

## Special Issue Article

# Parsing neurodevelopmental features of irritability and anxiety: Replication and validation of a latent variable approach

Elise M. Cardinale, Katharina Kircanski, Julia Brooks, Andrea L. Gold, Kenneth E. Towbin, Daniel S. Pine, Ellen Leibenluft and Melissa A. Brotman

Emotion and Development Branch, National Institute of Mental Health, Bethesda, MD, USA

### Abstract

Irritability and anxiety are two common clinical phenotypes that involve high-arousal negative affect states (anger and fear), and that frequently co-occur. Elucidating how these two forms of emotion dysregulation relate to perturbed neurodevelopment may benefit from alternate phenotyping strategies. One such strategy applies a bifactor latent variable approach that can parse shared versus unique mechanisms of these two phenotypes. Here, we aim to replicate and extend this approach and examine associations with neural structure in a large transdiagnostic sample of youth ( $N = 331$ ;  $M = 13.57$ ,  $SD = 2.69$  years old; 45.92% male). FreeSurfer was used to extract cortical thickness, cortical surface area, and subcortical volume. The current findings replicated the bifactor model and demonstrate measurement invariance as a function of youth age and sex. There were no associations of youth's factor scores with cortical thickness, surface area, or subcortical volume. However, we found strong convergent and divergent validity between parent-reported irritability and anxiety factors with clinician-rated symptoms and impairment. A general negative affectivity factor was robustly associated with overall functional impairment across symptom domains. Together, these results support the utility of the bifactor model as an alternative phenotyping strategy for irritability and anxiety, which may aid in the development of targeted treatments.

**Keywords:** anxiety, bifactor model, cortical structure, irritability, subcortical volume

(Received 15 November 2018; accepted 26 February 2019)

Emotion dysregulation occurs in many forms of pediatric psychopathology (Hyman, 2007; Morris & Cuthbert, 2012; Zald & Lahey, 2017), including irritability and anxiety, two common clinical phenotypes that involve high-arousal negative affect states (i.e., anger and fear; Brotman, Kircanski, Stringaris, Pine, & Leibenluft, 2017; Kircanski et al., 2018; Stoddard et al., 2014; Watson & Clark, 1984). Further, irritability and anxiety frequently co-occur in both clinically referred (Brotman et al., 2017) and community (Brotman et al., 2006; Copeland, Brotman, & Costello, 2015; Leadbeater & Homel, 2015; Savage et al., 2015; Stringaris & Goodman, 2009) samples. Evidence suggests that partly shared genetic mechanisms underlie irritability and anxiety (Savage et al., 2015). However, irritability and anxiety are also distinct; for example, they differ in the behavioral symptoms engaged (i.e., approach vs. avoidance of threats, respectively; Brotman et al., 2017; Kircanski et al., 2018; Leibenluft, 2017; Pine, 2007). Thus, attempts to relate these two forms of emotion dysregulation to perturbed neurodevelopment may benefit from alternate phenotyping strategies. Here, we replicated and extended a previously published bifactor model of irritability and anxiety (Kircanski

et al., 2018), investigating their common versus specific associations with brain structure, development, and clinical features.

There is limited research on brain structure across pediatric irritability and anxiety. One prior study compared neuroanatomical correlates of diagnostically defined anxiety (anxiety disorders) and severe irritability (disruptive mood dysregulation disorder; DMDD; Gold et al., 2016). Diagnosis-specific patterns emerged in prefrontal cortex (PFC) gray matter volume (GMV), such that larger GMV was associated with anxiety disorders, whereas smaller GMV was associated with DMDD. Studies of neural structure in either phenotype alone have reported abnormal ventromedial PFC thickness (Gold et al., 2017; Newman et al., 2015; Strawn et al., 2015) and hippocampal volume (Gold et al., 2017; Mueller et al., 2013) in anxiety, and abnormal dorsal lateral PFC GMV in irritability (Adleman et al., 2012; Dickstein et al., 2005). However, these studies also found decreased amygdala volume in both irritability (Dickstein et al., 2005) and anxiety (Milham et al., 2005; Newman et al., 2015; Strawn et al., 2015). Of note, these studies used categorical approaches and targeted youths with primary diagnoses involving either irritability or anxiety symptoms. Dimensional approaches have advantages when attempting to detect common versus specific neural correlates.

Together, the evidence for both shared and distinct abnormalities suggests that irritability and anxiety symptoms likely reflect both shared and unique underlying mechanisms of emotion dysregulation. Thus, transdiagnostic methods that facilitate such

**Author for Correspondence:** Elise M. Cardinale, National Institute of Mental Health, 9000 Rockville Pike, Building 15K, MSC 2670, Bethesda, MD 20892-2670; E-mail: [elise.cardinale@nih.gov](mailto:elise.cardinale@nih.gov).

**Cite this article:** Cardinale EM, Kircanski K, Brooks J, Gold AL, Towbin KE, Pine DS, Leibenluft E, Brotman MA (2019). Parsing neurodevelopmental features of irritability and anxiety: Replication and validation of a latent variable approach. *Development and Psychopathology* 31, 917–929. <https://doi.org/10.1017/S095457941900035X>

parsing of these mechanisms in early life may be crucial to better understanding emotion dysregulation as a transdiagnostic vulnerability. This requires strategies that allow researchers to assess irritability and anxiety dimensionally across diagnoses (Insel *et al.*, 2010) and then parse the phenotypes' shared versus unique features, an approach that has been used for other measures of psychopathology (Castellanos-Ryan *et al.*, 2014; Shanmugan *et al.*, 2016). Recently, we used such an approach; we performed a bifactor analysis to first quantify the unique and shared variances of dimensionally assessed irritability and anxiety before next relating them to neural function on a threat-orienting functional magnetic resonance imaging (MRI) task (Kircanski *et al.*, 2018). This approach revealed a double dissociation in which only irritability was associated with neural activation in multiple regions (e.g., PFC, striatum, and amygdala), whereas only anxiety was associated with functional connectivity of the amygdala. The shared component of irritability and anxiety—a general propensity to negative affectivity—was associated with increased activity in the thalamus. These distinct neural correlates were not found using a diagnostic approach.

These findings suggest broader use for this bifactor model in identifying differential early neurodevelopment of brain regions implicated in shared versus specific features of irritability and anxiety. However, it is unknown whether this model holds across key demographic variables and whether the estimated underlying constructs map onto external measures of irritability and anxiety symptoms (i.e., show convergent validity with clinician ratings). Furthermore, this approach has yet to be applied to the investigation of structural neural abnormalities associated with irritability and anxiety. Here, in a large, transdiagnostic sample of children and adolescents, we aim to replicate and extend the bifactor model of irritability and anxiety. First, we examine how the model performs with respect to important developmental variations: age and sex. Second, we relate scores derived from the bifactor model (irritability, anxiety, and general negative affectivity) to brain cortical thickness, cortical surface area, and subcortical volume. Consistent with the literature described above, we predicted distinct associations of anxiety with ventromedial PFC and hippocampal structure, irritability with dorsal lateral PFC structure, and negative affectivity with decreased amygdala volume. Third, we examine how factor scores relate to independent clinician assessments of symptoms and impairment. We predicted that irritability and anxiety scores would be specifically associated with clinician-rated irritability and anxiety, respectively, whereas negative affectivity scores would capture global impairment.

## Method

### Participants

Three hundred and thirty-one youth ( $M = 13.57$ ,  $SD = 2.69$  years old; 45.92% male) participated in the current study (Table 1). A transdiagnostic sample was recruited to capture full ranges of both irritability and anxiety symptoms. Specifically, the sample included participants with a primary diagnosis of DMDD (i.e., a mood disorder characterized by severe, chronic irritability;  $N = 70$ ; American Psychiatric Association, 2013; Leibenluft, 2017), an anxiety disorder (ANX;  $N = 95$ ; generalized, separation, and/or social anxiety disorders), attention-deficit/hyperactivity disorder (ADHD;  $N = 39$ ), or no psychiatric diagnosis (healthy volunteer; HV;  $N = 127$ ). We included individuals with a primary

diagnosis of ADHD because within this age group, ADHD is strongly associated with chronic irritability and often comorbid with DMDD (Shaw, Stringaris, Nigg, & Leibenluft, 2014). These groups were specifically recruited to collectively span youth with clinically significant irritability and/or anxiety, youth with sub-threshold levels of irritability and/or anxiety, and healthy youth with normative levels of irritability and/or anxiety. Diagnoses were made using the Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime version (KSADS-PL), administered by a doctoral- or master's-level clinician. All clinicians were trained to have high interrater reliability such that clinical determinations were reliable with other trained clinicians as assessed by a  $\kappa > .75$  (Wiggins *et al.*, 2016). Critically, all diagnoses were reviewed and confirmed by a licensed psychiatrist or senior clinical psychologist in a consensus meeting (K.E.T., D.S.P., E.L., or M.A.B.). These procedures are consistent with those used in previously published work (Gold *et al.*, 2016; Kircanski *et al.*, 2018; Wiggins *et al.*, 2016). Only patients who met full criteria for an anxiety, ADHD, or DMDD diagnosis were included. Exclusionary criteria included  $IQ < 70$ , pervasive developmental disorder, posttraumatic stress disorder, schizophrenia, bipolar disorder, substance use (within 3 months of participation), neurological disorder, or unstable medical illness. In addition, participants with a primary anxiety disorder had no history of posttraumatic stress disorder or current depression and were medication free. The sample included 20 sibling pairs (2 DMDD, 6 ANX, 8 ADHD, and 24 HV participants). All sibling pairs comprised participants within the same diagnostic group, with the exception of three sibling pairs (ANX & HV, ADHD & HV, and ANX & ADHD). Of the 331 youth, 90 participants overlap with the sample included in Gold *et al.* (2016), 106 participants overlap with the sample included in Gold *et al.* (2017), and 135 participants overlap with the sample included in Kircanski *et al.* (2018). For our confirmatory bifactor model analyses, we repeated all analyses in a fully independent sample ( $n = 196$ ; described below in the Results). Participants were recruited through advertisements in the community. Youth received monetary compensation for participation. Prior to participation, parents provided written informed consent and youth provided written assent. All study procedures were approved by the National Institute of Mental Health Institutional Review Board.

