

Submission Type: Original Article
Abstract: 248/250
Text Word Count: 3950/4000
Display Items: 4/4

The Addiction Genetic Factor $a_{(g)}$: A Unitary Genetic Vulnerability Characterizes Substance Use Disorders and Their Associations with Common Correlates

Alexander S. Hatoum^{1#}, Emma C. Johnson¹, Renato Polimanti^{2,3}, Hang Zhou^{2,3}, Raymond Walters^{4,5}, Substance Use Disorders Working Group of the Psychiatric Genomics Consortium, Joel Gelernter^{2,3,6,7}, Howard J. Edenberg^{8,9}, Ryan Bogdan^{10*} & Arpana Agrawal^{1*}

#Please send all correspondence to: Washington University School of Medicine, Department of Psychiatry, 660 S. Euclid, CB 8134, Saint Louis, MO 63110, USA; email: ashatoum@wustl.edu;

¹Washington University School of Medicine, Department of Psychiatry, Saint Louis, USA

²Department of Psychiatry, Division of Human Genetics, Yale School of Medicine, New Haven, CT, USA

³Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA

⁴Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

⁵Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁶Department of Genetics, Yale School of Medicine, New Haven, CT, USA

⁷Department of Neuroscience, Yale School of Medicine, New Haven, CT, USA

⁸Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA

⁹Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA

¹⁰Department of Psychological & Brain Sciences, Washington University in St. Louis

*These authors contributed equally to the work

35 **Abstract**

36

37 **Background.** Substance use disorders (SUDs) commonly co-occur with one another and with
38 other psychiatric disorders. They share common features including high impulsivity, negative
39 affect, and lower executive function. We estimate the shared genetic architecture across distinct
40 SUDs, its independence from genetic liability to substance use, and its relation to genetic liability
41 to impulsivity, negative affect, and executive function as well as non-substance
42 psychopathology.

43 **Methods.** We tested whether a common genetic factor undergirds liability to problematic alcohol
44 use (PAU), problematic tobacco use (PTU), cannabis use disorder (CUD), and opioid use
45 disorder (OUD) by applying Genomic structural equation modelling to genome-wide association
46 study statistics (Total N = **1,019,521**; substance specific Ns range: **82,707-435,563** of European
47 ancestry), while adjusting for the genetics of substance use (Ns = **184,765-632,802**). We tested
48 whether shared liability across SUDs is associated with behavioral constructs (risk taking,
49 executive function, neuroticism; Ns = **328,339-427,037**) and non-substance use psychopathology
50 (mood/psychotic, compulsive, and early neurodevelopmental disorders).

51 **Results.** Shared genetic liability to PAU, PTU, CUD, and OUD was characterized by a
52 unidimensional addiction factor (termed $a_{(g)}$), independent of substance use. OUD and CUD
53 demonstrated the largest loadings. $a_{(g)}$ was associated with risk taking, neuroticism, executive
54 function, and non-substance psychopathology, but retained specific variance (standardized
55 residual= .579).

56 **Interpretation.** A common genetic addictions factor partly explains susceptibility for alcohol,
57 tobacco, cannabis and opioid use disorder. $a_{(g)}$ has unique pathways that are not shared with
58 substance use or non-substance psychopathology, suggesting that addiction is not the linear
59 combination of substance use and psychopathology.

60

61

62
63 **Disclosures/COI:** JG is named as an inventor on PCT patent application #15/878,640 entitled:
64 "Genotype-guided dosing of opioid agonists," filed January 24, 2018.

65
66 **Funding.** This research was supported by MH109532 (AA, JG, HJE, ECJ) and T32DA007261
67 (ASH). AA acknowledges K02DA32573. ECJ was supported by F32AA027435. RP
68 acknowledges R21DA047527. RB acknowledges R21-AA027827.
69 The Substance Use Disorders Working Group of the Psychiatric Genomics Consortium (PGC-
70 SUD) is supported by funds from NIDA and NIMH to MH109532.

71 **Acknowledgments:** We gratefully acknowledge our contributing studies and the participants in
72 those studies without whom this effort would not be possible.

73 The MVP summary statistics were obtained via an approved dbGaP application
74 ([phs001672.v4.p1](#)). The authors thank Million Veteran Program (MVP) staff, researchers, and
75 volunteers, who have contributed to MVP, and especially participants who previously served
76 their country in the military and now generously agreed to enroll in the study. (For details, see
77 <https://www.research.va.gov/mvp/> and Gaziano, J.M. et al. Million Veteran Program: A mega-
78 biobank to study genetic influences on health and disease. *J Clin Epidemiol* 70, 214-23 (2016)).
79 This research is based on data from the Million Veteran Program, Office of Research and
80 Development, Veterans Health Administration, and was supported by the Veterans
81 Administration (VA) Cooperative Studies Program (CSP) award #G002.

82 This study included summary statistics of a genetic study on cannabis use (Pasman et al, 2018
83 *Nature Neuroscience*). We would like to acknowledge all participating groups of the
84 International Cannabis Consortium, and in particular the members of the working group
85 including Joelle Pasman, Karin Verweij, Nathan Gillespie, Eske Derks, and Jacqueline Vink.
86 Pasman et al, (2018) included data from the UK Biobank resource under application numbers
87 9905, 16406 and 25331.

88
89

90 **INTRODUCTION**

91 Substance use and use disorders (SUDs) represent large and growing international public health
92 problems that account for nearly 6% of global disease burden¹. SUDs, both licit and illicit,
93 commonly co-occur and also co-occur with non-substance psychopathology; comorbidity is
94 associated with increased symptom severity² and worse outcomes (e.g., less responsive to
95 treatment, greater socioeconomic costs³). However, the etiology underlying shared risk across
96 these disorders is poorly understood.

97

98 **Shared Genetic Liability**

99 According to twin studies, the moderate-large heritability (50-60%) of distinct SUDs (i.e., alcohol,
100 nicotine, cannabis, and other illicit drugs) is partly attributable to a shared latent genetic
101 vulnerability.⁴ Similarly, genetic correlations estimated from genome-wide association study
102 (GWAS) summary statistics support a shared genetic vulnerability across distinct SUDs (e.g.,
103 SNP-rG = .73 between problematic alcohol use and opioid use disorder),⁷ between SUDs and
104 substance use (e.g., SNP-rG = .78 between problematic alcohol use and drinks per week),⁸ and
105 between SUDs and other psychopathology (e.g., SNP-rG = .33 between cannabis use disorder and
106 major depressive disorder⁹). What remains unclear is the extent to which genetic liability across
107 substance use *disorders* is shared with *and distinct* from substance use and non-substance
108 psychopathology, and what putative intermediate phenotypes may link shared genetic liability
109 between SUDs and non-substance psychopathology.

