risks (i.e., minority race/ethnicity, and poverty status). Temper Loss scale scores predicted T2 ODD, ADHD, and SAD symptoms.

Next, we also examined the shape of the relation between Temper Loss scores and the composite *DSM-IV* mood and disruptive disorders outcome at T1 (Table 2). The probability of *DSM-IV* disorders increased substantially as Temper Loss scores increased above the mean (Figure 2). Importantly, this increase in risk for subsequent *DSM-IV* disorders occurred even across levels of Temper Loss that are considered to be normative. For example, the probability of having a *DSM-IV* mood or disruptive disorder at T1 was 67% for children who were 1 SD from the T0 Temper Loss population mean, a level typically viewed at the upper bound of normal. This risk increased linearly across the dimension, with those who fell 2 SDs above the mean at T0 having an 83% probability of a T1 *DSM-IV* disorder.

Incremental Validity. T0 Temper Loss scores predicted higher FLIS impairment in all domains, above and beyond the presence of *DSM* disorders. The specific shape of the relation between Temper Loss scores and impairment varied by domain and time point (Table 2; see Table S3, available online, for full models). For example, child functional impairment demonstrated a curvilinear relationship to T0 Temper Loss scores at both T1 and T2.

Patterns of Longitudinal Variation. Across the 3 time points, short-term longitudinal correlations of Temper Loss scores were approximately .70 ($r\_T0–T1 = 0.70$, $r\_T0–T2 = 0.69$, $r\_T1–T2 = 0.71$). Although this is typically considered "very stable," we underscore that this coefficient indicates that only about half ($0.70^2 = 0.49$) of the variance in Temper Loss at any measurement occasion is shared by Temper Loss at a previous occasion. Thus, approximately 50% of the variance in Temper Loss scores may be due to development, measurement error, or other factors.

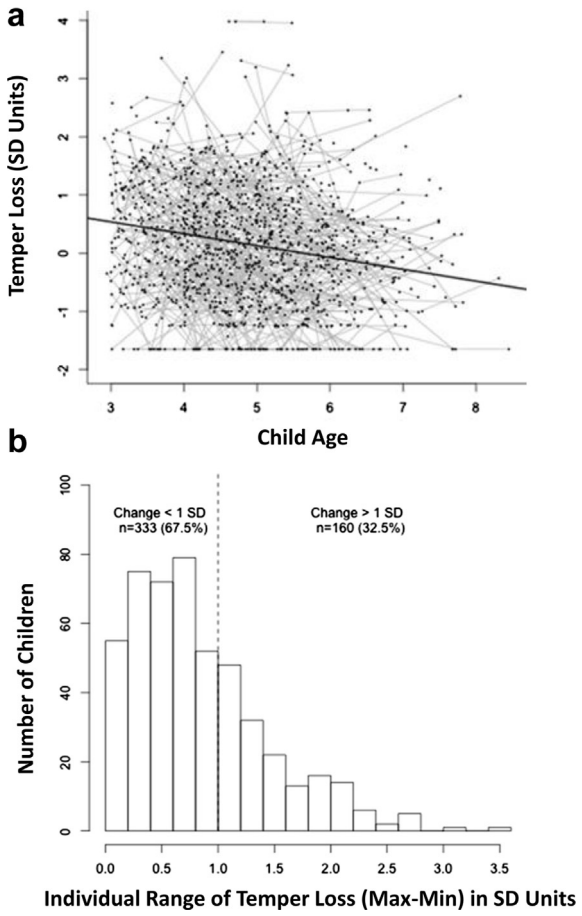
Longitudinal patterns for each individual child with data on at least 2 time points (n = 493) show considerable between- and within-subject variability around a small downward trend (0.2/SD per year), with a drop in variance at older ages (Figure 3a). Figure 3b shows the range (maximum–minimum score) for all children. Approximately one third (n = 160, 32.5%) of preschoolers have a set of scores that change more than 1 population SD across consecutive time points. (It should be noted that this reflects magnitude, not direction, of change.) Much of that change spans common thresholds for determining abnormality. For example, 6.5% (n = 32) of the sample exhibited Temper Loss levels >2 SD above the mean at least once, but nearly three-fourths of those children (23 of 32) were above this threshold at only 1 of 3 measurement occasions.

