Assessing General Versus Specific Liability for Externalizing Problems in Adolescence: Concurrent and Prospective Prediction of Symptoms of Conduct Disorder, ADHD, and Substance Use

Supplemental Material

Supplemental Method A:

Questionnaires that Served as Sources of Candidate Items for IMAGEN-Callousness Scale

Callousness-relevant items including in existing inventories administered to IMAGEN participants were identified via consensus ratings completed during the development of the IMAGEN-Disinhibition scale (Brislin et al., 2019). These items were used as the pool of candidate items from which the IMAGEN-Callousness scale was derived. To facilitate prospective analyses and future research with this sample, only items administered at both T1 and T2 were considered for inclusion in the scale. Existing inventories containing items previously identified as relevant to callousness are as follows.

NEO-FFI. The Neuroticism-Extraversion-Openness Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) is a 60-item questionnaire that indexes the five-factor model of personality. Participants responded using a five-point Likert scale ranging from "strongly disagree" to "strongly agree." For the purposes of scale development, responses were re-scaled as follows: 0 (strongly disagree) remained 0, 1 remained 1, 2 became 1.5, 3 became 2, and 4 (strongly agree) became 3 (see Brislin et al., 2019).

SDQ. The Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001) consists of 25 items assessing adolescent psychopathology in the domains of hyperactivity/inattention, behavior problems, peer problems, emotion dysregulation, and general psychosocial well-being. Participants responded on a three-point Likert scale ranging from "not true" to "certainly true." Items were re-scaled following Brislin et al. (2019), such that 0 (not true) remained 0, 1 became 1.5, and 2 (certainly true) became 3.

ROBVQ. The Revised Olweus Bully/Victim Questionnaire for Students (ROBVQ; Olweus, 1996) consists of 40 items assessing frequency of different types of bullying and victimization, including physical, verbal, and racial bullying, as well as settings, attitudes, and reactions

surrounding bullying. In the IMAGEN sample, a modified 12-item version was administered, consisting of eight items from the ROBVQ (e.g., “I was bullied at school”; “I took part in bullying another student at school”) as well as four additional items pertaining to non-peer bullying: “I have been bullied by a teacher”; “I have been bullied by a family member”; “I have bullied a teacher”; and “I have bullied a family member.” Participants responded to items according to frequency of these experiences over the last six months on a five-point Likert scale, ranging from “none” to “several times a week.” During re-scaling, 1 (none) became 0, 2 became 1, 3 became 2, 4 became 2.5, and 5 (several times a week) became 3. The scaling was performed asymmetrically for ROBVQ items due to the relative infrequency of higher-coded responses.

SURPS. The Substance Use Risk Profile Scale (SURPS; Woicik et al., 2009) consists of 23 items assessing four domains of personality associated with risk for substance use problems: anxiety sensitivity, hopelessness, sensation-seeking, and impulsivity. Participants responded on a four-point Likert scale ranging from “strongly disagree” to “strongly agree.” Following Brislin et al. (2019), items were re-scaled such that 1 (strongly disagree) became 0, 2 became 1, 3 became 2, and 4 (strongly agree) became 3.

TCI-R: NSS. The 35-item Novelty-Seeking Scale (NSS) of the Temperament and Character Inventory – Revised (TCI-R; Cloninger, 1999) consists of four subscales: exploratory impulsiveness, disorderliness, excitability, and extravagance. Participants responded on a five-point Likert scale ranging from “definitely false” to “definitely true.” Following Brislin et al. (2019), during re-scaling, 1 (definitely false) became 0, 2 became 1, 3 became 1.5, 4 became 2, and 5 (definitely true) became 3.

Supplemental Method B:

Detailed Description of Procedural Steps for Developing the IMAGEN-Callousness Scale

Scale construction occurred in two steps: (1) initial selection of candidate items for each scale, followed by (2) refinement of these initial scale item sets. This process has been used in several previous scale development efforts (e.g., Brislin et al., 2015, 2019; Drislane et al., 2015, 2018, 2019; Hall et al., 2014; Sellbom et al., 2016).

Candidate item selection. Candidate items were identified based on consensus ratings performed during the development of the IMAGEN-Disinhibition scale (Brislin et al., 2019). Five evaluators — two PhD-level psychologists and three clinical psychology graduate students — were provided with written definitions of callousness and disinhibition[1] and asked to rate the extent to which each item administered to the IMAGEN sample represented each construct on the following scale: unrelated to construct; strongly represents high level of construct; somewhat represents high level of construct; somewhat represents low level of construct; or strongly represents low level of construct. Items that were (1) identified by all five raters as representing high (i.e., extremely callous) or low (i.e., extremely affiliative/empathic) levels of callousness, (2) rated as more relevant to callousness than to disinhibition, and (3) not included in the IMAGEN-Disinhibition scale were included in the pool. The candidate item pool was further restricted to those from questionnaires administered at each time point to facilitate longitudinal research using the IMAGEN-Callousness scale, resulting in a pool of 26 candidate items. Items that were worded in the direction of higher empathy (e.g., "I try to be courteous to everyone I meet") were reverse-coded. Each item was re-scaled to a 0 to 3 scale for the purposes of computation of a total score, in line with Brislin et al. (2019). Specific information regarding recoding is provided in Supplemental Method A.

[1] The Construct Definition Form can be found in the online supplement to Hall et al. (2014), http://supp.apa.org/psycarticles/supplemental/a0035665/a0035665_supp.html.

Scale refinement. Next, the T1 sample was divided into two halves matched for site, sex, age, and disinhibition scores, allowing for a split-half cross-validation approach. Following prior work (e.g., Brislin et al., 2019), the scale was refined iteratively in the first half of the sample (n = 735) using the 26 candidate items. Items operating to reduce internal consistency reliability (Cronbach's alpha) as a function of low item-total correlations (i.e., less than r = .10) were removed. To ensure adequate content coverage, items rated as strongly representing high or low callousness by only three or four of five raters (n = 20) were also considered for inclusion based on strength of item-total correlations, benefit to internal consistency, and breadth of content coverage. These procedures were applied iteratively to the 46 candidate items, with effects of adding or removing items evaluated step-by-step, and priority for retention given to items that improved content coverage and item-polarity balance (i.e., similar representation of positively and negatively worded items). The final scale consisted of 17 items with acceptable to good internal consistency reliability (α = .75; Nunnally & Bernstein, 1994) and a mean inter-item correlation of r = .15, indicating coherence among items with representation of a broad range of content (see Brislin et al., 2019; Patrick, Kramer, et al., 2013). The final scale included eight items worded in the direction of higher empathy, which had been reverse-coded. Item numbers, adjusted item-total correlations, and content area for the final 17-item scale are presented in Supplemental Table A.

The second half of the sample (n = 736) was used to confirm the psychometric properties of the finalized scale. Cronbach's alpha and the mean inter-item correlation for this sample half were comparable to those for the first sample half (α = .77, mean inter-item r = .17; see Supplemental Table A).

