Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Polygenic contribution to cannabis use disorder shared with and independent of ADHD: a GWAS-by-subtraction approach

Marta Ribasés

marta.ribases@vhir.org

Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona https://orcid.org/0000-0003-1039-1116

Judit Cabana-Domínguez

Vall d'Hebron Research Institute (VHIR) https://orcid.org/0000-0002-4732-7284

Natalia Llonga

Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona https://orcid.org/0000-0003-4325-6201

Silvia Alemany

vhir https://orcid.org/0000-0002-7925-6767

Triinu Varvas

Estonian Genome Centre, Institute of Genomics, University of Tartu https://orcid.org/0009-0000-9321-9367

Laura Vilar-Ribó

UCSD https://orcid.org/0000-0001-8758-523X

Uxue Zubizarreta-Arruti

Vall d'Hebron Research Institute (VHIR) https://orcid.org/0009-0005-2267-6196

Pau Carabí-Gassol

Vall d'Hebron Research Institute (VHIR) https://orcid.org/0000-0001-9105-2946

Manuel D. Rodríguez-Romero

Vall d'Hebron Research Institut

Valeria Macias-Chimborazo

Vall d'Hebron Research Institut

Christian Fadeuilhe

Hospital Universitari Vall d'Hebron (IIS-IR-HUVH)

Corrales Montse

Department of Mental Health, Hospital Universitari Vall d'Hebron

Vanesa Richarte

Hospital Universitari Vall d'Hebron (IIS-IR-HUVH)

Silvia Amoretti
Bipolar and Depressive Disorders Unit, Institute of Neuroscience, Hospital Clinic, University of Barcelona, Barcelona, Spain https://orcid.org/0000-0001-6017-2734
Lara Grau-López
Hospital Universitari Vall d'Hebron
Daigre Constanza
Department of Mental Health, Hospital Universitari Vall d'Hebron
Ros-Cucurull Elena
Department of Mental Health, Hospital Universitari Vall d'Hebron
Palma-Alvarez Raúl
Department of Mental Health, Hospital Universitari Vall d'Hebron
Alex Segura
Vall d'Hebron Institut de Recerca https://orcid.org/0000-0002-9398-2183
Skojo Matea
Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona
Bru Cormand
Universitat de Barcelona https://orcid.org/0000-0001-5318-4382
Fernàndez-Castillo Noèlia
Universitat de Barcelona
Kadri Kõiv
University of Tartu https://orcid.org/0000-0001-5286-2014
Kelli Lehto
Estonian Genome Centre, Institute of Genomics, University of Tartu https://orcid.org/0000-0003-2735-7643
Josep Antoni Ramos-Quiroga
Hospital Universitari Vall d'Hebron. https://orcid.org/0000-0003-1622-0350
María Soler Artigas
Vall d'Hebron Research Institute https://orcid.org/0000-0002-3213-1107

Article

Keywords:

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11569749/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. CF has received fees to give talks for Takeda, Rubio, Casen-Recordati and Idorsia. JARQ was on the speakers' bureau and/or acted as consultant for Biogen, Janssen-Cilag, Idorsia, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice, Rubió, Uriach, Technofarma and Raffo in the last 3 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Idorsia, Janssen-Cilag, Rubió, Shire, Takeda, Shionogi, Bial and Medice. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 3 years: Janssen- Cilag, Shire, Oryzon, Roche, Psious, Idorsia, Neuraxpharm, Exeltis, and Rubió. MR received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Rubió. All other authors declare no biomedical financial interests or conflicts of interest.

Polygenic contribution to cannabis use disorder shared with and independent of ADHD: a GWAS-by-subtraction approach

Judit Cabana-Domínguez[1,2,3,4,&], Natalia Llonga[1,2,3,4,&], Silvia Alemany[1,2,3], Triinu Varvas[5], Laura Vilar-Ribó[6], Uxue Zubizarreta-Arruti[1,2,3,4], Pau Carabí-Gassol[1,2,3,4], Manuel Donato Rodríguez-Romero[1,2,3,4], Valeria Macias-Chimborazo[1,2,3], Christian Fadeuilhe[1,2,3,7], Montse Corrales[1,2,3,7], Vanesa Richarte[1,2,3,7], Silvia Amoretti[1,2,3], Lara Grau-López[2,3,8], Constanza Daigre[2,3,8], Elena Ros-Cucurull[2,8], Raúl Felipe Palma-Alvarez[2,3,8], Àlex Gozález-Segura[1,2,3], Matea Skojo[1,2,3], Bru Cormand[4,9,10,11], Noèlia Fernàndez-Castillo[4,9,10,11], Kadri Kõiv[5], Kelli Lehto[5], Estonian Biobank Research Team[5], Josep Antoni Ramos-Quiroga[1,2,3,7], María Soler Artigas[1,2,3,4,*], Marta Ribasés[1,2,3,4,*]

[1] Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Instituto de Investigación Sanitaria Acreditado Instituto de Investigación - Hospital Universitario Vall d'Hebron (IR-HUVH), Barcelona, Spain.

[2] Department of Mental Health, Hospital Universitari Vall d'Hebron, Barcelona, Catalonia, Spain.

[3] Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain.

[4] Department of Genetics, Microbiology, and Statistics, Faculty of Biology, Universitat de Barcelona (UB), Barcelona, Spain.

[5] Estonian Genome Centre, Institute of Genomics, University of Tartu, Estonia.

[6] Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

[7] Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona (UAB), Barcelona, Catalonia, Spain

[8] Addiction and Dual Diagnosis Unit, Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Catalonia, Spain

[9] Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III (ISCIII), 28029 Madrid, Spain.

[10] Sant Joan de Déu Research Institute (IRSJD), Esplugues de Llobregat, Catalonia, Spain

[11] Institute of Biomedicine of the University of Barcelona (IBUB), Barcelona, Catalonia, Spain.

& These authors contributed equally.

*Corresponding authors: Marta Ribasés and María Soler Artigas. Psychiatric Genetics Unit, Vall d'Hebron Research Institute (VHIR), Department of Psychiatry Hospital Universitari Vall d'Hebron, Passeig Vall d'Hebron, 119-129, 08035 Barcelona, Spain Phone: +34 934894162. marta.ribases@vhir.org and maria.soler@vhir.org.

ABSTRACT

Cannabis use disorder (CUD) is a heterogeneous condition characterized by different patterns of comorbidity and clinical presentation, yet the genetic contributors to this heterogeneity remain poorly understood. CUD frequently co-occurs with attention-deficit/hyperactivity disorder (ADHD), reflecting, in part, their shared genetic liability. Distinguishing the genetic component of CUD that operates through ADHD from the component independent of ADHD may help clarify whether these two components differ in their association with substance-use phenotypes, psychiatric traits, and broader health outcomes. We applied genomic structural equation modeling–based GWAS-by-subtraction to large-scale ADHD and CUD genome-wide association studies to characterize these two CUD components and investigate their genetic and phenotypic association profiles. The ADHD-independent component accounted for approximately 35% of CUD SNP-based heritability and included three genome-wide significant loci. Although both components showed genetic correlations with substance dependence and risk-taking behaviors, they exhibited distinct patterns of genetic correlation with substance use phenotypes, other mental disorders, and health outcomes. In a phenome-wide association study of 114,410 Estonian Biobank participants, both genetic components were associated with multiple health conditions, but the ADHD-related genetic component generally showed larger effects. Across two clinical samples, polygenic score generated with both components further revealed distinct associations with the comorbidity between ADHD and substance use disorders (SUD). Both scores were associated with SUD among individuals without ADHD. In contrast, only the polygenic score for the ADHD-related component (PGS_{adhd}) was associated with ADHD and comorbid ADHD and SUD, with the strongest association observed in the comorbid group. Mediation analysis further indicated that ADHD diagnosis accounted for approximately 20% of the association between PGS_{adhd} and SUD. Together, these findings demonstrate that ADHD-related and ADHD-independent components of CUD have distinct genetic and phenotypic association profiles, suggesting that genetic heterogeneity may contribute to the clinical heterogeneity of SUD. This framework may help guide further research into the etiology and clinical presentation of the disorder.

INTRODUCTION

Substance use disorders (SUD) and attention-deficit/hyperactivity disorder (ADHD) are both prevalent, co-occurring mental health conditions that significantly influence the course and severity of the disorders and long-term outcomes in affected individuals[1,2]. Cannabis is the most common illicit drug used by individuals with ADHD, who are up to two times more likely to consume cannabis compared to individuals without ADHD and have nearly three times greater risk for experiencing cannabis use disorder (CUD) than the general population[3]. Also, adolescents with problematic cannabis use, cannabis abuse and CUD receive a diagnosis of ADHD in more than 30% of the cases[4]. Furthermore, persistent ADHD has been associated with an increased likelihood of substance dependence in young adulthood, in a longitudinal study [5]. Together, these findings support the idea that ADHD may contribute to the development of CUD.

A major contributor to the high rates of comorbidity between ADHD and CUD is the shared genetic liability underlying both conditions[6]. ADHD and CUD are highly polygenic, with many variants exerting an individual small influence on risk and single nucleotide polymorphism-based heritability (h^2_{SNP}) around 14% and 6.7%, respectively[7,8]. Recent research has shown a strong positive genetic correlation between ADHD and CUD (rg = 0.57) and Genomic Structural Equation Modeling (GenomicSEM) suggests a latent genetic factor encompassing the two disorders[9,10]. Also, a multivariate genome-wide association study (GWAS) of ADHD and CUD identified 36 genome-wide significant loci, 17 of them novel for both disorders, and highlighted genes with increased expression across brain tissues and developmental stages[10]. Polygenic score (PGS) analyses further support these findings. Individuals with CUD exhibit higher PGS for ADHD compared to controls[11]. Also, individuals with comorbid ADHD and CUD show a greater genetic load of common variants for both disorders, in addition to an increased burden of rare variants, than individuals with ADHD without CUD[10]. Stratification of individuals with ADHD by their PGS for CUD also revealed that those in the highest quartile of polygenic risk for CUD have a 22% absolute risk of developing comorbid CUD versus 1.6% risk for controls[10].

Although ADHD and CUD show substantial genetic overlap, distinguishing the ADHD-related genetic component of CUD from the residual component independent of ADHD may help clarify whether these components contribute differently to the clinical heterogeneity of CUD. Here, we applied GWAS-by-subtraction to summary statistics from the largest GWAS datasets currently available[7,8,12] to derive an ADHD-shared component of CUD and a residual ADHD-independent component. We then characterized these components using complementary analyses (Supplementary Figure 1): genetic-correlation analyses to compare the similarity of their genome-wide association profiles, and a population-based phenome-wide association study

(PheWAS) to test their polygenic score (PGS) associations with a broad range of recorded diagnoses. Finally, we examined associations with ADHD, substance use disorders (SUD), and their co-occurrence in two clinical samples. Together, these analyses aimed to characterize the genetic architecture of CUD in relation to ADHD and determine whether the two components exhibit distinct genetic and phenotypic association profiles.

