



HHS Public Access

Author manuscript

Psychol Assess. Author manuscript; available in PMC 2020 February 01.

Published in final edited form as:

Psychol Assess. 2019 February ; 31(2): 236–247. doi:10.1037/pas0000664.

Do we need an irritable subtype of ADHD? Replication and extension of a promising temperament profile approach to ADHD subtyping

Sarah L. Karalunas^{1,2}, Hanna C. Gustafsson¹, Damien Fair^{1,2,3}, Erica D. Musser^{4,5}, and Joel T. Nigg^{1,2}

¹Psychiatry, Oregon Health & Science University, Portland, OR

²Behavioral Neuroscience, Oregon Health & Science University, Portland, OR

³Advanced Imaging Research Center, Oregon Health & Science University, Portland, OR

⁴Psychology, Florida International University, Miami, FL

⁵Center for Children and Families, Florida International University, Miami, FL

Abstract

Attention-deficit/hyperactivity disorder (ADHD) is emblematic of unresolved heterogeneity in psychiatric disorders—the variation in biological, clinical, and psychological correlates that impedes progress on etiology. One approach to this problem is to characterize subgroups using measures rooted in biological or psychological theory, consistent with NIMH’s RDoC initiative. Within ADHD, a promising application involves using emotion trait profiles that can address the role of irritability as a complicating feature for ADHD. Here, a new sample of 186 children with ADHD was evaluated using community detection analysis to determine if meaningful subprofiles existed and if they replicated those previously identified. The new sample and a prior sample were pooled for evaluation of a) method dependence, b) longitudinal assessment of the stability of classifications, and c) clinical prediction two years later. Three temperament profiles were confirmed within the ADHD group: one with normative emotional functioning (“Mild”), one with high surgency (“Surgent”), and one with high negative affect (“Irritable”). Profiles were similar across statistical clustering approaches. The Irritable group had the highest external validity: it was moderately stable over time and it enhanced prospective prediction of clinical outcomes beyond standard baseline indicators. The Irritable group was not reducible to ADHD+ODD, ADHD+DMDD, or other patterns of comorbidity. Among the negative affect domains studied, trait proneness to anger uniquely contributed to clinical prediction. Results extend our understanding of chronic irritability in psychiatric disorders and provide prospects for a fresh approach to assessing ADHD heterogeneity focused on the distinction between ADHD with and without anger/irritability.

Keywords

ADHD; irritability; temperament; longitudinal prediction; heterogeneity

Approaches to psychiatric nosology at present include those in the DSM-5, ICD10/11, and RDoC (Clark, Cuthbert, Lewis-Fernández, Narrow, & Reed, 2017). The DSM/ICD approaches rely on symptom profiles arrived at from clinical experience and subjected to tests of factor structure and reliability to identify diagnostic groups. The RDoC tactic employs biologically-validated, continuous dimensions of functioning as constituent elements of complex syndromes to develop a dimension-based characterization. Blending these approaches may prove particularly promising by building on the vast literature on reliability and validity of DSM/ICD categories, while integrating it with emerging knowledge of neurobiological bases of dimensions of psychological functioning. In this way, an integrated approach for improving assessment and, ultimately, nosology can emerge (Cuthbert, 2014; Karalunas et al., 2014; Lenzenweger, Clarkin, Yeomans, Kernberg, & Levy, 2008).

The first step in this blended approach is to use computational tools to identify hypothesized profiles (or clusters) of individuals within existing diagnostic populations (DSM), but using neurobiologically-validated dimensions of psychological function (as in RDoC) rather than symptoms. The next step focuses on external validation using multiple criteria, including a) biological correlates, b) stability of profile assignment over time, and c) clinical prediction. Replication of initial solutions in new samples is also of fundamental importance (Ioannidis, 2012). Karalunas et al. (2014) provided an illustration of this logic as applied to ADHD. Their approach grouped children with ADHD on biologically-supported dimensions of emotional functioning captured by well-validated temperament traits.

Temperament as used here refers to patterns of emotion response and emotion regulation measured as traits in children (Bates, Goodnight, & Fite, 2008; Rothbart, 2011). In this model, the trait structure is summarized empirically in terms of three broad, well-validated domains: (1) *negative affect*, encompassing emotions such as fear, sadness, and anger/frustration; (2) *positive affect (or surgency)*, reflecting tendency to express excitement and happiness, willingness to approach novel stimuli, and overall activity level; and (3) *effortful control*, related to top down self-regulatory capacities and actions (Nigg, 2006; Rothbart, 2011; Rothbart & Bates, 1998; Rothbart, Derryberry, & Posner, 1994). Each of these domains has a hypothesized neurobiological basis in the interaction of amygdala-circuitry, dopaminergic reward networks, and prefrontal-subcortical and cortical-cortical control networks, respectively (Posner & Rothbart, 2000; Whittle, Allen, Lubman, & Yücel, 2006). Karalunas et al. (2014) used a mathematical technique called community detection (Newman, 2006; Rubinov & Sporns, 2011) with temperament features as input to discover and initially validate three novel “types” of ADHD labeled: Mild, Surgent, and, Irritable. Of particular clinical significance was their identification of an Irritable group. Irritability is increasingly recognized as an insufficiently characterized influence in child psychopathology (Shaw, Stringaris, Nigg, & Leibenluft, 2014) that cuts across existing diagnostic categories. The definition includes “proneness to anger” (Vidal-Ribas, Brotman,

Valdivieso, Leibenluft, & Stringaris, 2016, p. 557). Although it is emphasized in the new syndrome of Disruptive Mood Dysregulation Disorder (DMDD) in DSM-5, irritability is seen as a feature of ADHD as well. In fact, most children who meet DMDD criteria also meet criteria for ADHD and/or ODD (Leibenluft, 2011; Leibenluft, Blair, Charney, & Pine, 2003; Leibenluft, Cohen, Gorrindo, Brook, & Pine, 2006; Stringaris, 2011). Thus, it is not surprising that irritability is associated with concurrent ADHD diagnosis (Kircanski et al., 2016), leading Shaw et al. (2014) to highlight the need to better characterize how it fits into ADHD nosology. If a subgroup of youth with ADHD can be reliably classified as an “irritable” subgroup, even though they may not meet full criteria for DMDD or ODD, and if this improves on clinical prediction versus existing use of comorbid diagnoses, then this would be an advance in nosology. If this hypothesized Irritable profile is valid, it should be related to the development of mood and anxiety disorders over time (Leibenluft et al., 2006; Stringaris, Cohen, Pine, & Leibenluft, 2009; Stringaris, Zavos, Leibenluft, Maughan, & Eley, 2012; Vidal-Ribas et al., 2016; Wakschlag et al., 2015) even after controlling for baseline indicators. Although Karalunas et al. (2014) reported a general relationship to pooled comorbid psychiatric disorders, their study was too small to examine specific outcomes.

The present study extends this promising line of work in several critical ways. First, we report an independent replication of Karalunas et al. (2014) in a new sample using the same statistical grouping approach (community detection). This is extremely important. Replication efforts have too often been relegated to second-tier importance or dismissed as insufficiently “novel” or “innovative.” It is now recognized that such dismissal is misguided. As extensive recent emphasis on reproducibility and replication in the literature attest (Asendorpf et al., 2013; Makel, Plucker, & Hegarty, 2012; Pashler & Wagenmakers, 2012), basic replication studies are at least as important as “novel” findings—which in fact often fail to replicate (Ioannidis, 2012). Second, we evaluate the reproducibility of emotion profiles using an alternative statistical grouping approach (latent profile analysis). Dozens of major clustering methods are available, with different advantages and weaknesses depending on study goals and no “best” method. Although all methods are not expected to yield identical results (Everitt, Landau, Leese, & Stahl, 2011), it is important to know whether findings are essentially method dependent.

Third, we evaluate reproducibility across two subsequent years of assessment, including testing the stability of individual children’s type assignments, which is crucial to clinical utility. Finally, because clinical prediction is the *sine qua non* for a clinically useful subtype profile in psychiatric nosology, we provide new data based on two-year follow up regarding clinical outcomes.

We hypothesized that a) the three emotion-based types of ADHD would reproduce in a new sample, b) that the Surgent and Irritable types would be stable over time, and c) profiles would be differentially associated with clinical outcome two years later. The Irritable profile, in particular, was expected to be prospectively associated with future onset of anxiety and depression. Dimensionally, elevated anger proneness in this group was expected to predict clinical outcomes based on the growing literature related to irritability in psychiatric disorders.