Supplemental Discussion:

Scale Development, Harmonization, and Neurobehavioral Traits: Setting the Stage for Research Using Consortium Studies

A noteworthy implication of the current study pertains to the possibility of applying similar scale development approaches in other large datasets. Pre-existing measures of disinhibition and callousness were not administered to IMAGEN participants. However, building on research explicating the psychological nature and correlates of these traits, we were able to effectively operationalize callousness and disinhibition using conceptually relevant items from the IMAGEN questionnaire battery. This approach allowed us to add to the literature showing that these traits longitudinally predict relevant forms of psychopathology (Elkins et al., 2006; Frick & White, 2008; Krueger, 1999). Using a similar method, it should be possible to operationalize other traits of interest in large, longitudinal, and genetically informative samples to examine whether they also operate as liability factors for general or specific forms of psychopathology.

An important consideration in future work of this kind is that not all traits are equally likely to function as liability factors as defined by Perkins et al. (2020a) — that is, as prospective predictors of psychopathology that share heritable variance with the outcome of interest. A notable feature of disinhibition and callousness is their robust association with neurophysiological and task-behavioral indices, which supports their characterization as neurobehavioral traits (Patrick et al., 2009, 2012). Although the current work did not include brain or task-behavioral measures, a previous paper by Brislin et al. (2019) evaluated the IMAGEN-Disinhibition scale in an American undergraduate sample and found that it covaried strongly (r~.8) with the Disinhibition factor scale (Patrick et al., 2013a) of the Externalizing Spectrum Inventory (ESI; Krueger et al., 2007), which shows replicable associations with brain and behavioral variables (e.g., Venables et al., 2018a; Yancey et al., 2013). Brislin et al. (2019) also found that IMAGEN-Disinhibition, like ESI-Disinhibition, correlated significantly with reduced

cognitive-brain response. Our finding of a comparably high correlation for IMAGEN-Callousness with the ESI's counterpart Callous-Aggression scale (r = .73) suggests that it would likely share neurophysiological and behavioral correlates observed for this pre-existing callousness scale (Brislin et al., 2018; Brislin & Patrick, 2019; Palumbo et al., 2020). Therefore, although developed in the self-report modality, the IMAGEN-Callousness scale may serve as an effective operationalization of the underlying neurobehavioral trait of callousness.

Neurobehavioral traits, including disinhibition and callousness, may hold distinct advantages for advancing etiological-developmental understanding of psychopathology. They are conceptualized as attributes that transcend particular assessment modalities and can be used to link together measures from brain and behavioral domains with report-based measures (e.g., questionnaire scores, clinician-generated ratings). Given their links to biological systems and measures, neurobehavioral traits — whether assessed via self-report, as in this case, or as a latent factor with indicators from multiple modalities — may prove more effective for indexing genetically based risk for psychopathology than conventional traits derived from experiential-report data alone (e.g., Five Factor Model traits; Joyner et al., 2020; Venables et al., 2018b). If so, they would be more likely to meet definitional criteria as indices of liability. Continued research on neurobehavioral traits — including disinhibition and callousness — can also facilitate the discovery of neural systems and processes underlying dispositional risk for psychopathology (Perkins et al., 2020b). Scale-development approaches similar to that employed in the current study will allow for this program of research to proceed in existing, large-scale datasets.

Supplemental Method C:

Characterization of Trait Scores and Externalizing Symptoms in the IMAGEN Sample

Trait Scores

The IMAGEN-Disinhibition and Callousness scales have not been administered as of yet to youth samples outside the IMAGEN project, limiting the ability to compare IMAGEN participants' levels of each trait to other samples. As an effort to characterize the distributions of these two traits in our sample, histograms for T1 and T2 scores are provided in Supplemental Figures B and C. Further, to clarify the meaning of score elevations on each trait, we examined the percentage of items endorsed by high scorers. At T1, participants scoring at or above the 25th percentile on disinhibition endorsed (e.g., selected 'somewhat true' or higher on) an average of 59.0% of the items composing that scale, keyed in the direction of higher disinhibition. Similarly, at T2, participants scoring at or above the 25th percentile endorsed an average of 54.5% of disinhibition items. For callousness, those scoring at or above the 25th percentile endorsed an average of 47.1% of callousness items at T1 and 42.1% at T2. These statistics provide insight into the meaning of high disinhibition and callousness scores in this community adolescent sample (namely, endorsement of 50-60% of disinhibition items and 40-50% of callousness items, with all items keyed in the direction of higher trait scores).

Externalizing Symptoms

About 3.4% of the T1 sample endorsed at least three of the 12 conduct disorder items included in our symptom count (i.e., full/threshold-level endorsement of at least three symptoms), and about 1.1% endorsed at least four. At T2, 5.6% endorsed at least three threshold symptoms and 2.3% endorsed four or more. Although the DAWBA items are not identical to those in DSM-IV/-5, and our symptom counts do not speak to impairment required for the categorical diagnosis, these statistics provide an estimate of the prevalence of diagnosable conduct disorder in our sample. Per prior epidemiological research in England (Maughan et al., 2004), the prevalence of conduct disorder

is about 3.3% at age 14, suggesting that our sample endorsed symptoms at a level typical of community adolescents.

Regarding ADHD, given the limited coverage of relevant symptoms in the DAWBA, we considered adolescents to be diagnosable with ADHD if they reported that both their parents and their teachers had expressed concerned about their hyperactivity/inattention, with at least one expressing "a lot" of concern. This was an attempt to approximate the DSM-IV/-5 criterion of impairment in 2+ settings. Using this metric, 5.1% of the T1 sample and 4.0% of the T2 sample would be diagnosed with ADHD. Prevalence rates for ADHD in Germany and worldwide are about 5% (Huss et al., 2008; Polanczyk et al., 2007), suggesting that our sample showed approximately typical levels of this diagnosis.

Finally, the AUDIT provides established clinical cutoffs for hazardous/harmful consumption of alcohol and diagnosable alcohol use disorder. Using these cutoffs, 4.2% of participants reported hazardous/harmful consumption at T1, and 0.5% of the sample met criteria for moderate-to-severe alcohol use disorder. At T2, 16.7% reported hazardous/harmful consumption, and 2.1% met criteria for moderate-to-severe alcohol use disorder. The results are roughly comparable to previous epidemiological research with European adolescents, of whom 5.9% engaged in frequent drinking (5+ occasions in the last month) at age 15 and 12.1% at age 16 (Soellner et al., 2014), with "heavy episodic drinking" (binge-drinking) occurring among 2.5% of all adolescents (ages 12 to 16; Bräker et al., 2015).