Contribution of Longitudinal Variation to Clinical Prediction. To test improvement in prediction when 2, rather than 1, time points of the Temper Loss scale were used, we tested the variance explained with only T0 temper loss in the model (as presented in Table 2), versus only T1 Temper Loss, versus both. In multivariate models, adding T1 Temper Loss to the models had significant effects on ODD, CD, and ADHD symptoms, above and beyond T0 effects (Table 3; for full model parameters, see Table S4, available online). In addition, T0 and T1 Temper Loss did not predict T2 GAD and SAD symptoms when considered individually, but were predictive when considered simultaneously. These findings indicate that accounting for change over time may enhance clinical prediction, perhaps by identifying periods with stronger effects and/or finding effects that cannot be detected at a single occasion because change per se is predictive.

DISCUSSION

Dimensional measurement of irritability has added value for short-term longitudinal prediction of clinical outcomes and

FIGURE 3 Longitudinal variation in Temper Loss score. Part 3a shows pattern of intraindividual change in Temper Loss score by child age. Part 3b shows variation in magnitude of intraindividual change in Temper Loss.



impairment in early childhood and is associated with theorized correlates. It is evident from these data that clinical risk is not an "either/or" phenomenon, and that children well below traditional clinical cut-points are at substantial clinical risk. Extreme (and sometimes arbitrary) cut-points may sacrifice important information by lumping together children below the cut-point who manifest substantial risk for clinical problems with children who are unlikely to develop problems. Consistent with the cross-cutting nature of irritability in psychopathology,[28,29] the Temper Loss scale and its longitudinal variation was strongly and uniquely predictive of symptoms of both mood and disruptive disorders.

What is the added value of assessing the dimensional spectrum of irritability in young children? First, Temper Loss provides unique information about emergent irritability relevant to cross-cutting syndromes. In addition, the use of continuous severity scores rather than symptoms provides unique information about variability. *DSM-IV* symptoms are designed to capture extreme manifestations of behavior that clearly demarcate clinical problems. This may be adequate for identifying children with severe problems, and for ruling out problems for children who are emotionally very well-regulated. However, there is increased consensus about the importance of identifying not just symptoms, but also prodromal patterns.[30] The Temper Loss scale identifies abnormality within a narrow developmental age band in terms of both unusual frequency (i.e., rare occurrence and high frequency of commonly occurring behaviors), as well as qualitatively atypical expression of behavior and emotions. In this way, dimensional patterns can identify children who have irritable tendencies that are not yet severely impairing but who have a significant probability of becoming impaired over time. Developmental specification of the boundaries between normal and abnormal is key for prodromal identification, particularly during early childhood, when the core behaviors that define irritability also occur normatively.[31]

These findings also highlight for the clinician the dynamic nature of irritability. Over only an average period of 16 months, approximately two-thirds of the children exhibited fairly stable irritability patterns, whereas even extreme irritability was transient in the other third. Characterizing such short-term variability lies at the heart of clinical prediction, and dimensional approaches provide a vital tool for addressing this need. The importance of this is further shown in models with two relatively closely spaced occasions of measurement, where effects varied across occasions or required multiple occasions to manifest. Even within a relatively short period of time, a substantial minority of children showed meaningful variation, and accounting for this variation enhanced clinical prediction. In some ways, examining changes over such relatively short periods is as important as examining longer time periods, because the relatively short time frame covered by the study establishes a short-term benchmark for the clinician. This short-term benchmark might define the boundaries of a meaningful "watch and wait" period for a clinician.