Method

Participants

For new sample replication, we recruited a sample of 186 children with ADHD who were screened and evaluated by the same procedures as in Karalunas et al. (2014). See Table S1 for comparison of original and replication samples. Following cross-sectional replication of groups, we considered longitudinal stability and prediction in more depth. To maximize statistical power, for this second set of analyses we pooled the two samples (as justified later) by combining these 186 children with 182 ADHD children who were included in Karalunas et al. (2014) for a total ADHD $N=368$ (69.2% male) at our baseline (Year 1) time point. Of these 368 children with baseline ADHD, 87% provided annual follow-up data for at least one annual assessment and 49% provided data at both annual assessments. (Data missing at only a single wave was due, in part, to a planned missingness design intended to conserve limited funding.) Overall, longitudinal follow-up data were available for 225 ADHD children at the first follow up (Year 2) and 271 ADHD children at the second follow up (Year 3). Additionally, 210 (50.4% male) typically-developing children provided a normative comparison group.

Enrollment and Diagnostic Procedures

Recruitment and diagnostic assessment procedures were identical for the ADHD and typically-developing groups. These procedures were the same as those described in Karalunas et al. (2014). All children were between the ages of 7–12 years at initial entry into the study. Ethics approval was obtained from the local Institutional Review Board. A parent/legal guardian provided written informed consent, and children provided written assent. Behavioral ratings data were collected and managed using REDCap electronic data capture tools hosted locally, which provide a secure web-based and intuitive interface and export capabilities (Harris et al., 2009).

All children were identified for the study via a best-estimate diagnostic confirmation procedure. A parent/guardian and teacher both completed standardized rating scales, including the Conners Rating Scales-Revised (CRS-R, Conners, 2003), Strengths and Difficulties Questionnaire (SDQ, Goodman, 2001), and the ADHD Rating Scale (ADHD-RS, DuPaul, Power, Anastopoulos, & Reid, 1998). The parent/guardian also completed a semi-structured clinical interview for DSM-IV diagnoses administered by a Master's-degree level clinician (Kiddie Schedule for Affective Disorders and Schizophrenia, Puig-Antich & Ryan, 1986). Inter-interviewer reliability was carefully evaluated by double coding of 20 videotape interviews to a gold-standard coding of $k>.7$ for all disorders with base rate $> 5\%$ in the sample. Interviewer fidelity to standardized procedures was checked by trainer review of taped interviews quarterly. IQ was estimated based on a reliable and valid three-subtest short form of the WISC-IV (Vocabulary, Block Design, and Information, Sattler & Dumont, 2004; Wechsler, 2003). Academic achievement was assessed using the subtests of the WIAT-II (Wechsler, 2002).

Final diagnoses (either “control” or “ADHD,” as well as all comorbid diagnoses) were made by a clinical diagnostic team (a board-certified child psychiatrist and licensed clinical child

psychologist), who took into account data from the parent *and* teacher ratings, parent KSADS interview, IQ, and achievement testing, and interviewer and tester observational notes. They formed initial diagnostic assignments blind to one another's ratings. Their blind agreement were acceptable for ADHD diagnosis ($\kappa=.88$) and for other disorders with $>5\%$ base rate in the sample (all $\kappa>.68$). Disagreements were conferenced. If consensus was not readily achieved, the participant was excluded.

DMDD criteria were not yet established when data collection for the current study began. However, following prior literature (Axelson et al., 2012), for secondary analyses examining DMDD comorbidity, children were assigned an estimated DMDD diagnosis if parents: 1) endorsed the ODD symptoms “angry/resentful,” “often loses temper,” and “easily annoyed/angered” on the structured clinical interview; 2) reported symptoms were present for at least 6 months; and 3) indicated symptoms caused moderate or severe impairment.

Exclusion Criteria.—Children in both diagnostic groups were excluded at baseline if they were prescribed long-acting psychotropic medications; had neurological impairment, seizure history, head injury with loss of consciousness, other major medical conditions, or substance abuse disorder; had prior diagnosis of intellectual disability, autism spectrum disorder, or psychosis; were experiencing a major depressive episode at the time of initial evaluation (deemed at the time a confound to other studies using this population); or had estimated $IQ<70$. Children with ADHD taking stimulant medications were included.

Temperament Ratings

A parent/guardian completed the Temperament in Middle Childhood Questionnaire at each time point (TMCQ, Simonds & Rothbart, 2004). The 157 TMCQ items combine into 16 scales based on prior factor analysis (Simonds & Rothbart, 2004): *Activity Level, Affiliation, Anger/Frustration, Assertiveness/Dominance, Attention Focusing, Discomfort, Fantasy/Openness, Fear, High Intensity Pleasure, Impulsivity, Inhibitory Control, Low Intensity Pleasure, Perceptual Sensitivity, Sadness, Shyness, and Soothability/Falling Reactivity*. Scale reliabilities (Cronbach's α) ranged from 0.74 to 0.94 at all years of the study, except for the *low intensity pleasure* scale, which ranged from 0.55 to 0.73. Note that some items on the inhibitory control, impulsivity, and attention focusing scales overlap closely with ADHD symptoms (Martel & Nigg, 2006), which was taken into account in interpretation of profiles. Items were not removed. The primary driver of profiles was expected to be the emotion subscales.

Longitudinal Follow-up

Clinical assessment was identical at Baseline (Year 1), first follow-up (Year 2), and second follow up (Year 3). Each included standardized parent and teacher questionnaires, semi-structured clinical interview (KSADS), and review by the best-estimate diagnostic team. However, IQ and academic achievement measures were not repeated at Years 2 and 3.

Analysis Plan

New sample replication.—First, we attempted to replicate the results reported by Karalunas et al. (2014) using the new sample of 186 children with ADHD and the same

method: community detection analysis with the 16 temperament scales from the TMCQ used as input features. Community detection is an optimization clustering method derived from graph theory (Newman, 2006; Rubinov & Sporns, 2011). Identical data handling procedures to those reported in Karalunas et al. (2014) were followed. A quality index (Q) is obtained for each community detection analysis that conceptually represents the overall segregation between identified subgroups (“communities”). Values above $\sim .3$ are customarily considered strong evidence of the existence of subgroups within the sample (Rubinov & Sporns, 2011). Multivariate ANOVA models were used to compare latent groups on the 16 temperament dimensions used as input features. In the case of a significant multivariate effect, effects for each individual variable were considered. Post-hoc tests for pairwise comparisons applied a Tukey LSD correction.

Pooled sample: Justification.—We next evaluated the justifiability of combining the original and new samples by comparing the samples on all 16 temperament scales covarying for age differences in a 3 (Profile: Mild, Surgent, Irritable) x 2 (Sample: Original or Replication) multivariate ANCOVA. In the case of a significant multivariate effect, the effects for each individual variable were considered. Post-hoc tests for pairwise comparisons applied a Tukey LSD correction.

Pooled sample: Justification.—We next evaluated the justifiability of combining the original and new samples by comparing the samples on all 16 temperament scales covarying for age differences in a 3 (Profile: Mild, Surgent, Irritable) x 2 (Sample: Original or Replication) multivariate ANCOVA. In the case of a significant multivariate effect, the effects for each individual variable were considered. Post-hoc tests for pairwise comparisons applied a Tukey LSD correction.

To further justify combining the samples, we also tested a two-group model in *Mplus* that constrained the correlations among the 16 TMCQ subscales in both the original and the replication samples to be equal across samples. The two-group model constraining correlations to be equal was chosen over a traditional factorial invariance approach because our research question was not related to how the scales loaded on higher order factors, and we did not use the higher-order factors as part of the latent grouping analyses.

Pooled Sample: Method dependence.—Based on the results of the two-group model and ANOVA replication analyses (results reported below), we then pooled the new sample of 186 children with an additional 182 participants from Karalunas et al. (2014), making the final $N = 368$. Community detection analyses were conducted on the Year 1 data to confirm the presence of the profiles in the full sample. One child whose parent did not rate enough items to generate all of the 16 TMCQ scale scores was removed from the community detection analyses.

Next, to determine whether the profiles identified were method dependent, we conducted latent profile analyses using the Year 1 data ($N = 368$ children with ADHD) using the *Mplus* 7.4 software package and the robust maximum likelihood estimator (Muthén & Muthén, 1998–2012). Full information maximum likelihood was used to handle missing data. Non-independence of observations (i.e., the nesting of children within families) was handled

using *Mplus*' cluster command. The 1- through 10-class models were considered and the best-fitting model was selected based on Bayesian information criterion (BIC), convergence (entropy), and the Vuong-Lo-Mendell-Rubin likelihood ratio test (VLMR LRT), which assesses whether the k-class model significantly improves on the k – 1 class model (Asparouhov & Muthén, 2012).

Pooled sample: Temporal reproducibility: (a) profile structure and (b) individual assignments over time.—A series of community detection analyses were

conducted separately at Years 1 (n = 367), 2 (n = 225), and 3 (n = 271) to evaluate reproducibility of profiles over time in the same children. Then, individual assignments were compared for those children who were in more than one of these analyses.