Supplemental Results:

Analyses of Rule-Breaking and Aggressive Subfactors of Conduct Disorder Symptoms

Conduct disorder symptoms are often grouped into subdimensions of rule-breaking and aggressive behaviors, which show distinct etiological influences and developmental courses (for a review, see Burt, 2012). Supplemental analyses were undertaken to better understand these subdimensions in the context of this study. See Supplemental Table B for specific items comprising each subdimension. Both subdimensions were moderately stable from ages 14 to 16, ϱs = .38 for rule-breaking and .25 for aggression, ps < .001. At each time point, the subdimensions were moderately intercorrelated, ϱs = .34 at age 14 (T1) and .29 at age 16 (T2), ps < .001. Both stability estimates and intercorrelations were somewhat lower than in most prior work (Burt, 2012).

All models paralleled those conducted for the main conduct disorder symptom count analyses, with only rule-breaking or only aggressive symptoms included as outcomes. Spearman's ϱ was used to quantify zero-order correlations among disinhibition, callousness, rule-breaking, and aggression. In the negative binomial regression models, disinhibition, callousness, age, sex, and assessment site were entered as predictors of each outcome. The p-values from all regression models were adjusted using Holm's step-down procedure (Holm, 1979), treating rule-breaking and aggressive models as the same family of tests; these adjusted values are denoted by p_H. See Supplemental Figures D and E for visual representations of regression results for rule-breaking and aggression, respectively.

Age 14 (T1). At the zero-order level, disinhibition was related to both rule-breaking (ϱ = .40) and aggressive symptoms (ϱ = .31, both p < .001) at T1. Callousness was also correlated with each subdimension (ϱs = .30 for rule-breaking and .19 for aggression, respectively, ps < .001). In the negative binomial regression model, the two traits were independently related to both rule-breaking and aggressive symptoms, over and above the effects of age, sex, and assessment site. The incidence rate ratios (IRRs) with 95% confidence intervals (CIs) for disinhibition and callousness with rule-

breaking were 1.54 [1.45, 1.64] and 1.24 [1.16, 1.32], respectively (p_Hs < .001; Z_{diff} = 4.08, p < .001). For aggression, the corresponding statistics were 1.87 [1.65, 2.12] and 1.33 [1.17, 1.51], respectively (p_Hs < .001; Z_{diff} = 3.20, p = .001). The adjusted pseudo-R^2 values for these models were .21 and .18, respectively.

Age 16 (T2). Disinhibition and callousness were moderately correlated with each conduct disorder symptom subdimension at T2, ϱs = .33 for rule-breaking and .19 for aggression with disinhibition, and ϱs = .25 for rule-breaking and .12 for aggression with callousness (all ps < .001). In the negative binomial model predicting rule-breaking, estimates were comparable to those at T1, with IRRs and 95% CIs of 1.73 [1.62, 1.85] (p_H < .001) and 1.13 [1.06, 1.21] (p_H = .001) for disinhibition and callousness, respectively (Z_{diff} = 7.29, p < .001). The adjusted pseudo-R^2 was .24. For aggressive symptoms, the IRRs and 95% CIs were 1.88 [1.64, 2.17] for disinhibition and 1.34 [1.17, 1.53] for callousness, p_Hs < .001, Z_{diff} = 2.83 (p = .004), with an adjusted pseudo-R^2 of .20.

Prospective prediction (T1 to T2). T1 disinhibition was associated with T2 conduct disorder symptom subdimensions at the zero-order level, ϱs = .33 for rule-breaking and .19 for aggression, ps < .001. Similar patterns were observed for callousness, ϱs = .25 and .12, respectively, ps < .001. In the prospective negative binomial model that included T1 rule-breaking symptoms as a covariate, the IRR for T1 disinhibition predicting rule-breaking was 1.26 [1.17, 1.35] (p_H < .001), whereas callousness was not a significant predictor (IRR = 1.07 [1.01, 1.15], p_H = .07; Z_{diff} = 2.76, p = .006). The adjusted pseudo-R^2 was .18. When the residualized T1 rule-breaking variable was used instead, prospective IRRs were 1.42 [1.33, 1.52] (p_H < .001) and 1.13 [1.05, 1.20] (p_H = .002; Z_{diff} = 4.09, p < .001), respectively, with an adjusted pseudo-R^2 of .19. For the prospective model with T1 aggressive symptoms as a covariate, disinhibition predicted T2 aggressive symptoms with an IRR of 1.36 [1.19, 1.57], p_H < .001, whereas callousness showed a nonsignificant IRR of 1.15 [1.01, 1.31], p_H = .07; however, the difference between these IRRs was nonsignificant (Z_{diff} = 1.43, p = .15). The adjusted

pseudo-R^2 was .14. In the residualized model, IRRs were 1.52 [1.33, 1.73], $p_H < .001$ for disinhibition and 1.22 [1.07, 1.39], $p_H = .01$ for callousness predicting aggression ($Z_{diff} = 1.88$, $p = .06$), with an adjusted pseudo-R^2 of .14.

Supplemental Table A. Items and Psychometric Properties of Final IMAGEN-Callousness Scale

Source & Item Number	Corrected Item-Total *r*		Item Content
	Sample Half 1	Sample Half 2	
NEO-FFI			
4 (–)	.36	.43	Courteous (–)
14	.43	.40	Selfish, Egotistical
19 (–)	.17	.28	Cooperative (–)
39	.35	.34	Cold, Calculating
49 (–)	.35	.40	Thoughtful, Considerate (–)
54	.21	.34	Impolite
59	.50	.54	Manipulative
SDQ			
1 (–)	.45	.42	Considerate (–)
4 (–)	.32	.31	Willing to Share (–)
9 (–)	.35	.38	Prosocial (–)
17 (–)	.36	.37	Kind (–)
20 (–)	.31	.33	Prosocial (–)
ROBVQ			
5	.30	.36	Bullying
6	.37	.36	Name-Calling
7	.24	.22	Socially Excluding
SURPS			
22	.36	.44	Manipulative
TCI-R			
51	.29	.24	Manipulative
	Sample Half 1: Cronbach's α = .75; mean inter-item r = .15		
	Sample Half 2: Cronbach's α = .77; mean inter-item r = .17		

Note. *N*s = 735 (Sample Half 1), 736 (Sample Half 2). All items were converted to common 4-point metric prior to analyses; see Supplemental Method A. NEO-FFI = Neuroticism-Extraversion-Openness Five-Factor Inventory (Costa & McCrae, 1992); SDQ = Strengths and Difficulties Questionnaire (Goodman, 2001); ROBVQ = Revised Olweus Bully/Victim Questionnaire for Students (Olweus, 1996); SURPS = Substance Use Risk Profile Scale (Woicik et al., 2009); TCI-R = Temperament and Character Inventory – Revised (Cloninger, 1999). (–) = item was reverse-scored.

