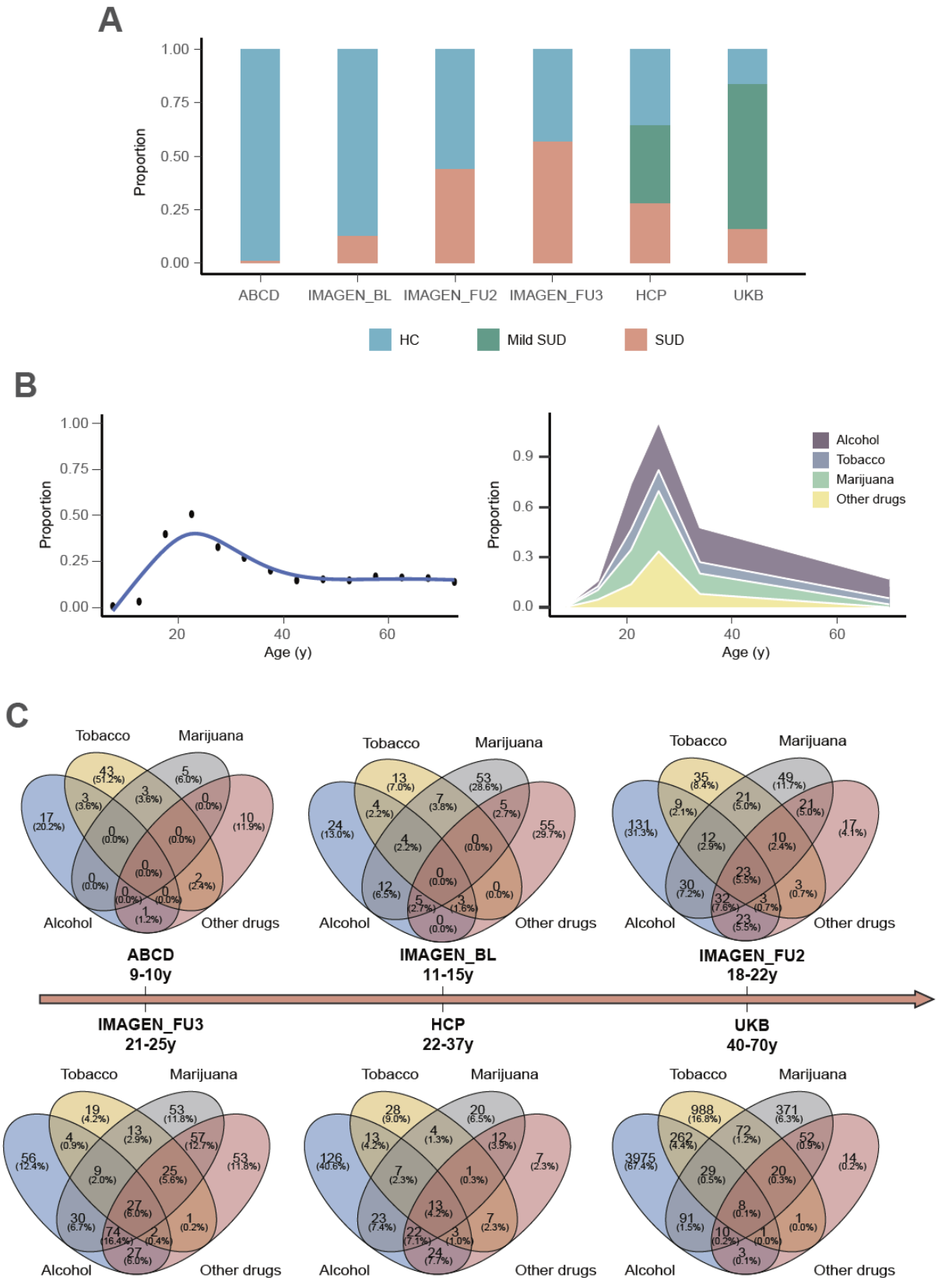
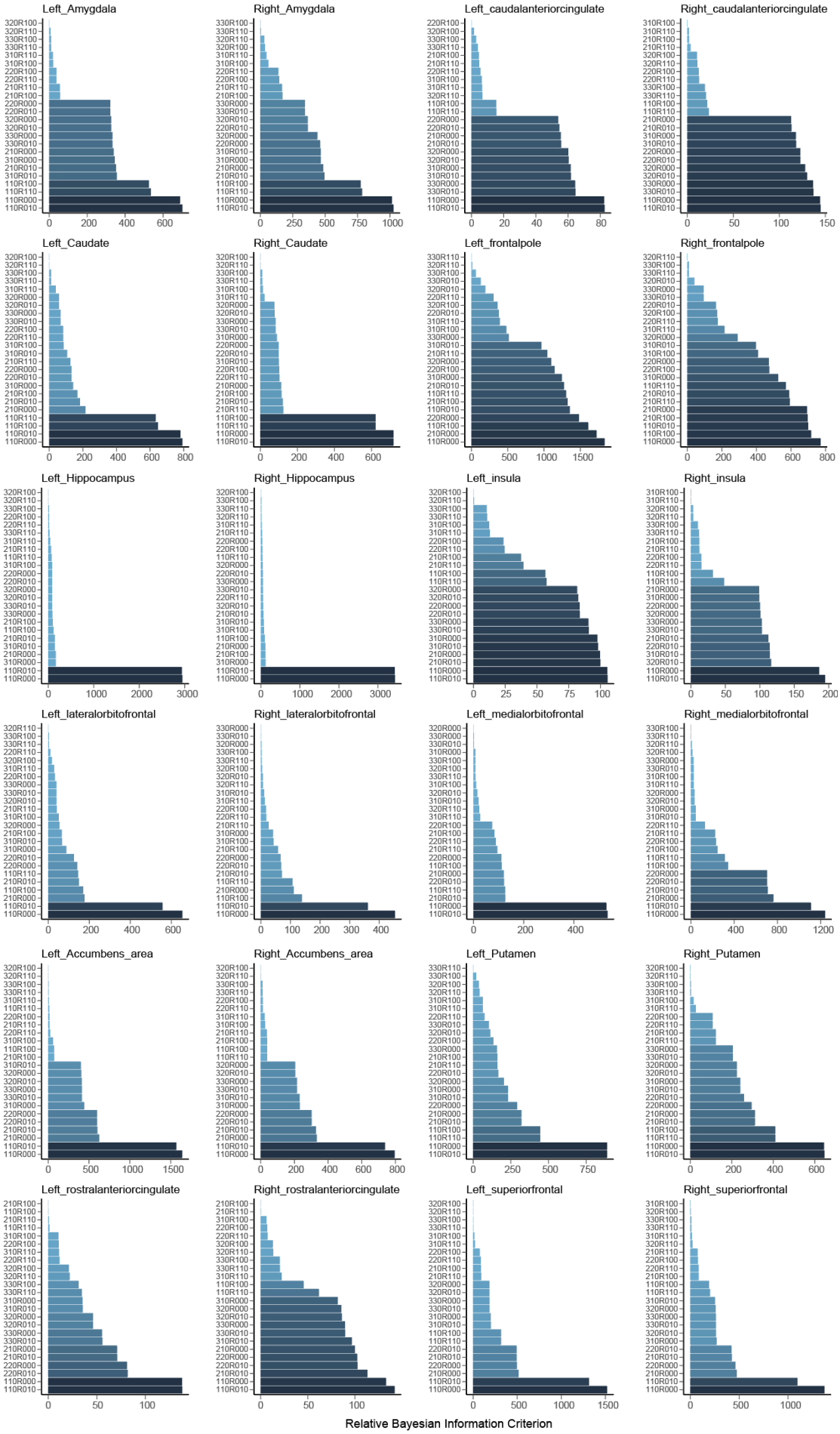
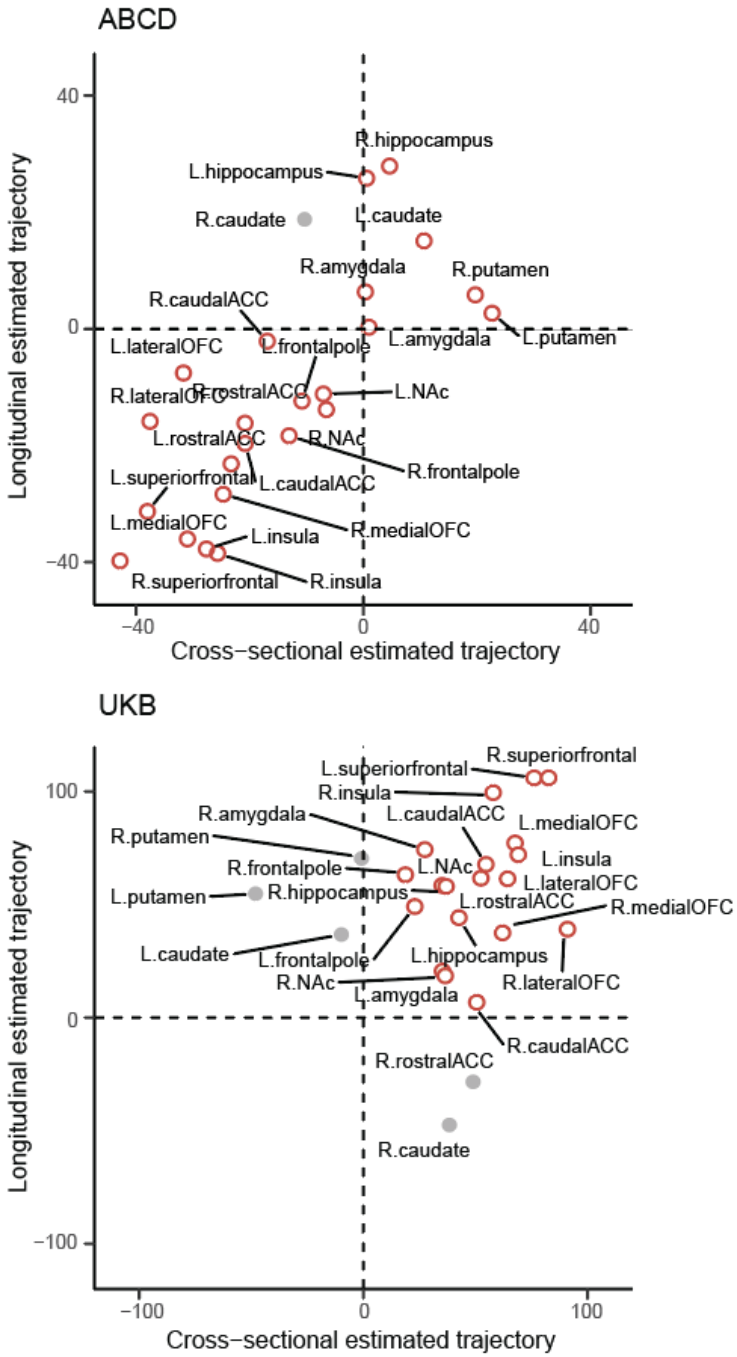