110 *Substance Use and Use Disorder.* Substance use and SUDs have substantial genetic
111 overlap; however, there remain genetic mechanisms that relate to SUD liability beyond normative
112 or frequently occurring substance use behaviors. For opioids¹⁴, alcohol^{15,16,17}, and cannabis⁹, the

113 use and use disorder dimensions show differing associations with psychopathology (e.g.
114 schizophrenia) and life outcomes (e.g. educational attainment). The genetic contributions to SUDs
115 can be confounded by genetic influences on earlier stages of substance use¹⁸, since the definition
116 of an SUD requires the individual to have initiated and escalated use. Explorations of shared
117 genetic variation across substance use disorders can potentially also address vulnerability to
118 initiation and use of substances.

119 ***Substance use and Psychopathology.*** Recently, Lee and colleagues¹⁰ identified 3 broad
120 clusters (mood/psychotic, compulsive, and early neurodevelopmental) representing shared and
121 distinct genetic liability to 8 non-substance psychiatric disorders. While there has been limited
122 integration of substance use phenotypes into psychopathology models, polygenic liability to cross-
123 psychiatric disorders is associated with general substance use and SUDs⁶. Further, emerging
124 evidence suggests partial overlap: tobacco *use* shares variance with ADHD¹¹, alcohol and cannabis
125 dependence load with antisocial behavior¹², and alcohol use and use disorder load together onto
126 an Externalizing factor¹³. Collectively, these data suggest that substance use and use disorders
127 share a common genetic liability with psychopathology.

128 ***Stage-Based Addiction Constructs and Substance Use Disorder Genetics.*** SUD
129 vulnerability has been conceptualized within a 3-stage neurobiological model consisting of
130 binge/intoxication, preoccupation/anticipation, and withdrawal/negative affect¹⁹. According to this
131 model, initial positive reinforcement associated with substance use is derived from stimulation of
132 neural reward circuitry that drives impulsive and risk-taking behaviors in the context of under-
133 developed physiological tolerance. With continued use and progression towards SUD, the
134 reinforcing properties of substances shift from positive to negative reinforcement; as use becomes
135 compulsive, it functions to return the body to drug-present homeostasis and alleviate low mood, a

136 predisposition to which is broadly indexed by neuroticism²⁰. Following repeated drug-reward and
137 drug-homeostasis pairings, cognitive preoccupation with the drug in expectation of reward/relief
138 emerges in the context of impaired executive functioning²¹. While GWAS support genetic
139 correlations between SUDs and risk-taking^{7,22}, cognition²³, and negative affect^{7,22}, the extent to
140 which common genetic liability across SUDs relates to these stage-based constructs has yet to be
141 examined. Given that risk-taking, neuroticism and executive functioning also share genetic
142 liability with non-SUD psychopathology, it is possible that these constructs may account for shared
143 genetic liability between SUDs and non-substance psychiatric disorders.

144

145 **The Current Study**

146 Given accumulating evidence of shared liability to SUDs and other forms of psychopathology,
147 understanding the shared and unique genetic contributions to SUDs and how these relate to stage-
148 based addiction constructs, non-substance psychopathology, and substance use may generate
149 etiologic insights that improve psychiatric nosology, prevention, and treatment. To this end, we
150 *first* estimate the shared and genetic structure across SUDs by applying genomic structural
151 equation modelling (gSEM)²⁴ to summary statistics generated by the largest GWAS of problematic
152 alcohol use (PAU)²², problematic tobacco use (PTU)^{25,26}, cannabis use disorder (CUD)⁹, and
153 opioid use disorder (OUD)⁷. We name the shared variance across SUDs the Addiction genetic
154 factor, $a_{(g)}$. *Second*, we relate $a_{(g)}$ to genetics of behavioral constructs underlying the stage-based
155 model of SUDs. We estimate the extent to which genetic liability to risk-taking (N = 328,339),
156 executive function (N = 427,037), and neuroticism (N = 390,278) are related to a common genetic
157 liability to SUDs. *Third*, we examine whether $a_{(g)}$ is associated with the 3 factors representing
158 genetic liability to non-substance psychopathology¹⁰ (i.e., mood/psychotic, compulsive, and early

159 neurodevelopment) and whether stage-based addiction constructs (i.e., risk-taking, executive
160 function, neuroticism) indirectly link $a_{(g)}$ to psychopathology. *Finally*, given that genetic liability
161 to substance use (e.g., ever using, quantity-frequency) and later stages of SUDs may be partially
162 distinct (e.g., risk-taking more closely related to use)⁸, we repeat all analyses while incorporating
163 genetic liability to substance use (i.e., alcohol drinks/week: N = 537,349²⁶; tobacco ever regularly
164 use: N = 632,802²⁶, cannabis ever use N = 162,082²⁷) as covariates at the indicator level.

165 We hypothesized that, similar to general intelligence (*g-factor*) and psychopathology (*p-*
166 *factor*), SUDs and problem substance use would be largely characterized by a common genetic
167 vulnerability (i.e., an addiction genetic factor [$a_{(g)}$]) with evidence of small but potentially
168 important substance-specific liability (e.g., metabolic and signaling pathways for a specific drug
169 such as *ADH1B* variants with alcohol²⁸). We hypothesized that (i) $a_{(g)}$ would be associated with all
170 3 non-substance psychiatric disorder clusters while retaining variance unique to itself, (ii) genetic
171 liability to behavioral phenotypes representing vulnerability to stage-based addiction constructs
172 (i.e., risk-taking, executive function, and neuroticism) would be associated with $a_{(g)}$ and account
173 for a proportion of the association between $a_{(g)}$, but also show specific associations with $a_{(g)}$ beyond
174 psychopathology, and (iii) after accounting for genetics of substance use, $a_{(g)}$ would retain unique
175 variance (i.e., we expect significant genetic correlations among SUDs) maintain similar patterns
176 with non-substance psychopathology and stage-based constructs.

177

178 **METHODS**

179 **Samples**

180 Summary statistics from the largest available discovery GWASs were used to represent genetic
181 risk for each construct (details and Ns are in **Supplemental Table 1**). These include: **i) 4 SUDs**
182 (problematic alcohol use²², problematic tobacco use²⁵, cannabis use disorder⁹, opioid use
183 disorder⁷); **ii) 3 *substance use phenotypes*** (alcohol drinks/week²⁶, lifetime ever smoking²⁶,
184 lifetime cannabis use²⁷); **iii) 3 *traits representing the stage-based model of SUDs*** (risk-taking,
185 executive function, neuroticism); and **iv) 9 *non-substance psychiatric disorders***. Unfortunately,
186 analyses were restricted to data from individuals of European ancestry because GWAS on these
187 constructs in other ancestral origins are not available or are underpowered, and cross-ancestry
188 analysis can confound genetic correlation estimates²⁹. All GWAS summary statistics were filtered
189 to retain variants with minor allele frequencies > 0.01 and INFO score > 0.90 for PGC and INFO
190 score > 0.70 for the MVP.