Supplemental Table B. Clinical Symptoms Used in Outcome Measures

Conduct Disorder (DAWBA; Goodman et al., 2000)

Rule-Breaking	Aggressive
• Lying to get ahead or avoid responsibility • Staying out late, past curfew • Running away from home or staying out all night • Skipping school • Deliberately starting a fire intended to cause damage • Breaking into a house, building, or car	• Starting fights (apart from siblings) • Using a dangerous weapon • Being physically cruel to a person • Being cruel to animals • Deliberately destroying others' property • Forcing unwanted sexual activity

Attention-Deficit/Hyperactivity Disorder (DAWBA; Goodman et al., 2000)

- Teachers complain about hyperactivity/inattention
- Family complains about hyperactivity/inattention
- Adolescent is concerned about their own hyperactivity/inattention

Alcohol Use Disorder (AUDIT; Saunders et al., 1993)

- Frequency of drinking
- Quantity of alcohol on a typical drinking day
- Frequency of binge-drinking (6+ drinks)
- Difficulty stopping drinking once started
- Drinking interfered with responsibilities
- Needing a drink the morning after heavy drinking
- Guilt/remorse after drinking
- Memory loss due to drinking
- Injury (self or someone else) due to drinking
- Others (relative, friend, healthcare worker) expressed concern about drinking

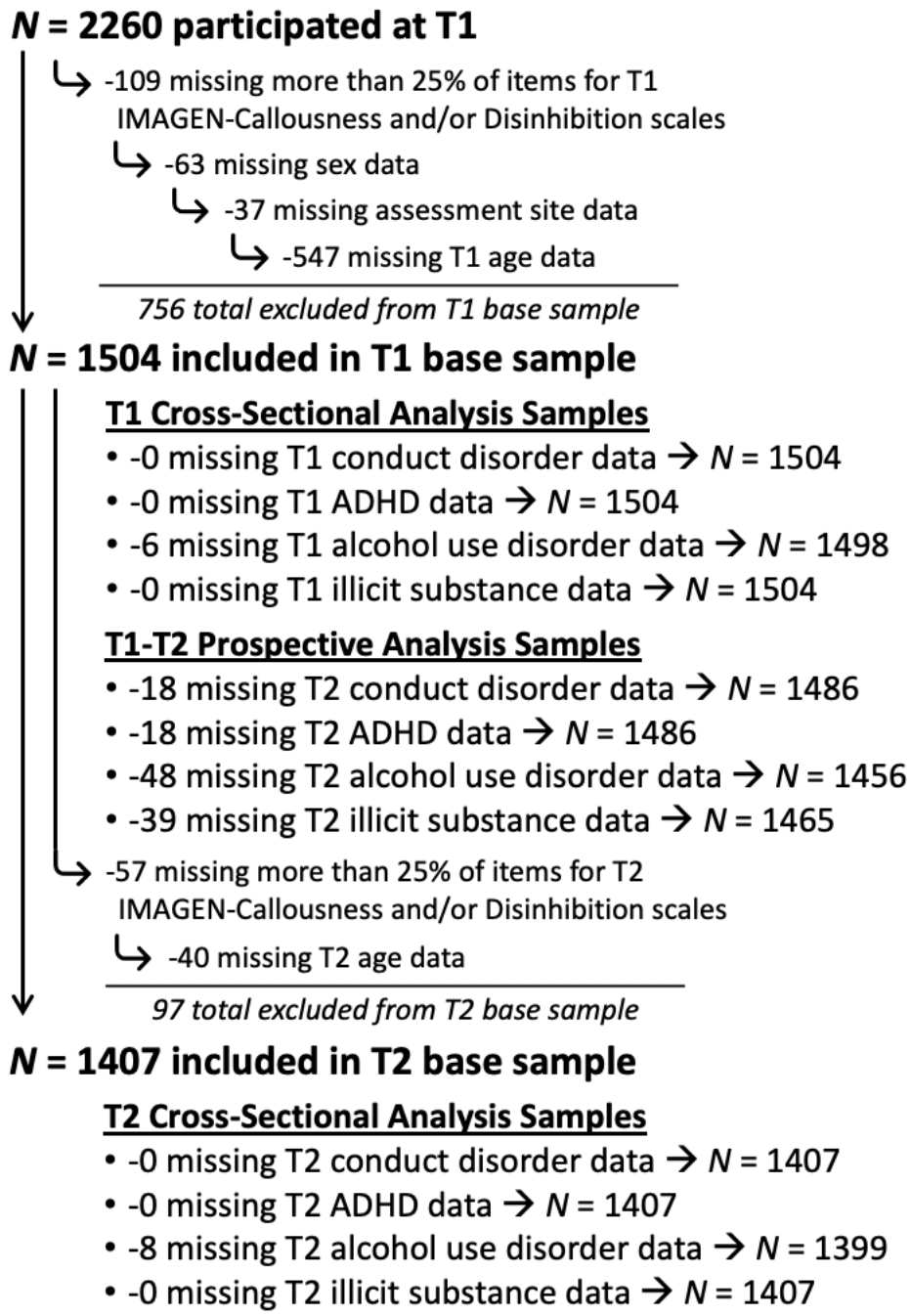
Note: Conduct disorder symptoms listed here and included in analyses were selected from the larger DAWBA set to avoid content overlap with trait disinhibition and callousness scales.

Supplemental Table C. Zero-Order Correlations and Descriptive Statistics for All Study Variables

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. T1 Disinhibition[a]	-													
2. T2 Disinhibition[a]	**.63**[b]	-												
3. T1 Callousness[a]	**.45**[b]	**.32**[b]	-											
4. T2 Callousness[a]	**.31**[b]	**.45**[b]	**.60**[b]	-										
5. T1 Conduct Disorder	**.44**[c]	**.40**[c]	**.30**[c]	**.23**[c]	-									
6. T2 Conduct Disorder	**.35**[c]	**.50**[c]	**.24**[c]	**.31**[c]	**.41**[c]	-								
7. T1 ADHD	**.38**[c]	**.32**[c]	.15[c]	.16[c]	**.27**[c]	**.20**[c]	-							
8. T2 ADHD	**.28**[c]	**.42**[c]	.12[c]	**.22**[c]	**.20**[c]	**.31**[c]	**.42**[c]	-						
9. T1 AUD	**.37**[c]	**.24**[c]	**.21**[c]	.14[c]	**.32**[c]	**.21**[c]	.10[c]	.06[c]	-					
10. T2 AUD	**.30**[c]	**.41**[c]	.13[c]	.16[c]	**.24**[c]	**.34**[c]	.12[c]	.16[c]	**.43**[c]	-				
11. T1 Externalizing	**.46**[b]	**.32**[b]	**.25**[b]	.16[b]	**.59**[c]	**.32**[c]	**.26**[c]	.15[c]	**.89**[c]	**.44**[c]	-			
12. T2 Externalizing	**.33**[b]	**.48**[b]	.17[b]	**.21**[b]	**.30**[c]	**.48**[c]	.17[c]	**.24**[c]	**.42**[c]	**.96**[c]	**.45**[b]	-		
13. T1 Illicit Drug	**.22**[d]	**.21**[d]	.13[d]	.11[d]	**.24**[d]	.16[d]	.06[d]	.06[d]	**.34**[d]	**.19**[d]	**.37**[d]	**.22**[d]	-	
14. T2 Illicit Drug	**.25**[d]	**.33**[d]	.15[d]	.16[d]	**.26**[d]	**.32**[d]	.08[d]	.16[d]	**.30**[d]	**.46**[d]	**.36**[d]	**.48**[d]	**.31**[e]	-
M (% Zeros)	1.28	1.21	.85	.78	1.49 (38.4)	1.67 (36.9)	.87 (59.1)	.74 (65.7)	1.53 (47.8)	3.97 (17.8)	-.06	-.07	.10 (90.0)	.28 (72.2)
SD	.30	.31	.30	.29	1.98	2.17	1.32	1.28	2.64	4.00	.27	.35	.30	.45
Skewness	.07	.11	.27	.27	2.62	2.22	1.61	1.97	3.10	1.48	1.19	.19	2.66	.99
Kurtosis	-.12	-.13	-.18	-.15	11.49	7.05	2.11	3.61	13.07	2.70	.64	-1.10	5.09	-1.02
Observed Range (Possible Range)	.48- 2.12 (0-3)	.38-2.02 (0-3)	.06-1.64 (0-3)	.06-1.52 (0-3)	0-20 (0-24)	0-18 (0-24)	0-6 (0-6)	0-6 (0-6)	0-24 (0-40)	0-30 (0-40)	-1.00-.85	-1.00-.98	0-1 (0-1)	0-1 (0-1)