### Symptom measures for bifactor model

#### Affective Reactivity Index

Irritability symptoms were assessed dimensionally using the parent- and youth-report versions of the Affective Reactivity Index (ARI; Stringaris *et al.*, 2012). Across all participants, both parent-report ( $M = 2.95$ ,  $SD = 3.42$ ) and youth-report ( $M = 2.31$ ,  $SD = 2.70$ ) scores ranged from 0 to 12, encompassing the full range of possible scores. Parent- and youth-report scores were significantly correlated,  $r(331) = .52$ ,  $p < .001$ .

#### Screen for Child Anxiety Related Emotional Disorders

Anxiety symptoms were assessed dimensionally using the parent- and youth-report versions of the Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher *et al.*, 1997). Across all participants, parent-report scores ranged from 0 to 75 ( $M = 16.33$ ,  $SD = 15.68$ ) and youth-report scores ranged from 0 to 64 ( $M = 17.48$ ,  $SD = 14.28$ ). Parent- and youth-report scores were significantly correlated,  $r(331) = .61$ ,  $p < .001$ .

**Table 1.** Parsing neurodevelopmental features of irritability and anxiety: Replication and validation of a latent variable approach

Sample characteristics	Total sample	Diagnostic group <sup>1</sup>				Omnibus Sig
		ADHD	ANX	DMDD	HV	
Demographic variables						
<i>N</i>	331	39 (11.78%)	95 (28.92%)	70 (20.78%)	127 (38.55%)	
Age (years)	13.56 (2.68)	13.66 (2.70) <sup>ab</sup>	12.86 (2.88) <sup>a</sup>	13.79 (2.69) <sup>ab</sup>	13.92 (2.88) <sup>b</sup>	$F(3, 327) = 3.14, p = .03$
Sex (male)	152 (45.92%)	27 (69.23%) <sup>a</sup>	34 (35.42%) <sup>b</sup>	42 (60.87%) <sup>a</sup>	49 (38.28%) <sup>b</sup>	$\chi^2(3) = 20.80, p < .001$
IQ	112.02 (12.56)	113.10 (10.69)	113.45 (14.18)	110.79 (13.58)	111.28 (11.14)	$F(3, 327) = 0.88, p = .45$
Clinical variables						
Symptom ratings						
ARI parent report	2.95 (3.42)	3.21 (3.24) <sup>a</sup>	3.36 (3.22) <sup>a</sup>	6.60 (2.95) <sup>b</sup>	0.55 (1.25) <sup>c</sup>	$F(3, 327) = 84.46, p < .001$
ARI youth report	2.31 (2.70)	2.79 (2.67) <sup>a</sup>	2.21 (2.32) <sup>ab</sup>	4.04 (3.18) <sup>b</sup>	1.03 (1.79) <sup>c</sup>	$F(3, 327) = 24.34, p < .001$
SCARED parent report	16.33 (15.68)	11.73 (12.78) <sup>a</sup>	30.29 (12.66) <sup>b</sup>	22.63 (14.66) <sup>a</sup>	3.81 (4.86) <sup>c</sup>	$F(3, 327) = 117.18, p < .001$
SCARED youth report	17.48 (14.28)	13.99 (9.74) <sup>a</sup>	29.14 (13.85) <sup>b</sup>	19.26 (14.65) <sup>c</sup>	8.84 (7.83) <sup>d</sup>	$F(3, 327) = 56.99, p < .001$
CDI youth report <sup>2</sup>	7.38 (7.49)	6.19 (5.52) <sup>ac</sup>	10.99 (8.24) <sup>b</sup>	9.92 (8.55) <sup>bc</sup>	3.70 (4.46) <sup>a</sup>	$F(3, 321) = 45.82, p < .001$
Medications <sup>3</sup>						
SSRI	35 (10.57%)	5 (12.82%) <sup>a</sup>	0 (0%) <sup>b</sup>	26 (37.14%) <sup>c</sup>	0 (0%) <sup>b</sup>	$p < .001^4$
Stimulant	52 (15.71%)	19 (48.72%) <sup>a</sup>	0 (0%) <sup>b</sup>	30 (42.86%) <sup>a</sup>	0 (0%) <sup>b</sup>	$p < .001^4$
SGA	11 (3.32%)	1 (2.56%) <sup>a</sup>	0 (0%) <sup>a</sup>	8 (11.43%) <sup>b</sup>	0 (0%) <sup>a</sup>	$p < .001^4$
AED	18 (5.44%)	0 (0%) <sup>a</sup>	0 (0%) <sup>a</sup>	18 (25.71%) <sup>b</sup>	0 (0%) <sup>a</sup>	$p < .001^4$

Note: Cells marked with different superscript letters indicate significant group differences across diagnostic groups. ADHD, attention-deficit/hyperactivity disorder. ANX, primary anxiety disorder. DMDD, disruptive mood dysregulation disorder. HV, healthy volunteer. ARI, Affective Reactivity Index. SCARED, Screen for Child Anxiety Related Disorders. CDI, Children's Depression Inventory. SSRI, selective serotonin reuptake inhibitor. SGA, second-generation antipsychotic. AED, antiepileptic drugs. <sup>1</sup>Primary diagnosis refers to the diagnosis for which the participant was referred. Participants could have multiple diagnoses in addition to their primary diagnosis. <sup>2</sup>CDI scores were missing for six participants (4 ANX and 2 HV). <sup>3</sup>Percentage of participants taking each medication type is reported. Participants could be taking more than one type of medication. <sup>4</sup>Omnibus significance calculated using Freeman Halton Test.

### Clinician-rated measures

#### Children's Global Assessment Scale

Overall severity of impairment due to psychiatric symptoms was assessed by clinicians using the Children's Global Assessment Scale (CGAS; Shaffer, 1983). CGAS scores range from 1 to 100, with lower scores reflecting greater impairment. CGAS scores were obtained on the DMDD, ANX, and ADHD groups only, including 176 participants of the 204 total participants in these groups. There was no significant association between missingness and diagnostic group,  $\chi^2(2) = 0.50, p = .78$ . Patients exhibited a wide range of CGAS scores, from 31 (*major impairment*) to 89 (*good functioning*;  $M = 55.29, SD = 11.49$ ).

#### Clinician-rated irritability

Clinician-rated irritability symptoms were assessed via the DMDD and severe mood dysregulation modules of the KSADS-PL. Using either module, DMDD symptom criteria were completed for 79 of the 108 total participants across the DMDD and ADHD groups. Specifically, temper outbursts ( $M = 2.28, SD = 0.80$ ) and irritable mood ( $M = 2.24, SD = 0.87$ ) were rated as 1 (*absent*), 2 (*exhibited at subthreshold levels*), or 3 (*exhibited at clinical levels*). Impairment in social ( $M = 2.23, SD = 1.09$ ), family ( $M = 3.01, SD = 1.20$ ), and school ( $M = 2.10, SD = 1.19$ ) contexts was also assessed using a scale of 1 (*absent*) to 4 (*severe*). A composite impairment score was created by averaging scores across social, family, and school contexts ( $M = 2.45, SD = 0.99$ ).

#### Pediatric Anxiety Rating Scale

Clinician-rated anxiety symptoms were assessed using the Pediatric Anxiety Rating Scale (PARS; Research Units On Pediatric Psychopharmacology Anxiety Study Group, 2002). PARS scores were completed for 89 of the 95 total participants in the ANX group. Scores ranged from 2 to 24 ( $M = 15.30, SD = 4.58$ ). Specific levels of anxiety-related impairment with family ( $M = 2.78, SD = 1.12$ ) and peers ( $N = 88; M = 2.80, SD = 1.23$ ) were assessed. A total impairment score was calculated by averaging the family and peer impairment scores ( $M = 2.80, SD = 1.02$ ).

### Imaging procedures

For each participant, a high-resolution T1-weighted magnetization-prepared rapid-acquisition gradient-echo scan (sagittal acquisition, repetition time = 7.7 ms, echo time = 3.42 ms, slices = 176  $1\text{ mm}^3$  isotropic voxels, matrix =  $256 \times 256$ , flip angle = 7 degrees) was acquired on a General Electric 3-T MR750 scanner (Waukesha, WI, USA) with a 32-channel head coil. All images were first processed using standard procedures of Version 5.3.0 of FreeSurfer image analysis software suite (Dale, Fischl, & Sereno, 1999; Fischl, 2004; Fischl et al., 2002). During the FreeSurfer cortical reconstruction process, nonbrain tissue is first removed (Ségonne et al., 2004). Images undergo intensity normalization (Sled, Zijdenbos, & Evans, 1998), the core of the brain white matter is segmented, and eventual holes are filled. For each exposed voxel face of the segmented white matter, two triangles are created, thus defining an initial mesh. This mesh is further smoothed, defects are corrected, and its final placement takes into consideration gradients in voxel intensities between gray and white matter (Dale et al., 1999; Fischl & Dale, 2000). The pial surface is defined by nudging outward this surface until maximal tissue contrast is found. Neuroanatomical labels are automatically assigned to each subcortical voxel based on

probabilistic information acquired through an a priori knowledge of spatial relationships acquired through a manually labeled training set (Fischl et al., 2002; Fischl, Salat, et al., 2004), and cortical regions are identified on the surfaces based on the gyral and sulcal structure of the cortical surface (Desikan et al., 2006; Fischl, van der Kouwe, et al., 2004). All processed images were visually inspected for image artifacts. A total of 462 images were processed and 129 were excluded following visual inspection, mostly due to motion artifacts or poor segmentation (17 ADHD patients, 39 anxious patients, 30 DMDD patients, and 43 HVs).