191

192 **Problematic Substance Use/Substance Use Disorder Summary Statistics**

193 *Problematic Alcohol Use:* Summary statistics for problematic alcohol use (PAU) were
194 derived from a meta-analysis of GWASs of DSM-IV alcohol dependence (AD) from the
195 Psychiatric Genomics Consortium (PGC; n = 11,569 case, 34,999 controls), ICD-9/10 based
196 diagnoses of alcohol use disorders from the Million Veteran Program phase 1 and 2 data (MVP;
197 n = 45,995 cases; 221,396 controls) and the Problem subscale score from the Alcohol Use
198 Disorders Identification Test (AUDIT-P) that was administered via an online health
199 questionnaire to a subset of participants in the UK Biobank (n = 121,604)²². The final GWAS
200 summary statistics included data on 435,563 participants.

201 *Problematic Tobacco Use:* We used summary statistics from the GWAS of the
202 Fagerström Test for Nicotine Dependence²⁵ (FTND). Because the genetic correlation between
203 FTND and cigarettes per day is high (calculated $r_G = 0.95^{30}$), we applied Multi-Trait Analysis of
204 Genome-wide association study summary statistics (MTAG³¹) to summary statistics generated
205 from the GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) GWAS of
206 cigarettes per day to create a combined phenotype²⁶. As this is the only observed genetic
207 correlation between a use and use disorder category that is sufficiently high to allow for MTAG
208 analysis, we used this procedure only for tobacco. The final GWAS summary statistics had an
209 effective sample size of $n = 270,120$ individuals.

210 *Cannabis Use Disorder:* Summary statistics were derived from a GWAS meta-analysis of
211 DSM-IV and DSM-III-R cannabis abuse and dependence from the Psychiatric Genomics
212 Consortium ($n = 5,289$ cases; $n = 10,004$ controls), ICD 10 cannabis use disorder from the
213 Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) ($n = 2,758$ cases;
214 $n = 53,326$ controls), and hospital-based diagnoses from deCODE ($n = 6,033$ cases; $n = 280,396$
215 controls)⁹. The final sample included 14,080 cases with CUD and 343,726 controls.

216 *Opioid Use Disorder:* Opioid use disorder (OUD) summary statistics were derived from
217 an meta-analytic GWAS of ICD-9/10-based opioid abuse or dependence from Yale-Penn, Study
218 of Addiction: Genetics and Environment, and the Million Veteran Program ($n = 10,544$ cases; n
219 $= 72,163$ opioid-exposed controls)⁷.

220

221 **Substance Use Summary Statistics**

222 *Alcohol Drinks/Week.* Alcohol use summary statistics were derived from the GSCAN
223 GWAS for current reported drinks/week ($n = 537,349$). There was a moderate correlation with
224 PAU (SNP-rG between drinks/week and PAU = 0.77 ± 0.02)²².

225 *Lifetime Tobacco Use.* Summary statistics came from the GSCAN GWAS of reported
226 ever/never regular cigarette smoking (ever $n = 301,524$, never $n = 331,278$). There was a
227 moderate correlation with PTU (SNP-rG = 0.28 ± 0.03).

228 *Lifetime Cannabis Use.* We used summary statistics from a meta-analysis of lifetime
229 cannabis ever-use from the International Cannabis Consortium and UK Biobank (ever $n =$
230 $43,380$; never $n = 118,702$). There was a moderate correlation with CUD (SNP-rG =
231 0.47 ± 0.05)⁹.

232

233 **Stage-based Behavioral Constructs**

234 The behavioral three-stage model of addiction focuses on “state” changes in substance
235 use behaviors. As GWAS measure individual differences in traits, we selected behaviors that (1)
236 are known to convey vulnerability to each stage, and (2) are heritable.

237 *Risk-taking and sensitivity to reward.* A GWAS of risk-taking derived from a single item
238 in the UK Biobank (“Would you describe yourself as someone who takes risks?”; data field
239 #2040; risk taker $n = 83,677$; non-risk taker $n = 244,662$)³² was used as a trait representing this
240 stage.

241 *Executive Function.* The “preoccupation/anticipation” stage is characterized by
242 maladaptive reward valuation and future planning. Recent work argues that this vulnerability is
243 captured by executive functioning³³. Summary statistics from a GWAS of a latent factor
244 representing general executive functioning were used ($N = 427,037$)²³.

245 *Negative Emotionality and Sensitivity to Stress*. The stage of withdrawal/negative affect
246 represents substance use functioning to mitigate aversive withdrawal symptoms, such as negative
247 affect. Neuroticism has been found to modify stress sensitivity and neural reward processing³⁴.
248 Neuroticism was chosen as a trait-based measure representing liability to negative affect as
249 opposed to depression because depression was included in the non-substance psychiatric disorder
250 factor generation and because neuroticism includes trans-diagnostic constructs such as negative
251 urgency (i.e., impulsive attempts to cope with negative affect) that may place individuals at risk
252 for the negative reinforcing aspects of SUDs. Thus, we selected the largest GWAS of
253 neuroticism to represent this stage (n = 390,278)²⁰.

254 **Non-Substance Summary Statistics**

255 Summary statistics from the PGC Cross-disorder GWAS on the 8 disorders that were
256 previously shown to fit a 3 factor confirmatory model were used¹⁰. These disorders included
257 Schizophrenia³⁵, Bipolar disorder³⁶, Major depressive disorder³⁷, Attention Deficit Hyperactivity
258 disorder³⁸, Obsessive Compulsive Disorder³⁹, Anorexia Nervosa⁴⁰, Tourette Syndrome⁴¹, and
259 Autism Spectrum Disorder⁴² (See **Supplemental Table 1 for details**).

260

261 **Statistical analysis**

262 *First*, we estimated the pairwise genetic correlation between PAU, PTU, CUD and OUD
263 using Linkage Disequilibrium Score Regression (LDSR)²⁹. After confirming that the four SUDs
264 were significantly genetically correlated (see **Results**), we applied confirmatory factor analysis
265 to the covariance matrix generated by LDSR using gSEM⁴³; a single latent factor (i.e., $a_{(g)}$ -
266 factor) was allowed to load freely on PAU, PTU, OUD, and CUD. Variance of this common
267 latent factor was scaled to 1.0. A residual correlation between PAU and OUD was also estimated

268 to account for measurement overlap, because the Million Veterans Project sample was contained
269 in both PAU and OUD GWAS.