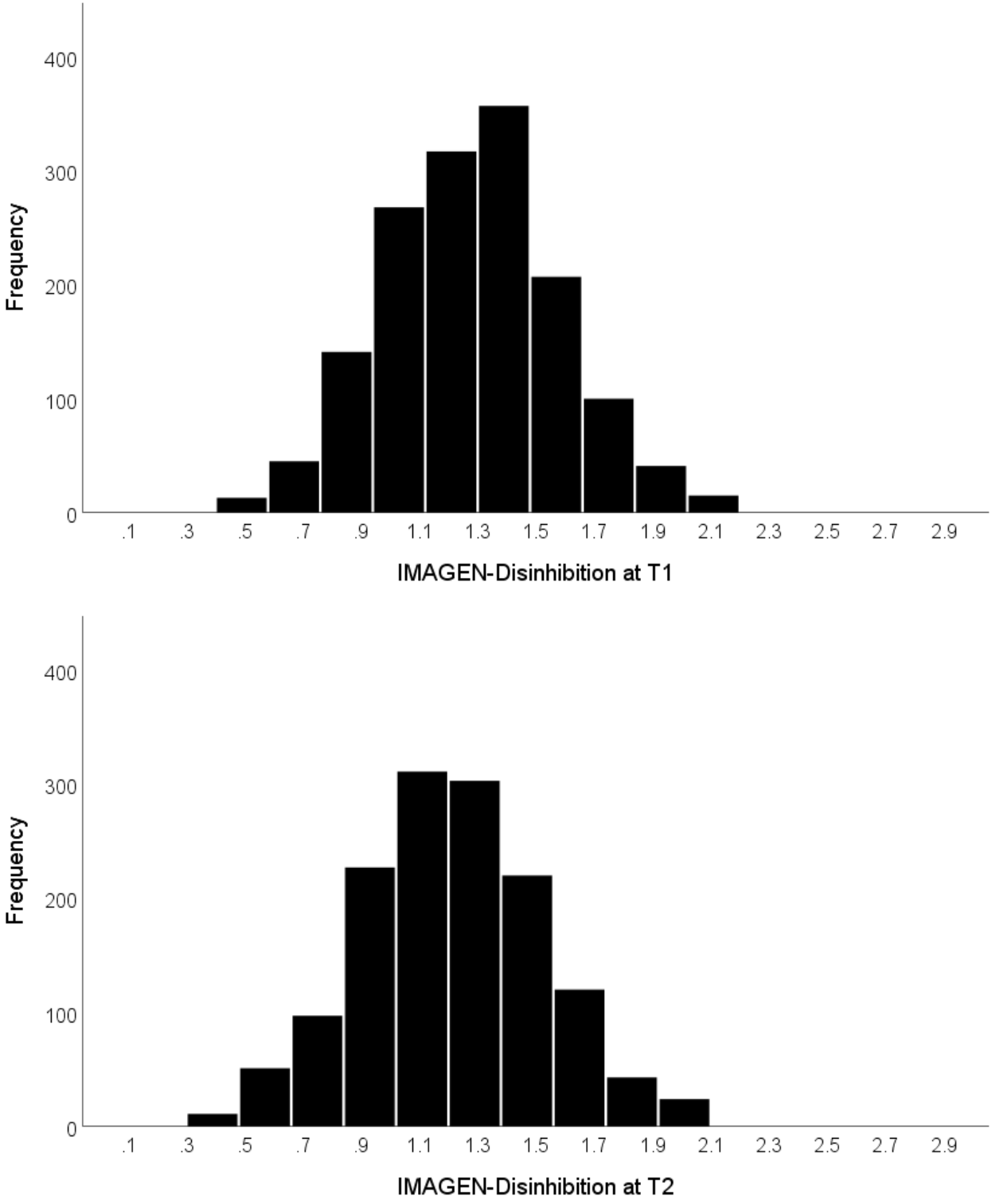
Note. *N*s = 1400 to 1504. All correlations ≥ .10 were significant at $p < .001$; correlations ≥ .20 are bolded for readability. T1 = Time 1 assessment (age 14); T2 = Time 2 assessment (age 16); ADHD = attention-deficit/hyperactivity disorder symptom count; AUD = alcohol use disorder symptom count; Externalizing = log-transformed estimated general factor scores from bifactor model. [a]Disinhibition and callousness were z-scored to facilitate ease of interpretation in all analyses (including correlations in the upper part of this table), but the descriptives in the lower part of this table refer to the raw (untransformed) versions of these variables. [b]Pearson's *r*; [c]Spearman's ϱ; [d]Pearson's point-biserial correlation (r_{pb}); [e]Pearson's φ.