For the current study, we focused on measures of cortical thickness, cortical surface area, and subcortical volume. Previous work examining FreeSurfer methods in the calculation of these measures has established strong reliability across groups, MRI scanners, and field strengths (Han et al., 2006; Morey et al., 2010; Reuter, Schmansky, Rosas, & Fischl, 2012) and strong validity through comparison with histological analyses (Rosas et al., 2002) and manual tracing (Grimm et al., 2015; Kuperberg et al., 2003; Morey et al., 2009; Schoemaker et al., 2016). These methods are consistent with prior psychiatric research examining cortical thickness, cortical surface area, and subcortical volume within pediatric samples (Cardinale et al., 2018; Gold et al., 2017; McLaughlin et al., 2014; Ostby et al., 2009; Sheridan, Fox, Zeanah, McLaughlin, & Nelson, 2012; Strawn et al., 2014; Sylvester et al., 2016). Following processing in FreeSurfer, all images were smoothed using a 20-mm full width at half maximum Gaussian filter and resampled to 10,242 vertices per hemisphere (based on 5 recursive subdivisions of a regular icosahedron). Resampling to a common grid uses the nearest neighbor method (Winkler, Greve, Bjuland, et al., 2018), which is the default in FreeSurfer. Finally, left and right hemispheres were merged so the analyses could be conducted across the whole brain. Images were masked to include in the analyses only the cortical surface. Subcortical regions were automatically segmented and labeled based on probabilistic information acquired through an a priori knowledge of spatial relationships acquired through a manually labeled training set (Fischl et al., 2002; Fischl, Salat, et al., 2004).

### Data analysis

#### Bifactor model of irritability and anxiety

**Confirmatory factor analysis.** Confirmatory factor analysis was employed to test the bifactor model of irritability and anxiety (Kircanski et al., 2018). Consistent with the original model, we included the six items each comprising the parent- and youth-report ARI scores and the five subscales each comprising the parent- and youth-report SCARED scores, for a total of 22 data points for each participant. ARI item scores were modeled as categorical variables and SCARED subscale scores were modeled as continuous variables. Data were analyzed using Mplus (Version 7.4). Analyses used the weighted least square mean and variance adjusted estimator given the inclusion of categorical variables in the model. There were no missing data.

**Measurement invariance.** Measurement invariance of the bifactor model as a function of participant age or sex was assessed using multiple group structural equation modeling in Mplus (Version 7.4). Age groups were defined using a median split (13.55 years old), resulting in a younger group ( $M = 11.28, SD = 1.38$ ) and an older group ( $M = 15.85, SD = 1.43$ ). Sex groups were defined as self/parent-identified male and female. Separately for age and

sex, we first assessed configural invariance in which the same pattern of fixed and free parameters was defined across groups, but there were no equality constraints for the loadings of observed variables on the latent factors across groups. These models assessed whether the overall structure of the bifactor model was similar across the groups. Next, we tested for weak factorial invariance, in which factor loadings were constrained to be equal across the groups. Absolute model fit statistics—comparative fit index (CFI); nonnormed fit index (NNFI); root mean square error of approximation (RMSEA); and a chi-square difference test (adjusted for the weighted least square mean and variance adjusted estimator)—were used to compare the absolute and relative fit, respectively, of the models testing configural invariance versus weak factorial invariance.

**Imaging analyses.** Associations between latent factor loadings extracted from the bifactor model and measures of brain structure were conducted using a nonparametric permutation approach as leveraged by the computation tool Permutation Analysis of Linear Models (PALM; Winkler, Ridgway, Webster, Smith, & Nichols, 2014). Using PALM, we tested for linear associations between each of the factor loadings resulting from the bifactor analyses with cortical thickness, cortical surface area, and subcortical volume. For analyses of cortical thickness and surface area, associations with factor loadings were assessed at each vertex. For analyses of subcortical volume, associations with factor loadings were assessed using measures of regional volume for each subcortical structure (including thalamus, caudate, putamen, pallidum, hippocampus, amygdala, and accumbens).

Initial analyses included factor loadings on each of the four factors entered as predictors of cortical thickness, surface area, and subcortical GMV with age, IQ, and sex as covariates. In order to probe for moderating effects of age and sex, we conducted a regression analysis that included the interaction between each of the four factor loadings with both age and sex as predictors of cortical thickness, surface area, and subcortical GMV. Significance was assessed based on the permutation distribution of maximum test statistic across the whole brain, across modalities (thickness and area), and across the associations tested (contrasts), thus allowing familywise error rate correction across subcortical regions and contrasts of the subcortical analyses and cortical vertices, modality, and contrasts for cortical analyses (Winkler et al., 2016).

PALM is a computational tool that leverages nonparametric methods to conduct classical uni- and multivariate analyses using permutation approaches. Whereas parametric methods require number of assumptions to be met (such that the data are independent and identically distributed, and following the normal distribution), the use of nonparametric methods requires fewer assumptions to be met (Holmes, Blair, Watson, & Ford, 1996; Nichols & Holmes, 2002).

For analyses of cortical thickness and area, the data were randomly shuffled 2,000 times, and inference was accelerated by the fitting of a generalized Pareto distribution to the tail of the permutation distribution (Winkler, Ridgway, Douaud, et al., 2016). For the analysis of subcortical volumes, speed was a lesser concern due to the lower dimensionality of the data, and 10,000 shufflings were performed. For each permutation, the general linear model was fit and a test statistic for each contrast was calculated. For thickness and area in the current study, we used the threshold free cluster enhancement on the surfaces as the test statistic for each vertex (Smith & Nichols, 2009). To control for

**Table 2.** Bivariate correlations among total scores on the ARI and SCARED

	1	2	3	4
1. Parent-report ARI	—			
2. Youth-report ARI	0.52***	—		
3. Parent-report SCARED	0.52***	0.41***	—	
4. Youth-report SCARED	0.32***	0.46***	0.61***	—

Note: ARI, Affective Reactivity Index. SCARED, Screen for Child Anxiety Related Disorders. \*\*\* $p < .001$ .

nonindependence of sibling measures of neural structure, sibling status was coded using the multilevel exchangeability blocks option in PALM such that during shuffling of data, sibling pairs were shuffled together to maintain the dependence across sibling pairs across permutations (Winkler, Webster, Vidaurre, Nichols, & Smith, 2015). In addition, during visual inspection of the FreeSurfer images, quality of the images that met our quality control inclusion criteria were coded using a 3-point Likert scale with 1 being the highest quality images and 3 being the lowest quality images. Image quality ratings were entered into our PALM analyses as a variance group whereby shuffling of the data also took into consideration potential distribution differences related to different image quality across the sample (Winkler et al., 2015).

#### Associations with clinician-rated measures

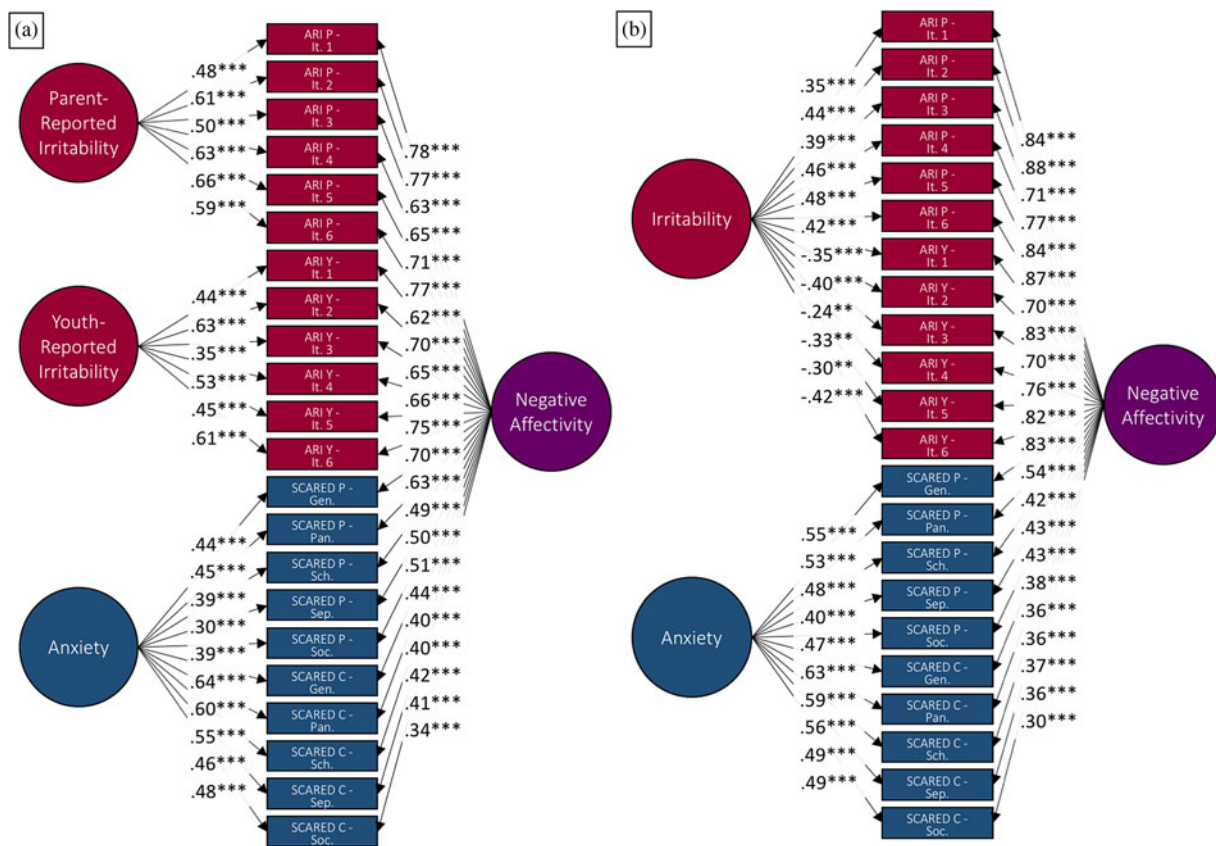
Associations with the independent clinician-rated measures were examined using multiple regression, with scores on the latent variables concurrently predicting clinician ratings. For all models, age, sex, and IQ were entered as covariates. Analyses of age and sex as moderators of associations with clinician ratings were conducted using the PROCESS macro in SPSS (Hayes, 2016). For each model, scores on the latent variables, and their interactions with age and sex, were predictors of each of the clinician ratings. IQ was included as a covariate. Post hoc interrogation of significant interactions was probed using the Johnson–Neyman procedure.