270 *Second*, we used a series of structural regression models to estimate the extent to which
271 genetic liability to stage-based constructs of addiction (i.e., risk-taking, executive function and
272 neuroticism) are related to $a_{(g)}$. Here, the $a_{(g)}$ variance was freed, and the OUD loading was set to
273 1.0 to scale the model. Intercorrelations were estimated between risk-taking, executive function
274 and neuroticism.

275 *Third*, we recapitulated the three factors from Lee et al¹⁰. (i.e., psychotic/mood disorders,
276 compulsive disorders, and early neuro-developmental disorders) and estimated their relationship
277 with $a_{(g)}$ while allowing for inter-factor correlations. This allowed us to estimate the unique
278 association between each of the 3 psychopathology factors and $a_{(g)}$ and to estimate variance that
279 was residual to $a_{(g)}$. We then examined whether stage-based addiction constructs (i.e., risk-
280 taking, executive function and neuroticism) indirectly linked $a_{(g)}$ to the 3 non-substance
281 psychopathology factors using a multiple mediator model. To estimate residual associations (i.e.,
282 direct paths) between the stage-based constructs and $a_{(g)}$, we re-structured the mediation model to
283 one in which the 3 non-SUD psychopathology factors served as “mediators” of the relationship
284 between risk-taking, executive functioning, neuroticism, and $a_{(g)}$.

285 To separate the genetics of SUD from the genetics of substance use, we ran a second
286 gSEM model where substance *use* GWAS summary statistics were endogenous predictors of all
287 measured variables in the model. For example, in the model estimating the association between
288 $a_{(g)}$ and psychiatric factors, the 8 psychiatric disorders and the 4 SUD disorder variables were
289 regressed on the 3 substance use variables. In this way, covariate effects were estimated robustly
290 and simultaneously to our associations of interest.

291 **RESULTS**

292 **The $a_{(g)}$ -Factor**

293 Genetic correlations between problematic alcohol use (PAU²²), problematic tobacco use
294 (PTU^{25,26}), cannabis use disorder (CUD⁹), and opioid use disorder (OUD⁷) ranged from 0.19(S.E.
295 = .04) to 0.78(.09) (**Supplemental Figure 1**). PTU showed the lowest SNP-rG with other SUD
296 phenotypes [i.e., PAU = 0.19(.04), CUD = 0.31(.05), OUD = 0.26(.08)] while OUD showed the
297 highest [PAU = 0.69(.07), CUD = 0.78(.09)]. A confirmatory factor model specifying a
298 unidimensional $a_{(g)}$ -factor underlying the genetic covariance between PAU, PTU, CUD and
299 OUD fit the data well [$X^2(1) = .017$, $p = .895$, CFI = 1, SRMR = .002; residual $r = .51$, $p = 0.016$;
300 **Figure 1A**]. Loadings were uniformly high except for PTU, where there was stronger evidence
301 for tobacco-specific genetic influences.

302 The inclusion of genetic liability to typical substance use did not modify the single factor
303 structure of $a_{(g)}$ (**Figure 1B**); all SUDs continued to load significantly on the factor. However,
304 factor loadings were lower for all substances, especially for PAU, which may be attributable to
305 the high genetic correlation between Drinks/Week and PAU.

306

307 **Shared Liability to Stage-based Behavioral Phenotypes**

308 Genetic liability to stage-based addiction constructs was shared with $a_{(g)}$ (**Figures 2**). As
309 expected, $a_{(g)}$ was positively associated with genetic liability to risk-taking ($\beta = 0.45$) and
310 neuroticism ($\beta = 0.25$), and negatively associated with executive function ($\beta = -0.165$; **Figure**
311 **2A**). Despite significant genetic overlap between $a_{(g)}$ and stage-based behavioral phenotypes, $a_{(g)}$
312 retained unique variance ($a_{(g)}$ residual = 0.68). When conditioning for genetic liability for
313 substance use, $a_{(g)}$ remained significantly associated with increased genetic liability to risk-taking

314 ($\beta = 0.22$) and neuroticism ($\beta = 0.18$) and decreased genetic liability to executive function ($\beta =$
315 0.28 ; **Figure 2B**). Accounting for genetic liability for substance use substantially reduced the
316 association between $a_{(g)}$ and risk-taking from 0.45 to 0.22 ($p_{(df = 1)} = 4e-09$) and accentuated the
317 negative association with executive function from $\beta = -0.17$ to -0.28 ($p_{(df = 1)} = 0.013$); there was
318 a smaller effect on the association with neuroticism (from $\beta = 0.25$ to 0.18 , $p_{(df = 1)} = 0.012$).

319

320 **Shared Liability to Non-substance Psychopathology**

321 Genetic liability to non-substance psychopathology (i.e., compulsive disorders, psychotic and
322 mood disorders, and neurodevelopmental disorders) was shared with $a_{(g)}$ (**Figure 3**).

323 Psychotic/mood disorders ($\beta = 0.45$) and neurodevelopmental disorders ($\beta = 0.74$) were
324 positively associated with $a_{(g)}$ while compulsive disorders showed a negative association ($\beta =$
325 0.32 ; **Figure 3A**). Due to the strong correlation between $a_{(g)}$ and early-onset neurodevelopmental
326 disorders (which includes ADHD) we allowed ADHD to load on $a_{(g)}$; here, a residual association
327 between $a_{(g)}$ and early-onset neurodevelopmental disorders remained, but was significantly
328 attenuated (from $\beta = 0.74$ to 0.43 , $p_{(df = 1)} = 5e-5$). When conditioning $a_{(g)}$ for substance use, only
329 the psychotic/mood and early neurodevelopmental disorder factors remained significantly
330 associated with $a_{(g)}$ (**Figure 3B**). Despite the significant genetic overlap with other psychiatric
331 disorder domains, $a_{(g)}$ retained unique variance representing genetic liability specific to SUDs
332 ($a_{(g)}$ residual = 0.30 , $p = 4.54e-3$). This unique variance was potentiated by modeling genetic
333 liability to substance use ($a_{(g)}$ residual = 0.58 , $p = 0.015$).

334 The specifications for the mediation models is shown in **Supplemental Figure 2**.

335 Genetic liability to risk taking accounted for a proportion of the associations between all non-
336 substance psychopathology domains and $a_{(g)}$ (**Table 1**). Executive function uniquely indexed an

337 indirect effect between mood/psychotic disorders and $a_{(g)}$ (**Table 1**). When conditioning $a_{(g)}$ for
338 genetic liability to substance use, risk-taking no longer accounted for a portion of the association
339 between any non-substance psychopathology domain and $a_{(g)}$, but executive function continued
340 to account for a proportion of the overlap (indirect effect of 0.048) between psychotic/mood
341 disorders and $a_{(g)}$ (**Table 1**). *Post-hoc* analyses revealed that executive function retained a unique
342 association with $a_{(g)}$ after accounting for genetic liability to both substance use and non-substance
343 psychopathology (**Supplementary Table 2**).