Supplemental Figure A. Participant Exclusions for the Current Analyses



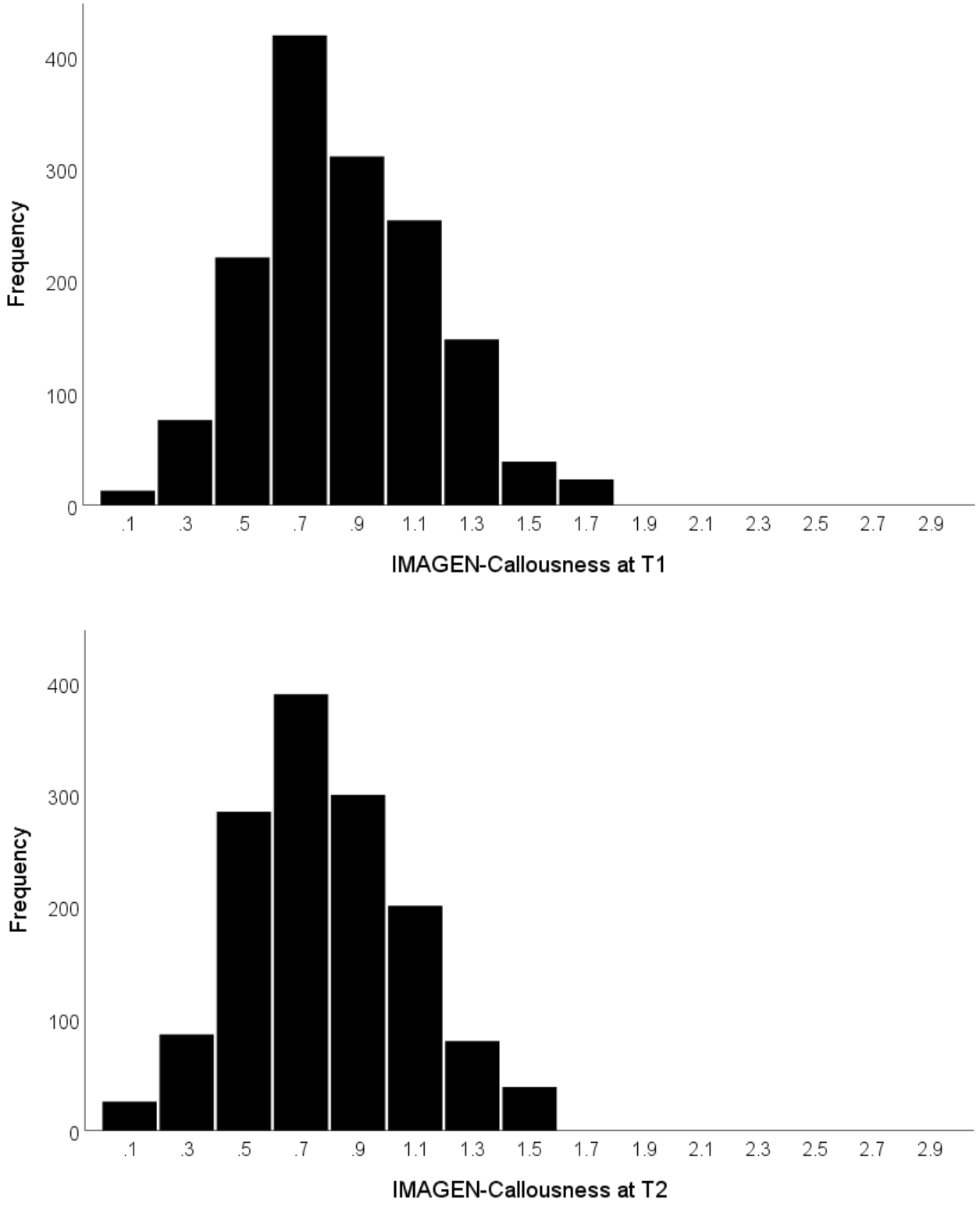
Note. T1 = Time 1 assessment (age 14); T2 = Time 2 assessment (age 16); ADHD = attention-deficit/hyperactivity disorder.

Supplemental Figure B. Distribution of IMAGEN-Disinhibition Scores at T1 and T2



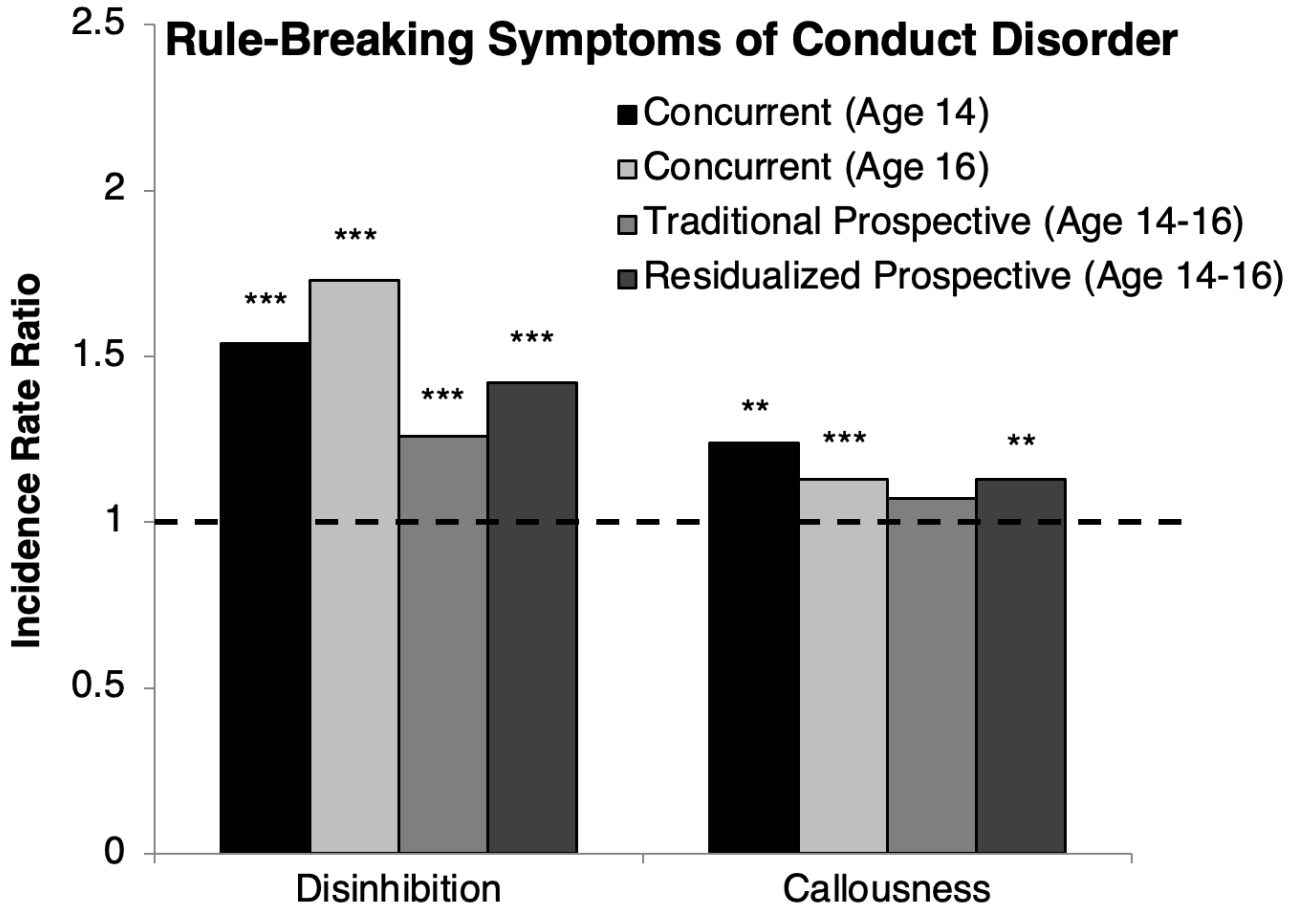
Note. Scores for IMAGEN-Disinhibition reflect mean item ratings, with a possible range of 0 to 3.