Pooled sample: Clinical prediction.—Clinical outcome was measured in two ways. First, we considered onset of new disorders (i.e., those not present at baseline) as an indicator of worsening clinical course because it has been associated with higher rates of service utilization in the National Survey of Children's Health (Larson, Russ, Kahn, & Halfon, 2011). Comorbid diagnoses were collapsed into the following yes/no categories: 1) "Mood Disorder" (Major Depressive Disorder or Dysthymia); 2) "Anxiety Disorder" (Generalized Anxiety Disorder, Separation Anxiety Disorder, Social Anxiety Disorder, Obsessive-Compulsive Disorder, or Panic Disorder); 3) "Disruptive Behavior Disorder" (Oppositional Defiant Disorder or Conduct Disorder); and 4) "Any Disorder" (encompassing the three previous groupings). A single "new onset" variable (yes/no) was created indicating whether the child had a new onset disorder at any year of longitudinal follow up.

Second, clinical outcome was evaluated using standardized scores of parent-rated functional impairment on the Impact supplement of the SDQ (range 0–2 with higher scores indicating greater impact) (Goodman, 2001). The Impact supplement score differentiates clinical from non-clinical groups (Goodman, 1999) and best predicts service utilization (Janssens & Deboutte, 2009) as compared to other SDQ scales.

Regression models predicting outcomes from temperament type and from individual TMCQ scales were computed using *Mplus* 7.4 (Muthén & Muthén, 1998–2012) using full information maximum likelihood approaches to handle missing data. To assess for incremental validity of the temperament groups over and above existing clinical indicators, prediction analyses were conducted with age, sex, baseline ADHD symptom severity and/or baseline SDQ Impairment as described for each analysis in the results below. To assess predictive power of categorical clinical groupings, ADHD presentation, ODD diagnosis, and temperament type were also simultaneously entered into a model prediction outcome. Next, given our specific hypotheses that an Irritable temperament type would predict outcomes, we tested an additional regression model that included 1) the covariates (age, sex, baseline ADHD symptom severity and/or baseline SDQ Impairment), 2) a predictor indicating whether a child was "ever Irritable" (i.e. assigned to an Irritable type at any year of the study), and 3) a predictor indicating whether a child was "stably Irritable" (i.e., assigned to the Irritable type at all years assessed). Finally, to examine which specific negative affect dimensions predicted outcomes, we examined a prediction model in which anger (due to its

prominent association with irritability) and fear (due to its relevance to anxiety disorders) were used as predictors of outcomes.

Results

New sample reproducibility of temperament trait profiles in ADHD

Community detection in the new sample of 186 children with ADHD identified three temperament profiles at Year 1 ($Q=.47$, Figure 1). The profiles were inspected and readily labeled as: Mild, Surgent, and Irritable. Details of the Q value range and groups across runs are provided in Table S3 in the supplement.

Focusing on the emotion-related TMCQ scales, the Mild ($n=41$, 21.9% of sample), was less assertive than non-ADHD controls ($p < .001$) and had lower scores on low intensity pleasure seeking and perceptual sensitivity (all $p < .05$); they were otherwise like typically-developing children in their temperament variation. The Surgent profile ($n=76$, 40.6% of sample) had more activity and high-intensity pleasure seeking than typically-developing children or the other ADHD groups (all $p < .001$). They also had lower shyness than all other groups (all $p < .001$). The Irritable profile ($n=69$, 36.9% of sample) had the highest anger, sadness, discomfort, fear, and soothability compared to all other groups (all $p < .001$). The Surgent profile showed intermediate levels of anger, discomfort, fear, and soothability but was significantly lower on these scales than the Irritable profile (all $p < .001$).

When considering effortful control scales, all three ADHD groups differed from typically-developing children in inhibitory control and attentional focus (all $p < .002$); this was unsurprising and consistent with their ADHD assignment as noted earlier. However, the Irritable and Surgent profiles had higher impulsivity scores than typically-developing children or the Mild profile (all $p < .001$); the Mild group did not differ from typically-developing children on impulsivity ($p = .10$). Scores on all temperament scales can be found in Table S2.

Secondary checks on reproducibility

Similarity of samples.—The original and replication samples differed on age (9.2 vs 9.8 years, $p < .01$), but did not differ reliably on sex ratio, IQ, rates of comorbid diagnoses, or teacher- or parent-rated hyperactivity or inattention (all $p > .05$). Descriptive information for the original and replication samples appears in Table S1 and those for the pooled sample appear in Table 1.

How similar were the replication profiles' scores to the original sample scores?

—The original and replication samples were compared in a 3 (Profile: Mild, Surgent, Irritable) \times 2 (Sample: Original or Replication) multivariate ANCOVA comparing all 16 temperament scores and covarying for age differences between the two samples. That analysis yielded a multivariate Profile by Sample interaction ($p < .001$, $\eta^2 = .09$). The multivariate interaction was driven by differences on Affiliation ($p = .034$, $\eta^2 = .02$), Impulsivity ($p = .023$, $\eta^2 = .02$), and Soothability ($p = .045$, $\eta^2 = .02$). Post-hoc tests between samples within each temperament type indicated that the replication Mild group was less soothable ($p = .041$, $\eta^2 = .05$) and less impulsive ($p < .001$, $\eta^2 = .15$) than in the original Mild

group; and the replication Surgent group was less affiliative ($p = .040$, $\eta^2 = .03$) than the original Surgent group. The replication and original Irritable samples did not differ on any scales (all $p > .164$). We considered the basic group structure to be replicated in the two samples.

We further evaluated the justifiability of combining the two samples into a single pooled sample with a two-group model in *Mplus* that constrained the correlations among the 16 TMCQ subscales in both the original and the replication samples to be equal across samples. That model fit the data adequately ($\chi^2(120) = 159.2$, $p = .010$, CFI = .98, TLI = .97, RMSEA = .04).

Taking into considerations the minimal clinical/demographic differences, replication of the three profiles, and two-group model fit, we concluded that it was reasonable to pool the samples to maximize power for longitudinal analyses. The three profiles remained clearly present in the pooled sample ($n = 367$) at Year 1 ($Q = .48$), including: 1) a Mild type ($n = 86$, 23.4%); 2) a Surgent type ($n = 142$, 38.7%); and 3) an Irritable type ($n = 139$, 37.9%). Profiles are shown in Figure 2. Details of the Q value range and groups across runs are provided in Table S3 in the supplement. The remaining analyses use this pooled sample.

Pooled sample: Method dependence

We next examined whether results depended on clustering method by attempting clustering using baseline data in the pooled sample using latent profile analysis (Oberski, 2016). Essentially the same clusters were obtained. The latent profile analysis supported a three-class solution, with groups that were readily labeled Mild, Surgent, and Irritable. Agreement between the clustering methods about which children were assigned to each group was substantial although not perfect. The percent in the same class as community detection when subjected to the LPA was: 74.4% for Mild, 97.2% for Surgent, and 84.2% for Irritable. Remaining details about these analyses and results appear in the online Supplement, Table S4, and Figure S1.

Pooled sample: Temporal reproducibility of temperament profiles across time

Community detection analyses were repeated separately for Years 2 and 3 for the pooled sample of youth with ADHD. Three temperament profiles were also observed at Year 2 (partially overlapping sample with Karalunas et al., 2014 Year 2 analysis) and Year 3 (completely new analysis) with strong separation of groups and similar fit to the data as in Year 1 (Year 2 $Q = .48$ and Year 3 $Q = .47$). The same 3 profiles observed at Year 1 were also present at Years 2 and 3 (Figure 2). Details of the Q value range and groups across runs are provided in Table S3 in the supplement. Temperament profiles thus were reproducible over at least three annual measurements.

Within child, temperament assignment was more stable than chance across all years of assessment (all $p < .001$); however, stability significantly differed between the groups ($\chi^2 [2] = 22.63$, $p < .001$). 66% of Mild children and 61% of Irritable children were stably assigned to the same profile at *all three* years they were assessed, whereas only 36% of Surgent children were stably assigned to the same profile. For comparison, stability of ADHD presentations based on “or” algorithm (parent/teacher) symptom counts over the same period

ranged from 38% (hyperactive presentation) to 50% (combined presentation). Thus, the Irritable and Mild profiles were more stable than the ADHD DSM presentations, but the Surgent profile was not. Children with ADHD with stable temperament type assignments at all years of assessment ($n = 170$) did not differ from unstably assigned children ($n = 148$) in age, FSIQ, rates of comorbid disorders, or symptoms severity (all $p > .124$), but females were more likely to have a stable type assignment than males (47% of males v. 64% of females, $p = .013$).

Clinical relevance of temperament profiles

Baseline symptoms and comorbidity.—In the pooled sample, temperament types differed in ADHD symptom severity. The Irritable type had more inattention symptoms (all $p < .001$) than either the Mild or Surgent types, who did not differ from each other ($p = .64$). The Irritable and Surgent types had more hyperactivity-impulsivity symptoms and more total symptoms than the Mild type (all $p < .001$); Irritable and Surgent types did not differ from each other (all $p > .09$).