Supplementary Fig. 1. Changes in the proportion of substance use disorder (SUD) across different ages. (A) Proportion of individuals with SUD across studies. Adults who did not meet the criteria for SUD but above defined HCs were defined as mild SUD. (B) Age-related changes in the proportion of individuals with SUD (top) and specific addictive substance use (bottom). The prevalence of SUD was calculated for each 5-year age interval, plotted as points with x-axis representing the average age of each group. A cubic spline was used to fit a smooth curve. The observed subgroup proportions are illustrated in the figure below. (C) Venn diagrams illustrating participants overlap in different substance use groups across studies.

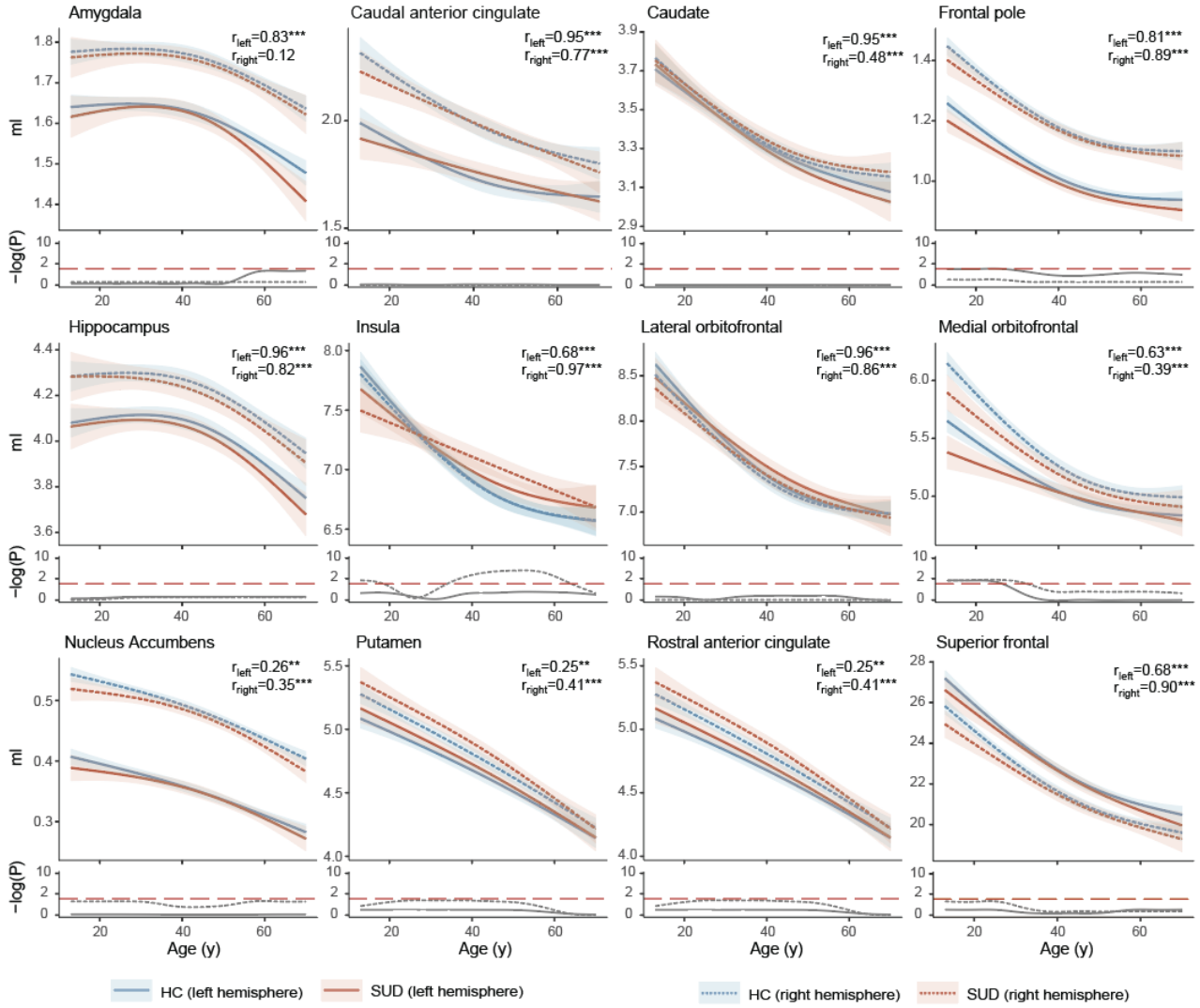


Supplementary Fig. 2. Optimal GAMLSS model for different ROIs based on the Bayesian information criterion (BIC). Relative BIC was provided relative to the lowest BIC of the best-fitting model. The y-axis

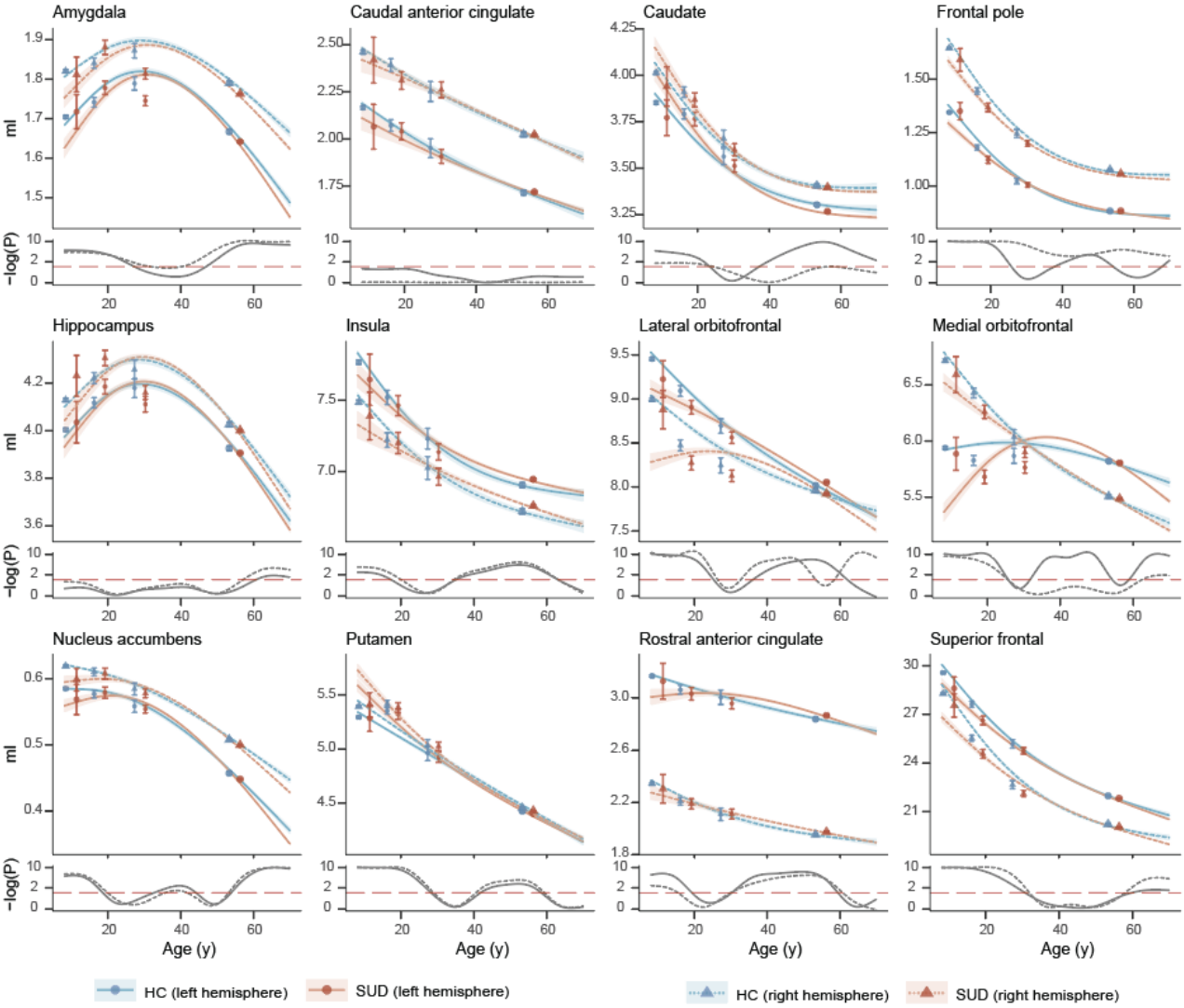
represents different model parameter settings, where the first three numbers and the last three numbers refer to the number of fractional polynomials and the inclusion of study random effect for the age term on μ, σ, and ν, respectively.