Findings suggest that a focus solely on the extreme end of the continuum will underestimate prodromal risk. However, a clinical challenge raised by identifying a dimensional spectrum of risk is that it brings to the fore a clinical gray area. This raises the question, "When is intervention warranted for young children at the boundaries of risk?" This murkiness is perhaps best highlighted by individual differences in the probability of having a *DSM-IV* disorder for preschoolers just a bit higher than average on the Temper Loss scale (1 SD). At this level of Temper Loss, two thirds of the children had a clinically significant disorder, but one third did not. Clearly, irritability is a complex trait that may or may not become impairing: What determines which way the clinical wind will blow for young children at this normal:abnormal boundary? This requires going beyond a single dimensional score for clinical decision making.

Empirically determining the factors that determine probabilistic risk for children at the mid-range of the irritability dimensional spectrum is the critical next step in research designed to advance clinical applications. This will require a neurodevelopmental profile approach[32] with key elements of dimensional clusters of behavior,

TABLE 3 Incremental Validity of Longitudinal Temper Loss Assessments (T0–T1) for Predicting T2 Symptoms and Impairment

	T0 Temper Loss (F)	T1 Temper Loss (F)	T0 and T1 Temper Loss (F)
T2 Clinical Symptoms			
ODD	0.51	9.27***	9.80***
CD	0.99	3.40*	2.90*
ADHD	0.87	3.37*	5.82***
Depression/ Dysthymia	1.35	1.75	1.36
GAD	1.89	2.53	2.70*
SAD	1.94	2.10	2.81*

Note: From three independent models controlling for child age, sex, race/ethnicity, poverty status, harsh physical discipline, child aggression, and time lag between visits. F values reflect Wald test with 2, 2, and 4 numerator degrees of freedom for tests of total T0, T1, and joint T0 and T1 tests, respectively. These Wald tests assess the total effect of Temper Loss (linear and quadratic) at any given time point (or set of time points), controlling for all other variables. For the T0 and T1 tests, each of these Temper Loss time points is also controlled. ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; GAD = generalized anxiety disorder; ODD = oppositional defiant disorder; SAD = separation anxiety disorder.
*$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

executive function, and longitudinal patterning. Clustering the co-variation of irritability with other salient dimensions of behavior (e.g., aggression, impulsivity, anxiety) and developmental competencies (e.g., social skills, language) is important for generating an integrated profile of developmental risk.[33,34] Atypicalities in prefrontal regions subserving executive function have been demonstrated in clinical populations of irritable youth,[28] and cognitive flexibility has been shown to buffer high irritable older youth from progressing to severe antisocial behavior (Hawes S, Perlman S, Byrd A, Raine A, Loeber R, Pardini D, unpublished data, 2014). In young children, conjoint consideration of irritable behavior and delays in maturation of executive function are theorized as a key explanatory factor for the clinical escalation of early irritability (Perlman S, Jones B, Wakschlag L, Axelson D, Birmaher B, Phillips M, unpublished data, 2014).[35] With regard to longitudinal variation, assessing irritability at a single time point is likely to contribute to both over- and under-identification. Empirical investigation is needed to determine the optimal number and spacing of time points to provide an adequate level of certainty in a manner that is also clinically feasible. This type of neurodevelopmental probabilistic assessment approach holds promise for differentiating children with moderate levels of temper loss at highest risk for clinical progression.