We next examined concurrent comorbidity patterns in each temperament type at Year 1. At baseline, children in the Irritable group were significantly more likely than other ADHD children to have comorbid anxiety ($\chi^2(2) = 25.9, p < .001$), ODD ($\chi^2(2) = 21.6, p < .001$), and DMDD ($\chi^2(2) = 9.5, p = .009$). Groups did not differ in rates of mood disorders, which have a very low prevalence in our sample at baseline ($\chi^2(2) = 0.03, p < .983$, see Table 1 for rates of comorbidity). Despite higher rates of comorbidity in the Irritable type, prevalence rates for comorbidities at baseline did not exceed 35% for any disorder, indicating that at least 65% of children in the Irritable profile had dysregulation of negative affect that did not reach diagnostic threshold for a comorbid disorder. Thus, while some differences in ADHD symptom severity and patterns of comorbidity were present, these could not fully account for temperament groups.

Clinical prediction.—Relationship of baseline temperament profile to specific clinical outcomes was examined by looking at new onsets of comorbid disorders across the 2-year follow-up period. See Table 1 for rates of new disorder onsets; data on rates of disorders for non-ADHD controls are also presented for comparison purposes. At future time points, the Irritable group had a greater chance of having a new comorbid disorder onset than either the Mild (Odds Ratio=3.3, $p < .001$) or Surgent types (Odds Ratio=1.9, $p < .001$). The Surgent and Mild types did not significantly differ in rates of new disorder onsets (Odds Ratio=1.7, $p = .186$). The Mild type did not differ from typically-developing controls, but the Surgent and Irritable types each did. When looking at onsets for mood, anxiety, and disruptive behavior disorders separately, children in the Irritable type had a higher rate of onset for anxiety disorders ($\chi^2 [2] = 8.70, p = .013$; 22% versus 10% and 11% in the Mild and Surgent types), but did not differ in rates of onsets for mood ($\chi^2 [2] = 2.38, p = .304$) or disruptive behavior disorders ($\chi^2 [2] = 1.66, p = .435$). Overall, the Irritable type predicted worse clinical outcome, supporting its validity; prediction was primarily driven by new onset of anxiety disorders.

We next examined whether prediction based on temperament profile was superior to existing prediction using baseline clinical features. Full details of all models can be found in Table S5. After controlling for age, sex, and baseline ADHD symptom severity, temperament type did not incrementally contribute to predicting rate of new disorder onsets between the Mild and Irritable types (Odds Ratio=1.9, $p = .181$), but did contribute significantly to predicting the increased rate of new onsets in the Irritable as compared to the Surgent group (Odds Ratio = 2.1, $p < .022$). See Figure 3 for a visual depiction of the models. In addition, when ADHD presentation, ODD diagnosis, and temperament type were all entered simultaneously into the prediction model, only temperament type uniquely predicted disorder onsets ($p = .006$). Baseline ODD ($p = .872$) and ADHD presentation ($p = .356$) did not contribute to prediction. See Figure 4 for a visual depiction.

Parent-reported clinical impairment on the SDQ Impact supplement served as a second measure of clinical outcome. Results here were similar. After controlling for age, sex, baseline impairment, and baseline ADHD symptom severity, temperament profile did not significantly contribute to predicting differences in overall impairment between the Mild and Irritable types at Year 3 ($p = .595$) but did predict more severe impairment at Y3 in the Irritable as compared to the Surgent profile ($p < .001$).

Given specific hypotheses that the Irritable type would be related to negative outcomes, we next probed the relative predictive power of *ever* being assigned to the Irritable type versus being *stably* assigned to the Irritable type by entering these predictors simultaneously (along with age, sex, baseline ADHD symptoms and/or baseline SDQ Impairment) into a regression model. Being *ever* Irritable and being *stably* Irritable each contributed unique prediction of SDQ Impairment (Ever Irritable $p = .004$, Stable Irritable $p = .014$). When predicting new disorder onsets, only the *ever* Irritable variable uniquely predicting outcomes (Ever Irritable $p = .006$, Stable Irritable $p = .393$).

Finally, we probed which dimensions of negative affect contributed to predicting impairment and new disorder onsets, focusing on anger (due to its prominent association with irritability in the literature as noted earlier) and fear (due to its relevance to anxiety disorders). In a model that included anger and fear controlling for age, sex, baseline impairment, and baseline ADHD symptoms, anger ($p = .030$) and fear ($p = .05$) predicted impairment at Y3 over and above other predictors. Onset of anxiety disorders was predicted by anger ($p = .012$) but not fear ($p = .892$) over and above other predictors.

Discussion

Within-group heterogeneity is recognized as a key problem for advancing assessment, diagnosis and treatment of psychiatric disorders. Recently, there has been interest in clarifying the role of emotion dysregulation generally, and chronic irritability specifically, in psychiatric nosology and its contribution to clinical prediction for ADHD (Leibenluft, 2011; Shaw et al., 2014; Stringaris, 2011; Vidal-Ribas et al., 2016). To do so, here we focus on individual differences in measures of child temperament. Although the conceptual framework for defining emotional functioning here differs somewhat from prior studies that focused solely on irritability, results dovetail with that literature and add to the picture that

dysregulation of negative affect in general, and proneness to anger in particular, are important predictors of future clinical impairment in children with ADHD. Although we focus on ADHD, the affective domains captured in temperament variation and the neurobiological systems that underlie them are relevant across many disorders (Cuthbert, 2014; Insel et al., 2010).

In an era when reproducibility is increasingly seen as vital to the field, we focused first on attempting to reproduce the results reported by Karalunas et al. (2014). The 3-group profile replicated well in a new sample. In addition, the same profiles reproduced in the pooled sample across three years, indicating stability in the overall group structure across middle childhood. Finally, the same 3 profiles were also identified using an alternative statistical grouping approach, confirming that findings are not method specific. These are reassuring findings for the potential validity of the temperament profiles.

When further examining the validity of each of the three profiles, the Irritable type stood out as more robust relative to the other two profiles. In particular, Irritable type assignment showed both respectable temporal stability (with 60% of children in this group assigned to the same profile at all three years of assessment) and superior prediction of clinical course. For perspective, it is helpful to consider that while 60% consistency in assignment is obviously below 100%, it is well above that seen for the current DSM subtyping scheme in the literature (Lahey, Pelham, Loney, Lee, & Willcutt, 2005; Todd et al., 2008) and exceeds subtype stability in our own sample when parallel methods (i.e., rating scales) are used to evaluate ADHD subtype. The Mild type showed similar stability to the Irritable type and those youth did not have worsening course, thus differentiating Mild from Irritable groups looks very promising clinically. The Surgent profile here was lower in stability and so would represent a group with inconclusive future predictive status.

Notably, the Irritable type identified here also appears more stable than DMDD. Longitudinal studies of DMDD yield only 19% stability over 2 years and as low as 1.4% over 4 years (Vidal-Ribas et al., 2016). Although we did not formally assess DSM-5 DMDD in our sample, we were able to estimate diagnoses based on previous published adaptation of DSM-IV ODD symptoms (Axelson et al., 2012); no children in our study met DMDD criteria at all years of assessment. This suggests that DMDD cutoffs may be more prone to transient or situational factors, whereas the profile identified here is more enduring. Thus, the results support use of dimensional measures to identify a moderately stable group of children with ADHD who experience high levels of irritability and other forms of negative affect.

Patterns of prospective prediction of comorbidity for the Irritable ADHD type were consistent with the longitudinal correlates of irritability measured using depression and ODD items from the DSM in population surveys (Leibenluft et al., 2006; Stringaris et al., 2009; Stringaris et al., 2012) and recent meta-analysis (Vidal-Ribas et al., 2016), which find that irritability most consistently predicts mood and anxiety disorders. We similarly found that the Irritable ADHD type specifically predicted anxiety, but not mood or disruptive behavior disorder onsets. Although the Irritable ADHD type did not predict higher rates of mood disorders in the present study, follow up as children progress to adolescence (only

20% of the sample was above age 12 years at the final assessment), when the onset of mood disorders normatively increases, may yield a different picture. We are undertaking that follow up now.

Interestingly, while the Irritable group was reasonably stable, being assigned to the Irritable group at a single time point added significantly to prediction of impairment beyond the prediction based on stable Irritable type assignments. When predicting new disorder onsets, stable assignment to the Irritable type did not add to prediction beyond being assigned even once to the Irritable type. Thus, it may be that even transient experience of mood dysregulation in the form of irritability is an important clinical risk indicator. Additional work to clarify the longitudinal stability of continuous traits underlying Irritable type assignment and the relationships to clinical outcomes will be important and we are undertaking this work now.