RESULTS

SNP-based heritability and genetic overlap between CUD and ADHD

The estimated SNP heritability (h^2_{SNP}) on the liability scale was 7.5% (s.e.=0.4) for CUD and 17.6% (s.e.=0.8) for ADHD, and there was a strong positive genetic correlation between them (rg=0.61, s.e.=0.027 P=8.2e-116) (Supplementary Table 1). This significant genetic correlation was consistent with the stratified cross-phenotype Q-Q plots, where p-values for one trait were plotted conditioning on different association strengths for the other trait (Supplementary Figure 2). A stronger leftward deflection from the null expectation was observed when conditioning by increased levels of association for both phenotypes.

Univariate MiXeR revealed that both CUD and ADHD are highly polygenic, with 6,097 variants (s.e.=442) and 7,459 variants (s.e.=311) explaining 90% of h^2_{SNP} of CUD and ADHD, respectively (Supplementary Table 2 and Supplementary Figure 2). We found that 95% of CUD-influencing variants are shared with ADHD (5,805 variants (s.e.=434); Dice coefficient=0.86), with a high proportion of variants showing concordant direction of the effect (76%) (Supplementary Table 2 and Supplementary Figure 2).

GWAS-by-subtraction

We dissected the genetic variation in CUD by the genetic variation in ADHD applying GWAS-by-subtraction with GenomicSEM on summary statistics from GWAS on ADHD and CUD[7,8] (Supplementary Table 3a, Supplementary Figure 3).

The SNP heritability for the non_*adhd* contribution to CUD ($h^2_{SNP\text{-}non_adhd}$) was estimated as 3.1% (s.e.=0.2), indicating that the non_*adhd* latent factor accounts for a substantial part of the total variance, around 35%, in CUD (Supplementary Table 1). There was no evidence of residual stratification or confounding leading to an inflation of test statistics, with a LD score regression intercept of 1.00 (s.e.=0.007) and a ratio of 0.0024 (s.e.=0.042), which indicates that most of the

deviation from the null in the distribution of the test statistics reflects polygenicity (Supplementary Table 1 and Supplementary Figure 4).

Overall, results were consistent after sensitivity analysis allowing for moderate causal effect of CUD on ADHD (Spearman's correlation coefficient r=0.93; P-value=5.4e-12) (Supplementary Table 3b and Supplementary Figure 5). The GWAS on *non_adhd* identified three independent loci that were previously associated with CUD with weaker effect and consistent direction of the association, but not with ADHD (Figure 1 and Supplementary Table 4). The functional annotation of variants associated with the non_*adhd* factor revealed over-representation of SNPs lying in introns (11.2%; P=1.7e-17), intergenic regions (69.2%; P=9.9e-12) as well as upstream (3.1%; P=0.012) and downstream of known genes (2.7%; P=0.044) (Supplementary Figure 6). Notably, 95.1% of the significant variants tagged eQTLs in brain, 86.6% were located in open chromatin regions (minimum chromatin state ≤ 7), 1.3% were predicted to affect the binding of transcription factors (RegulomeDB scores from 1b to 2c), and 4.9% were potentially deleterious (CADD score > 12.37) (Supplementary Table 5). The three loci identified were previously related (P<5e-08) with CUD, smoking (e.g., smoking status, smoking initiation, lifetime smoking and smoking cessation), risk-taking behavior (number of sexual partners) and mental disorders or related traits (major depressive disorder, bipolar disorder, schizophrenia, neuroticism and externalizing behavior) (Supplementary Table 6). Variants associated with the *non_adhd* factor mapped to 108 unique genes significant enriched in interesting gene-sets from the GWAScatalog including brain morphology and mental disorders (Supplementary Table 7). The gene-based analysis identified three significantly associated genes for the *non_adhd* factor, *CLU, TAF12* and *RHOA*s (Supplementary Table 8), while no significant gene-sets were found after multiple comparison correction.

Genetic correlations for adhd and non_adhd factors

The *adhd* and *non_adhd* factors showed significant differences in their pattern of genetic correlation with many disorders and traits (Supplementary Table 9 and 10).

SUD, risk taking and smoking behavior

Adhd and *non_adhd* genetics showed a similar relationship with risk-taking behaviors and substance use disorders, both factors being associated with higher risk tolerance, risk-taking, and number of sexual partners and increased risk for alcohol and opioid dependence. However, *non_adhd* showed positive and stronger genetic correlation with substance use than *adhd*, which showed genetic correlations either of smaller magnitude (e.g., lifetime cannabis use and alcohol

drinks per week) or in opposite direction (e.g., alcohol use). In contrast, *adhd* correlated with smoking-related traits (e.g., age of smoking initiation, ever smoker, smoking cessation, smoking initiation and cigarettes per day), whereas genetic correlations with *non_adhd* were of smaller magnitude for all of them, including for cigarettes per day that was non-significant (Figure 2 and Supplementary Table 10).

Mental disorders, personality, cognition and other traits

For most mental disorders and related traits (7 out of 10), *adhd* and *non_adhd* displayed significantly different pattern of genetic correlation. The genetic correlations for *adhd* were similar, but generally of higher magnitude, to genetic correlations identified for CUD. *Adhd* showed a negative genetic correlation with anorexia nervosa but positively correlated with other mental disorders and related traits including antisocial behavior, anxiety disorder, autism spectrum disorder (ASD), bipolar disorder, depression, loneliness, schizophrenia and suicide attempt. *Non_adhd* genetic correlations were either of smaller magnitude than *adhd* (e.g., antisocial behavior, depression and suicide attempt), in the opposite direction (e.g., ASD) or non-significant (e.g., anorexia nervosa, anxiety disorder and loneliness). Notably, schizophrenia displayed a stronger positive genetic correlation with *non_adhd*, while bipolar disorder showed similar positive genetic correlation with both addiction factors (Figure 3 and Supplementary Table 10).

Both *adhd* and *non_adhd* factors displayed a negative correlation with cognitive outcomes (e.g., cognitive performance, educational attainment or intelligence quotient), as well as with subjective well-being, neuroticism, longevity, and age at first intercourse, with genetic correlations for *adhd* being of higher magnitude than for *non_adhd*. Also, BMI and overweight and obesity showed positive correlation with *adhd*, but displayed opposite pattern of correlation or no genetic correlation with *non_adhd*, respectively. Finally, number of children showed a positive correlation with both genetic factors, but of higher magnititude for *adhd* than for *non-adhd* (Figure 3 and Supplementary Table 10).

Phenome-wide association study in the Estonian Biobank (EstBB)

We constructed genome-wide polygenic scores for the two genetic factors estimated by GWAS-by-subtraction (PGS_{adhd} and PGS_{non_adhd}) and performed a PheWAS to test their association with 1,092 electronic health record–based ICD-10 codes in a total of 114,410 individuals from the EstBB.

The association of both PGS_{adhd} and PGS_{non_adhd} with ADHD diagnosis (F90; 0.01% of the individuals with ADHD; n=1,174) revealed that only PGS_{adhd} was associated with the disorder (OR= 1.33, 95% CI 1.29-1.37; P=2.72e-21), whereas no association was observed for PGS_{non_adhd}. In addition to ADHD diagnosis, the PheWAS revealed evidence of association of PGS_{adhd} and/or PGS_{non_adhd} with 192 additional ICD-10 codes (21 mental health, 123 general health, and 48 external causes) after Bonferroni correction for multiple testing (Supplementary Table 11). Of those, a total of 15 were associated with PGS_{non_adhd}, which included mental and behavioral disorders due to alcohol and multiple drug use; emotional or behavioral reaction to severe stress; self-harm as intentional self-poisoning by exposure to drugs, external causes such as accidents related to poisoning (including alcohol and other drugs) or fractures; as well as other health and lifestyle-related problems and infectious diseases such as chronic hepatitis (Supplementary Table 11).

When we examine ICD-10 codes associated to either PGS_{adhd} and/or PGS_{non_adhd} in the PheWAS to assess differences in effect sizes, we identify significant differences for 81 diagnostic codes after Bonferroni correction for multiple testing (0.05/193=2.59e-04), reflecting distinct patterns of association between the two genome-wide polygenic score profiles. In all cases PGS_{non_adhd} showed associations in the same direction as PGS_{adhd}, but with smaller effect sizes. The only exception was abnormalities of breathing (R06), which exhibited an opposite pattern of association between PGS_{adhd} and PGS_{non_adhd}. Overall, the medical conditions showing the strongest evidence of differential effect sizes were obesity (E66; P=1.56e-38), dorsalgia (M54; P=1.48e-21), gastro-oesophageal reflux disease (K21; P1.14e-19), mononeuropathies of upper limb (G56; P=2.89e-19), headache syndromes (G44; P=4.46e-18) and type 2 diabetes mellitus (E11; P=5.37e-15) (Figure 4 and Supplementary Table 12a).

Focusing on mental health conditions, we found that, in addition to ADHD (F90), nicotine dependence (F17), other mental disorders due to known physiological condition (F06), and eating disorders (F50) were only associated with PGS_{adhd}. In contrast, depression-related traits (F32 and F33), anxiety disorders (F41), and emotional or behavioral reaction to severe stress (F43) showed association with both PGSs in the same direction but with stronger effects for PGS_{adhd} (Figure 4 and Supplementary Table 12a).

Polygenic score analyses in ADHD and/or SUD clinical samples

To investigate the contribution of the *adhd* and *non_adhd* genetic components in the comorbidity between SUD and ADHD, we conducted PGS analyses constructed on CUD (PGS_{CUD}) and the two

genetic components estimated through GWAS-by-subtraction (PGS_{adhd} and PGS_{non_adhd}) in two independent clinical datasets comprising individuals with SUD and/or ADHD, as well as controls.

In the SUD clinical sample, all three PGS were positively associated with SUD, with similar effect sizes in the overall SUD sample and the SUD-only subgroup. In contrast, among individuals with comorbid SUD and ADHD (SUD+ADHD), SUD was positively associated with PGS_{CUD} and $PGS_{adhd,}$, but not with PGS_{non_adhd}. Notably, the effect of PGS_{adhd} was approximately 60% larger in comorbid cases than in either of the other SUD groups. Consistently, PGS_{adhd} explained a greater proportion of the phenotypic variance in the comorbid group, with liability R^2 increasing from 2.1% in the overall SUD sample and 1.7% in the SUD-only subgroup to 3.07% in the comorbid group (Figure 5a and Supplementary Table 13a).

In the ADHD clinical sample, PGS_{CUD} was positively associated with ADHD in both the overall ADHD sample and in individuals with comorbid SUD (ADHD+SUD), but not in the ADHD-only group. PGS_{adhd} was positively associated with ADHD across all groups, with an effect approximately twice as large in individuals with comorbid SUD than in those with ADHD alone. PGS_{adhd} explained approximately 3.2 times more phenotypic variance in the comorbid group than in the ADHD-only group (2.3% vs. 0.7%). In contrast, PGS_{non_adhd} was not associated with ADHD in any of the groups (Figure 5b and Supplementary Table 13b).