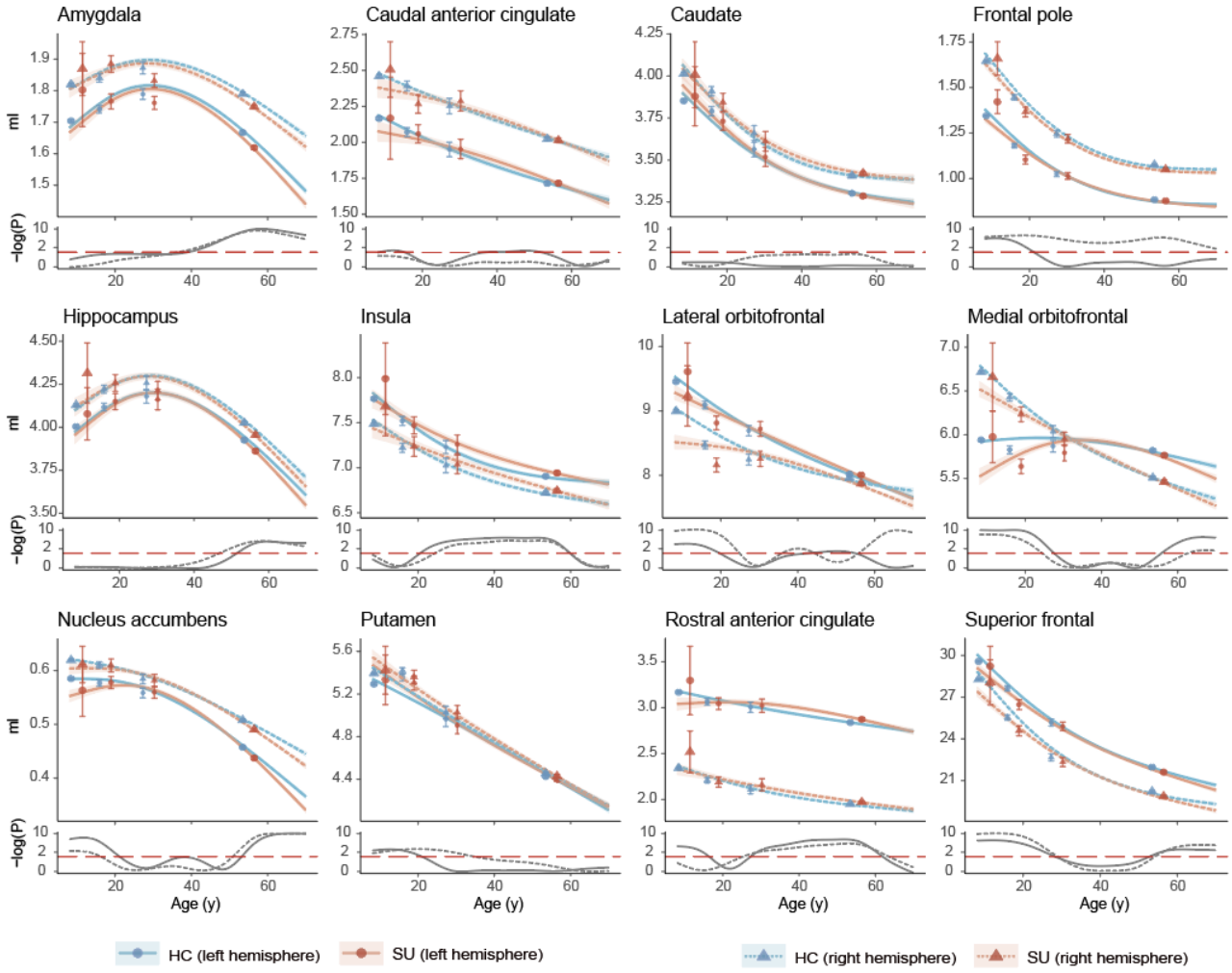
Supplementary Fig. 3. Replication of lifespan volumetric patterns between SUDs and HCs using ABCD and UKB longitudinal data.



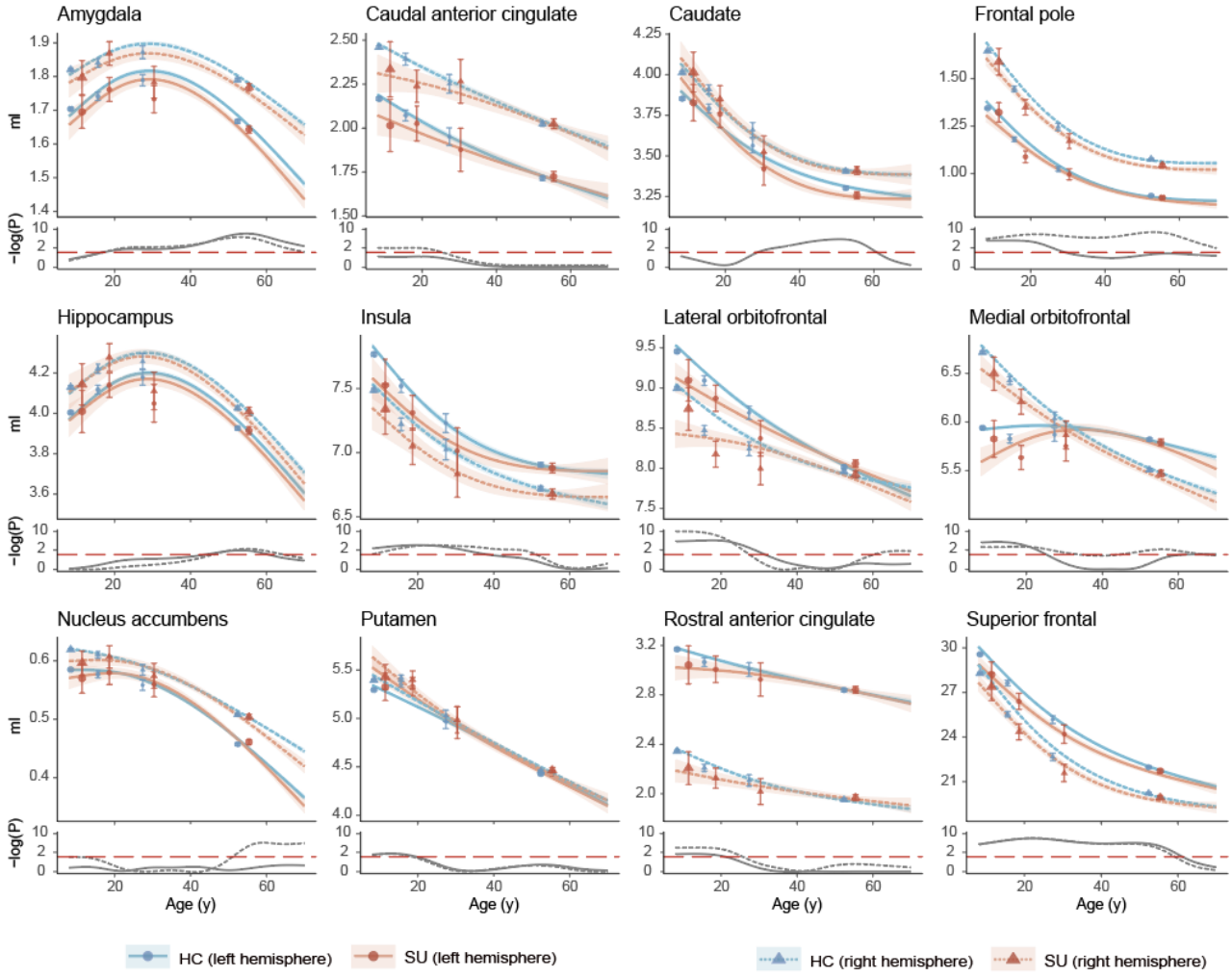
Supplementary Fig. 4. Replication of lifespan volumetric patterns between SUDs and HCs using validation samples. The comparison curves were only displayed for the age range of 13 - 70 years, due to the age limitations in the SUD group and also a clearer comparison with the exploratory results. Correlations of Z-values comparing SUD and HC between the discovery set and the validation set were provided. * P<0.05; ** P<0.01; *** P<0.001.