Temperament profile did not add to clinical prediction when comparing the Mild and Irritable types, in part, because they had relatively large differences in ADHD symptom severity at baseline. However, temperament type did contribute to predicting differences between the Surgent and Irritable types, who started out with similar levels of total ADHD symptoms at baseline. Findings suggest that while symptom severity remains an important predictor of clinical course, the addition of temperament or other emotional-related measures to assessment, particularly in the case of relatively more severe ADHD symptoms, adds incremental validity. Together, results suggest the possibility of identifying an ADHD symptom cutoff beyond which knowing a child's temperament profile is particularly clinically informative.

In cases where symptoms and temperament measures are equally predictive, a focus on temperament profiles may still provide a clearer mechanistic perspective for understanding individual differences in prognosis. When considering specific dimensions of affect, anger uniquely predicted clinical outcomes for the Irritable group, suggesting that proneness to anger may specifically drive risk for worsening clinical course. Anger was more predictive than fear, despite anxiety disorders being the most common new disorder onset in our sample. Together, findings suggest that a differentiation of ADHD youth with and without "anger/irritability" has promise to powerfully reflect differential prognosis.

The profiles did not show some other patterns that might have been expected. The "Irritable" profile included 65% males, similar to the overall gender distribution in our sample and not significantly different than the Mild temperament type. Although the Inattentive ADHD presentation was more likely to be represented in the Mild than other types, the Irritable ADHD type nonetheless included children from all three DSM-5 ADHD symptom presentations. (Nearly 30% of children with the Inattentive presentation were in the Irritable group.)

Critically, although the Irritable type had higher rates of ODD and DMDD than the other groups, 65% of Irritable children did not meet diagnostic criteria for these disorders. Further, 27% of children with DMDD were *not* in the Irritable type and 42% of children with ODD were *not* in the Irritable type. These results suggest this Irritable type is not reducible to a

simplified ADHD+ODD or ADHD+DMDD comorbidity pattern. Instead it may be a more precise refinement of the clinical structure. This is consistent with prior work identifying many children who experience chronic and impairing irritability, even when DMDD criteria are not met (Deveney et al., 2015). Overall, results converge on the importance of the subgroup of children with ADHD and high levels of negative affect and irritability, which is distinct from diagnosis of ODD or other comorbid conditions.

Several additional areas for continued study are also notable. In choosing a statistical grouping algorithm, we attempted to be consistent with our prior work (Fair, Bathula, Nikolas, & Nigg, 2012; Karalunas et al., 2014), but our secondary analysis showed the results are not method dependent; additional computational approaches may enable further refinement (Bassett et al., 2013). In addition, alternative subtyping schemes based on latent symptoms subgroups have also been proposed (Elia et al., 2009; Freitag, Rohde, Lempp, & Romanos, 2010; Neuman et al., 1999; Rasmussen et al., 2004; Volk, Todorov, Hay, & Todd, 2009) and need to be directly compared to temperament types. Finally, we focused here on whether reproducible, stable groups could be identified at single time points. These analyses mirror the common scenario in which a child presents for diagnostic assessment at a clinic, but ratings from the past are not available and future outcome is not yet known. Thus, assessing temporal reproducibility of groups and stability group assignments using models that consider only information from a single time is a conservative approach and important for understanding clinical applicability. Alternative models that simultaneously consider information from multiple time points, such as latent class growth analyses or multi-year LPA models, will also be important for fully elucidating the relationship between temperament traits and clinical outcomes. We are undertaking this work now.

Summary

Current results add to evidence that emotional functioning is a critical domain of heterogeneity in ADHD. Children with ADHD and an Irritable temperament profile appear to reflect a moderately stable subgroup in the ADHD population with increased risk for negative clinical outcomes. Results build the case for chronic irritability as a clinically-relevant trait across multiple forms of psychopathology, and further highlight one approach to translating the stability of measures of trait negative affect into categorical groupings for clinical application. Together, work suggests that revision of psychiatric nosology based on neurobiologically-informed psychological dimensions as envisioned in RDoC (Cuthbert, 2014; Insel et al., 2010), appears tractable. Inclusion of an ADHD specifier based on the presence or absence of irritability warrants further examination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

This project was supported by R37 MH059105 (PI: Nigg). The corresponding author's time was also supported by K23 MH108656 (PI: Karalunas). Dr. Gustafsson's time was supported, in part, by National Center for Advancing Translational Sciences of the National Institutes of Health under award number TL1TR000129. Dr. Musser's time was supported, in part by R03 MH110812 (PI: Musser).

References

- Asendorpf JB, Conner M, De Fruyt F, De Houwer J, Denissen JJ, Fiedler K, ... Nosek BA (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27(2), 108–119.
- Asparouhov T, & Muthén B (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes. *Mplus Web Notes*, 14, 22.
- Axelson D, Findling RL, Fristad MA, Kowatch RA, Youngstrom EA, Horwitz SM, ... Demeter C (2012). Examining the proposed disruptive mood dysregulation disorder diagnosis in children in the Longitudinal Assessment of Manic Symptoms study. *The Journal of clinical psychiatry*, 73(10), 1342. [PubMed: 23140653]
- Bassett DS, Porter MA, Wymbs NF, Grafton ST, Carlson JM, & Mucha PJ (2013). Robust detection of dynamic community structure in networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(1), 013142.
- Bates JE, Goodnight JA, & Fite JE (2008). *Temperament and Emotion*. In Lewis M, Haviland-Jones JM & Barrett LF (Eds.), *Handbook of Emotion* (3 ed.). New York, NY: The Guilford Press.
- Clark LA, Cuthbert B, Lewis-Fernández R, Narrow WE, & Reed GM (2017). Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest*, 18(2), 72–145. [PubMed: 29211974]
- Conners CK (2003). *Conners' rating scales: Revised technical manual*. New York, NY: Multi-Health Systems.
- Cuthbert BN (2014). The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, 13(1), 28–35. [PubMed: 24497240]
- Deveney CM, Hommer RE, Reeves E, Stringaris A, Hinton KE, Haring CT, ... Leibenluft E (2015). A prospective study of severe irritability in youths: 2- and 4-year follow-up. *Depress Anxiety*, 32(5), 364–372. doi:10.1002/da.22336. [PubMed: 25504765]
- DuPaul G, Power T, Anastopoulos A, & Reid R (1998). *ADHD Rating Scale—IV: Checklists, Norms, and Clinical Interpretation*. NY, NY: Guilford Press.
- Elia J, Arcos-Burgos M, Bolton KL, Ambrosini PJ, Berrettini W, & Muenke M (2009). ADHD latent class clusters: DSM-IV subtypes and comorbidity. *Psychiatry Research*, 170(2), 192–198. [PubMed: 19900717]
- Everitt BS, Landau S, Leese M, & Stahl D (2011). *Cluster Analysis* (5th ed.). West Sussex, United Kingdom: John Wiley & Sons, Ltd.
- Fair D, Bathula D, Nikolas M, & Nigg JT (2012). Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *Proceedings of the National Academy of Sciences of the United States of America*, 10.1073/pnas.1115365109.
- Freitag CM, Rohde LA, Lempp T, & Romanos M (2010). Phenotypic and measurement influences on heritability estimates in childhood ADHD. *European Child & Adolescent Psychiatry*, 19(3), 311–323. [PubMed: 20213230]
- Goodman R (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *J Child Psychol Psychiatry*, 40(5), 791–799. [PubMed: 10433412]
- Goodman R (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337–1345. [PubMed: 11699809]
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, & Conde JG (2009). Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–381. [PubMed: 18929686]
- Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, ... Wang P (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The*

American Journal of Psychiatry, 167(7), 748–751. doi:10.1176/appi.ajp.2010.09091379. [PubMed: 20595427]