Because dimensional approaches do not provide clear thresholds for clinical decision making, a stepped framework may be needed to incorporate dimensional information clinically. We have previously suggested the following heuristic[12]: Level 1, Well-Regulated Emotions and Behaviors (with annual assessments); Level 2, Watch (longitudinally) and Wait; Level 3, Clinically at Risk (this level would be targeted to children with mildly elevated scores, e.g., 0.5–1 SD from the mean, to assess whether patterns escalate, remain stable, or diminish over the course of a year); and Level 4, low-intensity intervention: clinically prodromal (a level targeted for children with moderately elevated, e.g., >1 SD above the mean, scores at least 3 [e.g., bi-monthly] time points). Our data suggest that this subgroup of children have more than two thirds of a chance of having a clinical disorder. Thus, developmentally promoting, low-intensity interventions (e.g., improving self-regulation skills) may be warranted, as the benefits of preventing frank disorder are likely to far outweigh the costs. Finally, Level 5 would be Treatment: clinically significant. Highly elevated scores (e.g., ≥2 SD above the mean) at 2 time-points indicates clearly abnormal patterns that warrant in-depth assessment and treatment. The use of computer-adaptive test (CAT) approaches may provide a brief, efficient method of longitudinal monitoring for this purpose.[36,37] The importance of proximal family context as a buffer against clinical progression is well documented.[38,39] Although it is beyond the scope of the present paper, this will be another important future direction to incorporate into a multifaceted framework for probabilistic clinical decision making.

Our findings must be considered within the limitations of the present data set. First, shared method variance may have inflated associations, as mothers were reporters for most outcomes. However, this is representative of what typically happens in clinical assessments of preschoolers, in which the mother's report is often the only source of information. Second, to explicate the "shape" of dimensional patterns and clinical risk, we focused centrally on a single dimension, that is, irritability, although clearly such behaviors cannot be considered in isolation for clinical purposes. Third, our longitudinal follow-up was over a relatively short period (~16 months on average), and children's age at baseline and intervals across time points were not uniform. All of these limitations in longitudinal measurement impeded our ability to specify when change was due to development versus other factors (e.g., contextual changes).

The present findings speak to the need for an empirically validated multi-level dimensional assessment toolkit, aligned with the framework articulated by the National Institute of Mental Health Research Domain Criteria (RDoC).[40] With this in hand, clinicians will have an integrated way to assess and interpret dimensional patterns in conjunction with directly observed behavior and neurocognition. Studies conducted within this type of clinical–developmental framework will provide crucial data for determining how to meaningfully incorporate information on variations across the normal:abnormal spectrum to enhance early identification and clinical decision making. Such an approach holds promise for advancing a truly developmental understanding of clinical phenomenology, ontogeny, and course. &

Accepted June 8, 2015.

This article was reviewed under and accepted by ad hoc editor Shelli Avenevoli, PhD.

Dr. Wakschlag is with Department of Medical Social Sciences and the Institute for Policy Research, Northwestern University, IL. Drs. Estabrook, Petitclerc, and Voss, and Mr. Burns are with Northwestern University, Chicago. Dr. Henry is with Institute for Health Research and Policy, University of Illinois at Chicago. Dr. Perlman is with University of Pittsburgh, Pittsburgh, PA. Drs. Pine and Leibenluft are with the National Institute of Mental Health, Bethesda, MD. Dr. Briggs-Gowan is with University of Connecticut, Farmington.

Drs. Wakschlag, Estabrook, Henry, Voss, and Briggs-Gowan were supported by National Institute of Mental Health grants R01MH082830, UO1MH090301, and 2U01MH082830, and Dr. Perlman was supported by K01MH094467. Dr. Wakschlag was also supported by the Walden and Jean Young Shaw Foundation.

Dr. Estabrook served as the statistical expert for this research.

Drs. Wakschlag and Estabrook contributed equally to the article.

The authors thank Alice Carter, PhD, University of Massachusetts-Boston, Patrick Tolan, PhD, University of Virginia, Carri Hill, PhD, Jewish Children and Family Services, and Barbara Danis, PhD, Family Institute, for their contributions to the Multidimensional Assessment Profile of Disruptive Behavior (MAP-DB). This study would not have been possible without the outstanding efforts of Jacqueline Kestler, MPH, Erica Anderson, PhD, of Northwestern University, and their dedicated team, and the generous participation of the Multidimensional Assessment of Preschoolers (MAPS) Study families. The authors also thank the pediatric clinics and participants from Rush University, the University of Illinois at Chicago, NorthShore University HealthSystem, North Suburban Pediatrics, and the following pediatric research group practices for their participation: Healthlinc in Valparaiso, IN, Healthlinc in Michigan City, IN, and Associated Pediatricians in Valparaiso, IN.