## Results

### Bifactor analyses

#### Bifactor model of irritability and anxiety

**Confirmatory factor analysis.** In the full sample, total scores on the ARI and SCARED were correlated (Table 2;  $r_s = .32-.61$ ,  $p_s < .001$ ). Consistent with our previous findings (Kircanski et al., 2018), an initial bifactor model including one unique irritability factor resulted in negative loadings for the ARI youth-report items and positive loadings for the ARI parent-report items on this factor (Figure 1). Therefore, the final model included separate unique factors for parent- and youth-reported irritability, a unique factor for anxiety (both parent and youth reported), and a general factor, again termed *negative affectivity*. Results indicated adequate fit of this model to the data (factor loading  $p_s < .001$ ; CFI = .973; NNFI = .966; RMSEA = .072; 90% confidence interval, CI [.065, .080]; Figure 1). Given that a proportion of our sample overlapped with the sample examined in our previously published work, we repeated these analyses in a fully independent sample ( $n = 196$ ). Results indicated adequate fit of the



**Figure 1.** (a) Initial bifactor model including only one unique irritability factor and (b) best fit bifactor model with unique factors of parent-reported irritability, youth-reported irritability, and anxiety, as well as a common factor of negative affectivity. Each path includes the factor loadings for each observed variable on each latent variable. ARI, Affective Reactivity Index. It, item. SCARED, Screen for Child Anxiety Related Emotional Disorders. Gen, generalized anxiety disorder subscale. Pan, panic disorder subscale. Sch, school avoidance subscale. Sep, separation anxiety disorder subscale. Soc, social anxiety disorder subscale. Y, youth report. P, parent report. \*\*\* $p < .001$ .

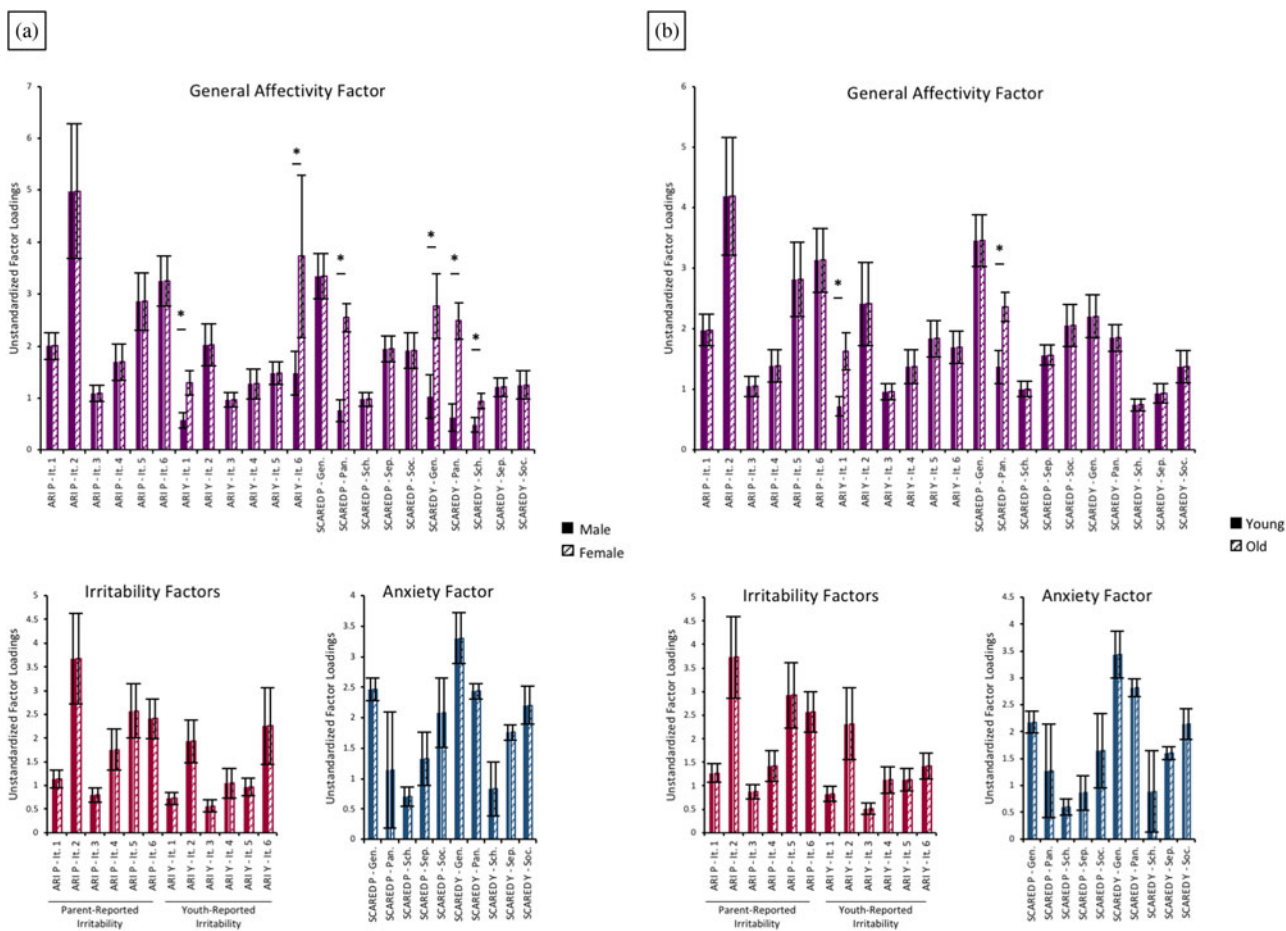
bifactor model to the data even within this restricted and fully independent sample (factor loading  $ps < .001$ ; CFI = .976; NNFI = .971; RMSEA = .067; 90% CI [.056, .078]). Finally, participants' scores on the factors were extracted to examine in relation to brain structure and independent clinical measures.

*Measurement invariance*

**Age.** Analyses first tested for configural invariance in which there were no equality constraints for the loadings of observed variables on the latent variables across age groups. Results indicated adequate fit (CFI = .960; NNFI = .953; RMSEA = .082; 90% CI [.074, .090]), such that for both younger and older participants the bifactor model fit the data well. Analyses next tested for weak factorial invariance in which factor loadings were constrained to be equal across age groups. Similarly, results indicated adequate fit, and absolute model fit statistics supported strong fit of this more constrained model (CFI = .983; NNFI = .982; RMSEA = .051; 90% CI [.042, .060]). However, a chi-square difference test comparing the two models was significant,  $\chi^2(44) = 67.08, p = .02$ , indicating that the weak factorial model had poorer relative fit based on this index. Supplementary analyses utilizing model modification indices indicated that removing constraints for two factor loadings (SCARED parent-report panic subscale and ARI youth-report Item 1 [“I am easily annoyed by others”] on the negative affectivity latent factor) resulted in a nonsignificant chi-square difference test (CFI = .984; NNFI = .983;

RMSEA = .049; 90% CI [.039, .058]);  $\chi^2(42) = 55.48, p = .08$ . In both the younger and older age groups, the two factor loadings in question remained significant, but were even higher in older participants than younger participants (Figure 2). Thus, overall the results were interpreted to support similar factor loadings across age groups.

**Sex.** As with age groups, analyses first tested for configural invariance across sexes. Results revealed adequate model fit (CFI = .964; NNFI = .958; RMSEA = .080; 90% CI [.072, .088]) such that for both male and female participants the bifactor model fit the data well. Analyses next tested for weak factorial invariance. Similarly, results indicated adequate fit, and absolute model fit statistics supported strong fit of this more constrained model (CFI = .972; NNFI = .970; RMSEA = .067; 90% CI [.059, .075]). However, the chi-square difference test was significant,  $\chi^2(44) = 121.82, p < .001$ , indicating that the weak factorial model had poorer relative fit. Supplementary analyses utilizing model modification indices indicated that removing constraints for six factor loadings (SCARED parent- and youth-report panic subscale, SCARED youth-report generalized anxiety subscale, SCARED youth-report school anxiety subscale, and ARI youth-report Items 1 [“I am easily annoyed by others”] and 6 [“I lose my temper easily”] on the negative affectivity latent factor) resulted in a nonsignificant chi-square difference test (CFI = .983; NNFI = .981; RMSEA = .053; 90% CI [.044, .062]);  $\chi^2(38) = 51.45,$



**Figure 2.** Unstandardized factor loadings resulting from weak factorial model analyses of the effects of (a) sex and (b) gender. \*Indicate loadings that when freed resulted in nonsignificant chi-square difference test between configural and weak factorial models.

$p = .07$ ). In both the male and female groups, the five factor loadings in question remained significant, but were even higher in female participants than male participants (Figure 2). Again, these changes were not substantive, and thus we interpreted the results to support similar factor loadings across sexes.

**Associations with neural structure**

*Cortical thickness and surface area*

We observed no significant associations between scores on any of the four latent factors and measures of cortical thickness and surface area. Similarly, our analyses of age and sex as potential moderators revealed no significant interactions between age and sex with scores on the four factors in predicting cortical thickness and surface area. All results were consistent across correction methods, including the least stringent correction that did not correct for the number of contrasts and modalities tested.

*Subcortical volume*

Consistent with analyses of cortical thickness and surface area, we observed no significant associations between scores on any of the four latent factors with any measure of subcortical volume. Results of analyses examining age and sex as potential moderators again revealed no significant interactions between age and sex with scores on the four factors in predicting subcortical volume. All results were consistent across correction methods, including

our least stringent correction, which did not correct for the number of contrasts and modalities tested.