- Ioannidis JP (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7(6), 645–654. [PubMed: 26168125]
- Janssens A, & Deboutte D (2009). Screening for psychopathology in child welfare: the Strengths and Difficulties Questionnaire (SDQ) compared with the Achenbach System of Empirically Based Assessment (ASEBA). *Eur Child Adolesc Psychiatry*, 18(11), 691–700. doi:10.1007/s00787-009-0030-y. [PubMed: 19462154]
- Karalunas SL, Fair D, Musser ED, Aykes K, Iyer SP, & Nigg JT (2014). Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: toward biologically based nosologic criteria. *JAMA Psychiatry*, 71(9), 1015–1024. [PubMed: 25006969]
- Kircanski K, Zhang S, Stringaris A, Wiggins JL, Towbin KE, Pine DS, ... Brotman MA (2016). Empirically derived patterns of psychiatric symptoms in youth: A latent profile analysis. *J Affect Disord*. doi:10.1016/j.jad.2016.09.016.
- Lahey BB, Pelham WE, Loney J, Lee SS, & Willcutt E (2005). Instability of the DSM-IV subtypes of ADHD from preschool through elementary school. *Archives of General Psychiatry*, 62(8), 896–902. [PubMed: 16061767]
- Larson K, Russ SA, Kahn RS, & Halfon N (2011). Patterns of comorbidity, functioning, and service use for US children with ADHD, 2007. *Pediatrics*, peds. 2010–0165.
- Leibenluft E (2011). Severe mood dysregulation, irritability, and the diagnostic boundaries of bipolar disorder in youths. *The American Journal of Psychiatry*, 168(2), 129. [PubMed: 21123313]
- Leibenluft E, Blair RJR, Charney DS, & Pine DS (2003). Irritability in pediatric mania and other childhood psychopathology. *Annals of the New York Academy of Sciences*, 1008(1), 201–218. [PubMed: 14998886]
- Leibenluft E, Cohen P, Gorrindo T, Brook JS, & Pine DS (2006). Chronic Versus Episodic Irritability in Youth: A Community-Based, Longitudinal Study of Clinical and Diagnostic Associations. *Journal of Child & Adolescent Psychopharmacology*, 16(4), 456–466. [PubMed: 16958570]
- Lenzenweger MF, Clarkin JF, Yeomans FE, Kernberg OF, & Levy KN (2008). Refining the borderline personality disorder phenotype through finite mixture modeling: Implications for classification. *Journal of Personality Disorders*, 22(4), 313. [PubMed: 18684047]
- Makel MC, Plucker JA, & Hegarty B (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7(6), 537–542. [PubMed: 26168110]
- Martel MM, & Nigg JT (2006). Child ADHD and personality/temperament traits of reactive and effortful control, resiliency, and emotionality. *Journal of Child Psychology and Psychiatry*, 47(11), 1175–1183. doi:10.1111/j.1469-7610.2006.01629.x. [PubMed: 17076757]
- Muthén LK, & Muthén BO (1998–2012). *MPLUS User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Neuman RJ, Todd RD, Heath AC, Reich W, Hudziak JJ, Bucholz KK, ... Kuperman S (1999). Evaluation of ADHD typology in three contrasting samples: a latent class approach. *Journal of the American Academy of Child & Adolescent Psychiatry*, 38(1), 25–33. [PubMed: 9893413]
- Newman MEJ (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8577–8582. [PubMed: 16723398]
- Nigg JT (2006). Temperament and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 47(3–4), 395–422. [PubMed: 16492265]
- Oberski D (2016). Mixture models: Latent profile and latent class analysis *Modern statistical methods for HCI* (pp. 275–287): Springer, Cham.
- Pashler H, & Wagenmakers EJ (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528–530. [PubMed: 26168108]
- Posner MI, & Rothbart MK (2000). Developing mechanisms of self-regulation. *Development and Psychopathology*, 12(03), 427–441. [PubMed: 11014746]
- Puig-Antich J, & Ryan N (1986). *Kiddie schedule for affective disorders and schizophrenia*. Pittsburgh, PA: Western Psychiatric Institute.

- Rasmussen ER, Neuman RJ, Heath AC, Levy F, Hay DA, & Todd RD (2004). Familial clustering of latent class and DSM-IV defined attention-deficit/hyperactivity disorder (ADHD) subtypes. *Journal of Child Psychology and Psychiatry*, 45(3), 589–598. [PubMed: 15055377]
- Rothbart MK (2011). *Becoming who we are: Temperament and personality in development*. New York, NY: Guilford Press.
- Rothbart MK, & Bates JE (1998). *Temperament Handbook of Child Psychology*.
- Rothbart MK, Derryberry D, & Posner MI (1994). A psychobiological approach to the development of temperament In Bates JE & Wachs TD (Eds.), *Temperament: Individual differences at the interface of biology and behavior* (pp. 83–116). Washington, DC: American Psychological Association.
- Rubinov M, & Sporns O (2011). Weight-conserving characterization of complex functional brain networks. *Neuroimage*, 56(4), 2068–2079. [PubMed: 21459148]
- Sattler J, & Dumont R (2004). *Assessment of Children: WISC-IV and WPPSI-III Supplement*. San Diego: Jerome Sattler Publisher, Inc.
- Shaw P, Stringaris A, Nigg J, & Leibenluft E (2014). Emotion dysregulation in attention deficit hyperactivity disorder. *American Journal of Psychiatry*, 171(3), 276–293. [PubMed: 24480998]
- Simonds J, & Rothbart MK (2004). The Temperament in Middle Childhood Questionnaire (TMCQ): A computerized self-report measure of temperament for ages 7–10. Paper presented at the Occasional Temperament Conference, Athens, GA.
- Stringaris A (2011). Irritability in children and adolescents: a challenge for DSM-5. *European Child & Adolescent Psychiatry*, 20(2), 61–66. [PubMed: 21298306]
- Stringaris A, Cohen P, Pine DS, & Leibenluft E (2009). Adult outcomes of youth irritability: a 20-year prospective community-based study. *The American Journal of Psychiatry*, 166(9), 1048–1054. [PubMed: 19570932]
- Stringaris A, Zavos H, Leibenluft E, Maughan B, & Eley TC (2012). Adolescent irritability: phenotypic associations and genetic links with depressed mood. *American Journal of Psychiatry*, 169(1), 47–54. [PubMed: 22193524]
- Todd RD, Huang H, Todorov AA, Neuman RJ, Reiersen AM, Henderson CA, & Reich WC (2008). Predictors of stability of attention-deficit/hyperactivity disorder subtypes from childhood to young adulthood. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(1), 76–85. [PubMed: 18174828]
- Vidal-Ribas P, Brotman MA, 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*.
- Volk HE, Todorov AA, Hay DA, & Todd RD (2009). Simple identification of complex ADHD subtypes using current symptom counts. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4), 441–450. [PubMed: 19318883]
- Wakschlag LS, Estabrook R, Petittclerc A, Henry D, Burns JL, Perlman SB, ... Briggs-Gowan ML (2015). Clinical Implications of a Dimensional Approach: The Normal:Abnormal Spectrum of Early Irritability. *J Am Acad Child Adolesc Psychiatry*, 54(8), 626–634. doi:10.1016/j.jaac.2015.05.016. [PubMed: 26210331]
- Wechsler D (2002). *Wechsler Individual Achievement Test, 2nd Ed (WIAT-II) Examiner's Manual*. San Antonio: Harcourt Brace.
- Wechsler D (2003). *Wechsler Intelligence Scale for Children, 4th Ed (WISC-IV) Technical and Interpretive Manual*. San Antonio: Harcourt Brace.
- Whittle S, Allen NB, Lubman DI, & Yücel M (2006). The neurobiological basis of temperament: Towards a better understanding of psychopathology. *Neuroscience & Biobehavioral Reviews*, 30(4), 511–525. [PubMed: 16289282]

Public Significance Statement.

This work identifies a group of children with ADHD and irritable mood who are at increased risk for negative outcomes. Addition of an specifier in the ADHD diagnostic criteria based on the presence or absence of irritability could be considered.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