In comorbid cases, we found no significant differences in the effects of any of the PGS depending on whether the primary diagnosis for treatment seeking was SUD (SUD+ADHD) or ADHD (ADHD+SUD) (Supplementary Table 14), and both datasets were therefore meta-analyzed and considered together in subsequent analyses (n=461). Upon meta-analysis, PGS_{CUD} and PGS_{adhd} were positively associated with the comorbid condition compared with controls (P<6.42e-08), whereas PGS_{non_adhd} was not (Supplementary Table 14). Figure 6 summarizes the contrasting profiles of the two CUD components across clinical presentations: PGS_{adhd} was associated with every presentation, with the strongest effect in comorbid cases, whereas PGS_{non_adhd} was associated with SUD-only cases but not with ADHD-only or comorbid cases. When compared with ADHD-only cases, comorbid cases displayed higher genetic load for PGS_{CUD} and PGS_{adhd}, but not for PGS_{non_adhd} (Supplementary Table 15).

To further investigate whether ADHD diagnosis mediated the association between genetic liability to ADHD and SUD, we conducted a causal mediation analysis in the pooled clinical datasets. PGS_{adhd}, was associated with ADHD diagnosis (OR=1.22, 95% CI 1.13–1.32; P=2.8e-07; Supplementary Table 16a), and ADHD diagnosis was associated with higher odds of SUD

(OR=3.83, 95% CI 3.03–4.84; P=4.2e-29; Supplementary Table 16b). In the causal mediation analysis (Supplementary Table 16c), each 1-SD increase in PGS_{adhd} was associated with a 3.6 percentage-point higher probability of SUD (total effect=0.036, 95% CI 0.025–0.046; P<2e-16). Of this, 0.70 percentage points were mediated by ADHD diagnosis (ACME=0.0070, 95% CI 0.0025–0.0115; P=2e-03) and 2.9 percentage-points were direct (ADE=0.0286, 95% CI 0.019–0.037; P<2e-16), corresponding to a proportion mediated of 19.8% (95% CI 7.7–31.7; P=2e-03).

DISCUSSION

Strengthening results of observational and genetic studies, we confirm substantial genetic overlap between ADHD and CUD, and, for the first time, characterize the genetic influences on CUD that are shared with, as well as independent of, the genetic liability of ADHD.

In addition to evidence that meeting diagnostic criteria for ADHD increases the risk of SUD, we found that individuals with SUD exhibit a higher polygenic burden for ADHD compared to controls, regardless of the presence of comorbid ADHD. These results extend previous findings of substantial genetic overlap between the two disorders and further support the contribution of an ADHD-dependent genetic component to SUD risk, independently of a formal ADHD diagnosis.

Focusing on the genetic influences on CUD shared with ADHD, we found that this genetic component was associated with both SUD and ADHD across the clinical samples, with the strongest effect observed in the comorbid condition. This pattern of association suggests that the ADHD-related genetic component may be particularly relevant to SUD risk among individuals with ADHD, extending prior research highlighting the substantial shared genetic liability underlying the increased risk of SUD in this population[1,13]. These findings also align with previous research suggesting that the effect of the genetic liability to ADHD on CUD is largely direct, with only a small proportion mediated by genetic liability to cannabis use itself[10]. Furthermore, mediation analysis indicated that the ADHD diagnosis accounted for approximately 20% of the association between PGS_{adhd} and SUD, consistent with the idea that the ADHD-related polygenic burden may, at least in part, increase SUD risk through the manifestation of ADHD.

Beyond the contribution of an ADHD-shared genetic component, we identified genetic influences of CUD that are independent of the genetic liability of ADHD. This ADHD-independent genetic component is associated with SUD particularly in the absence of comorbid ADHD, and shows no significant association with ADHD, regardless of SUD status, confirming that GWAS-by-subtraction effectively isolated genetic variance in CUD that is independent of ADHD. It accounts for approximately 35% of the genetic variance in CUD and comprises three distinct loci previously

identified in the original GWAS on CUD[8]. They include a lead signal nearby *CHRNA2*, which encodes the nicotinic acetylcholine receptor alpha-2 subunit that has been widely associated with CUD[14,15], as well as a locus near *RPSAP2*, a gene associated with intelligence[16] and major depressive disorder[17]. Additionally, this ADHD-independent component shows a stronger correlation with substance use-related phenotypes than the ADHD-shared genetic component, which aligns with prior findings of a lower genetic correlation between ADHD and cannabis use compared to that between ADHD and CUD[10].

Notably, both ADHD-shared and independent genetic components influencing CUD displayed similar genetic correlations with risk-taking behaviors and substance dependence. However, the ADHD-related genetic influences show significantly higher genetic correlation with other mental disorders and adverse health outcomes compared to genetic influences that are independent of ADHD. This distinct pattern of association between the two genetic components was also observed in the PheWAS, in which the ADHD-related component exhibited larger effect sizes than the ADHD-independent load in ICD-10 codes spanning both mental and non-mental health conditions, as well as external causes of morbidity and mortality. Overall, these results align with evidence from GenomicSEM of shared liability encompassing ADHD, CUD and other psychiatric disorders, and suggest a substantial portion of the negative health outcomes associated with CUD are influenced by the genetic architecture shared with ADHD. These findings also provide further evidence of genetic effects contributing, at least in part, to the increased risk of CUD in individuals with ADHD [18], as well as to the higher rates of comorbidity between ADHD, CUD and other co-occurring medical conditions [1,6,19,20].

Several considerations are noteworthy. First, our PGS analyses did not account for substance-specific SUD, ADHD symptoms or other comorbid conditions. Given that there are substance-specific patterns of genetic overlap with ADHD[21] and that specific ADHD symptom profiles, along with co-occurring disorders, influence substance use outcomes[22], further research focusing specifically on CUD, while considering clinical heterogeneity, are warranted. Second, comorbidity patterns in individuals with ADHD change across the lifespan[23]. As the ADHD dataset includes children and adults, the group that did not meet diagnostic criteria for SUD at the time of recruitment may include individuals at risk of developing SUD later in life. Further studies considering the age of onset for both ADHD and SUD, as well as longitudinal designs are needed to further investigate genetic effects contributing to maladaptive cannabis use and the development of CUD in individuals with ADHD. Third, our results may not be generalizable to non-

European ancestries due to the lack of large datasets for non-European individuals, particularly for ADHD.

CONCLUSION

In conclusion, we characterized the genetic influences on CUD that are shared with, as well as independent of, the genetic liability of ADHD. While both genetic components of CUD are associated with risk-taking behaviors and substance dependence, they contribute differently to the comorbidity between ADHD and SUD, and exhibited different patterns of association with substance use, other mental disorders, and overall health outcomes. Together, these findings provide new insights into the polygenic architecture of CUD, suggest that genetic heterogeneity may contribute to its clinical heterogeneity and offer a framework to further investigating the etiology and clinical presentation of the disorder.

MATERIALS & METHODS

GWAS summary statistics

Summary statistics from the largest GWAS meta-analysis to date on CUD in individuals with European ancestry (N effective=161,053 individuals)[8] and ADHD[7] were used. Because of sample overlap between the CUD and ADHD datasets, we removed the effect of two studies from the ADHD GWAS meta-analysis (Barcelona: 572 cases and 425 controls and Yale-Penn: 182 cases and 1,315 controls), which provided a restricted PGC + iPSYCH ADHD sample of 37,937 cases and 185,103 controls (N effective= 101,521 individuals) that was used in all analyses. This restriction was implemented to reduce potential dependence between the discovery summary statistics used to derive the GWAS-by-subtraction components and the clinical datasets used in subsequent analyses.

Clinical datasets: sample description, clinical assessment and genotyping

Two independent clinical datasets were considered to assess the polygenic contribution of CUD to ADHD and/or SUD. All participants were of European ancestry, which was confirmed through principal component analysis (PCA) using genetic data. The study was approved by the Clinical Research Ethics Committee (CREC) of Hospital Universitari Vall d'Hebron and methods were performed in accordance with the relevant guidelines and regulations. Written informed consent was obtained from all participants or their parents/caregivers before inclusion in the study.

Substance use disorder clinical sample:

A total of 1,426 individuals with SUD for at least one of the following drugs: cocaine, opiates, cannabis and sedatives (23.5% female, mean age=45.2 years, s.d.=10.2) were recruited at the Addiction and Dual Diagnosis Unit in the Hospital Univeritari Vall d'Hebron in Barcelona, Spain. Inclusion criteria were age over 18 years old and abuse or dependence according to the DSM-IV criteria. The Conners' Adult ADHD Diagnostic Interview for DSM-IV (CAADID) was used to assess ADHD comorbidity in 1,037 participants (73%), 223 of which fulfilled criteria for ADHD (21.5%). Full description of the clinical assessment can be found in Vilar-Ribó et al.[24]. The control sample consisted of 2,309 unrelated blood donors recruited from the Blood and Tissues Bank of Barcelona (37.5% females, mean age=45.8 years-old, s.d.=13.7). Chronic alcoholism, intravenous drugs use, and recent cocaine consumption (past four months) were exclusion criteria for blood donation. Genomic DNA was isolated from whole blood by the salting-out procedure and genotyped with the Illumina Infinium Global Screening Array-24 version 2 (GSA v2) (Illumina, CA, San Diego, USA) in three different waves.

Attention-deficit/hyperactivity disorder clinical sample:

An in-house sample of 1,789 subjects with ADHD (31.5% female, mean age=27.8 years, s.d.=13.3) was recruited at the ADHD Program in the Hospital Univeritari Vall d'Hebron in Barcelona, Spain. Children were interviewed using clinical history and the Diagnostic Interview Schedule for Affective Disorders and Schizophrenia for School-Age Children 6-18 years–Lifetime Version DSM-IV (K-SADS-PL). The assessment of ADHD diagnosis in adults was based on symptomatology using the Conner's Adult ADHD Diagnostic Interview for DSM-IV (CAADID-II). The presence of comorbid disorders was assessed with the Structured Clinical Interview for DSM-IV Axis I and II Disorders (SCID-I and SCID-II). Exclusion criteria was IQ<70; pervasive developmental disorders; a history or the current presence of a condition or illness, including neurologic, metabolic, cardiac, liver, kidney, or respiratory disease; a chronic medication of any kind; birth weight ≤ 1.5 kg; and other neurological or systemic disorders that might explain ADHD symptoms. Information about lifetime substance use was collected by a psychiatrist with the SCID-I from half of the adult participants (48.4%; n=867), 235 of which fulfilled criteria for SUD (27.4%). More information regarding the clinical assessment can be found in Rovira et al. [25]. The control sample consisted of 3,002 unrelated blood donors recruited from the Blood and Tissues Bank of Barcelona (48.9% female, mean age=52.3 years, s.d.=16.3). ADHD symptomatology was excluded under the following criteria: 1) having been diagnosed with ADHD previously and 2) answering positively to the life-time presence of the following ADHD symptoms: a) often has trouble in keeping attention

on tasks, b) usually loses things needed for tasks, c) often fidgets with hands or feet or squirms in seat, and d) often gets up from seat when remaining in seat is expected. Genomic DNA was isolated from whole blood by the salting-out procedure or from saliva samples or buccal swabs collected using Oragene DNA OG-500 or OC-175 kits, respectively (DNA Genotek). Genotyping was run in seven different genotyping waves with the Illumina HumanOmni1-Quad BeadChip for wave 1, Illumina HumanOmni2.5-8 BeadChip for wave 2, Illumina Infinium PsychArray-24 BeadChip for wave 3, GSA_v2 for waves 4 to 6 or GSA_v3 for wave 7 (Illumina).