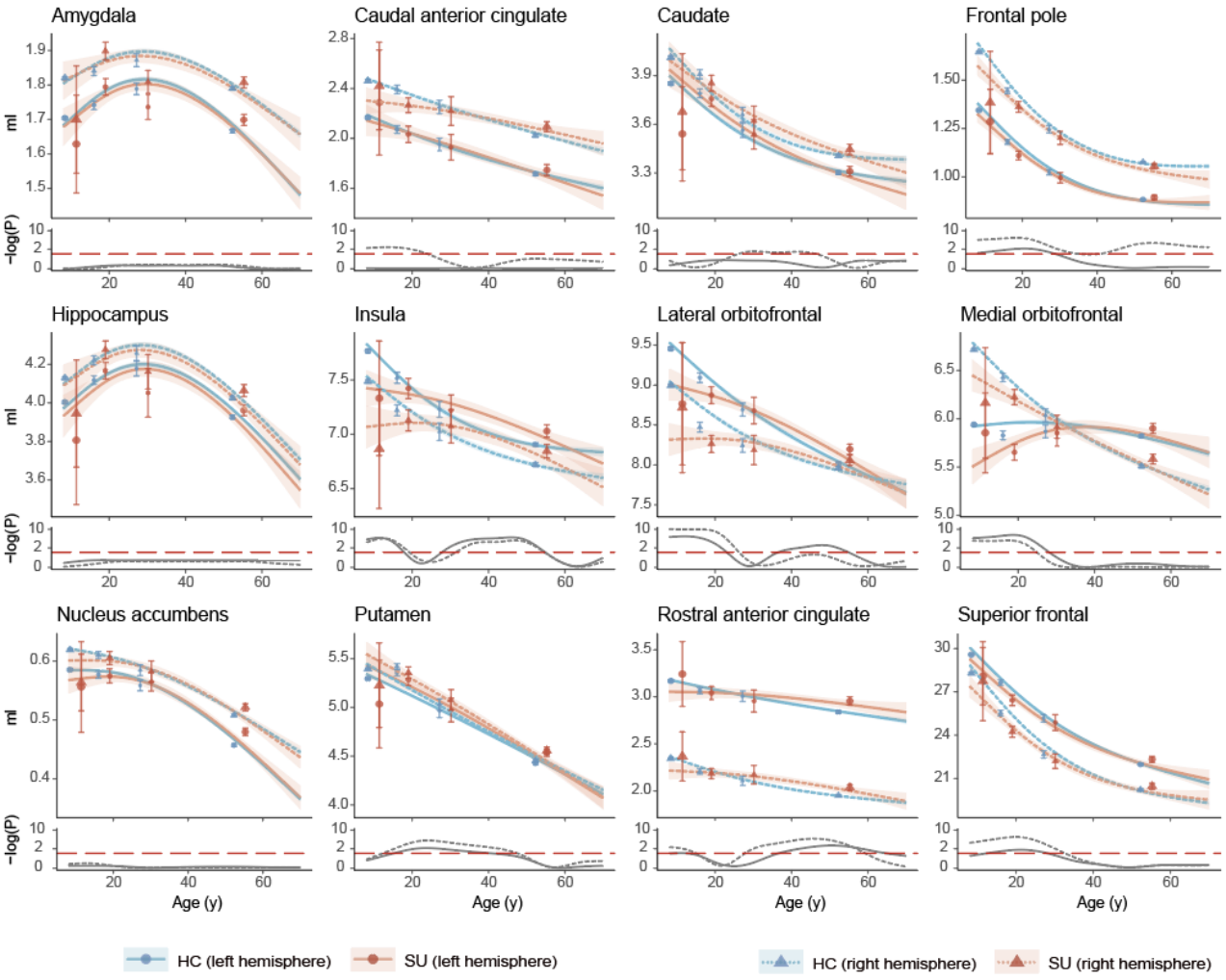
Supplementary Fig. 5. Lifespan volumetric patterns between SUDs and HCs with a relaxed definition of SUD in adults to the standards used for adolescents.



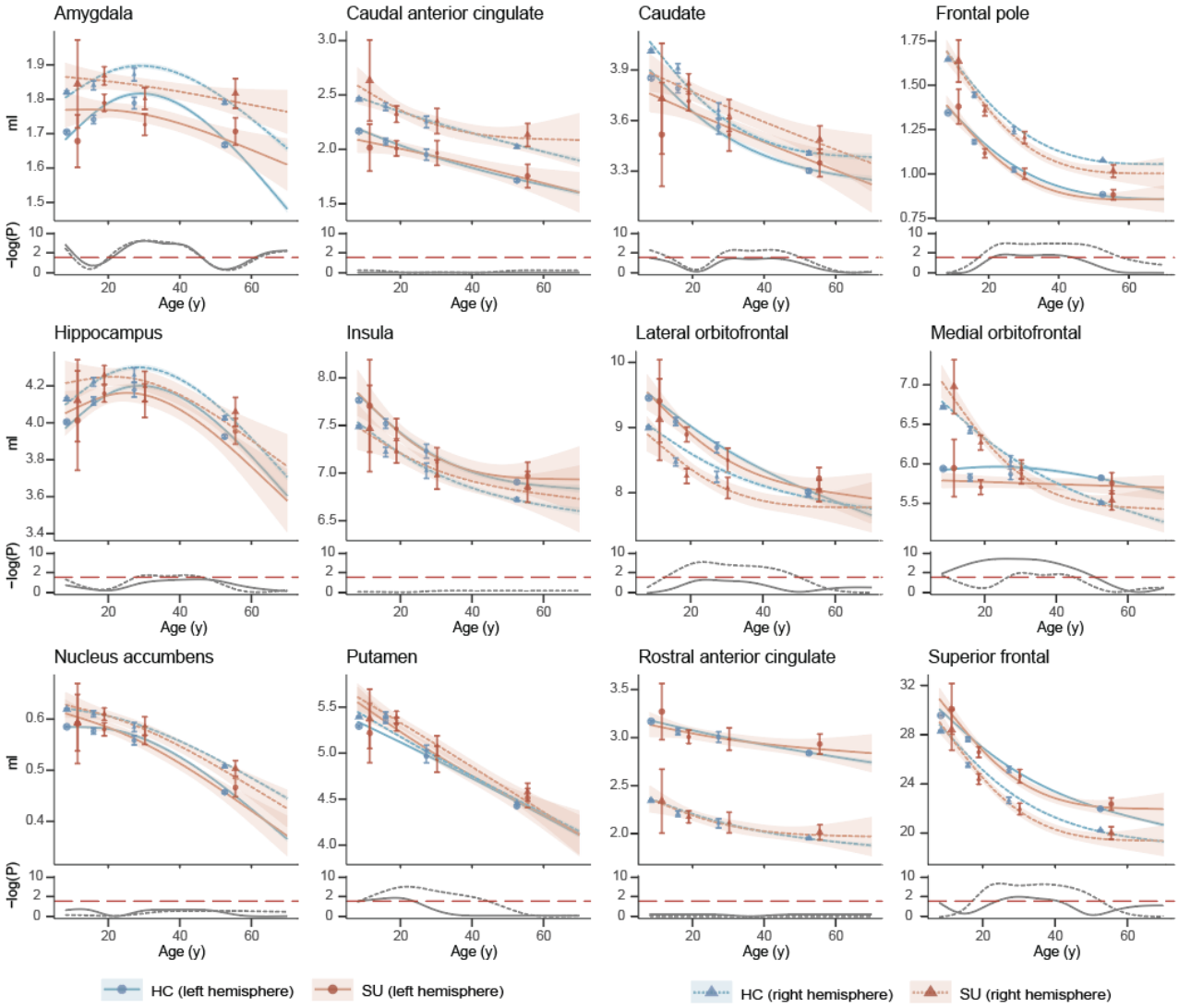
Supplementary Fig. 6. Lifespan volumetric patterns between alcohol use disorder participants and HCs.



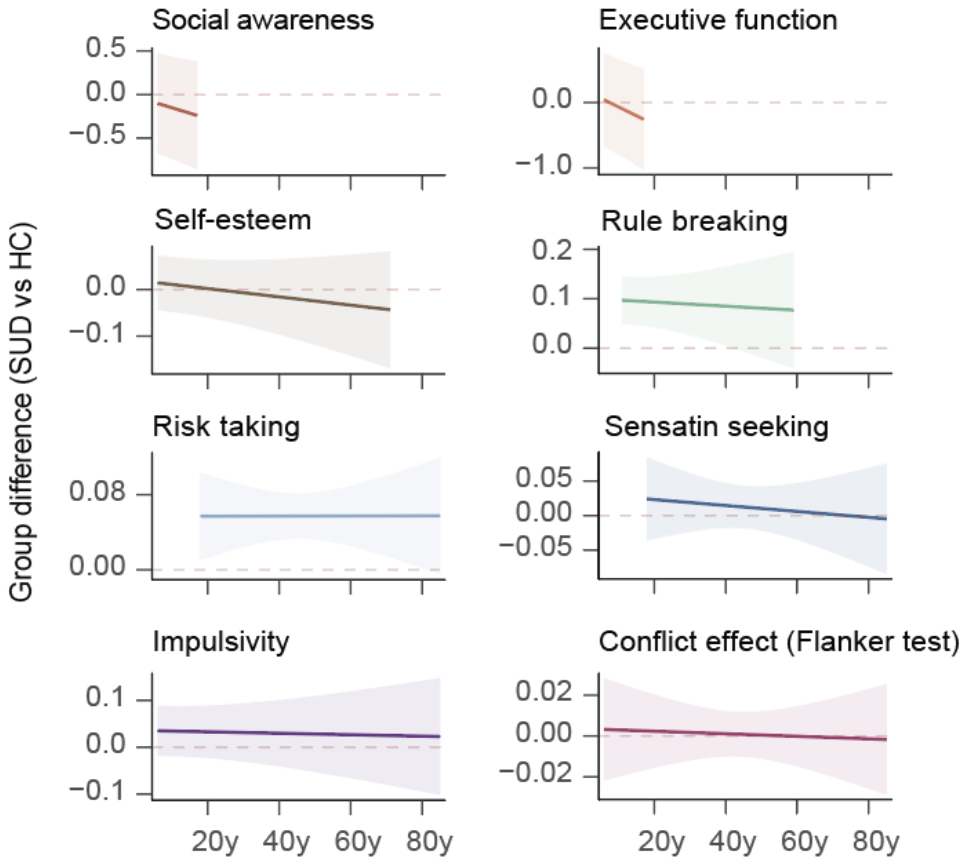
Supplementary Fig. 7. Lifespan volumetric patterns between tobacco use disorder participants and HCs.