344

345 **DISCUSSION**

346 We applied genomic structural equation modeling (gSEM)²⁴ to GWAS summary statistics to
347 characterize the genetic influences shared across SUDs and estimated how common genetic
348 liability is related to trait conceptualizations of a theoretical stage-based SUD model as well as to
349 non-substance psychopathology. Three primary findings emerged. *First*, genetic risk for specific
350 SUD phenotypes (i.e., PAU²², PTU^{25,26}, CUD⁹, and OUD⁷) was largely attributable to a single
351 addiction genetics factor, $a_{(g)}$ (**Figure 1**). *Second*, $a_{(g)}$ was associated with genetic liability to trait
352 representations of stage-based facets of addiction (risk taking [binge/intoxication], executive
353 function [preoccupation/anticipation], neuroticism [negative affect]¹⁹; **Figure 2**). It was also
354 associated with non-substance psychopathology factors (compulsive disorders, psychotic and
355 mood disorders, neurodevelopmental disorders; **Figure 3**). Trait representations of stage-based
356 facets of addiction partially accounted for the shared genetic liability between non-substance
357 psychopathology and $a_{(g)}$. *Third*, associations between $a_{(g)}$ and stage-based constructs and non-
358 substance psychopathology were largely independent of genetic liability to substance *use*
359 phenotypes (i.e., tobacco use, cannabis use, alcoholic drinks/week). However, consistent with the
360 stage-based model of addiction, accounting for substance use attenuated associations between
361 risk taking and $a_{(g)}$ while potentiating associations with executive functioning. Collectively, our
362 findings suggest that SUDs are characterized by a common genetic factor, $a_{(g)}$. This shared
363 genetic etiology may be leveraged to refine nosology and increase statistical power for research
364 in treatment and prevention.

365

366 **$a_{(g)}$ retains variance that is not shared with other psychopathology**

367 Even after accounting for genetic liability to substance use as well as the commonality between
368 $a_{(g)}$ and non-substance psychopathology, $a_{(g)}$ retained a substantial proportion of specific genetic
369 variance. These data suggest that $a_{(g)}$ may be characterized by unique pathways that are not
370 shared with substance use or non-substance psychopathology, i.e., addiction is not the linear
371 combination of substance use and psychopathology.

372 A single latent factor, $a_{(g)}$, fit these data well, but specific SUDs showed varying degrees
373 of association. The illicit SUDs (CUD and OUD; **Figure 1**) were almost entirely captured by the
374 common latent factor, unlike PAU and PTU, suggesting that heightened $a_{(g)}$ may confer liability
375 for potentially illicit substance use disorder. Notably, the loading for PTU on $a_{(g)}$ was lower than
376 other substances. One potential contributor to the residual genetic variance of PTU may be the
377 use of FTND and cigarettes/day as indices of PTU. Unlike the Diagnostic and Statistical Manual
378 (DSM) criteria which index psychological and physiological aspects of tobacco use disorder, the
379 FTND, which includes amount smoked, is an index of biochemical dependence and shows only
380 moderate agreement with DSM-defined nicotine dependence ($r = 0.50$; kappa = 0.3)⁴⁴.

381

382 **Stage-based Behavioral Constructs and $a_{(g)}$**

383 Behavioral stage-based models of SUD posit a cyclical relationship between positive
384 reinforcement, negative reinforcement, and incentive salience¹⁹ that we hypothesized can be
385 (partially) captured by individual differences in genetic liability to risk-taking, executive
386 functioning, and negative emotionality. The strongest association with $a_{(g)}$ was for risk-taking.
387 Consistent with this, when substance use was included in the model the shared genetic loading
388 between $a_{(g)}$ and both risk-taking and neuroticism was attenuated while the association with
389 executive function increased. The reduction in the association with neuroticism is counter to

390 expectations from the stage-based model which posits a more prominent role of negative affect
391 for SUD relative to substance use. We speculate that neuroticism, which represents an amalgam
392 of negative affect traits, may be too broad a construct when considering SUD-specific negative
393 affect; large-scale studies of domains of negative affectivity (e.g., negative urgency) are needed.

394

395 **Non-substance Psychopathology and $a_{(g)}$**

396 We found that the 3 non-substance psychopathology clusters, derived from 8 psychiatric
397 disorders¹⁰, were genetically associated with $a_{(g)}$. The association with early neurodevelopmental
398 disorders, which include ADHD, was the strongest. Cross-loading ADHD on $a_{(g)}$ attenuated the
399 loading but it remained high, possibly because Major Depressive Disorder also loaded on this
400 factor. Associations between $a_{(g)}$ and the psychopathology clusters were greater in magnitude
401 than associations with trait representations of behavioral stages of addiction (with the possible
402 exception of risk-taking). For instance, the genetic association between $a_{(g)}$ and the two disorder
403 clusters that included Major Depressive Disorder (i.e., mood/psychotic disorders and early
404 neurodevelopmental disorders) was greater in magnitude than the $a_{(g)}$ -neuroticism association.
405 These findings reinforce the idea that from a genetic perspective, SUDs align closely with
406 pathophysiological states.

407 Of the 3 behavioral correlates, risk-taking was the most prominent contributor to the
408 association between $a_{(g)}$ and all non-substance psychopathology factors. Executive function also
409 contributed to the association between $a_{(g)}$ and psychotic/mood disorders. After accounting for
410 substance use, risk-taking and executive function only mediated $a_{(g)}$ associations with the
411 psychotic/mood disorder factor. Executive function maintained the only direct association with
412 $a_{(g)}$ after accounting for genetics of substance use and genetics of non-substance

413 psychopathology. Thus, we speculate that while risk-taking may characterize the genetic overlap
414 between substance use and other psychopathology, executive function impairment is a risk factor
415 that not only shapes the overlap between addiction and non-substance psychopathology but also
416 explains variance in addiction above and beyond that overlap. Thus, our analyses highlight the
417 prominence of frontal brain executive function impairments in SUDs.

418

419 **Limitations**

420 There are several limitations of note. First, we had to restrict our analyses to those of European
421 descent due to the lack of well-powered discovery GWAS informative of other ancestry groups.
422 Second, we focused on common genetic variants, so it is unclear if our identified factor structure
423 will generalize to rare genetic variation; because our findings align with evidence from twin
424 studies, which aggregate effects of common and rare variants, we suspect it may. Third, to
425 maximize sample size of discovery GWASs, our alcohol and tobacco use GWAS incorporated
426 measures of “problematic” use that, while genetically highly correlated with AUD and ND, may
427 include behavioral patterns that are less severe than those represented by use disorder. Finally,
428 the analyses contain an over-representation of men, in part because the MVP samples contributed
429 most of OUD and half of PAU and this sample is ~90% male. Studies with larger numbers of
430 women would allow stratified analyses to explore the differences between sexes observed in
431 epidemiological studies.