Supplemental Figure C. Distribution of IMAGEN-Callousness Scores at T1 and T2



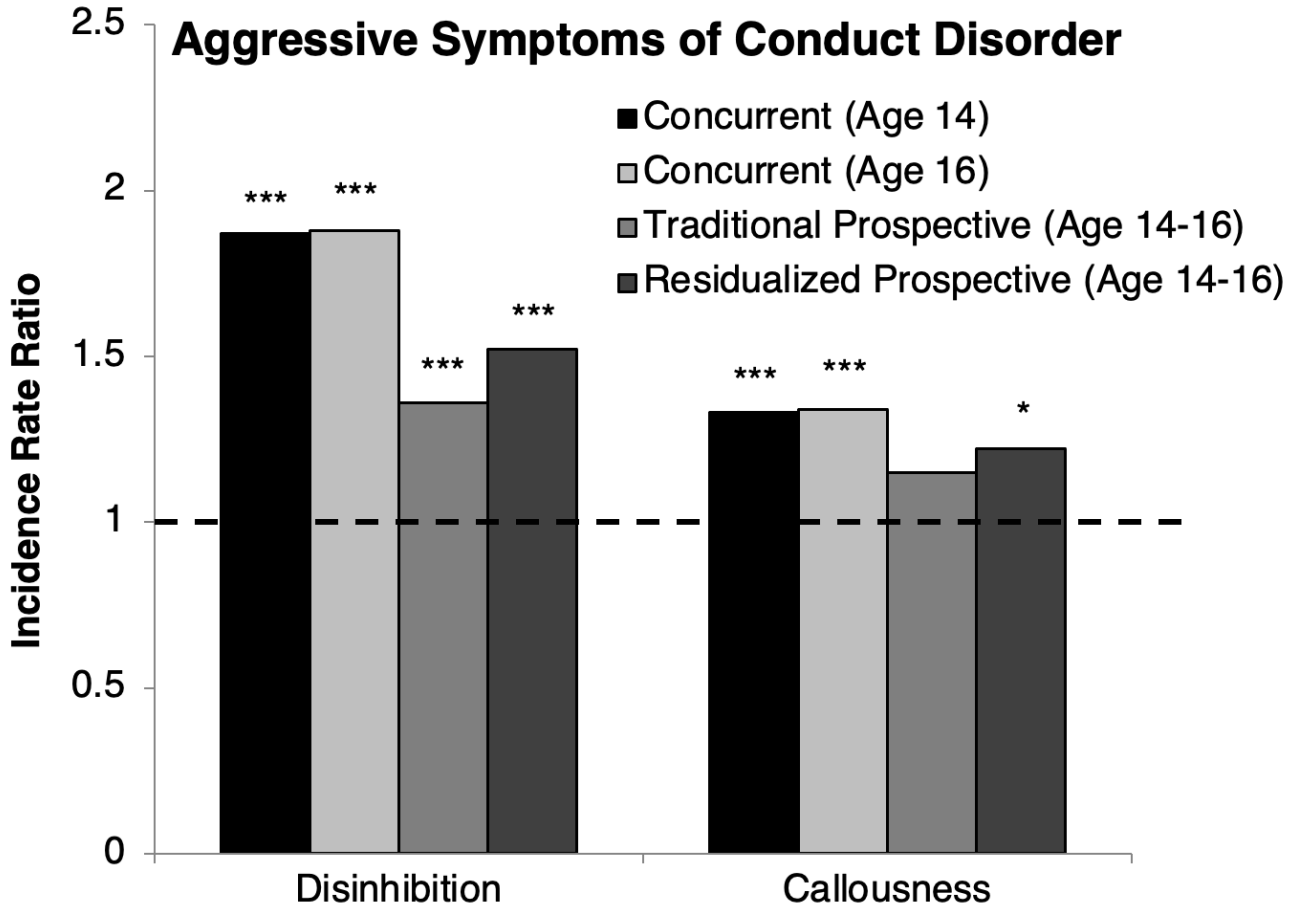
Note. Scores for IMAGEN-Callousness reflect mean item ratings, with a possible range of 0 to 3.

Supplemental Figure D. Incidence Rate Ratios for Disinhibition and Callousness as Predictors of Rule-Breaking Symptoms of Conduct Disorder in Negative Binomial Regression Models



Note. Traditional prospective models included age 14 rule-breaking symptoms as a covariate; residualized prospective models instead included as a covariate a residualized score that represented variance in age 14 rule-breaking symptoms that was independent from age 14 disinhibition and callousness. All models included the other trait, age, sex, and assessment site as covariates. ***p_H < .001, **p_H < .01, *p_H < .05.

Supplemental Figure E. Incidence Rate Ratios for Disinhibition and Callousness as Predictors of Aggressive Symptoms of Conduct Disorder in Negative Binomial Regression Models



Note. Traditional prospective models included age 14 aggressive symptoms as a covariate; residualized prospective models instead included as a covariate a residualized score that represented variance in age 14 aggressive symptoms that was independent from age 14 disinhibition and callousness. All models included the other trait, age, sex, and assessment site as covariates. ***p_H < .001, **p_H < .01, *p_H < .05.

References

Bräker, A.-B., Göbel, K., Scheithauer, H., & Soellner, R. (2015). Adolescent alcohol use patterns from 25 European countries. *Journal of Drug Issues*, *45*(4), 336–350.

Brislin, S. J., Drislane, L. E., Smith, S. T., Edens, J. F., & Patrick, C. J. (2015). Development and validation of triarchic psychopathy scales from the Multidimensional Personality Questionnaire. *Psychological Assessment*, *27*(3), 838–851. https://doi.org/10.1037/pas0000087

Brislin, S. J., & Patrick, C. J. (2019). Callousness and affective face processing: Clarifying the neural basis of behavioral-recognition deficits through the use of brain event-related potentials. *Clinical Psychological Science*, *7*(6), 1389-1402.

Brislin, S. J., Patrick, C. J., Flor, H., Nees, F., Heinrich, A., Drislane, L. E., … Foell, J. (2019). Extending the construct network of trait disinhibition to the neuroimaging domain: Validation of a bridging scale for use in the European IMAGEN project. *Assessment*, *26*(4), 567–581. https://doi.org/10.1177/1073191118759748

Brislin, S. J., Yancey, J. R., Perkins, E. R., Palumbo, I. M., Drislane, L. E., Salekin, R. T., Fanti, K. A., Kimonis, E. R., Frick, P. J., Blair R. J. R., & Patrick, C. J. (2018). Callousness and affective face processing in adults: Behavioral and brain-potential indicators. *Personality Disorders*, *9*(2), 122-132.