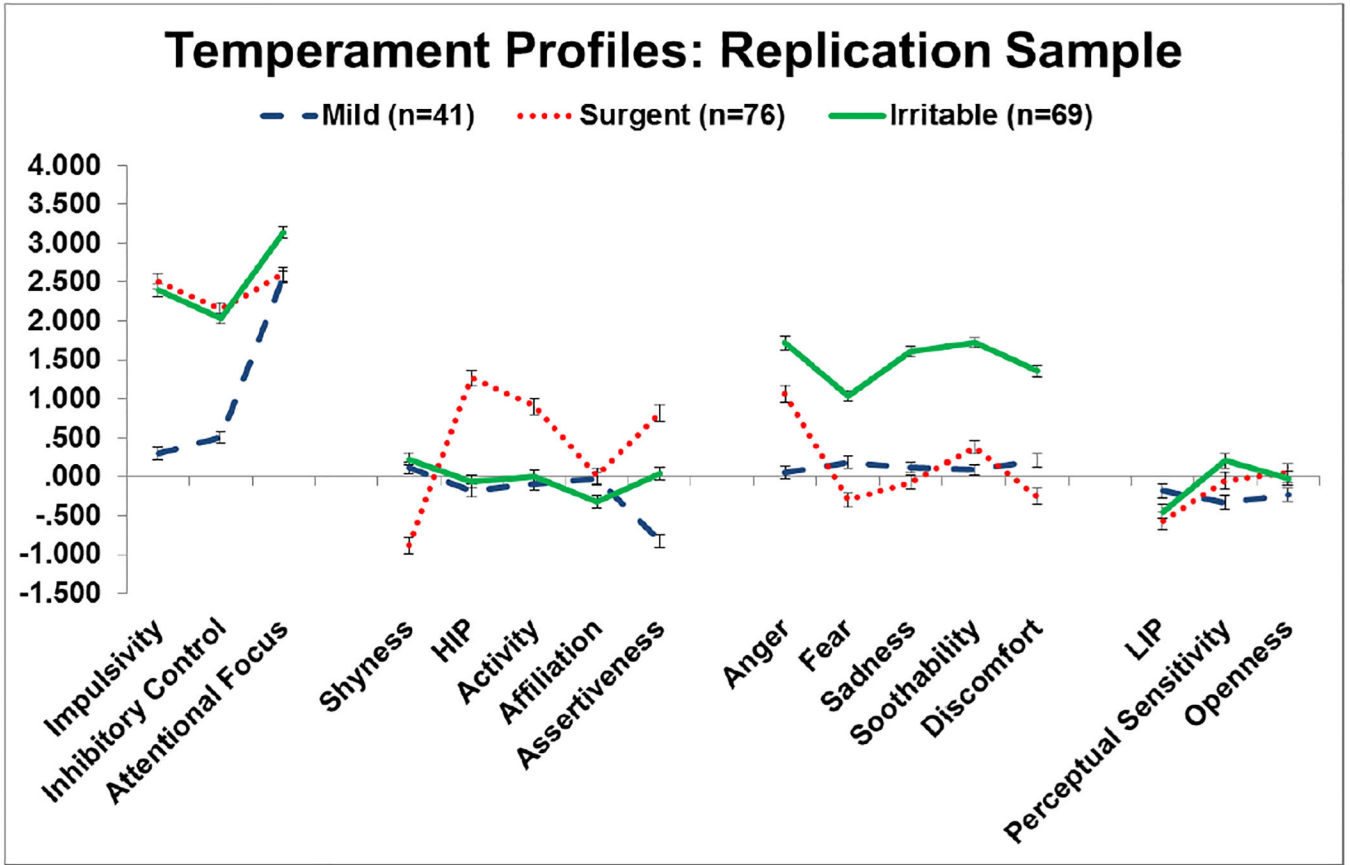


Figure 1. depicts the three temperament profiles in the replication sample. Scores are presented as z-scores relative to the non-ADHD group mean. The Soothability, Inhibition, and Attentional Focus scales are recoded such that higher scores indicate less of these traits (i.e., more impairment in these areas). HIP= High Intensity Pleasure, LIP= Low Intensity Pleasure. Error bars indicate standard error.

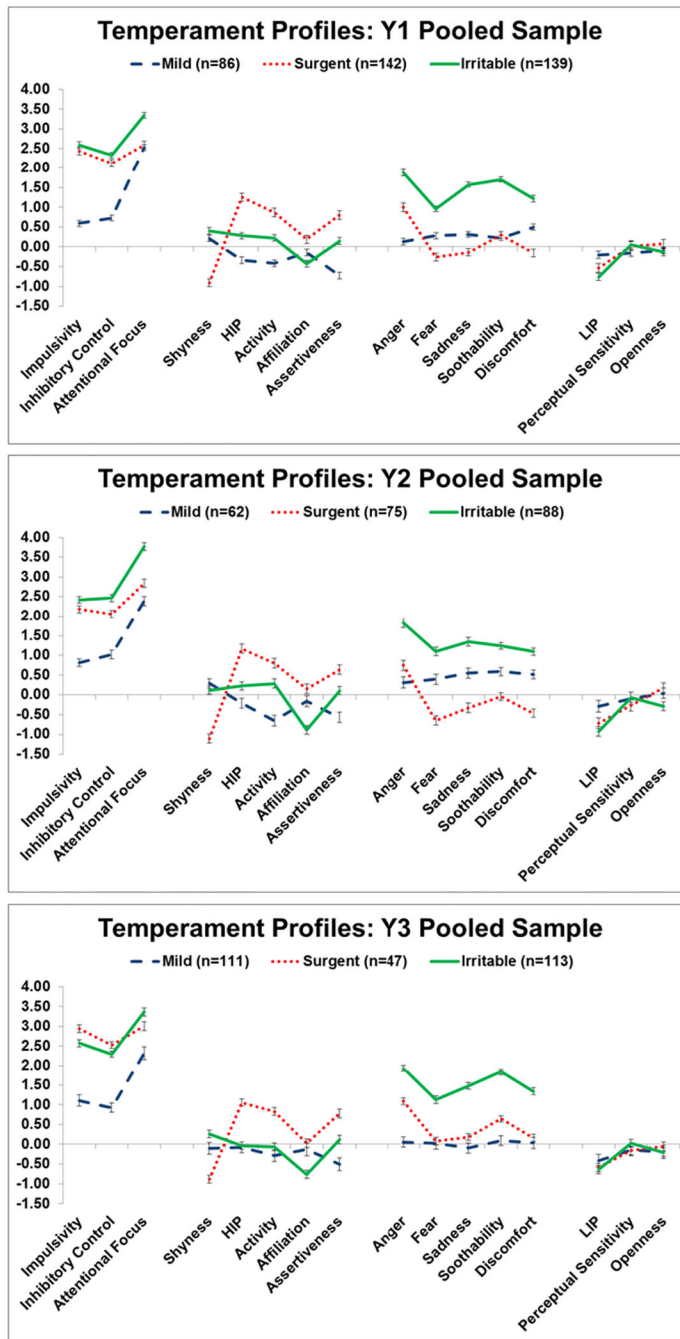


Figure 2. depicts the three temperament profiles in the pooled sample at Y1 (A), Y2 (B), and Y3 (C). Scores are presented as z-scores relative to the non-ADHD group mean. The Soothability, Inhibition, and Attentional Focus scales are recoded such that higher scores indicate less of these traits (i.e., more impairment in these areas). HIP= High Intensity Pleasure, LIP= Low Intensity Pleasure. Error bars indicate standard error.

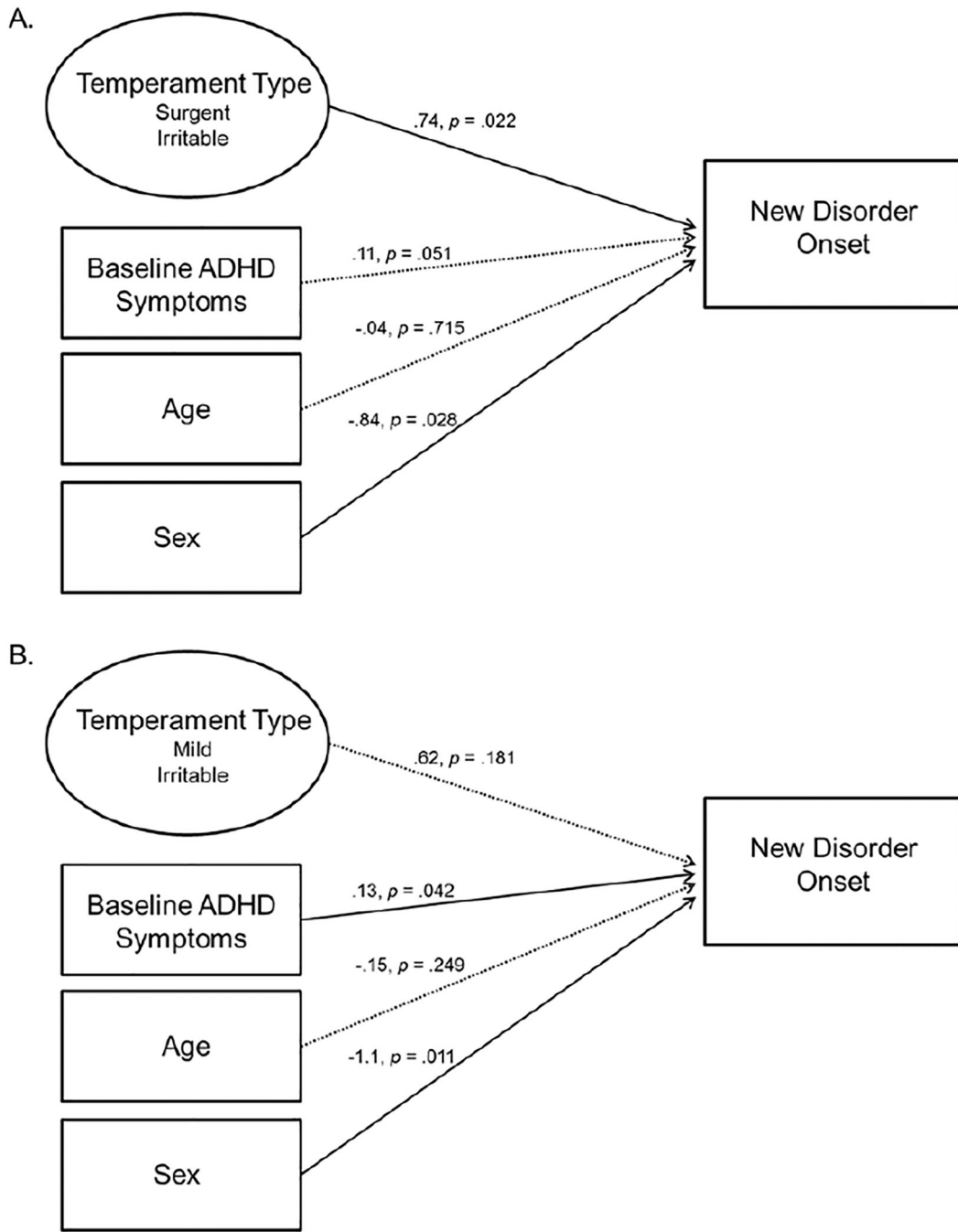


Figure 3. depicts the clinical prediction model with temperament type and baseline ADHD symptom severity.