**Associations with clinician-rated symptoms and impairment**

*Irritability symptoms and impairment*

Associations between factor scores and clinician-rated irritability symptoms and impairment were assessed via the DMDD and severe mood dysregulation modules of the KSADS-PL. Higher negative affectivity factor scores and higher parent-reported irritability factor scores were associated with higher clinician-rated temper outbursts: negative affectivity,  $\beta = 0.60, t(71) = 5.65, p < .001$ ; parent-reported irritability:  $\beta = 0.35, t(71) = 3.45, p = .001$ , and irritable mood: negative affectivity,  $\beta = 0.68, t(71) = 6.50, p < .001$ , parent-reported irritability,  $\beta = 0.31, t(71) = 3.06, p = .002$ . There were no significant associations between anxiety or youth-reported irritability factor scores and temper outbursts. In addition, higher negative affectivity scores were associated with higher total impairment scores,  $\beta = 0.79, t(65) = 6.94, p < .001$ , whereas higher youth-reported irritability scores were associated with lower total impairment scores,  $\beta = -0.27, t(65) = -2.67, p = .01$ . A similar pattern of findings emerged for associations with impairment in social, school, and family context. Additional associations between the parent-reported irritability and the anxiety factor with impairment in the family context were observed. Higher negative affectivity factor scores were associated with increased impairment due to

irritability symptoms in social and school contexts: social,  $\beta = 0.66$ ,  $t(65) = 5.72$ ,  $p < .001$ ; school,  $\beta = 0.57$ ,  $t(65) = 4.03$ ,  $p < .001$ . Conversely, lower youth-reported irritability factor scores were associated with increased impairment due to irritability symptoms in social and school contexts: social,  $\beta = -0.27$ ,  $t(65) = -2.69$ ,  $p = .01$ ; school,  $\beta = -0.27$ ,  $t(65) = -2.13$ ,  $p = .04$ . For impairment in the family context, scores on the negative affectivity factor,  $\beta = 0.81$ ,  $t(65) = 7.40$ ,  $p < .001$ , and parent-reported irritability factor,  $\beta = 0.22$ ,  $t(65) = 2.26$ ,  $p = .03$ , were associated with increased impairment whereas scores on the anxiety factor were associated with decreased impairment,  $\beta = -0.24$ ,  $t(65) = -2.22$ ,  $p = .03$ .

#### Anxiety symptoms and impairment

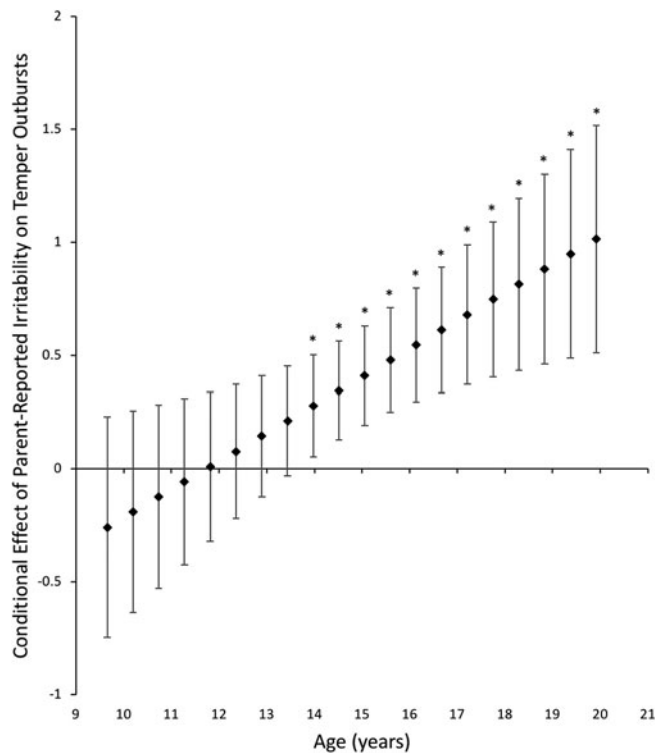
Associations between factor scores and clinician-rated anxiety symptoms and impairment were assessed via the PARS. Higher negative affectivity factor scores,  $\beta = 0.30$ ,  $t(81) = 2.72$ ,  $p = .01$ , and higher anxiety factor scores,  $\beta = 0.23$ ,  $t(81) = 2.06$ ,  $p = .04$ , were associated with higher PARS total scores. Results were similar when examining PARS subscale scores (Supplementary Table S.3). There were no significant associations with parent- or youth-reported irritability factor scores. In addition, higher negative affectivity factor scores,  $\beta = 0.30$ ,  $t(80) = 2.77$ ,  $p = .01$ , and anxiety factor scores,  $\beta = 0.30$ ,  $t(81) = 2.67$ ,  $p = .01$ , were associated with higher PARS total impairment scores. Again, associations with impairment across contexts were consistent with the exception of family setting in which only scores on the anxiety factor and not the negative affectivity factor were associated with impairment in the family setting. Higher negative affectivity factor scores were associated with an increased number of anxiety symptoms,  $\beta = 0.35$ ,  $t(81) = 3.36$ ,  $p = .001$ . Higher anxiety factor scores were associated with an increased number,  $\beta = 0.34$ ,  $t(81) = 3.21$ ,  $p = .002$ , frequency,  $\beta = 0.28$ ,  $t(81) = 2.47$ ,  $p = .02$ , and severity,  $\beta = 0.24$ ,  $t(81) = 2.06$ ,  $p = .04$ , of anxiety symptoms. Higher negative affectivity,  $\beta = 0.25$ ,  $t(80) = 2.25$ ,  $p = .03$ , and anxiety,  $\beta = 0.27$ ,  $t(80) = 2.33$ ,  $p = .02$ , factor scores were associated with increased impairment in the social context. Higher negative affectivity,  $\beta = 0.24$ ,  $t(81) = 2.15$ ,  $p = .03$ , but not anxiety,  $\beta = 0.17$ , factor scores,  $t(81) = 1.46$ ,  $p = .15$ , were associated with impairment in the family context.

#### Global impairment

Higher negative affectivity factor scores were associated with lower CGAS scores (i.e., greater functional impairment),  $\beta = -0.42$ ,  $t(168) = -5.39$ ,  $p < .001$ . There were no significant associations with parent- or youth-reported irritability or anxiety factor scores.

#### Age and sex as moderators

Finally, we investigated whether age and sex moderated the above observed main effects for the associations between factor scores and clinician ratings. Only one significant interaction emerged. For ratings of irritable symptoms and impairment on the DMDD and severe mood dysregulation modules of the KSADS-PL in the ADHD and DMDD groups, there was an interaction between parent-reported irritability and age in predicting temper outbursts,  $\beta = 1.69$ ,  $t(63) = 2.87$ ,  $p = .01$ . At older ages ( $>13.63$ ), higher parent-reported irritability was associated with increased ratings of temper outbursts. At younger ages, there was no significant association between parent-reported irritability and temper outburst ratings (Figure 3).



**Figure 3.** Conditional effect of the relationship between parent-reported irritability factor scores and clinician-rated temper outbursts across values of the moderator, age. Error bars represent the 95% confidence intervals. \* $p < .05$ .