Cloninger, C. R. (1999). *The temperament and character inventory—Revised*. St. Louis, MO: Center for Psychobiology of Personality, Washington University.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.

Drislane, L. E., Brislin, S. J., Jones, S., & Patrick, C. J. (2018). Interfacing five-factor model and triarchic conceptualizations of psychopathy. *Psychological Assessment, 30*(6), 834–840. https://doi.org/10.1037/pas0000544

Drislane, L. E., Brislin, S. J., Kendler, K. S., Andershed, H., Larsson, H., & Patrick, C. J. (2015). A triarchic model analysis of the youth psychopathic traits inventory. *Journal of Personality Disorders, 29*(1), 15–41. https://doi.org/10.1521/pedi_2014_28_144

Drislane, L. E., Sellbom, M., Brislin, S. J., Strickland, C. M., Christian, E., Wygant, D. B., … Patrick, C. J. (2019). Improving characterization of psychopathy within the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), alternative model for personality disorders: Creation and validation of Personality Inventory for DSM-5 Triarchic scales. *Personality Disorders, 10*(6), 511–523. https://doi.org/10.1037/per0000345

Elkins, I. J., King, S. M., McGue, M., & Iacono, W. G. (2006). Personality traits and the development of nicotine, alcohol, and illicit drug disorders: Prospective links from adolescence to young adulthood. *Journal of Abnormal Psychology*, *115*(1), 26-39.

Frick, P. J., & White, S. F. (2008). Research Review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of Child Psychology and Psychiatry*, *49*(4), 359-375.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015

Hall, J. R., Drislane, L. E., Patrick, C. J., Morano, M., Lilienfeld, S. O., & Poythress, N. G. (2014). Development and validation of triarchic construct scales from the Psychopathic Personality Inventory. *Psychological Assessment, 26*(2), 447–461. https://doi.org/10.1037/a0035665

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65-70.

Huss, M., Hölling, H., Kurth, B.-M., & Schlack, R. (2008). How often are German children and adolescents diagnosed with ADHD? Prevalence based on the judgment of health care professionals: results of the German health and examination survey (KiGGS). *European Child & Adolescent Psychiatry*, *17*(S1), 52–58.

Joyner, K. J., Yancey, J. R., Venables, N. C., Burwell, S. J., Iacono, W. G., & Patrick, C. J. (2020). Using a co-twin control design to evaluate alternative trait measures as indices of liability for substance use disorders. *International Journal of Psychophysiology, 148*, 75-83.

Krueger, R. F. (1999). Personality traits in late adolescence predict mental disorders in early adulthood: A prospective-epidemiological study. *Journal of Personality, 67*(1), 39-65.

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology*, *116*(4), 645-666.

Maughan, B., Rowe, R., Messer, J., Goodman, R., & Meltzer, H. (2004). Conduct disorder and oppositional defiant disorder in a national sample: Developmental epidemiology. *Journal of Child Psychology and Psychiatry*, *45*(3), 609–621.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Olweus, D. (1996). *Revised Olweus bully/victim questionnaire*. Bergen, Norway: Research Center for Health Promotion (HEMIL Center), University of Bergen.

Palumbo, I. M., Perkins, E. R., Yancey, J. R., Brislin, S. J., Patrick, C. J., & Latzman, R. D. (2020). Toward a multi-modal measurement model for the neurobehavioral trait of affiliative capacity. *Personality Neuroscience, 3,* e11.

Patrick, C. J., Durbin, C. E., & Moser, J. S. (2012). Reconceptualizing antisocial deviance in neurobehavioral terms. *Development and Psychopathology*, *24*(3), 1047-1071.

Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and Psychopathology*, *21*(3), 913-938.

Patrick, C. J., Kramer, M. D., Krueger, R. F., & Markon, K. E. (2013). Optimizing efficiency of psychopathology assessment through quantitative modeling: Development of a brief form of the Externalizing Spectrum Inventory. *Psychological Assessment*, *25*(4), 1332–1348. https://doi.org/10.1037/a0034864

Perkins, E. R., Joyner, K. J., Patrick, C. J., Bartholow, B. D., Latzman, R. D., DeYoung, C. G., Kotov, R., Reininghaus, U., Cooper, S. E., Afzali, M. H., Docherty, A. R., Dretsch, M. N., Eaton, N. R., Goghari, V. M., Haltigan, J. D., Krueger, R. F., Martin, E. A., Michelini, G., Ruocco, A. C., … Zald, D. H. (2020a). Neurobiology and the Hierarchical Taxonomy of Psychopathology: Progress toward ontogenetically informed and clinically useful nosology. *Dialogues in Clinical Neuroscience*, *22*(1), 51-63.

Perkins, E. R., Latzman, R. D., & Patrick, C. J. (2020b). Interfacing neural constructs with the Hierarchical Taxonomy of Psychopathology: 'Why' and 'how'. *Personality and Mental Health*, *14*(1), 106-122.

Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J., & Rohde, L. A. (2007). The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *American Journal of Psychiatry*, *164*, 942–948.

Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, *88*(6), 791–804.

Sellbom, M., Drislane, L. E., Johnson, A. K., Goodwin, B. E., Phillips, T. R., & Patrick, C. J. (2016). Development and validation of MMPI-2-RF scales for indexing triarchic psychopathy constructs. *Assessment, 23*(5), 527–543. https://doi.org/10.1177/1073191115590853

Soellner, R., Göbel, K., Scheithauer, H., & Bräker, A.-B. (2014). Alcohol use of adolescents from 25 European countries. *Journal of Public Health, 22*(1), 57–65.

Venables, N. C., Foell, J., Yancey, J. R., Kane, M. J., Engle, R. W., & Patrick, C. J. (2018a). Quantifying inhibitory control as externalizing proneness: A cross-domain model. *Clinical Psychological Science, 6*(4), 561-580.

Venables, N. C., Yancey, J. R., Kramer, M. D., Hicks, B. M., Krueger, R. F., Iacono, W. G., Joiner, T. E., & Patrick, C. J. (2018b). Psychoneurometric assessment of dispositional liabilities for suicidal behavior: Phenotypic and etiological associations. *Psychological Medicine, 48*(3), 463-472.

Woicik, P. A., Stewart, S. H., Pihl, R. O., & Conrod, P. J. (2009). The Substance Use Risk Profile Scale: A scale measuring traits linked to reinforcement-specific substance use profiles. *Addictive Behaviors, 34*(12), 1042–1055. https://doi.org/10.1016/j.addbeh.2009.07.001

Yancey, J. R., Venables, N. C., Hicks, B. M., & Patrick, C. J. (2013). Evidence for a heritable brain basis to deviance-promoting deficits in self-control. *Journal of Criminal Justice, 41*(5), 309-317.