Supplementary Fig. 8. Lifespan volumetric patterns between marijuana use disorder participants and HCs.

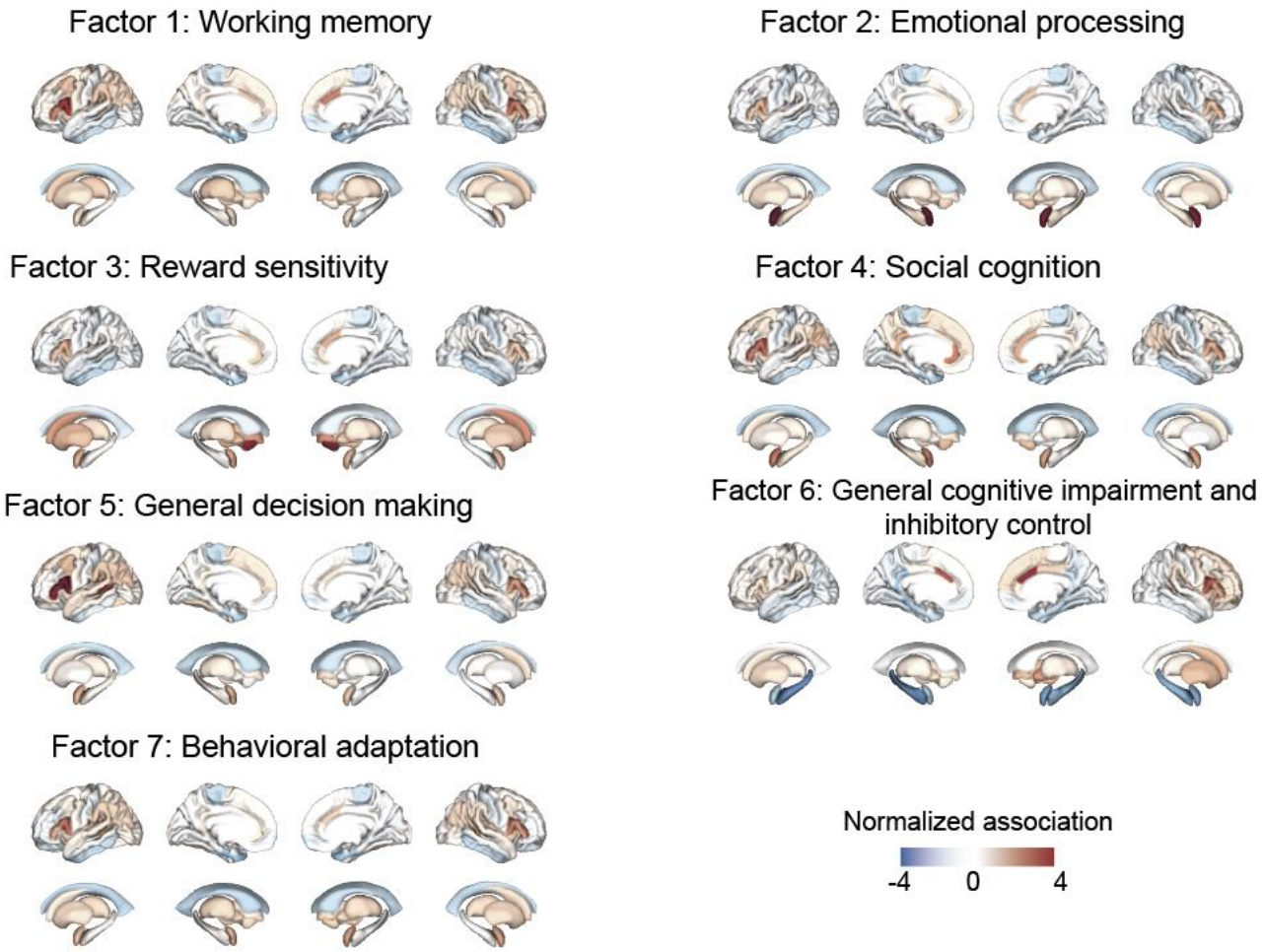


Supplementary Fig. 9. Lifespan volumetric patterns between drug use disorder participants and HCs.

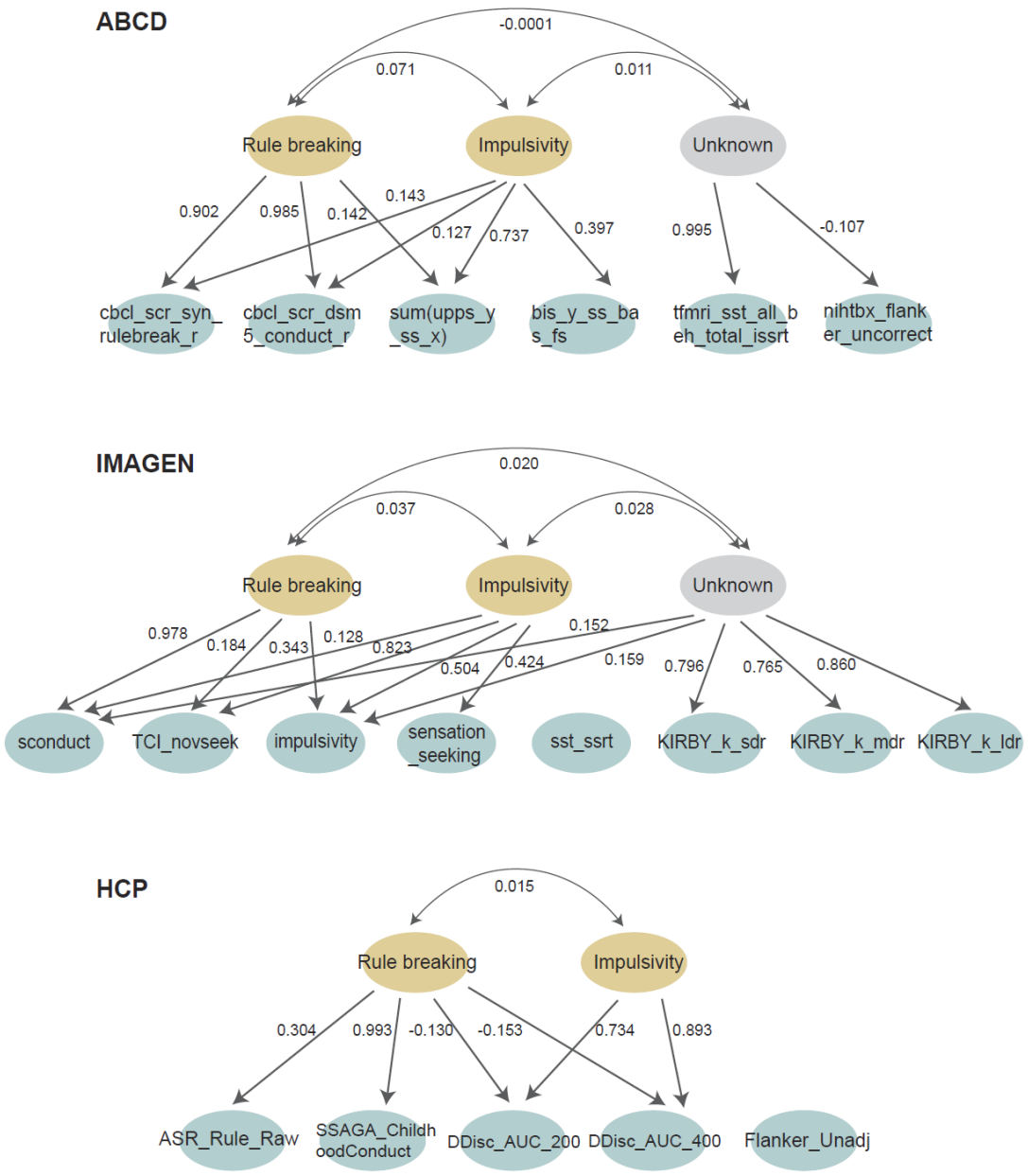


Supplementary Fig. 10. Validation of behavioral and neurocognitive comparisons between SUDs and HCs in NKI-RS.