432

433 **Conclusions**

434 Common genetic liability, the $a_{(g)}$ -factor largely undergirds distinct SUDs and shares variance
435 with putative behavioral intermediary phenotypes/SUD risk factors and non-substance

436 psychopathology. This genetic factor is more than a linear combination of substance use and
437 psychopathology, but represents a unique addiction dimension that is partially captured by
438 executive functions.

439

440

441 **Reference Cited**

- 442 1. Degenhardt, L. *et al.* The global burden of disease attributable to alcohol and drug use in
443 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of
444 Disease Study 2016. *The Lancet Psychiatry* **5**, 987–1012 (2018).
- 445 2. Bhalla, I. P., Stefanovics, E. A. & Rosenheck, R. A. Clinical Epidemiology of Single
446 Versus Multiple Substance Use Disorders. *Med. Care* **55**, S24–S32 (2017).
- 447 3. Merikangas, K. R. & Kalaydjian, A. Magnitude and impact of comorbidity of mental
448 disorders from epidemiologic surveys. *Current Opinion in Psychiatry* vol. 20 353–358
449 (2007).
- 450 4. Palmer, R. H. C. *et al.* Genetic etiology of the common liability to drug dependence:
451 evidence of common and specific mechanisms for DSM-IV dependence symptoms. *Drug*
452 *Alcohol Depend.* **123 Suppl**, S24–32 (2012).
- 453 5. Fu, Q. *et al.* Shared Genetic Risk of Major Depression, Alcohol Dependence, and
454 Marijuana Dependence. *Arch. Gen. Psychiatry* **59**, 1125 (2002).
- 455 6. Carey, C. E. *et al.* Associations between polygenic risk for psychiatric disorders and
456 substance involvement. *Front. Genet.* **7**, (2016).
- 457 7. Zhou, H. *et al.* Association of OPRM1 Functional Coding Variant with Opioid Use
458 Disorder: A Genome-Wide Association Study. *JAMA Psychiatry* (2020)
459 doi:10.1001/jamapsychiatry.2020.1206.
- 460 8. Sanchez-Roige, S., Palmer, A. A. & Clarke, T. K. Recent Efforts to Dissect the Genetic
461 Basis of Alcohol Use and Abuse. *Biological Psychiatry* vol. 87 609–618 (2020).
- 462 9. Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of
463 cannabis use disorder. *The Lancet Psychiatry* **0**, (2020).
- 464 10. Lee, P. H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across
465 Eight Psychiatric Disorders. *Cell* **179**, 1469–1482.e11 (2019).
- 466 11. Jang, S. K. *et al.* Genetic correlation, pleiotropy, and causal associations between
467 substance use and psychiatric disorder. *Psychol. Med.* 1–11 (2020)
468 doi:10.1017/S003329172000272X.
- 469 12. Waldman, I. D., Poore, H. E., Luningham, J. M. & Yang, J. Testing structural models of
470 psychopathology at the genomic level. *World Psychiatry* **19**, 350–359 (2020).
- 471 13. Karlsson Linnér, R. *et al.* Multivariate genomic analysis of 1.5 million people identifies
472 genes related to addiction, antisocial behavior, and health. *Henry R. Kranzler* **17**, 20
473 (2020).
- 474 14. Polimanti, R. *et al.* Leveraging genome-wide data to investigate differences between
475 opioid use vs. opioid dependence in 41,176 individuals from the Psychiatric Genomics
476 Consortium. *Mol. Psychiatry* **25**, 1673–1687 (2020).
- 477 15. Kranzler, H. R. *et al.* Genome-wide association study of alcohol consumption and use
478 disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10**, (2019).
- 479 16. Sanchez-Roige, S. *et al.* Genome-wide association study meta-analysis of the alcohol use
480 disorders identification test (AUDIT) in two population-based cohorts. *Am. J. Psychiatry*
481 **176**, 107–118 (2019).
- 482 17. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common
483 genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
- 484 18. Neale, M. C. & Kendler, K. S. Models of comorbidity for multifactorial disorders. *Am. J.*
485 *Hum. Genet.* (1995).
- 486 19. Goldstein, R. Z. & Volkow, N. D. Drug addiction and its underlying neurobiological

- 487 basis: Neuroimaging evidence for the involvement of the frontal cortex. *American Journal*
488 *of Psychiatry* vol. 159 1642–1652 (2002).
- 489 20. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & van der Sluis, S. Item-level
490 analyses reveal genetic heterogeneity in neuroticism. *Nat. Commun.* **9**, 905 (2018).
- 491 21. Koob, G. F. & Volkow, N. D. Neurobiology of addiction: a neurocircuitry analysis. *The*
492 *Lancet Psychiatry* vol. 3 760–773 (2016).
- 493 22. Zhou, H. *et al.* Genome-wide meta-analysis of problematic alcohol use in 435,563
494 individuals yields insights into biology and relationships with other traits. *Nat. Neurosci.*
495 1–10 (2020) doi:10.1038/s41593-020-0643-5.
- 496 23. Hatoum, A. *et al.* GWAS of Over 427,000 Individuals Establishes GABAergic and
497 Synaptic Molecular Pathways as Key for Cognitive Executive Functions. *GWAS Over*
498 *427,000 Individ. Establ. GABAergic Synaptic Mol. Pathways as Key Cogn. Exec. Funct.*
499 674515 (2019) doi:10.1101/674515.
- 500 24. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the
501 multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
- 502 25. Hancock, D. B. *et al.* Genome-wide association study across European and African
503 American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence.
504 *Mol. Psychiatry* **23**, 1–9 (2018).
- 505 26. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into
506 the genetic etiology of tobacco and alcohol use. *Nature Genetics* vol. 51 237–244 (2019).
- 507 27. Pasmán, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap
508 with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.* **21**, 1161–
509 1170 (2018).
- 510 28. Edenberg, H. J. & McClintick, J. N. Alcohol Dehydrogenases, Aldehyde Dehydrogenases,
511 and Alcohol Use Disorders: A Critical Review. *Alcoholism: Clinical and Experimental*
512 *Research* vol. 42 2281–2297 (2018).
- 513 29. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
514 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 515 30. Quach, B. C. *et al.* Expanding the genetic architecture of nicotine dependence and its
516 shared genetics with multiple traits. *Nat. Commun.* **11**, 1–13 (2020).
- 517 31. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using
518 MTAG. *Nat. Genet.* **50**, 229–237 (2018).
- 519 32. Strawbridge, R. J. *et al.* Genetics of self-reported risk-taking behaviour, trans-ethnic
520 consistency and relevance to brain gene expression. *Transl. Psychiatry* **8**, (2018).
- 521 33. Volkow, N. D., Koob, G. F. & McLellan, A. T. Neurobiologic Advances from the Brain
522 Disease Model of Addiction. *N. Engl. J. Med.* **374**, 363–371 (2016).
- 523 34. Bondy, E. *et al.* Neuroticism and reward-related ventral striatum activity: Probing
524 vulnerability to stress-related depression. (2020) doi:10.31234/osf.io/5wd3k.
- 525 35. Schizophrenia Working Group of the Psychiatric Genomics Consortium, S. W. G. of the
526 P. G. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature*
527 **511**, 421–7 (2014).
- 528 36. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with
529 bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
- 530 37. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102
531 independent variants and highlights the importance of the prefrontal brain regions. *Nat.*
532 *Neurosci.* **22**, 343–352 (2019).