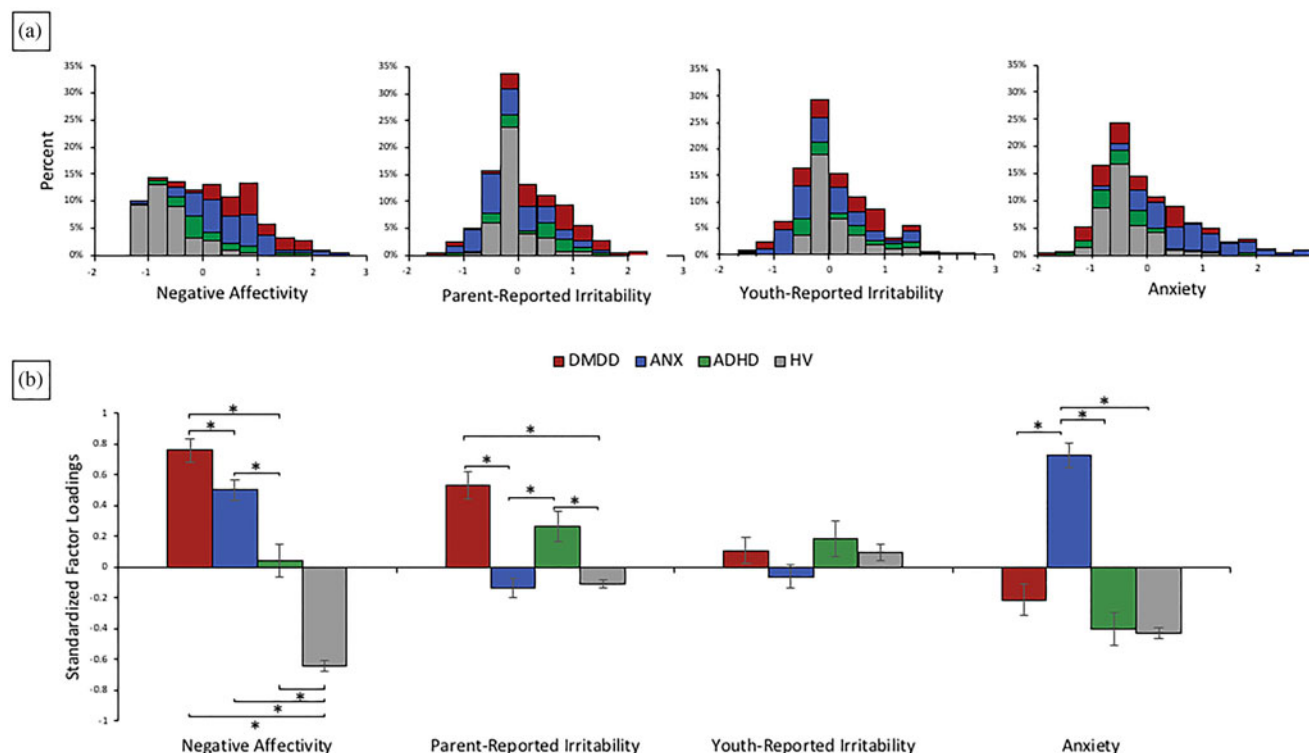
#### Factor loadings and diagnoses

Participants' scores on each of the four latent variables exhibited variability both within and across diagnostic groups (Figure 4). To investigate associations between factor scores and primary diagnoses, four separate analyses of covariance were run with primary diagnosis predicting factor scores. Age, sex, and IQ were covariates. Post hoc comparisons were conducted when omnibus significance was  $p < .008$  (in order to correct for multiple comparison). Results revealed that negative affectivity,  $F(3, 327) = 119.68$ ,  $p < .001$ , parent-reported irritability,  $F(3, 327) = 25.47$ ,  $p < .001$ , and anxiety factor scores,  $F(3, 327) = 26.89$ ,  $p < .001$ , but not youth-reported irritability factor scores,  $F(3, 327) = 1.77$ ,  $p = .15$ , differed significantly by diagnostic group. Specifically, parent-reported irritability factor scores were highest for participants in the DMDD and ADHD groups, and anxiety factor scores were highest for participants in the ANX group. Negative affectivity factor scores exhibited a unique pattern, with the highest scores in the DMDD group, followed by the ANX group and then the ADHD group. All clinical groups had significantly higher negative affectivity factor scores than HVs (Figure 4).

#### Discussion

In the present study, we replicate the bifactor model as a phenotypic strategy to parse irritability and anxiety in a large transdiagnostic sample (Kircanski et al., 2018). First, we extend previous findings by using a multiple group structural equation modeling approach to establish measurement invariance as a function of youth age and sex. Second, associations between factor scores and clinician ratings documented strong convergent and divergent validity with respect to the parent-reported irritability and





**Figure 4.** (a) Histogram of factor loadings across diagnostic groups and (b) average standardized factor loadings for each of the four latent factors across the four diagnostic groups: disruptive mood dysregulation disorder (DMDD), anxiety disorder (ANX), attention-deficit/hyperactivity disorder (ADHD), and no psychiatric disorder (HV). \*Bonferroni-corrected  $p < .05$ .

anxiety factors, but not the youth-reported irritability factor. In addition, negative affectivity factor scores were associated with both global and specific domains of clinician-rated impairment. Third, investigation of cortical thickness, surface area, and subcortical volume revealed no significant associations with any of the four factors.

The lack of significant associations between latent factors and measures of brain structure may seem inconsistent with previous work finding neuroanatomical differences across diagnostic groups characterized by irritability and anxiety symptoms. However, previous findings have been inconsistent. While some regions are consistently implicated across the literature, the direction of the findings are often mixed. For example, an anxiety diagnosis has been associated with both larger (De Bellis et al., 2000; Gold et al., 2017; Qin et al., 2014) and smaller (Newman et al., 2015; Strawn et al., 2015) PFC thickness or volume, with other studies failing to find any association between anxiety diagnosis and PFC structure (De Bellis et al., 2002; Qin et al., 2014; Strawn et al., 2014). The same is true for subcortical amygdala and hippocampal volume: studies report both larger (De Bellis et al., 2000; Qin et al., 2014) and smaller (Gold et al., 2017; Mueller et al., 2013; Strawn et al., 2015) subcortical volumes in patients with an anxiety diagnosis, while others find no association with subcortical volume (De Bellis et al., 2002; Liao et al., 2013; Newman et al., 2015; Strawn et al., 2013). Few studies examine associations between dimensionally assessed symptoms and neural structure; those that have failed to find associations between dimensionally assessed anxiety or irritability with cortical or subcortical structure (Gold et al., 2016; Merz, He, & Noble, 2018).

The current study is the largest to date to investigate associations between dimensionally assessed irritability and anxiety

symptoms and neural structure. These findings provide further evidence that the dimensionally assessed unique and shared features of irritability and anxiety are not associated with neuroanatomical abnormalities. Of note, our previous work using the bifactor model revealed both shared and unique associations with functional neural activity and connectivity during an attention orienting to threat task (Kircanski et al., 2018). This suggests that neural abnormalities in irritability and anxiety may be more functional than structural in nature. Whereas in the current study we investigate more general structural feature abnormalities across cortical and subcortical regions, functional neuroimaging analyses investigate neural function in specific contexts (i.e., orienting to threat; Kircanski et al., 2018). As such, it is possible that neural correlates of irritability and anxiety may be characterized more by functional responses underlying specific psychological processes than by general structural abnormalities.

Associations between the latent factors and clinician-rated measures provide psychometric support for measures derived in the bifactor model. The parent-reported irritability and anxiety factors were associated specifically with clinician ratings of irritability and anxiety, respectively. Furthermore, the negative affectivity factor captured more global symptoms (e.g., showing associations with ratings of global impairment on the CGAS). Of note, while child- and parent-rated anxiety loaded on the same factor, this was not the case for irritability. Moreover, the youth-reported irritability factor failed to demonstrate associations with clinician ratings of irritability; higher youth-reported irritability scores were related to decreased clinician-rated impairment due to irritability. Consistent with our previous findings (Kircanski et al., 2018), in the current sample, patients with a DMDD diagnosis were fairly evenly distributed across youth-reported irritability scores, suggesting that some

DMDD youths may be underreporting irritability symptoms relative to parent and clinician ratings. While discrepancy between parent and child report of psychopathology is common (De Los Reyes & Kazdin, 2005), the current results emphasize the need for improved assessments of youth-reported irritability.

The bifactor model demonstrated invariance across both age and sex, confirming that this phenotyping strategy is robust across males and females from late childhood through adolescence. Further, we observed only one significant interaction, between a factor and age, in which the association between parent-reported irritability and clinician-rated temper outbursts was significant only in adolescence. Future research is needed to replicate this finding to determine if this association is driven by convergence of parent and clinician reports of developmentally inappropriate temper outbursts into adolescence. Alternatively, this finding may instead capture a unique developmental trajectory for temper outbursts as they relate to clinical levels of irritability.

The current study has several limitations. First, irritability and anxiety are symptoms of emotion dysregulation that are present across a large number of childhood psychiatric disorders (Cornacchio, Crum, Coxe, Pinuc, & Comer, 2016; Stoddard *et al.*, 2014). While the current study includes a sample composed of several childhood psychiatric diagnoses, it does not include a truly transdiagnostic, more broadly representative sample. Moreover, the recruitment procedures could have influenced statistical model fitting. As such, the current study is limited in our ability to draw conclusions regarding the performance of the bifactor model as a phenotypic strategy across a wider range of youth and symptom expressions. For example, the current study does not include youth with a primary diagnosis of depression. Both cross-sectional and longitudinal work link depression to symptoms of both irritability and anxiety (Stringaris, Maughan, Copeland, Costello, & Angold, 2013; Vidal-Ribas, Brotman, Valdivieso, Leibenluft, & Stringaris, 2016), and shared functional neural substrates (Zisner & Beauchaine, 2016). Because we did not include youths with primary unipolar or bipolar depression, depressive symptoms in our sample were restricted. Inclusion of depressive symptoms and a wider range of childhood psychiatric disorders could affect the model structure given the possibility that depression may partially explain the observed links between irritability and anxiety symptoms (Zisner & Beauchaine, 2016). Within the context of the current study, we find it unlikely that depression explains links between irritability and anxiety given that a current diagnosis of major depressive disorder was an exclusionary criterion. However, within a broader transdiagnostic framework, such a link may emerge. In order to address these important limitations, future work should examine how the current bifactor model of irritability and anxiety performs within a more representative transdiagnostic sample. Furthermore, given associations among irritability, anxiety, and depressive symptoms, future work should consider adding parent- and youth-reported depressive symptoms to the bifactor model to further parse these symptoms of negative affective states into their unique and shared features. Second, due to the severity of their impairment, a portion of our patients were taking psychotropic medications (Table 1). Psychotropic medications may impact our ability to detect structural neural abnormalities associated with psychiatric symptoms. However, medication appears to increase between-group neuroanatomical differences in MRI studies and appears to be related to increases in neuroanatomical differences (Hafeman, Chang, Garrett, Sanders, & Phillips, 2012; Navari & Dazzan, 2009; Smieskova *et al.*, 2009). Given the lack of

between-group differences observed in the current study, it is unlikely that medication can fully account for our results.

Third, the aim of the current study was to replicate the bifactor model developed in previous work by our group (Kircanski *et al.*, 2018) and examine its associations with brain structure and clinician-rated symptoms and impairment. There has been growing critique of bifactor models with respect to data overfitting and statistical advantages that bias fit indices as compared to first- and second-order models (Morgan, Hodge, Wells, & Watkins, 2015; Murray & Johnson, 2013). Adequate fit for a bifactor model may emerge in the absence of meaningful latent factors to support conclusions regarding structure of psychopathology (Bonifay, Lane, & Reise, 2017). Thus, it is critical to examine associations between derived factors and external validators such as neural function (Kircanski *et al.*, 2018). In the current study, we found no significant associations between derived factor scores and measures of neural structure as a potential external validator. As a result, the current results provide limited data concerning the validity of the current bifactor model. However, prior research demonstrates associations with neural function (Kircanski *et al.*, 2018), and inconsistencies exist in the literature examining structural neural abnormalities associated with anxiety and irritability. It is plausible that no robust, significant associations exist between neural structure and symptoms related to either anxiety or irritability. Thus, future studies should examine associations between the bifactor model and both neural function and other potential external validators. Furthermore, the clinician ratings of symptoms and impairment are derived from related or overlapping symptoms with those in the bifactor model. Our findings that factor scores showed strong convergent and divergent associations with clinician ratings provide some confidence that the specific latent factors for anxiety and irritability are capturing unique variability in anxiety and irritability symptoms, respectively, while the general latent factor of negative affectivity relates to global impairment. Future work should examine associations of factor scores with independent, clinically relevant validators (e.g., treatment response or outcome).