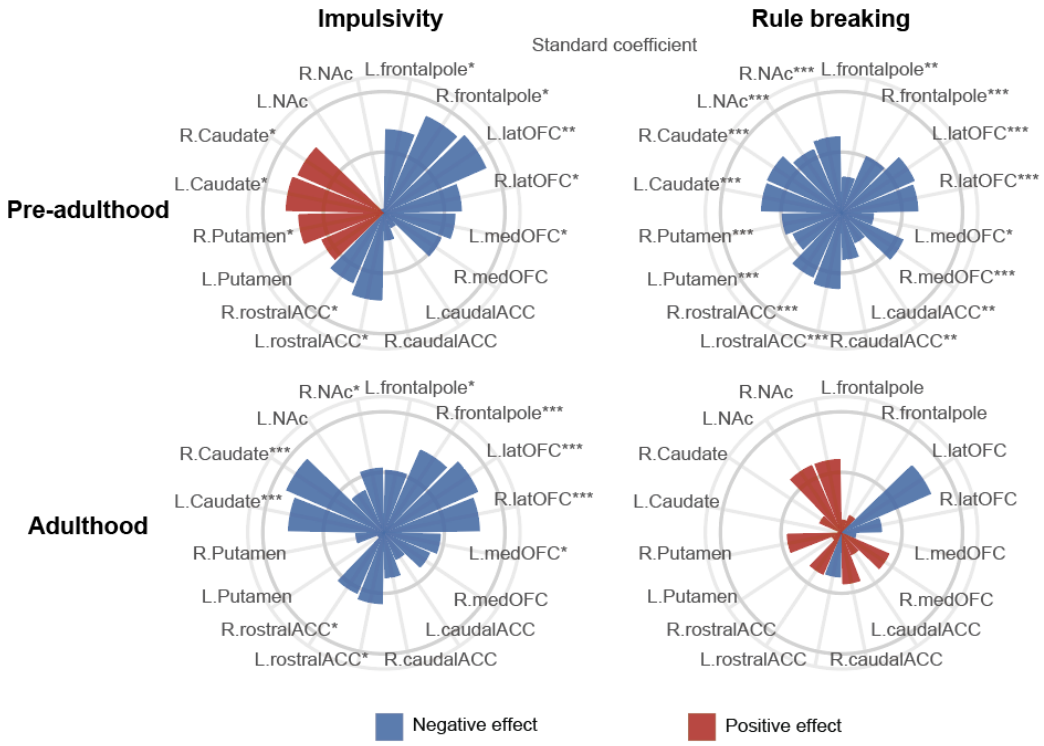
Supplementary Fig. 11. Word cloud visualizing cognitive terms related to dimensionality-reduced cognitive factors. Word size indicates the absolute strength of the loading.



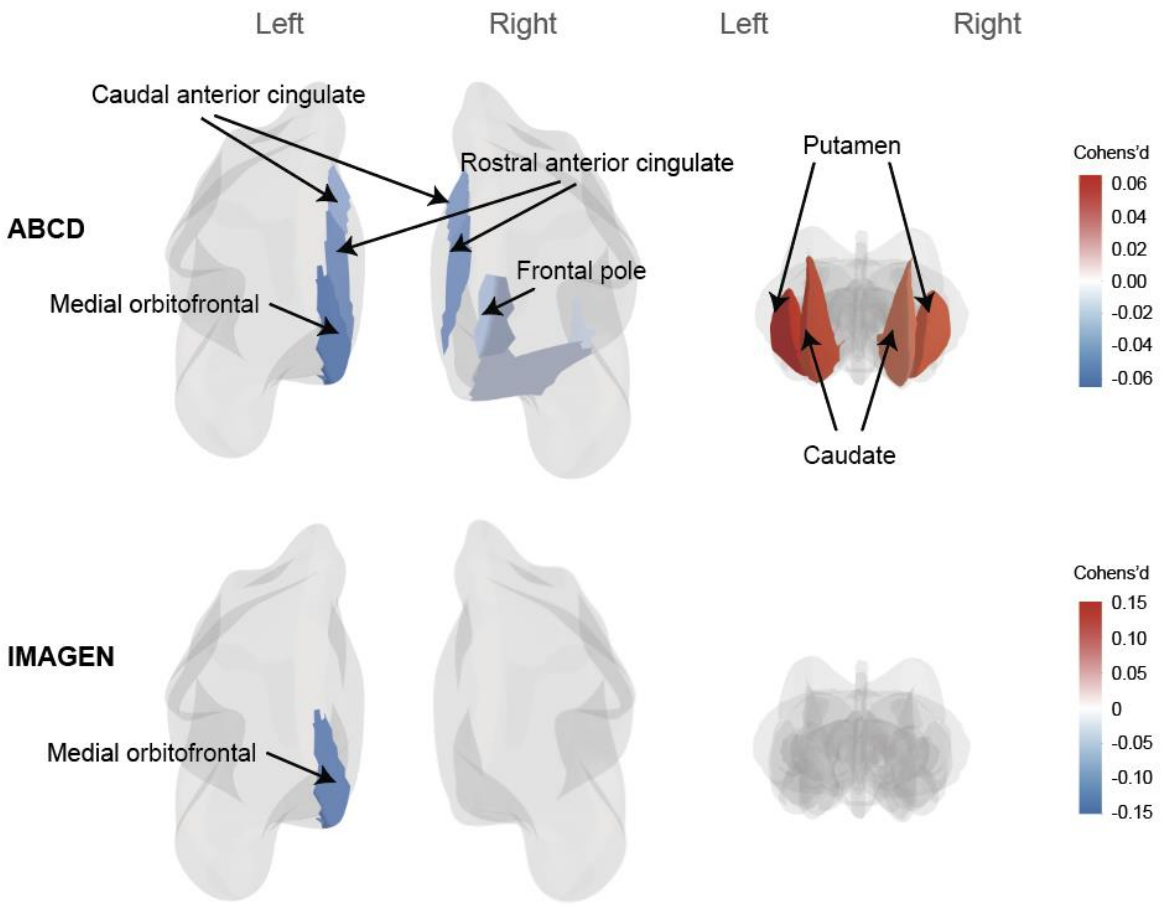
Supplementary Fig. 12. Spatial distribution of cognitive factors on parcellated brain atlases. Colors represent normalized parcellated regional associations.



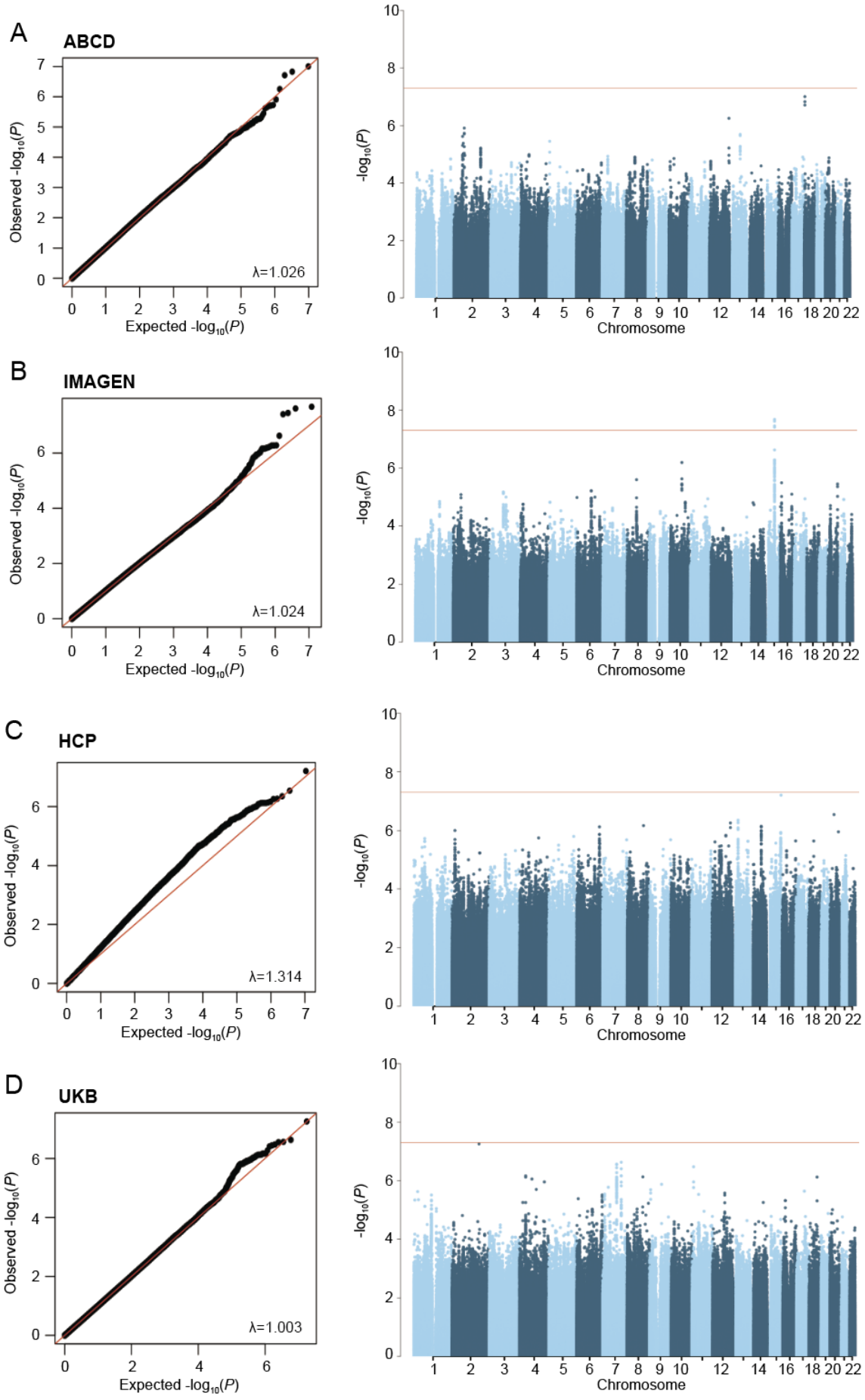
Supplementary Fig. 13. Explorative factor analysis for generalized rule-breaking and impulsivity scores extraction across studies. The number of factors were decided by the number of eigen values larger than 1. The absolute loadings lower than 0.1 were not shown.