Estonian Biobank (EstBB): sample description and genotyping

EstBB is a volunteer-based biobank covering approximately 20% of the adult population in Estonia (N=212,000; 66% female; mean year of birth 1971)[26]. The activities of the EstBB are regulated by the Human Genes Research Act, which was adopted in 2000 specifically for the operations of the EstBB. Individual level data analysis in the EstBB was carried out under ethical approval 1.1-12/3455 from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application 6-7/GI/17033 from the EstBB. EstBB participants were genotyped at the Core Facility of Genomics, University of Tartu, using four revisions of the Illumina GSA array (GSAMD-24v1, GSAMD-24v2, ESTchip-1_GSAv2, ESTchip-2_GSAv3) across 17 genotyping batches, each containing a minimum of ~1,000 samples.

Imputation and quality control

For the clinical datasets, pre-imputation quality control (QC) was done with the PLINK 2.0 [27] and included individual and variant filtering based on the following parameters: variant call rate >0.95 (before individual filtering), individual call rate >0.98, autosomal heterozygosity deviation (|Fhet|<0.2), variant call rate >0.98 (after individual filtering), SNP Hardy-Weinberg equilibrium (HWE; P>1e−10 in cases and P>1e-06 in controls) and minor allele frequency (MAF)>0.01. Genetic outliers were identified by PCA using PLINK 2.0 and the mixed ancestry 1000G reference panel [28]. Ancestry outliers were identified separately within each genotyping wave during pre-imputation quality control. Depending on the genotyping wave, individuals were excluded if their principal component (PC) values for PC1 or PC2 were greater than 1, 2, or 3 SD from the corresponding wave-specific mean.. Related and duplicated samples were identified by the "KING-robust kinship estimator" analysis in PLINK 2.0 [29] and one individual from each pair of subjects with kinship coefficient >0.0884 was excluded. Imputation was done with McCarthy tools, for data preparation, and the Michigan Imputation Server [30], using the Haplotype Reference Consortium (HRC_Version_r1.1_2016) reference panel (GRCh37/hg19). After imputation, we only kept SNPs

with an imputation INFO score >0.8, MAF >0.01 and SNP call rate >0.95. Then, samples from the five GWAS waves were merged and only SNPs present in all of them, with MAF >0.01, SNP call rate >0.95 and differential case/control missingness rate <0.2 in the overall sample were considered for subsequent analyses.

For the EstBB, QC procedures were performed for each batch separately. At the sample level, individuals with a call rate <0.95 and/or a sex mismatch between genotype and phenotype data were excluded. At the variant level, genotypes were removed if they failed one of the following filters: call rate >0.95, HWE P >1e-4, Illumina cluster separation >0.4, and Illumina GenTrain score >0.6. Genotyping data was processed using the GRCh37 human genome reference. To control for potential genotyping batch effects, variants showing inconsistent allele frequencies across batches were removed. Specifically, variant allele frequencies were calculated for each genotyping batch containing more than 10,000 samples (9 batches in total). The mean allele frequency across these batches was then computed, and variants with frequencies deviating by more than 5% from the mean in any batch were excluded from the combined dataset. Only single-nucleotide variants (SNVs) were retained for imputation. Among these, AT, GC, and rare variants (MAF<0.01) were removed. For SNVs typed by more than one probe, genotype concordance between different probes within each sample was assessed, and variants with discordant calls were removed on a per-sample basis. This was followed by another round of a variant-level 95% call rate filter. After QC, all genotyping batches were merged into a single cohort dataset. Approximately 307K SNVs passed all QC filters and were used for phasing and imputation. The dataset was phased reference-free using Eagle v2.4.1. Subsequent imputation of the pre-phased data was performed with Beagle v5.4 (beagle.22Jul22.46e.jar), using a population-specific haplotype reference panel derived from whole-genome sequencing data of 2,056 Estonian individuals[31]. SNPs with MAF <0.01, palindromic and HWE p <1e-04 were removed for subsequent analyses.

Statistical analyses

SNP-based heritability and genetic overlap

We estimated the SNP-based heritability (h^2_{SNP}) and the genetic correlation between CUD and ADHD using Linkage Disequilibrium Score Regression (LDSC_v1.0)[32]. All analyses were restricted to Hapmap3 SNPs and conversion of h^2_{SNP} estimates for CUD and ADHD from observed scale to the liability scale was done using a population prevalence of 5% for CUD and for ADHD[7,33]. MiXeR (https://github.com/precimed/mixer) was used to calculate the number of trait-influencing SNPs

with the univariate model, and the genetic overlap between traits with the bivariate model using default settings[34]. The Dice coefficient explains the proportion of shared SNPs between traits, and the Akaike Information Criterion (AIC) was used to determine the model fit (https://github.com/precimed/mixer?tab=readme-ov-file#aic-bic-interpretation)

GWAS-by-subtraction

We dissected genetic variation in CUD by genetic variation in ADHD applying GWAS-by-subtraction[12] with Genomic Structural Equation Modeling (GenomicSEM)[35] on summary statistics from GWAS on CUD[8] and ADHD[7] using default settings. We constructed a two-factor model, where CUD and ADHD were regressed on the first latent factor (*adhd*), representing genetic variance in ADHD, and CUD was further regressed on the second factor (*non_adhd*), representing the residual genetic variance in CUD (Supplementary Figure 2). By construction, the two latent factors were uncorrelated, so the *non_adhd* factor represented genetic variation in CUD that is not accounted for by the *adhd* factor. Then, the two latent factors (*adhd* and *non_adhd*) were regressed on individual SNPs, which resulted in two paths of association with CUD: (i) SNP effects on CUD that are fully mediated by *adhd* and (ii) SNP effects on CUD that are independent of *adhd*, and generated two new GWAS: ADHD (*adhd*) and non-ADHD (*non_adhd*) contributions to CUD.

The effective sample size (Neffective) of *adhd* and *non_adhd* GWAS was estimated using the method described by Demange et al. [12]. The main analysis model assumed a causal effect of the genetic liability of ADHD on CUD, which was followed by a sensitivity analysis allowing for moderate bidirectional causality, considering a standardized effect size of 0.2 of CUD on ADHD. The correlation between Z-statistics of genome-wide significant hits for *adhd* and *non_adhd* in the main versus the alternative model was assessed with a Spearman correlation test.

For loci with more than two secondary lead variants (P-value<5e-08) in less than 1Mb, we further checked independency. Independent lead variants were identified through clumping (r2 = 0.05, kb = 5000) using PLINK 1.09 and the 1,000 Genomes Project Phase 3 European reference panel (http://www.internationalgenome.org/) and conditional analyses using COJO implemented in Genome-wide Complex Trait Analysis (GCTA)[36]. Overlap between *non_adhd* lead variants and genome-wide significant variants identified in the original GWAS on ADHD or CUD was assessed according to distance (previous lead SNPs within +/-500kb from any of the *non_adhd* lead SNPs) and linkage disequilibrium (r2>0.05), and further confirmed using conditional analysis using COJO.

The functional annotation of genome-wide significant variants (lead SNPs or SNPs in LD (r2≥0.6) with a lead SNP and a P-value <0.05) was conducted with FUMA (https://fuma.ctglab.nl/)

considering CADD scores, RegulomeDB scores and chromatin stages [37–40]. Variants were annotated to known brain eQTLs using information in the genotype tissue expression (GTEx v8) portal [41], BRAINEAC [42], PsychENCODE [43] and CommonMind Consortium (CMC/CMC) [44]. Traits and disorders associated with the annotated SNPs in previous GWAS (P<5e-08) were identified using information from the GWAS catalog [45]. SNPs were annotated to genes based on physical proximity (using the default parameters), brain eQTL (based on GTEx v8, BRAINEAC, PsychENCODE and CMC/CMC) and chromatin interaction using FUMA. These genes were used in gene-set enrichment analyses using the GENE2FUNC module of FUMA. Enrichment was tested among the predefined sets of differentially expressed genes in GTEx v8 (54 tissue types) and Brainspan (29 different ages of samples and 11 general developmental stages) using hypergeometric test with protein coding genes as background genes. Genes were also tested for enrichment in gene-sets from NHGRI-EBI GWAS Catalog. P-values were corrected for multiple comparisons using Benjamini–Hochberg False Discovery Rate (FDR) method (adj-P-value<0.05). Gene-based and gene-set analyses were performed using MAGMA v1.08 implemented in FUMA considering the full summary statistics from *adhd* and *non_adhd*. Genome-wide significance was assessed through Bonferroni correction for the number of genes (adj-P-value=0.05/18,305 genes=2.7e-06) and FDR for gene-sets tested, including Gene ontology biological process (GO), KEGG, Reactome (adj-P-value <0.05).

Genetic correlations

We explored genome-wide genetic correlations between *adhd* and *non_adhd* and 35 phenotypes across nine categories using LDSC_v1.0[32] (Supplementary Table 9). We used GWAS summary statistics including both males and females of European ancestry. If several GWAS were available for the same disorder, we chose the study with the largest effective sample size (N effective=4/(1/Ncases+1/Ncontrols)). We used GenomicSEM to test whether genetic correlations with these traits differed between *adhd* and *non_adhd*. Models tested the null hypothesis that trait genetic correlations with *adhd* and *non_adhd* could be constrained to be equal using a χ^2 test following the method described by Demange et al.[12]. P-values were corrected for multiple comparisons using the FDR method (adj-P-value<0.05).

Phenome-wide association study in the Estonian Biobank (EstBB)

PGS-based phenome-wide association study (PheWAS) was conducted in the EstBB cohort[46]. From linkage with national databases, all ICD-10 main codes derived from electronic health records (EHRs) over a 20-year period (2004–2023; N=1,942) were considered for the PheWAS

analysis. After excluding sex-specific codes and codes with <100 cases, 1,092 ICD-10 phenotypes remained for the association testing. PGS for *adhd* (PGS_{adhd}) and *non_adhd* (PGS_{non_adhd}) were computed using PRS-CS with the 1,000 Genomes European reference panel for allele weighting and PLINK 2.0 for allele scoring. PGSs were standardized to a mean of 0 and a standard deviation of 1.