Disclosure: Drs. Wakschlag, Estabrook, Petitclerc, Henry, Perlman, Voss, Pine, Leibenluft, Briggs-Gowan, and Mr. Burns report no biomedical financial interests or potential conflicts of interest.

Correspondence to Lauren Wakschlag, PhD, Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, 633 St. N. Clair Street, 19th Floor, Chicago, IL 60611; e-mail: lauriew@northwestern.edu

http://dx.doi.org/10.1016/j.jaac.2015.05.016

REFERENCES

1. Leibenluft E. Severe mood dysregulation, irritability, and the diagnostic boundaries of bipolar disorder in youths. Am J Psychiatry. 2011;168:129-142.
2. Stringaris A, Goodman R. Longitudinal outcome of youth oppositionality: irritable, headstrong, and hurtful behaviors have distinctive predictions. J Am Acad Child Adolesc Psychiatry. 2009;48:404-412.
3. Stringaris A, Zavoa H, Leibenluft E, Maughan B, Eley T. Adolescent irritability: phenotypic associations and genetic links with depressed mood. Am J Psychiatry. 2012;169:47-54.
4. Markon K, Krueger R, Watson D. Delineating the structure of normal and abnormal personality: an integrative hierarchical approach. J Pers Soc Psychol. 2005;88:137-157.
5. Deater-Deckard K, Wang Z. Anger and irritability. In: Zentner M, Shiner RL, eds. Handbook of Temperament. New York: Guilford Press; 2012:124-144.
6. Wakschlag L, Tolan P, Leventhal B. "Ain't misbehavin": towards a developmentally-specified nosology for preschool disruptive behavior. J Child Psychol Psychiatry. 2010;51:3-22.
7. Wakschlag L, Henry D, Tolan P, Carter A, Burns J, Briggs-Gowan M. Putting theory to the test: added value of a multidimensional, developmentally-sensitive approach to disruptive behavior in early childhood. J Am Acad Child Adolesc Psychiatry. 2012;51:593-604.
8. Belden A, Thompson N, Luby J. Temper tantrums in healthy versus depressed and disruptive preschoolers: defining tantrum behaviors associated with clinical problems. J Pediatr. 2008;152:117-122.
9. Dougherty LR, Smith VC, Bufferd SJ, *et al*. Preschool irritability: longitudinal associations with psychiatric disorders at age 6 and parental psychopathology. J Am Acad Child Adolesc Psychiatry. 2013;52:1304-1313.
10. Wakschlag L, Choi S, Carter AS, *et al*. Defining the developmental parameters of temper loss in early childhood: implications for developmental psychopathology. J Child Psychol Psychiatry. 2012;53:1099-1108.
11. Wakschlag L, Briggs-Gowan M, Choi S, *et al*. Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. J Am Acad Child Adolesc Psychiatry. 2014;53:82-96.
12. Bufferd S, Dyson M, Hernandez I, Wakschlag L. Explicating the "developmental" in preschool psychopathology. In: Cicchetti D, ed. Developmental Psychopathology. 3rd ed. Hoboken, NJ: Wiley; in press.
13. Kochanska G, Coy K, Murray K. The development of self-regulation in the first four years of life. Child Dev. 2001;72:1091-1111.
14. Wiggins J, Mitchell C, Stringaris A, Leibenluft E. Developmental trajectories of irritability and bidirectional associations with maternal depression. J Am Acad Child Adolesc Psychiatry. 2014;53:1191-1205.
15. Nichols S, Briggs-Gowan M, Estabrook R, *et al*. Punishment insensitivity in early childhood: a developmental, dimensional approach. J Abnormal Child Psychology. 2014 [Epub ahead of print].
16. Bentler PM. Comparative fit indexes in structural models. Psychology. 1990;107:238-246.
17. Tucker L, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Pschometrika. 1973;23:111-136.
18. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, eds. Testing Structural Equation Models. Beverly Hills, CA: Sage; 1993:136-162.