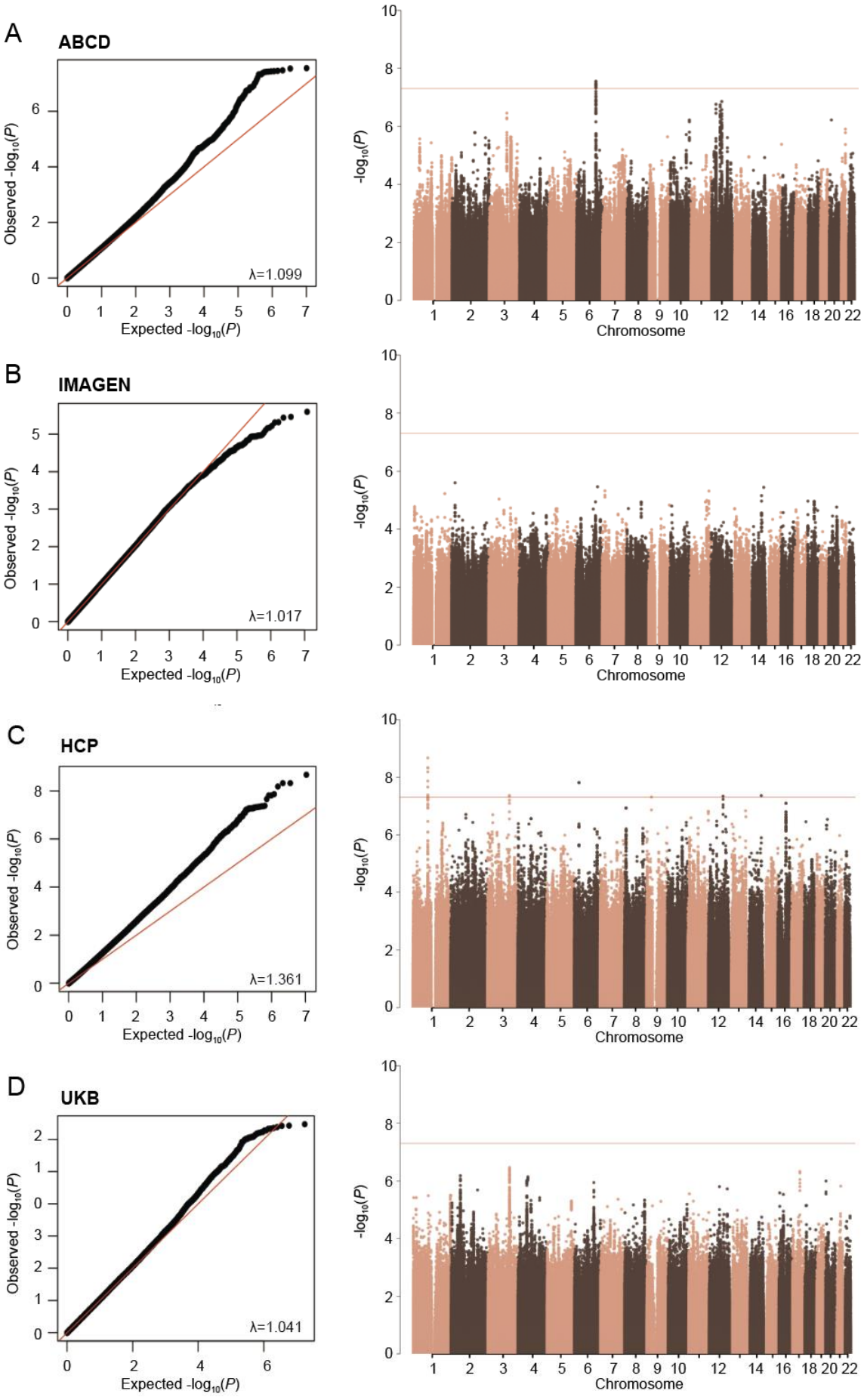
Supplementary Fig. 14. Correlation patterns between rule-breaking/impulsivity scores and GMVs in ROIs after adjusting for the correlation between rule-breaking and impulsivity. BH-FDR method was used for multiple testing.



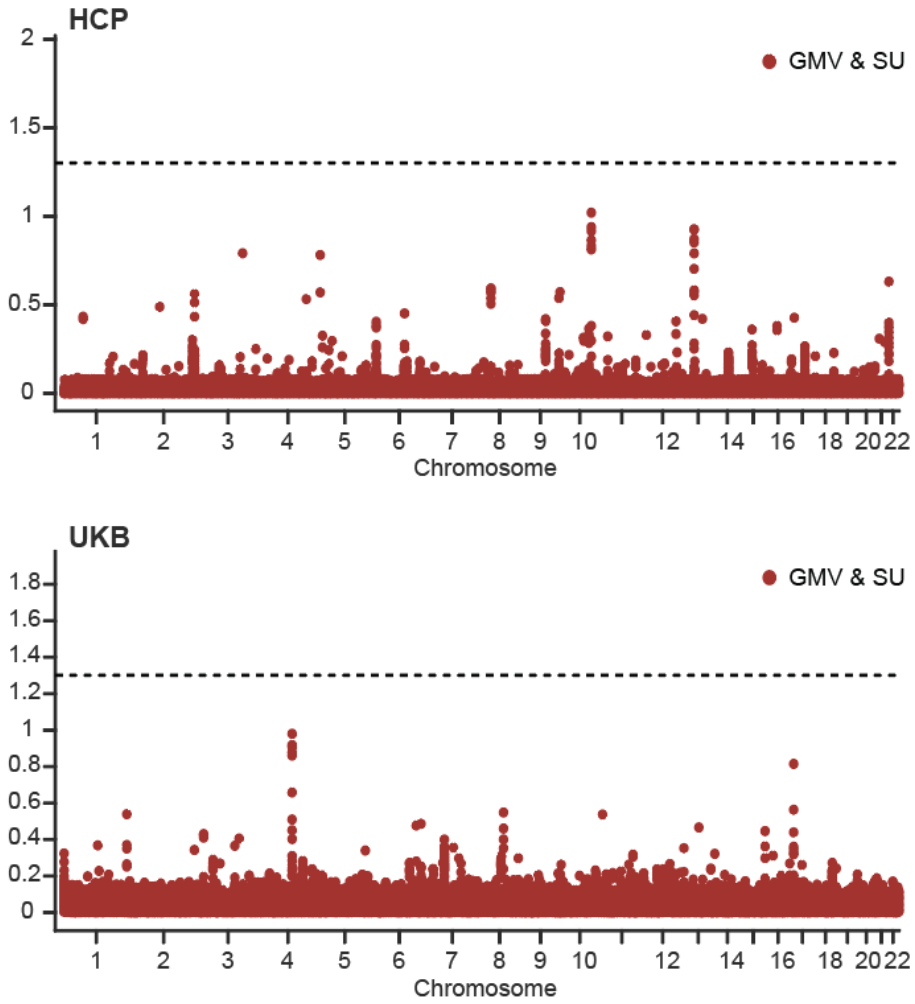
Supplementary Fig. 15. Longitudinal investigation of GMV changes leading to SUD. Two-tailed t-test was used and BH-FDR method was used for multiple testing.



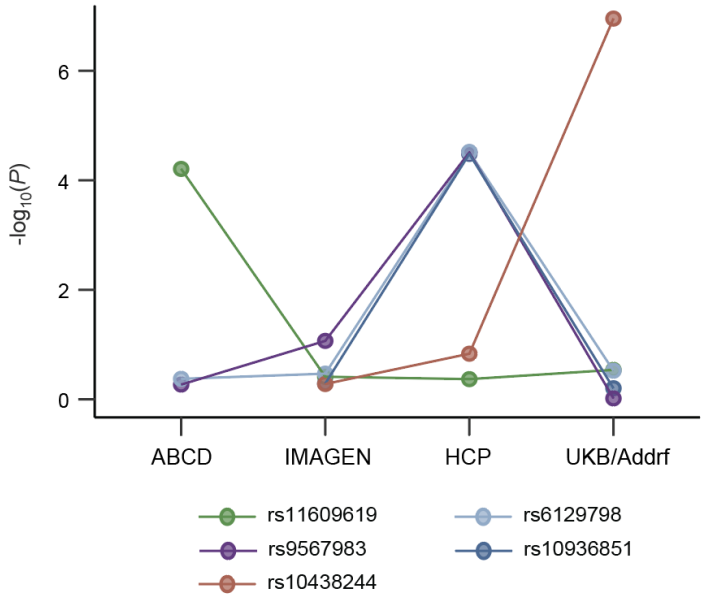
Supplementary Fig. 16. GWAS results for SUD across studies. SUD was used as a binary phenotype in GWAS for ABCD (A), IMAGEN (B), HCP (C) and UKB (D), respectively.



Supplementary Fig. 17. GWAS results for wGMV across studies. wGMV was used as a continuous phenotype in GWAS for ABCD (A), IMAGEN (B), HCP (C) and UKB (D), respectively.



Supplementary Fig. 18. conjFDR results in IMAGEN and UKB.