- 533 38. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention
534 deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
- 535 39. Arnold, P. D. *et al.* Revealing the complex genetic architecture of obsessive-compulsive
536 disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
- 537 40. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and
538 implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214
539 (2019).
- 540 41. Yu, D. *et al.* Interrogating the genetic determinants of Tourette’s syndrome and other tic
541 disorders through genome-wide association studies. *Am. J. Psychiatry* **176**, 217–227
542 (2019).
- 543 42. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum
544 disorder. *Nat. Genet.* **51**, 431–444 (2019).
- 545 43. Grotzinger, A. D. *et al.* Genomic SEM Provides Insights into the Multivariate Genetic
546 Architecture of Complex Traits. *bioRxiv* 305029 (2018) doi:10.1101/305029.
- 547 44. Agrawal, A. *et al.* A latent class analysis of DSM-IV and fagerström (FTND) criteria for
548 nicotine dependence. *Nicotine Tob. Res.* **13**, 972–981 (2011).
- 549

Table 1. Behavioral Liabilities Mediate the Association Between $a_{(g)}$ and Psychiatric Factors

	F1 Indirect %	F1 Indirect P	F2 Indirect %	F2 Indirect P	F3 Indirect %	F3 Indirect P
No covariates						
executive function	0.007	0.463	0.051*	0.001*	0.003	0.486
Neuroticism	0.005	0.842	0.003	0.836	0.013	0.777
Risk-taking	-0.090*	0.001*	0.152*	1.47E-08*	0.043*	0.013*
Controlling for Genetics of Substance Use						
executive function	0.013	0.233	0.048*	0.047*	-0.007	0.491
Neuroticism	0.011	0.687	-0.022	0.693	-0.040	0.689
Risk-taking	0.001	0.509	0.019	0.446	-0.002	0.741

Table 2. Behavioral Liabilities Mediate the Association Between $a_{(g)}$ and Psychiatric Factors

Indirect associations from a mediation model (see Figure 4) where stage-based constructs link non-substance psychopathology (3 factors from Lee et al.,) and $a_{(g)}$. F1 = compulsive disorders, F2 = Psychotic disorders, F3 = Neurodevelopmental disorders. The proportion accounted for by the indirect association (%) and the significance of the indirect association are shown. * $P < .05$ for the indirect association pathway.

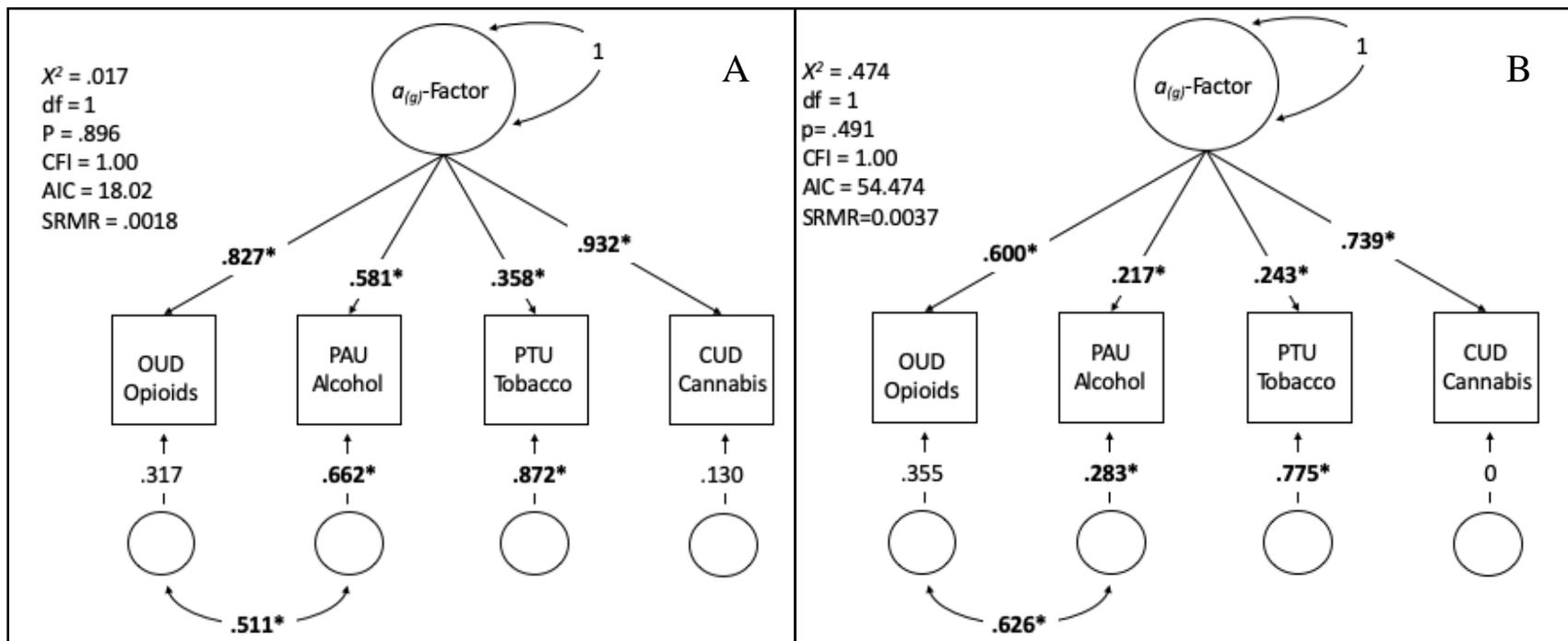


Figure 1. Factor Structure of 4 SUD GWAS. Panel A: the model, loadings, and fit for a model that allowed all 4 SUD categories to load on a latent factor. A residual correlation was added between OUD and PAU to account for large sample overlap (both used the Million Veterans Project data; models without residual correlations also fit well: Supplemental Figure 1). Panel B: the same model, but accounting for common substance use (ever smoke, ever use marijuana, and drinks per week) as covariates at the indicator level, i.e. the three substance use measures are exogenous to all indicators in this model and the model represents the residual associations after accounting for substance use. Both models provided excellent fit to the data. **Bold*** represents significance at $p < .05$.