19. Reise S, Waller N. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009;5:25-46.
20. Egger HL, Erkanli A, Keeler G, Potts E, Walter B, Angold A. The test-retest reliability of the Preschool Age Psychiatric Assessment (PAPA). J Am Acad Child Adolesc Psychiatry. 2006;45:538-549.
21. Cella D, Yount S, Rothrock N, *et al*. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap Cooperative Group during its first two years. Med Care. 2007;45:S3-S11.
22. Crowley M, Mays LC. The Candy Reversal Learning Task. New Haven, CT: Yale Child Study Center; 2004.
23. Gorrindo T, Blair R, Budhani S, Dickstein D, Pine D, Leibenluft E. Deficits on a probabilistic response-reversal task in patients with pediatric bipolar disorder. Am J Psychiatry. 2005;1975-1977.
24. Gresham F, Elliot S. Social Skills Improvement System: Rating Scales Manual. Minneapolis, MN: Pearson; 2008.
25. LeBuffe PA, Nagalieri JA. The Devereux Early Childhood Assessment. Lewisville, NC: Kaplan Press Publishing; 1999.
26. Elliott C. Differential Abilities Scales: Introductory and Technical Handbook. New York: Psychological Corporation; 1983.
27. Gadow KD, Sprafkin J, Nolan EE. DSM-IV symptoms in community and clinic preschool children. J Am Acad Child Adolesc Psychiatry. 2001;40:1383-1392.
28. Stringaris A, Rowe R, Maughan B. Mood dysregulation across developmental psychopathology—general concepts and disorder specific expressions. J Child Psychol Psychiatry. 2012;53:1095-1097.
29. Leibenluft E, Stoddard D. The developmental psychopathology of irritabiltiy. Dev Psychopathol. 2013;25:1473-1487.
30. Insel T. Disruptive insights in psychiatry: transforming a clinical discipline. J Clin Investig. 2009;119:700-705.
31. Wakschlag LS, Choi S, Carter A, *et al*. Defining the developmental parameters of temper loss in young children: implications for developmental psychopathology. J Child Psychol Psychiatry. 2012;53:1099-1108.
32. Karalunas S, Fair DA, Musser E, Aykes K, Iyer S, Nigg J. Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: towards biologically based nosologic criteria. JAMA Psychiatry. 2014;71:1015-1024.
33. Beitchman JH, Wilson B, Brownlie EB, Walters H, Inglis A, Lancee W. Long-term consistency in speech/language profiles: II. Behavioral, emotional, and social outcomes. J Am Acad Child Adolesc Psychiatry. 1996;35:815-825.

34. Cote S, Vaillancourt T, Barker E, Nagin D, Tremblay R. The joint development of physical and indirect aggression: predictors of continuity and change during childhood. Dev Psychopathol. 2007;19:37-55.
35. Healey DM, Brodzinsky L, Bernstein B, Rabinovitz B, Halperin J. Moderating effects of neurocognitive abilities on the relationship between temperament and global functioning. Child Neuro Psychol. 2009; 16:20-31.
36. Choi SW, Reise SP, Pilkonis PA, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. Qual Life Res. 2010;19:125-136.
37. Wakschlag LS, Briggs-Gowan M, Choi S, *et al*. Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. J Am Acad Child Adolesc Psychiatry. 2014; 53:82-96.
38. Wakschlag L, Hans S. Relation of maternal responsiveness during infancy to the development of behavior problems in high risk youths. Dev Psychol. 1999;37:569-579.
39. Hill C, Maskowitz K, Danis B, Keenan K, Burns J, Wakschlag L. Validation of a clinically sensitive observational coding system for parenting behaviors: the Parenting Clinical Observation Schedule (P-COS) Parent Sci Pract. 2008;8:53–85.
40. Cuthbert BN. The RDoC framework: facilitiating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry. 2014;13:28-35.