Association testing was undertaken separately for PGS_{adhd} and $PGS_{non_adhd,}$ considering year of birth, sex and first 10 genetic PCs as covariates in all the logistic regression models. The Bonferroni adjusted P-value across 1,092 phenotypes and two PGSs was used as a significance threshold to correct for multiple testing (0.05/1,092*2=2.29e-05). ICD-10 codes significantly associated with at least one of the two PGSs were then selected for an additional analysis to assess whether PGS_{adhd} and PGS_{non_adhd} effect sizes differed. A mutually adjusted logistic regression model was run including both PGSs and the same covariates described above and then a linear hypothesis test was run to assess whether the coefficients of both PGSs were equal using the car R package. The Bonferroni adjusted P-value considering the number of ICD-10 codes tested was used as a significance threshold to correct for multiple testing.

Genome-wide Polygenic Score Analysis in the clinical datasets

PGS analyses were conducted in our in-house clinical samples of SUD (1,426 individuals with SUD and 2,309 controls) and ADHD (1,789 individuals with ADHD and 3,002 controls). Cases were further stratified according to the presence of comorbid condition as SUD-only (n=814), SUD+ADHD, ADHD-only (n=629) and ADHD+SUD (n=238). Subgroup classifications were restricted to cases with available information on comorbid diagnoses. When PGS effect estimates were consistent across the two comorbid groups, regardless of the primary diagnosis prompting treatment seeking, results from the SUD+ADHD and ADHD+SUD subgroups were combined using meta-analysis.

PGS were constructed using PRS-CS [47], PLINK 2.0 and summary statistics for CUD[8], *adhd* and *non_adhd.* All PGS were computed and standardized to a mean of 0 and a standard deviation of 1. The association between PGS and each case-control phenotype was undertaken using the full set of controls and the different case subsets for ADHD and SUD samples separately. Logistic regression models were used considering age, sex, genotyping waves and 10 genetic PC as covariables. The percentage of variance attributable to each PGS was reported as the mean percentage of variance calculated as the increase in Nagelkerke's pseudo-R2 between models

with and without the PGS. PGS P-values were corrected for multiple comparisons using the FDR method (adj-P-value<0.05).

Mediation Analysis

When a significant association between PGS_{adhd} and SUD was observed, we further investigated the potential mediating role of ADHD diagnosis in this relationship. Causal mediation analysis was conducted using the R package mediation in the pooled sample of unrelated individuals from the two clinical datasets (N = 7,173). SUD diagnosis was specified as the outcome, ADHD diagnosis as the mediator, and standardized PGS_{adhd} (mean 0, SD 1) as the exposure. Two logistic regression models were fitted: a mediator model (ADHD diagnosis ~ PGS_{adhd} + sex + age + genotyping batch + 10 genetic principal components) and an outcome model (SUD diagnosis ~ PGS_{adhd} + ADHD diagnosis + sex + age + genotyping batch + 10 genetic principal components). These models were used to estimate the average causal mediation effect (ACME; the effect of the PGS_{adhd} on SUD operating through ADHD diagnosis), the average direct effect (ADE; the effect not operating through ADHD diagnosis), the total effect (ACME + ADE) and the proportion mediated (ACME/total effect). Estimates and 95% confidence intervals were obtained with 1,000 quasi-Bayesian Monte Carlo simulations. The effects estimated in this causal mediation analysis are expressed on the probability scale, as the change in the probability of SUD per 1-SD increase in PGS_{adhd}. Mediation was inferred when the 95% CI of the ACME excluded zero.

Acknowledgements

The authors are grateful to the patients and controls who kindly participated in this research. The genotyping service was carried out at the Genotyping Unit-CEGEN in the Spanish National Cancer Research Centre (CNIO), supported by Instituto de Salud Carlos III (ISCIII), Ministerio de Ciencia e Innovación. CEGEN is part of the initiative IMPaCTGENóMICA (IMP/00009) cofunded by ISCIII and the European Regional Development Fund (ERDF). We thank all participants of the Estonian Biobank. Estonian Biobank research team refers to Andres Metspalu, Lili Milani, Tõnu Esko, Reedik Mägi, Mait Metspalu, Mari Nelis and Georgi Hudjashov. The Estonian Genome Centre analyses were partially carried out in the High Performance Computing Center of University of Tartu.

Funding

This work was supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, 2017SGR-1461, 2021SGR-00840, 2021SGR-01093), the Instituto de Salud Carlos III (P19/01224,

PI20/00041, PI22/00464, PI23/00404, PI24/00195, PI24/00671, PI25/00343, FI23/00152 to P.C.G, FI24/00198 to M.D.R.R, CP22/00128 to M.S.A and CP22/00026 to S.A), the Network Center for Biomedical Research (CIBER) to J.C.D. and U.Z.A; the European Regional Development Fund (ERDF); "la Marató de TV3" (202228-30, 202228-31 and 202218-31), Servei Català de la Salut (CatSalut), Ministerio de Sanidad, Servicios Sociales e Igualdad/Plan Nacional Sobre Drogas' (PNSD-2020I042 and PNSD-2024I056), the Spanish 'Ministerio de Ciencia, Innovación y Universidades' (PID2021-127760B-I100 and PID2024-158634OB-I00 funded by MICIU/AEI/10.13039/501100011033/ and FEDER-EU), and ICREA Academia 2021 to BC. The research in EstBB was conducted using the Estonian Center of Genomics/Roadmap II funded by the Estonian Research Council (project number TT17), and the Estonian Research Council's grant No. PSG615 and was supported by Estonian Centre of Excellence for Well-Being Sciences (EstWell), funded by grant TK218 from the Estonian Ministry of Education and Research. This project has also received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 965381 (TIMESPAN) and European Union's Horizon Europe research and innovation programme under Grant Agreement no 101222412 (AT-TENSION).

Conflict of interest

CF has received fees to give talks for Takeda, Rubio, Casen-Recordati and Idorsia. JARQ was on the speakers' bureau and/or acted as consultant for Biogen, Janssen-Cilag, Idorsia, Novartis, Shire, Takeda, Bial, Shionogi, Sincrolab, Novartis, BMS, Medice, Rubió, Uriach, Technofarma and Raffo in the last 3 years. He also received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Idorsia, Janssen-Cilag, Rubió, Shire, Takeda, Shionogi, Bial and Medice. The Department of Psychiatry chaired by him received unrestricted educational and research support from the following companies in the last 3 years: Janssen- Cilag, Shire, Oryzon, Roche, Psious, Idorsia, Neuraxpharm, Exeltis, and Rubió. MR received travel awards (air tickets + hotel) for taking part in psychiatric meetings from Rubió. All other authors declare no biomedical financial interests or conflicts of interest.

Data availability

GWAS summary data for adhd and non_adhd have been deposited in the GWAS catalog with accession no. GCP001881. ADHD GWAS summary statistics from the Psychiatric Genomics Consortium (PGC) are available for download at https://www.med.unc.edu/pgc/download-results/. CUD GWAS summary statistics are available through the Gelernter Lab website

(https://medicine.yale.edu/lab/gelernter/). GWAS summary statistics used in the genetic correlation analyses are accessible through the links provided in Supplementary Table 9. Individual-level EstBB data are not publicly available due to data protection regulations. Access is subject to ethical approval and managed through the EstBB. Further information is available at https://urldefense.com/v3/__https://genomics.ut.ee/en/content/estonian-biobank__;!!D9dNQwwGXtA!S-7Mtf6Ii1bfrGxbW_d0XB7Tmq8Vbz04CXalaxEFYS3muWStRNU7z0RBeDGhyP11cKAlPXAstCtYbW4YJm-b$. The clinical datasets analyzed in the current study are not publicly available because participants did not provide consent for public data sharing and the relevant ethical approvals do not permit public release.

Code availability

Code used to run the analyses is available at https://github.com/PerlineDemange/non-cognitive. A tutorial on how to perform GWAS-by-subtraction is available at http://rpubs.com/MichelNivard/565885. All additional software used to perform analyses are available online.

BILBIOGRAPHY

1. Hartman, C. A. *et al.* Anxiety, mood, and substance use disorders in adult men and women with and without attention-deficit/hyperactivity disorder: A substantive and methodological overview. *Neurosci. Biobehav. Rev.* **151**, 105209 (2023).
2. Molina, B. S. G. *et al.* Substance use through adolescence into early adulthood after childhood-diagnosed ADHD: findings from the MTA longitudinal study. *Journal of Child Psychology and Psychiatry* **59**, 692–702 (2018).
3. Froude, A. M. *et al.* The prevalence of cannabis use disorder in attention-deficit hyperactivity disorder: A clinical epidemiological meta-analysis. *J. Psychiatr. Res.* **172**, 391–401 (2024).
4. Zaman, T., Malowney, M., Knight, J. & Boyd, J. W. Co-Occurrence of Substance-Related and Other Mental Health Disorders Among Adolescent Cannabis Users. *J. Addict. Med.* **9**, 317–321 (2015).
5. Breyer, J. L., Lee, S., Winters, K. C., August, G. J. & Realmuto, G. M. A longitudinal study of childhood ADHD and substance dependence disorders in early adulthood. *Psychology of Addictive Behaviors* **28**, 238–246 (2014).

6. Ribasés, M. *et al.* Genetic architecture of ADHD and overlap with other psychiatric disorders and cognition-related phenotypes. *Neurosci. Biobehav. Rev.* **153**, 105313 (2023).
7. Demontis, D. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat. Genet.* **55**, 198–208 (2023).
8. Levey, D. F. *et al.* Multi-ancestry genome-wide association study of cannabis use disorder yields insight into disease biology and public health implications. *Nat. Genet.* **55**, 2094–2103 (2023).
9. Abdellaoui, A., Smit, D. J. A., van den Brink, W., Denys, D. & Verweij, K. J. H. Genomic relationships across psychiatric disorders including substance use disorders. *Drug Alcohol Depend.* **220**, 108535 (2021).
10. Nielsen, T. T. *et al.* Shared genetics of ADHD, cannabis use disorder and cannabis use and prediction of cannabis use disorder in ADHD. *Nature Mental Health* **2**, 1071–1083 (2024).
11. Demontis, D. *et al.* Genome-wide association study implicates CHRNA2 in cannabis use disorder. *Nat. Neurosci.* **22**, 1066–1074 (2019).
12. Demange, P. A. *et al.* Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nat. Genet.* **53**, 35–44 (2021).
13. Gujska, J. H., Silczuk, A., Madejek, R. & Szulc, A. Exploring the Link Between Attention-Deficit Hyperactivity Disorder and Cannabis Use Disorders: A Review. *Medical Science Monitor* **29**, (2023).
14. Demontis, D. *et al.* Genome-wide association study implicates CHRNA2 in cannabis use disorder. *Nat. Neurosci.* **22**, 1066–1074 (2019).
15. Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry* **7**, 1032–1045 (2020).
16. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 2098 (2018).
17. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* **50**, 920–927 (2018).
18. Lee, S. S., Humphreys, K. L., Flory, K., Liu, R. & Glass, K. Prospective association of childhood attention-deficit/hyperactivity disorder (ADHD) and substance use and abuse/dependence: A meta-analytic review. *Clin. Psychol. Rev.* **31**, 328–341 (2011).
19. Soler Artigas, M. *et al.* Mendelian randomization analysis for attention deficit/hyperactivity disorder: studying a broad range of exposures and outcomes. *Int. J. Epidemiol.* **52**, 386–402 (2023).