The co-occurrence of emotion dysregulation symptoms across a wide range of psychopathology in youth presents a unique challenge to uncovering shared versus unique neurodevelopmental mechanisms. The current paper extended our previous work using a bifactor analysis to quantify the unique and shared variances of dimensionally assessed irritability and anxiety, two common emotion dysregulation symptoms that frequently co-occur across diagnostic groups. We found strong convergent associations between factor scores and clinician ratings as well as invariance across sex and age. Results also propose that the unique and shared features of irritability and anxiety may be unrelated to common measures of brain structure. The application of latent variable models to co-occurring symptom dimensions could ultimately lead to improvements in targeted treatments across various domains of child and adolescent psychopathology.

**Financial support.** Supported by the NIMH Intramural Research Program, conducted under NIH Clinical Study Protocols 15-M-0182 (ClinicalTrials.gov identifier: NCT02531893), 02-M-0021 (ClinicalTrials.gov identifier: NCT00025935), and 00-M-0198 (ClinicalTrials.gov identifier: NCT00006177) and 01-M-0192 (ClinicalTrials.gov identifier: NCT00018057).

## References

- Adleman, N. E., Fromm, S. J., Razdan, V., Kayser, R., Dickstein, D. P., Brotman, M. A., ... Leibenluft, E. (2012). Cross-sectional and longitudinal abnormalities in brain structure in children with severe mood dysregulation

- or bipolar disorder: Cross-sectional and longitudinal volumetric abnormalities in SMD and BD. *Journal of Child Psychology and Psychiatry*, 53, 1149–1156. doi:10.1111/j.1469-7610.2012.02568.x
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: Author.
- Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36, 545–553. doi:10.1097/00004583-199704000-00018
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5, 184–186. doi:10.1177/2167702616657069
- Brotman, M. A., Kircanski, K., Stringaris, A., Pine, D. S., & Leibenluft, E. (2017). Irritability in youths: A translational model. *American Journal of Psychiatry*, 174, 520–532. doi:10.1176/appi.ajp.2016.16070839
- Brotman, M. A., Schmajuk, M., Rich, B. A., Dickstein, D. P., Guyer, A. E., Costello, E. J., ... Leibenluft, E. (2006). Prevalence, clinical correlates, and longitudinal course of severe mood dysregulation in children. *Biological Psychiatry*, 60, 991–997. doi:10.1016/j.biopsych.2006.08.042
- Cardinale, E. M., O'Connell, K., Robertson, E. L., Meena, L. B., Breeden, A. L., Lozier, L. M., ... Marsh, A. A. (2018). Callous and uncaring traits are associated with reductions in amygdala volume among youths with varying levels of conduct problems. *Psychological Medicine*. Advance online publication. doi:10.1017/S0033291718001927
- Castellanos-Ryan, N., Struve, M., Whelan, R., Banaschewski, T., Barker, G. J., Bokde, A. L. W., ... IMAGEN Consortium. (2014). Neural and cognitive correlates of the common and specific variance across externalizing problems in young adolescence. *American Journal of Psychiatry*, 171, 1310–1319. doi:10.1176/appi.ajp.2014.13111499
- Copeland, W. E., Brotman, M. A., & Costello, E. J. (2015). Normative irritability in youth: Developmental findings from the Great Smoky Mountains Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 54, 635–642. doi:10.1016/j.jaac.2015.05.008
- Cornacchio, D., Crum, K. I., Cox, S., Pincus, D. B., & Comer, J. S. (2016). Irritability and severity of anxious symptomatology among youth with anxiety disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55, 54–61. doi:10.1016/j.jaac.2015.10.007
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9, 179–194. doi:10.1006/nimg.1998.0395
- De Bellis, M. D., Casey, B. J., Dahl, R. E., Birmaher, B., Williamson, D. E., Thomas, K. M., ... Ryan, N. D. (2000). A pilot study of amygdala volumes in pediatric generalized anxiety disorder. *Biological Psychiatry*, 48, 51–57. doi:10.1016/S0006-3223(00)00835-0
- De Bellis, M. D., Keshavan, M. S., Shifflett, H., Iyengar, S., Dahl, R. E., Axelson, D. A., ... Ryan, N. D. (2002). Superior temporal gyrus volumes in pediatric generalized anxiety disorder. *Biological Psychiatry*, 51, 553–562.
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131, 483–509. doi:10.1037/0033-2909.131.4.483
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31, 968–980. doi:10.1016/j.neuroimage.2006.01.021
- Dickstein, D. P., Milham, M. P., Nugent, A. C., Drevets, W. C., Charney, D. S., Pine, D. S., & Leibenluft, E. (2005). Frontotemporal alterations in pediatric bipolar disorder: Results of a voxel-based morphometry study. *Archives of General Psychiatry*, 62, 734. doi:10.1001/archpsyc.62.7.734
- Fischl, B. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14, 11–22. doi:10.1093/cercor/bhg087
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the USA*, 97, 11050–11055. doi:10.1073/pnas.200033797
- Fischl, B., Salat, D. H., Busa, E., Albert, M. S., Dieterich, M., Haselgrove, C., ... Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33, 341–355. doi:10.1016/S0896-6273(02)00569-X
- Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., & Dale, A. M. (2004). Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23(Suppl. 1), S69–S84. doi:10.1016/j.neuroimage.2004.07.016
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., ... Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14, 11–22.
- Gold, A. L., Brotman, M. A., Adleman, N. E., Lever, S. N., Steuber, E. R., Fromm, S. J., ... Leibenluft, E. (2016). Comparing brain morphometry across multiple childhood psychiatric disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55, 1027–1037. doi:10.1016/j.jaac.2016.08.008
- Gold, A. L., Steuber, E. R., White, L. K., Pacheco, J., Sachs, J. F., Pagliaccio, D., ... Pine, D. S. (2017). Cortical thickness and subcortical gray matter volume in pediatric anxiety disorders. *Neuropsychopharmacology*. Advance online publication. doi:10.1038/npp.2017.83
- Grimm, O., Pohlack, S., Cacciaglia, R., Winkelman, T., Plichta, M. M., Demirakca, T., & Flor, H. (2015). Amygdalar and hippocampal volume: A comparison between manual segmentation, Freesurfer and VBM. *Journal of Neuroscience Methods*, 253, 254–261. doi:10.1016/j.jneumeth.2015.05.024
- Hafeman, D. M., Chang, K. D., Garrett, A. S., Sanders, E. M., & Phillips, M. L. (2012). Effects of medication on neuroimaging findings in bipolar disorder: An updated review: Medication effects on neuroimaging in bipolar disorder. *Bipolar Disorders*, 14, 375–410. doi:10.1111/j.1399-5618.2012.01023.x
- Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., ... Fischl, B. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, 32, 180–194. doi:10.1016/j.neuroimage.2006.02.051
- Hayes, A. F. (2016). *The PROCESS macro for SPSS and SAS*. Retrieved from <http://www.processmacro.org/index>
- Holmes, A. P., Blair, R. C., Watson, J. D. G., & Ford, I. (1996). Nonparametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow and Metabolism*, 16, 7–22.
- Hyman, S. E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience*, 8, 725–732. doi:10.1038/nrn2218
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., ... Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167, 748–751. doi:10.1176/appi.ajp.2010.09091379
- Kircanski, K., White, L. K., Tseng, W.-L., Wiggins, J. L., Frank, H. R., Sequeira, S., ... Brotman, M. A. (2018). A latent variable approach to differentiating neural mechanisms of irritability and anxiety in youth. *JAMA Psychiatry*, 75, 631. doi:10.1001/jamapsychiatry.2018.0468
- Kuperberg, G. R., McGuire, P. K., Ozawa, F., Goff, D., West, W. C., & Williams, S. C. R. (2003). Regionally localized thinning of the cerebral cortex in schizophrenia. *Archives of General Psychiatry*, 60, 11.
- Leadbeater, B. J., & Homel, J. (2015). Irritable and defiant sub-dimensions of ODD: Their stability and prediction of internalizing symptoms and conduct problems from adolescence to young adulthood. *Journal of Abnormal Child Psychology*, 43, 407–421. doi:10.1007/s10802-014-9908-3
- Leibenluft, E. (2017). Pediatric irritability: A systems neuroscience approach. *Trends in Cognitive Sciences*, 21, 277–289. doi:10.1016/j.tics.2017.02.002
- Liao, M., Yang, F., Zhang, Y., He, Z., Song, M., Jiang, T., ... Li, L. (2013). Childhood maltreatment is associated with larger left thalamic gray matter volume in adolescents with generalized anxiety disorder. *PLOS ONE*, 8, e71898. doi:10.1371/journal.pone.0071898
- McLaughlin, K. A., Sheridan, M. A., Winter, W., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2014). Widespread reductions in cortical thickness following severe early-life deprivation: A neurodevelopmental pathway to attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 76, 629–638. doi:10.1016/j.biopsych.2013.08.016
- Merz, E. C., He, X., & Noble, K. G. (2018). Anxiety, depression, impulsivity, and brain structure in children and adolescents. *NeuroImage: Clinical*, 20, 243–251. doi:10.1016/j.nicl.2018.07.020
- Milham, M. P., Nugent, A. C., Drevets, W. C., Dickstein, D. S., Leibenluft, E., Ernst, M., ... Pine, D. S. (2005). Selective reduction in amygdala volume in