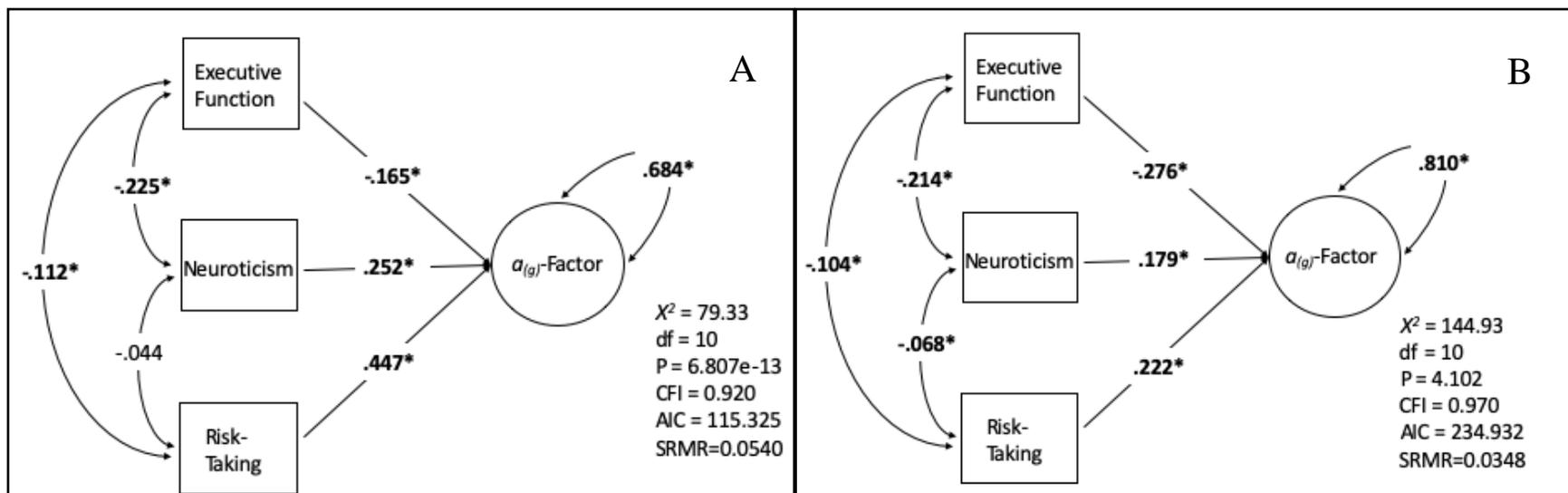


Figure 2. Genetic associations between $a_{(g)}$ and behavioral traits from the behavioral model of Addiction: Executive Function, Neuroticism and Risk-taking. Panel A the model, fit, and regression pathways without accounting for common substance use i.e. the three substance use measures are exogenous to all indicators in this model and the model is the residual associations after accounting for substance use (model was scaled by setting the Opioid Use Disorder loading to 1). Panel B is the same model, but accounting for common substance use (ever smoke, ever use marijuana, and drinks per week) as covariates at the indicator level (regressed on all measured variables/GWAS). **Bold*** represents significance at $p < .05$.

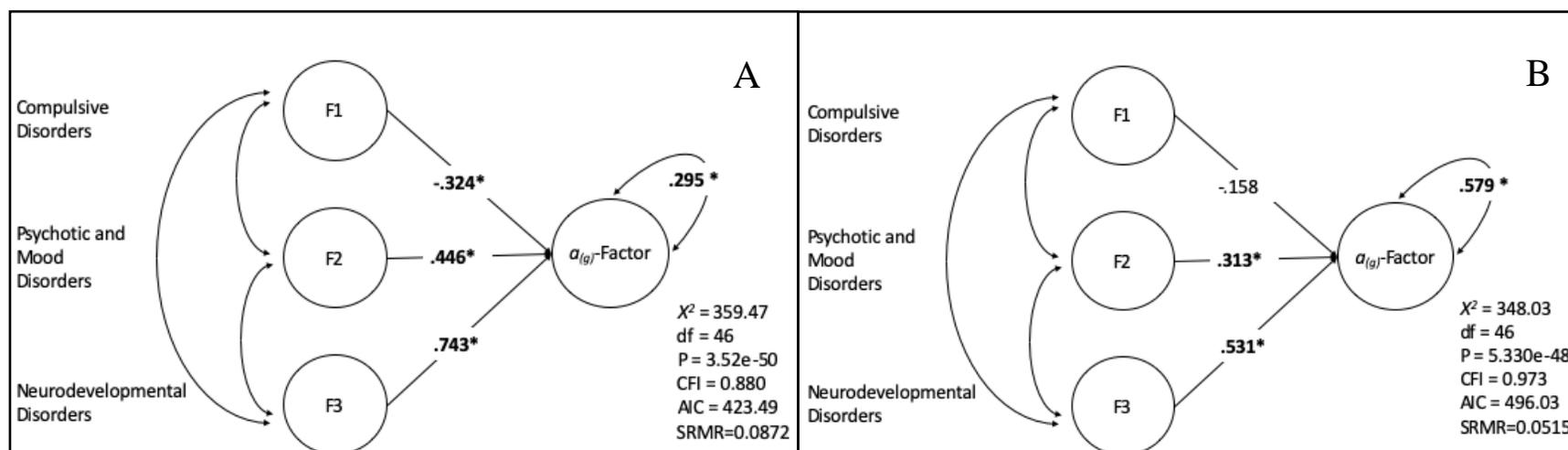


Figure 3. Genetic associations between $a_{(g)}$ and Latent Psychopathology Factors: Compulsive disorders (F1; Tourette's syndrome, Obsessive compulsive disorder, and Eating Disorders), Psychotic Disorders (F2; Major Depressive Disorder, Schizophrenia, and Bipolar Disorder) and neurodevelopmental dysfunction (F3; ADHD, Autism, and Major Depressive Disorder). Panel A the model, fit, and regression pathways without accounting for common substance use (model was scaled by setting the Opioid Use Disorder loading to 1). Panel B is the same model, but accounting for common substance use (ever smoke, ever use marijuana, and drinks per week) as covariates at the indicator level (regressed on all measured variables/GWAS), i.e. the three substance use measures are exogenous to all indicators in this model and the model is the residual associations after accounting for substance use. **Bold*** represents significance at $p < .05$.