20. Kittel-Schneider, S. *et al.* Non-mental diseases associated with ADHD across the lifespan: Fidgety Philipp and Pippi Longstocking at risk of multimorbidity? *Neurosci. Biobehav. Rev.* **132**, 1157–1180 (2022).
21. Koller, D. *et al.* Genetic contribution to the comorbidity between attention-deficit/hyperactivity disorder and substance use disorders. *Psychiatry Res.* **333**, 115758 (2024).
22. Barbuti, M. *et al.* Challenges of Treating ADHD with Comorbid Substance Use Disorder: Considerations for the Clinician. *J. Clin. Med.* **12**, 3096 (2023).
23. Franke, B. *et al.* Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *European Neuropsychopharmacology* **28**, 1059–1088 (2018).
24. Vilar-Ribó, L. *et al.* Disentangling heterogeneity in substance use disorder: Insights from genome-wide polygenic scores. *Transl. Psychiatry* **14**, 221 (2024).
25. Rovira, P. *et al.* Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *Neuropsychopharmacology* **45**, 1617–1626 (2020).
26. Milani, L. *et al.* The Estonian Biobank's journey from biobanking to personalized medicine. *Nature Communications* vol. 16 Preprint at https://doi.org/10.1038/s41467-025-58465-3 (2025).
27. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, (2015).
28. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
29. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
30. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
31. Mitt, M. *et al.* Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* **25**, 869–876 (2017).
32. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
33. Browne, K. C. *et al.* Prevalence and Correlates of Cannabis Use and Cannabis Use Disorder Among U.S. Veterans: Results From the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC-III). *American Journal of Psychiatry* **179**, 26–35 (2022).
34. Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat. Commun.* **10**, 2417 (2019).

35. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
36. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
37. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
38. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
39. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
40. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
41. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (1979).* **369**, 1318–1330 (2020).
42. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418–1428 (2014).
43. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science (1979).* **362**, (2018).
44. Hoffman, G. E. *et al.* CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci. Data* **6**, 180 (2019).
45. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
46. Milani, L. *et al.* The Estonian Biobank's journey from biobanking to personalized medicine. *Nature Communications* vol. 16 Preprint at https://doi.org/10.1038/s41467-025-58465-3 (2025).
47. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).

FIGURE LEGENDS

Figure 1. GWAS-by-subtraction results. (a) Miami plot showing results of the *non_adhd* (top) and *adhd* (bottom) GWAS. Dotted lines indicate statistical significance at p<5e-08.

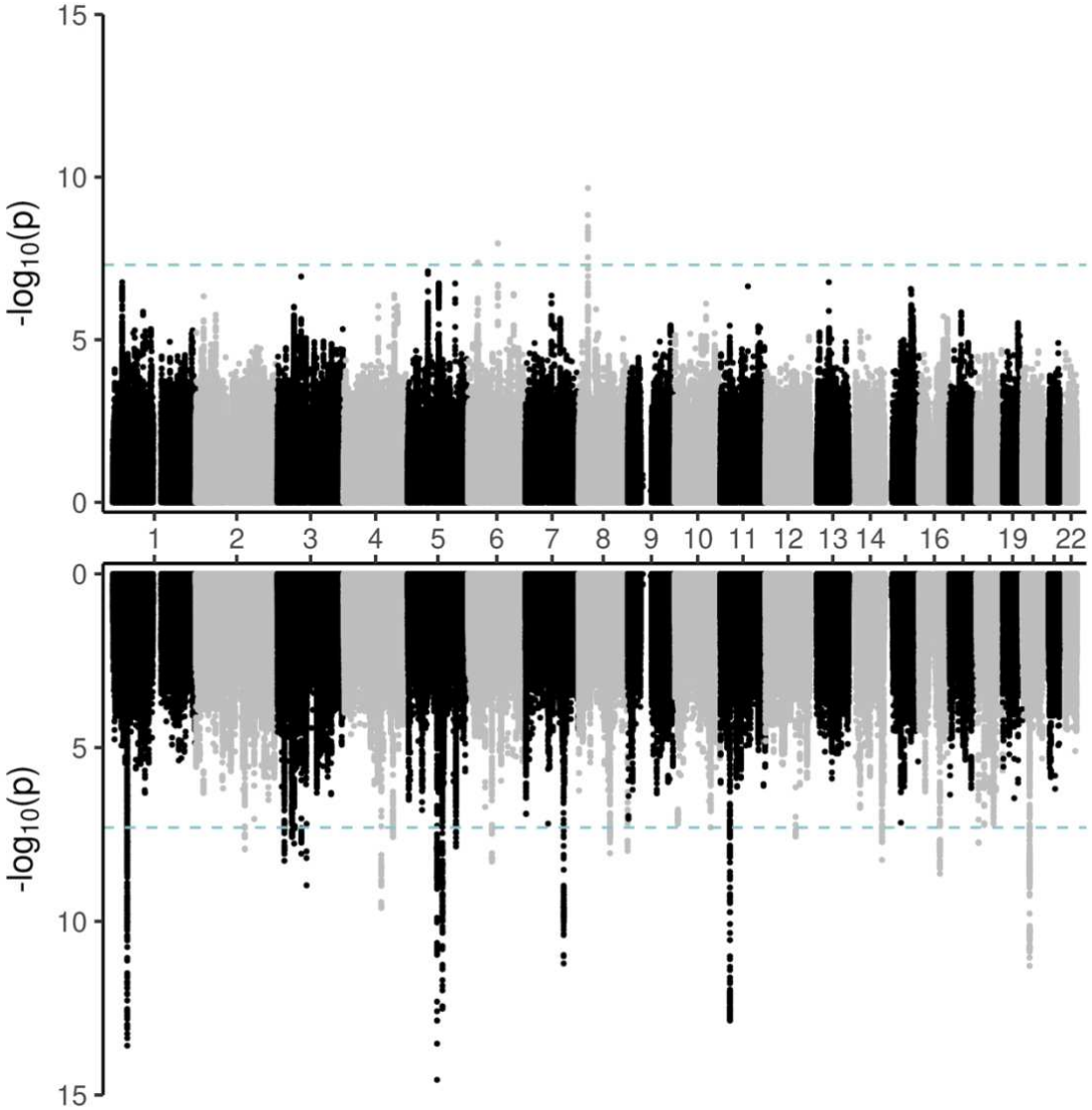
Figure 2. Genetic correlation of CUD, *adhd* and *non_adhd* factors with SUD, risk taking and smoking behavior. Significant differences (P_{FDR}<0.05) in the magnitude of the correlations between *adhd* and *non_adhd factors* are indicated with an asterisk. Dots represent genetic correlation estimates and error bars represent 95% CI.

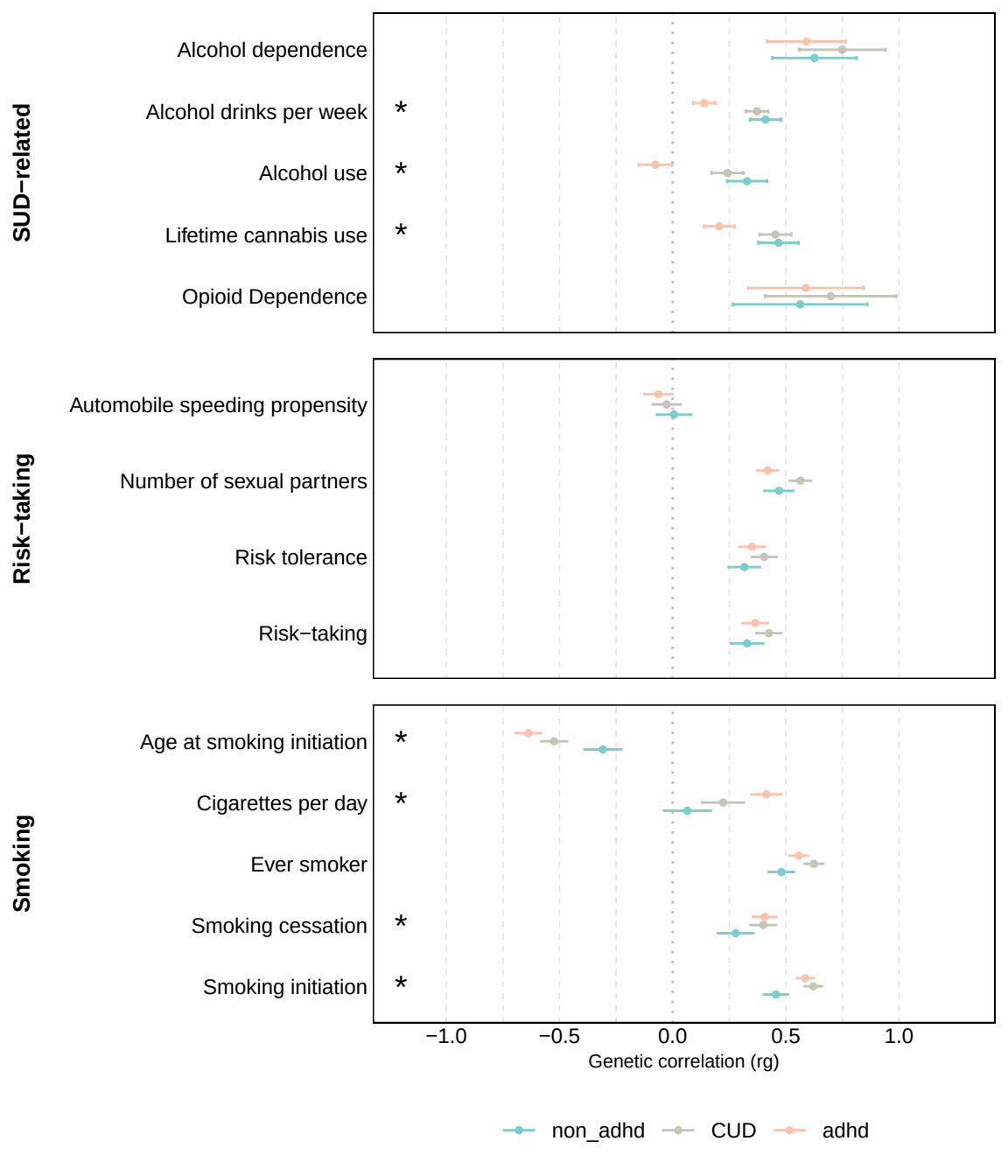
Figure 3. Genetic correlation of CUD, *adhd* and *non_adhd* factors with mental disorders, personality, cognition and related-traits. Significant differences (P_{FDR}<0.05) in the magnitude of the correlations between *adhd* and *non_adhd* factors are indicated with an asterisk. Dots represent genetic correlation estimates and error bars represent 95% CI.