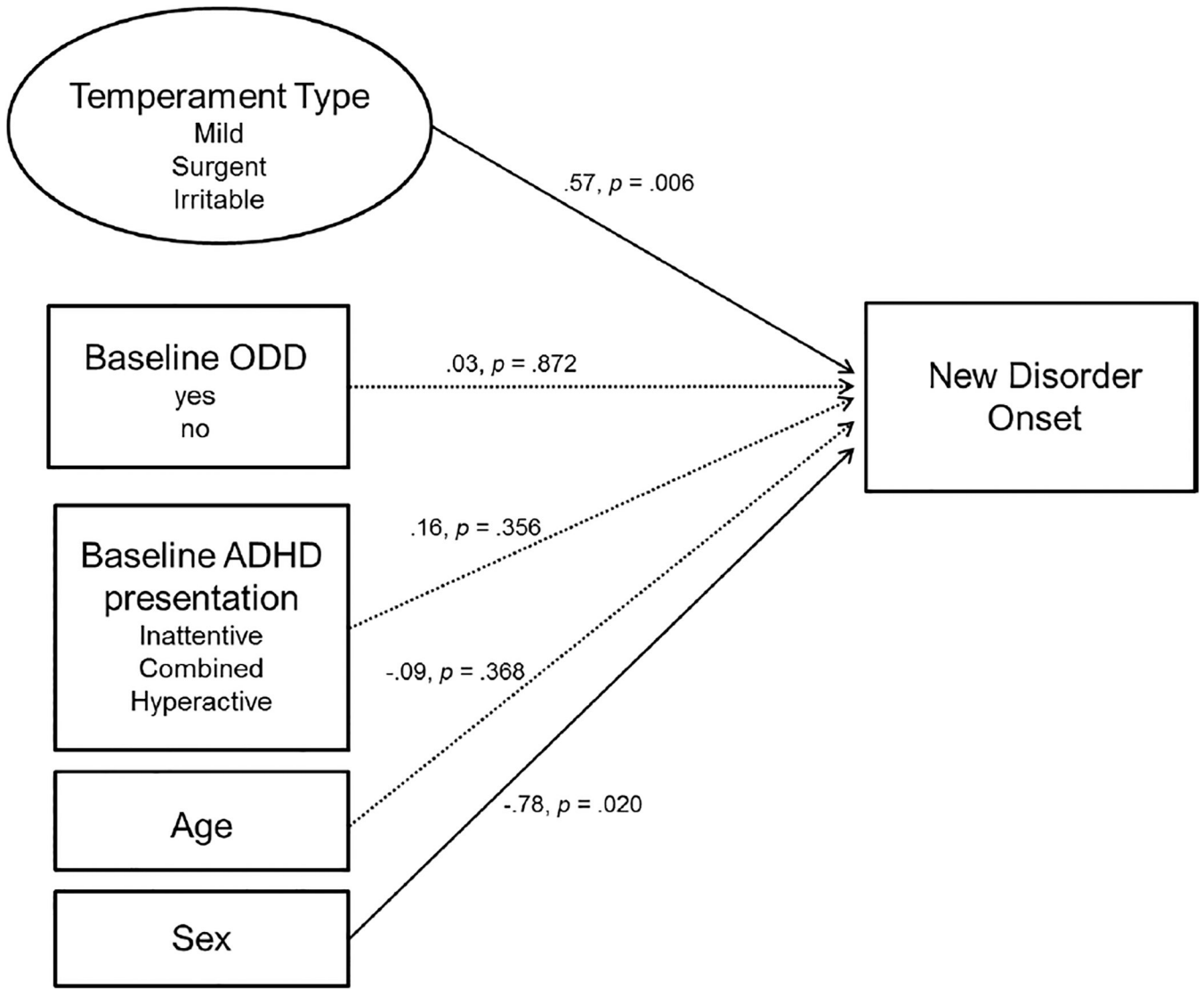


Figure 4. depicts the clinical prediction model including temperament type, ADHD presentation, and baseline ODD diagnosis.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Descriptive Information and Longitudinal Outcomes for the Pooled Sample

	Control	ADHD	Mild	Surgent	Irritable	P comparing ADHD subtypes	partial η^2	Post-hoc
Year 1								
Gender (% boys)	50.5%	69.4%	60.5%	78.9%	64.7%	0.005	---	M, I < S
DSM-5 Presentations (H:FC)	---	12:97:264	2:54:30	8:15:119	2:27:110	<.001	---	S, I < M ^a
Full Scale IQ	115.8 (12.2)	108.1 (14.3)	110.0 (14.1)	109.0 (13.0)	107.4 (15.1)	0.373	0.01	---
Age (years)	8.8 (1.4)	9.5 (1.5)	9.7 (1.6)	9.0 (1.5)	9.3 (1.3)	0.003	0.03	S, I < M
% on Stimulants	---	36.7%	24.4%	31.0%	50.4%	<.001	---	M, S < I
"OR" Algorithm Int Symptoms	0.3 (0.7)	7.5 (1.9)	7.2 (2.0)	7.0 (2.1)	8.1 (1.4)	<.001	0.07	M, S < I
"OR" Algorithm Hyp Symptoms	0.4 (0.9)	5.9 (2.7)	3.4 (2.5)	6.8 (2.2)	6.4 (2.4)	<.001	0.27	M < S, I
Mood Disorder (%)	0.0%	1.3%	1.2%	1.4%	1.4%	0.983	---	---
Anxiety Disorder (%)	8.1%	20.6%	19.8%	9.2%	33.8%	<.001	---	S < M < I
ODD (%)	1.0%	19.3%	4.7%	17.6%	29.5%	<.001	---	M < S < I
CD (%)	0.0%	1.9%	0.0%	3.5%	0.7%	0.071	---	---
DMDD (%)	0.5%	4.0%	0.0%	2.8%	8.0%	0.009	---	M < I
Year 2								
"OR" Algorithm Int Symptoms	0.8 (1.6)	6.9 (2.5)	5.5 (3.0)	6.7 (2.5)	7.9 (1.7)	<.001	0.13	M < S < I
"OR" Algorithm Hyp Symptoms	0.6 (1.5)	5.4 (2.9)	2.6 (2.5)	6.5 (2.5)	5.9 (2.6)	<.001	0.28	M < S, I
Mood Disorder (%)	1.4%	2.4%	3.8%	1.1%	3.2%	0.498	---	---
Anxiety Disorder (%)	5.2%	22.8%	15.4%	9.5%	39.4%	<.001	---	M, S < I
ODD (%)	2.1%	11.8%	3.8%	10.5%	17.0%	0.054	---	M < I
CD (%)	0.0%	0.8%	0.0%	1.1%	1.1%	0.758	---	---
DMDD (%)	0.0%	2.8%	1.9%	1.0%	4.3%	0.348	---	---
Year 3								
"OR" Algorithm Int Symptoms	1.0 (1.7)	6.2 (2.8)	5.3 (2.9)	6.2 (2.8)	6.8 (2.6)	0.019	0.04	M < S, I
"OR" Algorithm Hyp Symptoms	0.7 (1.5)	4.1 (3.0)	2.1 (2.3)	4.6 (3.0)	4.8 (2.9)	<.001	0.15	M < S, I
Mood Disorder (%)	0.0%	2.9%	2.9%	1.0%	4.9%	0.261	---	---
Anxiety Disorder (%)	5.7%	15.6%	17.4%	7.0%	21.6%	0.013	---	---
ODD (%)	2.4%	9.8%	4.3%	7.0%	15.7%	0.026	---	---
CD (%)	0.0%	0.7%	0.0%	0.0%	2.0%	0.188	---	---

	Control	ADHD	Mild	Surgent	Irritable	P comparing ADHD subtypes	partial η^2	Post-hoc
<i>DMDD (%)</i>	1.1%	1.1%	0.0%	2.0%	1.1%	0.471	---	
New Onsets								
<i>Mood Disorder Onsets (%)</i>	1.1%	3.1%	2.5%	1.7%	5.0%	0.304	---	
<i>Anxiety Disorder Onsets (%)</i>	7.1%	15.7%	10.0%	10.8%	22.7%	0.013	---	
<i>ODD/CD Onsets (%)</i>	2.9%	8.6%	5.0%	9.0%	10.0%	0.435	---	
<i>DMDD Onsets (%)</i>	0.5%	2.2%	1.3%	2.5%	1.7%	0.807	---	
<i>Any Onset (%)</i>	10.5%	23.0%	12.5%	19.7%	31.9%	0.004	---	

^aPost-hoc indicates that Inattentive presentation children were more likely to be in the Mild type