- pediatric anxiety disorders: A voxel-based morphometry investigation. *Biological Psychiatry*, 57, 961–966. doi:10.1016/j.biopsych.2005.01.038
- Morey, R. A., Petty, C. M., Xu, Y., Hayes, J. P., Wagner, H. R., Lewis, D. V., ... McCarthy, G. (2009). A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *NeuroImage*, 45, 855–866. doi:10.1016/j.neuroimage.2008.12.033
- Morey, R. A., Selgrade, E. S., Wagner, H. R., Huettel, S. A., Wang, L., & McCarthy, G. (2010). Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Human Brain Mapping*, 31, 1751–1763. doi:10.1002/hbm.20973
- Morgan, G., Hodge, K., Wells, K., & Watkins, M. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, 3, 2–20. doi:10.3390/jintelligence3010002
- Morris, S. E., & Cuthbert, B. N. (2012). Research Domain Criteria: Cognitive systems, neural circuits, and dimensions of behavior. *Dialogues in Clinical Neuroscience*, 14, 29–37.
- Mueller, S. C., Aouidad, A., Gorodetsky, E., Goldman, D., Pine, D. S., & Ernst, M. (2013). Grey matter volume in adolescent anxiety: An impact of the brain-derived neurotrophic factor Val66Met polymorphism? *Journal of the American Academy of Child & Adolescent Psychiatry*, 52, 184–195. doi:10.1016/j.jaac.2012.11.016
- Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, 41, 407–422. doi:10.1016/j.intell.2013.06.004
- Navari, S., & Dazzan, P. (2009). Do antipsychotic drugs affect brain structure? A systematic and critical review of MRI findings. *Psychological Medicine*, 39, 1763. doi:10.1017/S0033291709005315
- Newman, E., Thompson, W. K., Bartsch, H., Hagler, D. J., Chen, C.-H., Brown, T. T., ... Jernigan, T. L. (2015). Anxiety is related to indices of cortical maturation in typically developing children and adolescents. *Brain Structure and Function*, 221, 3013–3025. doi:10.1007/s00429-015-1085-9
- Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15, 1–25. doi:10.1002/hbm.1058
- Ostby, Y., Tamnes, C. K., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2009). Heterogeneity in subcortical brain development: A structural magnetic resonance imaging study of brain maturation from 8 to 30 years. *Journal of Neuroscience*, 29, 11772–11782. doi:10.1523/JNEUROSCI.1242-09.2009
- Pine, D. S. (2007). Research Review: A neuroscience framework for pediatric anxiety disorders. *Journal of Child Psychology and Psychiatry*, 48, 631–648. doi:10.1111/j.1469-7610.2007.01751.x
- Qin, S., Young, C. B., Duan, X., Chen, T., Supekar, K., & Menon, V. (2014). Amygdala subregional structure and intrinsic functional connectivity predicts individual differences in anxiety during early childhood. *Biological Psychiatry*, 75, 892–900. doi:10.1016/j.biopsych.2013.10.006
- Research Units On Pediatric Psychopharmacology Anxiety Study Group. (2002). The Pediatric Anxiety Rating Scale (PARS): Development and psychometric properties. *Journal of the American Academy of Child & Adolescent Psychiatry*, 41, 1061–1069. doi:10.1097/00004583-200209000-00006
- Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, 61, 1402–1418. doi:10.1016/j.neuroimage.2012.02.084
- Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., ... Fischl, B. (2002). CME regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology*, 58, 695–701.
- Savage, J., Verhulst, B., Copeland, W., Althoff, R. R., Lichtenstein, P., & Roberson-Nay, R. (2015). A genetically informed study of the longitudinal relation between irritability and anxious/depressed symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, 54, 377–384. doi:10.1016/j.jaac.2015.02.010
- Schoemaker, D., Buss, C., Head, K., Sandman, C. A., Davis, E. P., Chakravarty, M. M., ... Pruessner, J. C. (2016). Hippocampus and amygdala volumes from magnetic resonance images in children: Assessing accuracy of FreeSurfer and FSL against manual segmentation. *NeuroImage*, 129, 1–14. doi:10.1016/j.neuroimage.2016.01.038
- Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 22, 1060–1075. doi:10.1016/j.neuroimage.2004.03.032
- Shaffer, D. (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry*, 40, 1228. doi:10.1001/archpsyc.1983.01790100074010
- Shanmugan, S., Wolf, D. H., Calkins, M. E., Moore, T. M., Ruparel, K., Hopson, R. D., ... Satterthwaite, T. D. (2016). Common and dissociable mechanisms of executive system dysfunction across psychiatric disorders in youth. *American Journal of Psychiatry*, 173, 517–526. doi:10.1176/appi.ajp.2015.15060725
- Shaw, P., Stringaris, A., Nigg, J., & Leibenluft, E. (2014). Emotion dysregulation in attention deficit hyperactivity disorder. *American Journal of Psychiatry*, 171, 276–293. doi:10.1176/appi.ajp.2013.13070966
- Sheridan, M. A., Fox, N. A., Zeanah, C. H., McLaughlin, K. A., & Nelson, C. A. (2012). Variation in neural development as a result of exposure to institutionalization early in childhood. *Proceedings of the National Academy of Sciences of the USA*, 109, 12927–12932. doi:10.1073/pnas.1200041109
- Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17, 87–97. doi:10.1109/42.668698
- Smieskova, R., Fusar-Poli, P., Allen, P., Bendfeldt, K., Stieglitz, R., Drewe, J., ... Borgwardt, S. (2009). The effects of antipsychotics on the brain: What have we learnt from structural imaging of schizophrenia?—A systematic review. *Current Pharmaceutical Design*, 15, 2535–2549. doi:10.2174/138161209788957456
- Smith, S., & Nichols, T. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44, 83–98. doi:10.1016/j.neuroimage.2008.03.061
- Stoddard, J., Stringaris, A., Brotman, M. A., Montville, D., Pine, D. S., & Leibenluft, E. (2014). Irritability in child and adolescent anxiety disorders. *Depression and Anxiety*, 31, 566–573. doi:10.1002/da.22151
- Strawn, J. R., Chu, W.-J., Whitsel, R. M., Weber, W. A., Norris, M. M., Adler, C. M., ... DelBello, M. P. (2013). A pilot study of anterior cingulate cortex neurochemistry in adolescents with generalized anxiety disorder. *Neuropsychobiology*, 67, 224–229. doi:10.1159/000347090
- Strawn, J. R., Hamm, L., Fitzgerald, D. A., Fitzgerald, K. D., Monk, C. S., & Phan, K. L. (2015). Neurostructural abnormalities in pediatric anxiety disorders. *Journal of Anxiety Disorders*, 32, 81–88. doi:10.1016/j.janxdis.2015.03.004
- Strawn, J. R., John Wegman, C., Dominick, K. C., Swartz, M. S., Wehry, A. M., Patino, L. R., ... DelBello, M. P. (2014). Cortical surface anatomy in pediatric patients with generalized anxiety disorder. *Journal of Anxiety Disorders*, 28, 717–723. doi:10.1016/j.janxdis.2014.07.012
- Stringaris, A., & Goodman, R. (2009). Three dimensions of oppositionality in youth. *Journal of Child Psychology and Psychiatry*, 50, 216–223. doi:10.1111/j.1469-7610.2008.01989.x
- Stringaris, A., Goodman, R., Ferdinando, S., Razdan, V., Muhrer, E., Leibenluft, E., & Brotman, M. A. (2012). The Affective Reactivity Index: A concise irritability scale for clinical and research settings. *Journal of Child Psychology and Psychiatry*, 53, 1109–1117. doi:10.1111/j.1469-7610.2012.02561.x
- Stringaris, A., Maughan, B., Copeland, W. S., Costello, E. J., & Angold, A. (2013). Irritable mood as a symptom of depression in youth: Prevalence, developmental, and clinical correlates in the Great Smoky Mountains Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52, 831–840. doi:10.1016/j.jaac.2013.05.017
- Sylvester, C. M., Barch, D. M., Harms, M. P., Belden, A. C., Oakberg, T. J., Gold, A. L., ... Pine, D. S. (2016). Early childhood behavioral inhibition predicts cortical thickness in adulthood. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55, 122–129. doi:10.1016/j.jaac.2015.11.007
- 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, 556–570. doi:10.1016/j.jaac.2016.04.014
- Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, 96, 465–490.

- Wiggins, J. L., Brotman, M. A., Adleman, N. E., Kim, P., Oakes, A. H., Reynolds, R. C., ... Leibenluft, E. (2016). Neural correlates of irritability in disruptive mood dysregulation and bipolar disorders. *American Journal of Psychiatry*, *173*, 722–730. doi:10.1176/appi.ajp.2015.15060833
- Winkler, A. M., Greve, D. N., Bjostrand, K. J., Nichols, T. E., Sabuncu, M. R., Häberg, A. K., ... Rimol, L. M. (2018). Joint Analysis of Cortical Area and Thickness as a Replacement for the Analysis of the Volume of the Cerebral Cortex. *Cerebral Cortex*, *28*(2), 738–749. <https://doi.org/10.1093/cercor/bhx308>
- Winkler, A. M., Ridgway, G. R., Douaud, G., Nichols, T. E., & Smith, S. M. (2016). Faster permutation inference in brain imaging. *NeuroImage*, *141*, 502–516.
- Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397. doi:10.1016/j.neuroimage.2014.01.060
- Winkler, A. M., Webster, M. A., Brooks, J. C., Tracey, I., Smith, S. M., & Nichols, T. E. (2016). Non-parametric combination and related permutation tests for neuroimaging: NPC and related permutation tests for neuroimaging. *Human Brain Mapping*, *37*, 1486–1511. doi:10.1002/hbm.23115
- Winkler, A. M., Webster, M. A., Vidaurre, D., Nichols, T. E., & Smith, S. M. (2015). Multi-level block permutation. *NeuroImage*, *123*, 253–268. doi:10.1016/j.neuroimage.2015.05.092
- Zald, D. H., & Lahey, B. B. (2017). Implications of the hierarchical structure of psychopathology for psychiatric neuroimaging. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *2*, 310–317. doi:10.1016/j.bpsc.2017.02.003
- Zisner, A., & Beauchaine, T. P. (2016). Neural substrates of trait impulsivity, anhedonia, and irritability: Mechanisms of heterotypic comorbidity between externalizing disorders and unipolar depression. *Development and Psychopathology*, *28*(4, Pt. 1), 1177–1208. doi:10.1017/S0954579416000754