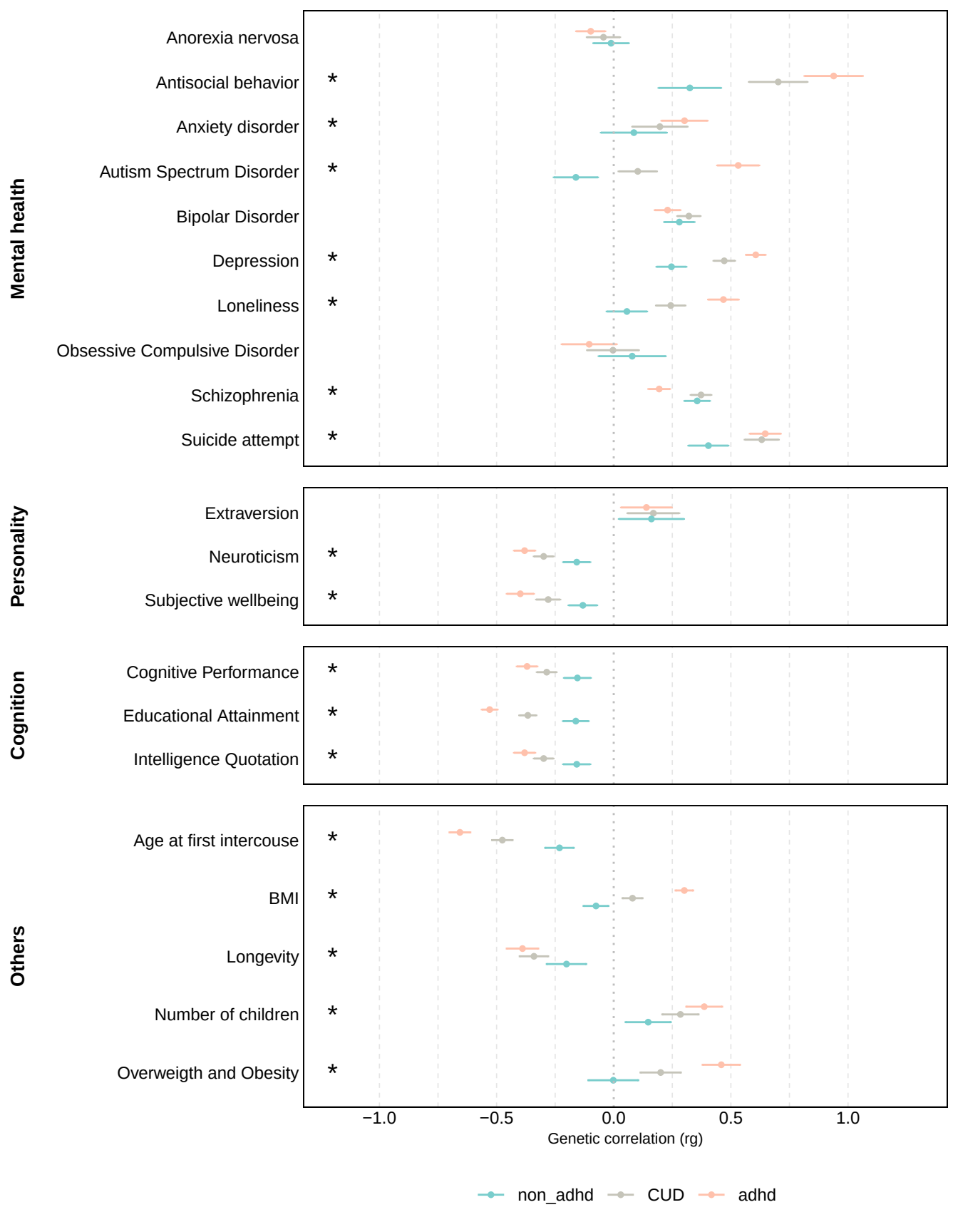
Figure 4. Medical conditions showing differential associations with PGS_{adhd} and $PGS_{non\text{-}adhd}$ in phenome-wide association analysis considering 114,410 individuals from the Estonian Biobank. Forest plot displaying odds ratios (ORs) and 95% confidence intervals (CIs) for the associations between selected diagnoses and PGS_{adhd} (orange) and PGS_{non_adhd} (blue). Only conditions showing statistically significant difference in effect between the two polygenic scores are shown.

Figure 5. Comparison of polygenic scores (PGS) profiles across (a) SUD (1426 cases (814 SUD-only and 223 SUD+ADHD) and 2309 controls) and (b) ADHD (1789 cases (629 ADHD-only and 238 ADHD+SUD) and 3002 controls). Bars represent regression coefficients from logistic regression models considering age, sex, GWAS wave and the 10 principal components as covariates and error bars are 95% confidence intervals centered on the point estimate. Significant comparisons (P_{FDR}<0.05) are indicated with an asterisk.

Figure 6. Effect-size distributions of PGS_{adhd} and PGS_{non_adhd} in SUD-only, ADHD-only and comorbid subtypes. PGS_{adhd} and PGS_{non_adhd} were tested separately in a logistic regression model comparing cases with controls (814 SUD-only cases and 2,309 controls; 629 ADHD-only cases and 3,002 controls; 461 comorbid cases and 5311 controls), adjusting for age, sex, genotyping wave and the first 10 principal components. Effects are expressed as standardized regression coefficients (log-odds per 1-SD increase in PGS). Densities were obtained by randomly generating normal distributions in which the estimate was the mean and the standard error was the standard deviation; dashed lines indicate the point estimates. For comorbid cases, results from the meta-analysis of SUD+ADHD and ADHD+SUD are shown.


15
10
5
0
-log10(p)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
16
19
22
0
5
10
15
-log10(p)

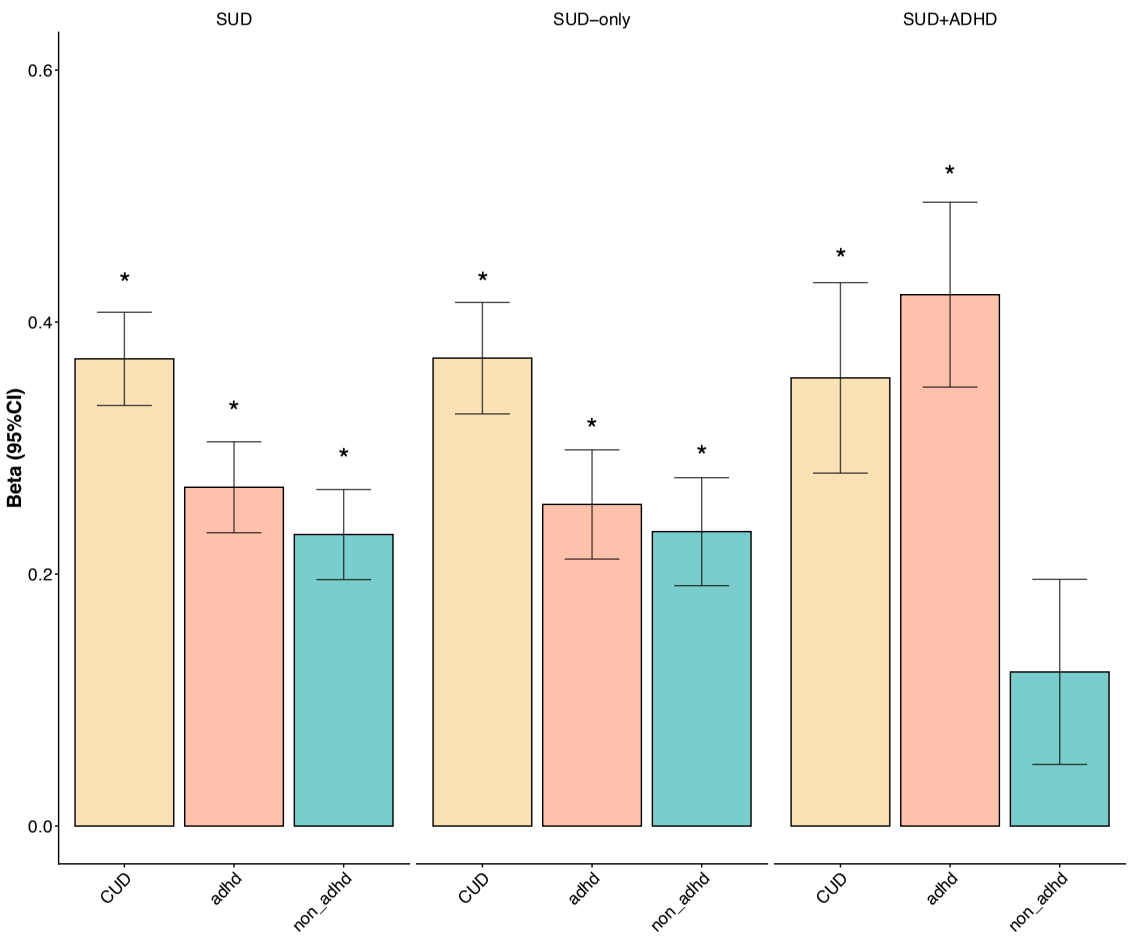

SUD−related
Alcohol dependence
Alcohol drinks per week
Alcohol use
Lifetime cannabis use
Opioid Dependence
*
*
*
Risk−taking
Automobile speeding propensity
Number of sexual partners
Risk tolerance
Risk−taking
Smoking
Age at smoking initiation
Cigarettes per day
Ever smoker
Smoking cessation
Smoking initiation
*
*
*
−1.0
−0.5
0.0
0.5
1.0
Genetic correlation (rg)
non_adhd
CUD
adhd

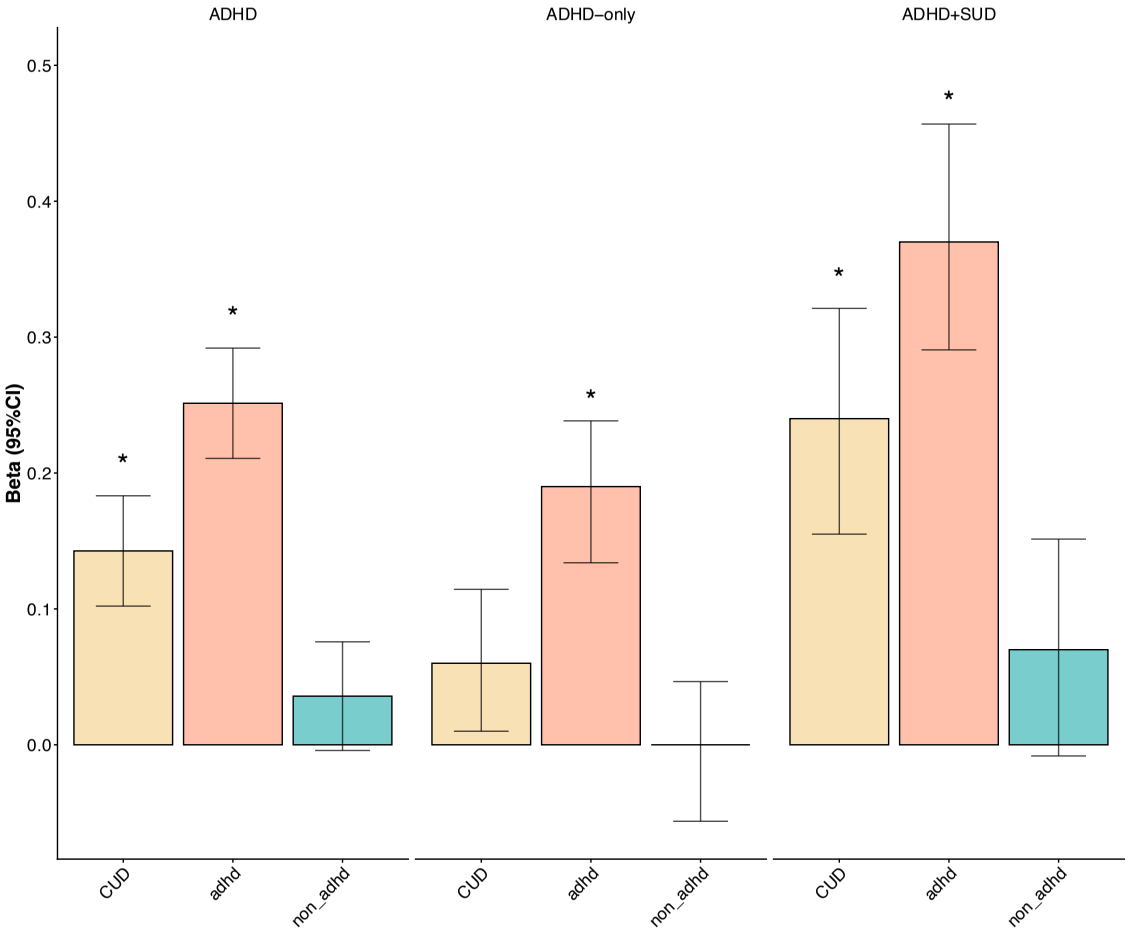

Mental health
Anorexia nervosa
Antisocial behavior
Anxiety disorder
Autism Spectrum Disorder
Bipolar Disorder
Depression
Loneliness
Obsessive Compulsive Disorder
Schizophrenia
Suicide attempt
*
Personality
Extraversion
Neuroticism
Subjective wellbeing
Cognition
Cognitive Performance
Educational Attainment
Intelligence Quotation
Others
Age at first intercouse
BMI
Longevity
Number of children
Overweigth and Obesity
−1.0
−0.5
0.0
0.5
1.0
Genetic correlation (rg)
non_adhd
CUD
adhd

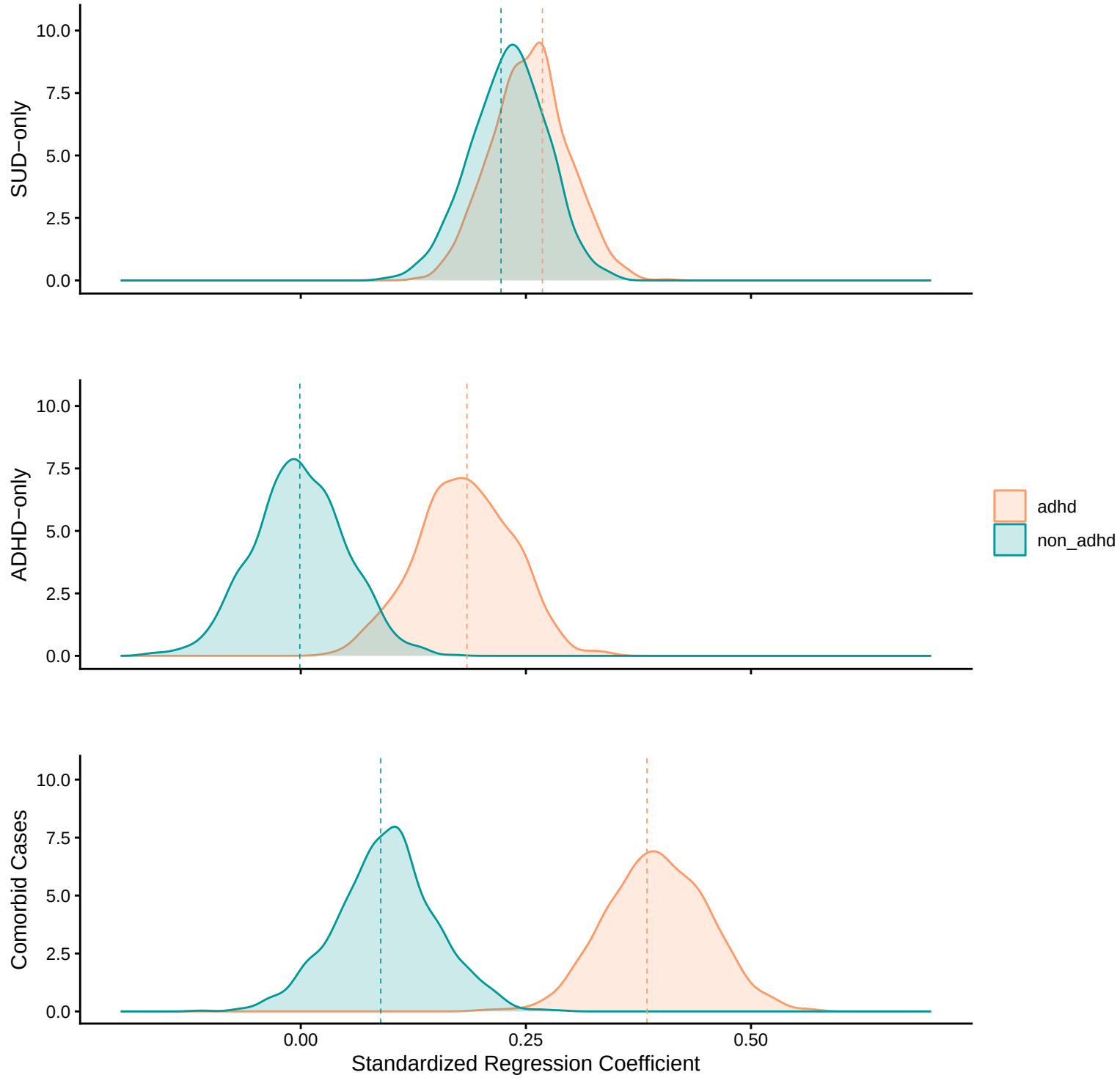
Melanocytic naevi; D22
Type 2 diabetes mellitus; E11
Obesity; E66
Disorders of lipoprotein metabolism and other lipidaemias; E78
Disorders of purine and pyrimidine metabolism; E79
Other mental disorders due to brain damage and dysfunction and to physical disease; F06
Mental and behavioural disorders due to use of tobacco; F17
Depressive episode; F32
Recurrent depressive disorder; F33
Other anxiety disorders; F41
Reaction to severe stress, and adjustment disorders; F43
Eating disorders; F50
Hyperkinetic disorders; F90
Migraine; G43
Other headache syndromes; G44
Sleep disorders; G47
Mononeuropathies of upper limb; G56
Other mononeuropathies; G58
Hemiplegia; G81
Other paralytic syndromes; G83
Essential (primary) hypertension; I10
Hypertensive heart disease; I11
Angina pectoris; I20
Acute myocardial infarction; I21
Heart failure; I50
Cerebral infarction; I63
Acute tonsillitis; J03
Acute laryngitis and tracheitis; J04
Acute bronchitis; J20
Chronic diseases of tonsils and adenoids; J35
Other chronic obstructive pulmonary disease; J44
Asthma; J45
Diseases of pulp and periapical tissues; K04
Gastro-oesophageal reflux disease; K21
Gastric ulcer; K25
Gastritis and duodenitis; K29
Functional dyspepsia; K30
Diverticular disease of intestine; K57
Other diseases of liver; K76
Cholelithiasis; K80
Allergic contact dermatitis; L23
Seborrhoeic keratosis; L82
Gout; M10
Polyarthrosis; M15
Gonarthrosis [arthrosis of knee]; M17
Other arthrosis; M19
Other joint disorders, not elsewhere classified; M25
Spondylosis; M47
Other spondylopathies; M48
Cervical disc disorders; M50
Other intervertebral disc disorders; M51
Dorsalgia; M54
Synovitis and tenosynovitis; M65
Soft tissue disorders related to use, overuse and pressure; M70
Shoulder lesions; M75
Other enthesopathies; M77
Other soft tissue disorders, not elsewhere classified; M79
Other disorders of urinary system; N39
Haemorrhage from respiratory passages; R04
Abnormalities of breathing; R06
Pain in throat and chest; R07
Abdominal and pelvic pain; R10
Nausea and vomiting; R11
Headache; R51
Pain, not elsewhere classified; R52
Elevated blood glucose level; R73
Superficial injury of head; S00
Superficial injury of thorax; S20
Superficial injury of wrist and hand; S60
Open wound of wrist and hand; S61
Superficial injury of ankle and foot; S90
Dislocation, sprain and strain of joints and ligaments at ankle and foot level; S93
Foreign body on external eye; T15
Need for immunization against COVID-19; U11
Fall on same level from slipping, tripping and stumbling; W01
Other fall from one level to another; W17
Struck by thrown, projected or falling object; W20
Caught, crushed, jammed or pinched in or between objects; W23
Overexertion and strenuous or repetitive movements; X50
Need for immunization against other single viral diseases; Z25
Persons encountering health services in other circumstances; Z76

PGS
PGSadhd
PGSnon_adhd

0.9
1.0
1.1
1.2
1.3
1.4
Odds Ratio (95% CI)


SUD
SUD−only
SUD+ADHD
Beta (95%CI)
0.6
0.4
0.2
0.0
*
CUD
adhd
non_adhd


ADHD
ADHD−only
ADHD+SUD
Beta (95%CI)
0.5
0.4
0.3
0.2
0.1
0.0
*
CUD
adhd
non_adhd


SUD−only
ADHD−only
Comorbid Cases
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
Standardized Regression Coefficient
adhd
non_adhd

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfigures.docx
- SupplementaryTables.xlsx
- mr.pdf