Supplementary Fig. 19. Log *P*-values of significant loci identified in Fig. 4C across different life stages in the study cohorts.

Supplementary Table 1. Demographic characteristics of the samples in ABCD, IMAGEN, HCP and UKB.

	ABCD （*n*=11,301）			IMAGEN (*n*=1,504)									HCP (*n*=1,113)			UKB (*n*=37,549)		
				BL			FU2			FU3								
	HC	SU	*P*	HC	SU	*P*	HC	SU	*P*	HC	SU	*P*	HC	SU	*P*	HC	SU	*P*
n	11,217	84		1,298	185		539	419		345	450		396	310		6,243	5,897	
Age	9.92 (0.62)	10.04 (0.67)	0.101	14.39 (0.41)	14.48 (0.37)	0.003	18.95 (0.74)	19.09 (0.72)	0.002	22.93 (0.74)	22.93 (0.65)	0.937	28.76 (3.70)	28.69 (3.42)	0.800	54.14 (7.75)	55.05 (7.44)	<0.001
Gender			0.003			0.023			<0.001			<0.001			<0.001			<0.001
Male	5,843 (52.1%)	58 (69.0%)		680 (52.4%)	114 (61.6%)		240 (44.5%)	273 (65.2%)		137 (39.7%)	281 (62.4%)		151 (38.1%)	195 (62.9%)		2,023 (32.4%)	3,802 (64.5%)	
Female	5,374 (47.9%)	26 (31.0%)		618 (47.6%)	71 (38.4%)		299 (55.5%)	146 (34.8%)		208 (60.3%)	169 (37.6%)		245 (61.9%)	115 (37.1%)		4,220 (67.6%)	2,095 (35.5%)	
Handedness			0.579			1.00			0.700			0.473			0.702			0.221
Right	8,920 (79.5%)	63 (75.0%)		1,156 (89.1%)	165 (89.2%)		481 (89.2%)	378 (90.2%)		305 (88.4%)	405 (90.0%)		358 (90.4%)	281 (90.6%)		5,601 (89.7%)	5,245 (88.9%)	
Left	813 (7.2%)	7 (8.3%)		142 (10.9%)	20 (10.8%)		58 (10.8%)	41 (9.8%)		40 (11.6%)	45 (10.0%)		37 (9.3%)	27 (8.7%)		541 (8.7%)	563 (9.5%)	
Mixed	1,484 (13.2%)	14 (16.7%)		-	-		-	-		-	-		1 (0.3%)	2 (0.6%)		101 (1.6%)	89 (1.5%)	
Ethnics			0.139			0.595			0.648			0.162			0.069			<0.001
White	7,083 (67.2%)	46 (56.8%)		1,122 (91.7%)	153 (89.5%)		453 (91.3%)	350 (92.6%)		237 (88.1%)	325 (92.1%)		298 (76.0%)	251 (82.6%)		5,731 (92.0%)	5,747 (97.8%)	
Black	1,764 (16.7%)	18 (22.2%)		13 (1.1%)	2 (1.2%)		4 (0.8%)	4 (1.1%)		3 (1.1%)	5 (1.4%)		56 (14.3%)	36 (11.8%)		99 (1.6%)	32 (0.5%)	
Other	1,694 (16.1%)	17 (21.0%)		88 (7.2%)	16 (9.4%)		39 (7.9%)	24 (6.3%)		29 (10.8%)	23 (6.5%)		38 (9.7%)	17 (5.6%)		396 (6.4%)	96 (1.6%)	
BMI	18.84 (4.19)	20.03 (4.96)	0.011	20.61 (3.21)	21.15 (3.24)	0.024	22.35 (3.84)	23.07 (3.96)	0.025	23.33 (4.50)	23.45 (3.74)	0.783	26.44 (5.38)	26.57 (4.57)	0.933	25.93 (4.65)	25.73 (3.81)	0.005
SES	7.22 (2.42)	6.09 (2.47)	<0.001	0.64 (1.00)	0.80 (0.98)	0.046	0.54 (0.87)	0.62 (0.90)	0.136	0.52 (0.83)	0.60 (0.87)	0.288	5.31 (2.09)	4.88 (2.20)	0.002	2.73 (1.12)	3.09 (1.15)	<0.001
NLE	2.45 (2.29)	3.48 (2.75)	<0.001	4.97 (2.44)	5.30 (2.76)	0.096	2.52 (1.76)	2.74 (1.94)	0.009	1.22 (1.05)	1.34 (1.26)	0.095	47.76 (8.64)	49.37 (9.61)	0.010	2.06 (2.74)	2.46 (2.94)	<0.001

IQ	86.29 (9.08)	83.18 (11.06)	<0.001	176.04 (22.31)	172.20 (24.91)	0.015	-	-		-	-		17.15 (4.83)	17.05 (4.67)	0.240	10.22 (2.05)	9.70 (2.07)	0.031
Substance misuse																		
Alcohol	-	21 (25.0%)		-	52 (28.1%)		-	263 (62.8%)		-	229 (50.9%)		-	231 (74.5%)		-	4,379 (74.3%)	
Tobacco	-	51 (60.7%)		-	31 (16.8%)		-	116 (27.7%)		-	100 (22.2%)		-	76 (24.5%)		-	1,381 (23.4%)	
Marijuana	-	8 (9.5%)		-	86 (46.5%)		-	198 (47.3%)		-	288 (64.0%)		-	102 (32.9%)		-	653 (11.1%)	
Other Drugs	-	13 (15.5%)		-	68 (36.8%)		-	132 (31.5%)		-	266 (59.1%)		-	89 (28.7%)		-	109 (0.02%)	

Continuous variables were provided as mean(s.d.) and discrete variables were provided as n(%).

Supplementary Table 2. Comparisons in neurobehavioral performances between SUDs and HCs in ABCD.

Measurements	Cohen's d	P_{adj}
CBCL - rule breaking	0.64	1.02E-08
CBCL - conduct score	0.73	1.12E-10
CBCL - stress score	0.34	0.004
BIS/BAS - fun seeking score	0.27	0.023
UPPS-P total impulsivity score	0.44	1.35E-04
Picture vocabulary test score	-0.22	0.046
Flanker test score	-0.15	0.161
List sorting test score	-0.20	0.071
Card sort test score	-0.20	0.072
Pattern comparison test score	0.07	0.582
Picture sequence memory test score	-0.24	0.033
Reading recognition test score	-0.31	0.008
Fluid intelligence	-0.26	0.028
Crystal intelligence	-0.40	0.001
Overall intelligence	-0.25	0.032
SST - stop signal reaction time (integration estimation)	0.01	0.927
MID - Positive rate in small reward trails	0.13	0.602
MID - Positive rate in large reward trails	-0.44	0.046

Two-sided t-test was used for comparison, adjusting for age, sex and handedness. BH-FDR was used for multiple testing. CBCL, Child Behavior Check List; BIS/BAS, Behavioral inhibition and Behavioral Activations scales; SST, Stop Signal Test; MID, Monetary Incentive Delay task. All the rate calculation in reward trails was adjusted for the corresponding rate in neural trails.

Supplementary Table 3. Comparisons in neurobehavioral performances between SUDs and HCs in IMAGEN.

Measurements	Cohen's d	P_{adj}
Conduct problem	0.25	5.88E-06
TCI - disorderliness vs. regimentation	0.33	1.23E-09
TCI - novelty seeking total score	0.48	3.49E-18
CGT - delay aversion	0.11	0.154
CGT - deliberation time	-0.01	0.910
CGT - overall proportion bet	0.09	0.205
CGT - quality of decision making	-0.03	0.725
CGT - risk adjustment	-0.01	0.910
CGT - risk taking	0.08	0.223
SST - stop signal reaction time	-0.06	0.363
SWM - between errors	-0.07	0.363
SWM - strategy	-0.01	0.910
AGN - mean correct latency positive	0.03	0.781
AGN - mean correct latency negative	0.01	0.910
AGN - total omissions positive	-0.07	0.454
AGN - total omissions negative	-0.06	0.495
SURPS - sensation seeking	0.41	4.40E-14
SURPS - impulsivity	0.36	7.25E-11
MID - Positive rate in small reward trails	0.03	0.725
MID - Positive rate in large reward trails	0.06	0.363
PRM - Percent correct	-0.18	0.034
RVP - A	-0.15	0.053
KIRBY - k	0.09	0.191
KIRBY - k for small LDRs	0.13	0.045
KIRBY - k for Medium LDRs	0.13	0.039
KIRBY - k for Large LDRs	0.10	0.154

Two-sided t-test was used for comparison, adjusting for age, sex and handedness. BH-FDR was used for multiple testing. TCI, Temperament and Character Inventory; CGT, Cambridge Guessing Task; SST, Stop Signal Test; SWM, Spatial Working Memory; AGN, Affective Go-No Go; SURPS, Substance Use Risk Profile Scale; MID, Monetary

Incentive Delay task; PRM, Pattern Recognition Memory; RVP, Rapid visual Information Processing; KIRBY, Monetary-Choice Questionnaire. All the rate calculation in reward trails was adjusted for the corresponding rate in neural trails.

Supplementary Table 4. Comparisons in neurobehavioral performances between SUDs and HCs in HCP.

Measurements	Cohen's d	P_{adj}
ASR - rule breaking	0.89	9.39E-28
SSAGA - conduct problem	0.42	3.07E-07
Picture vocabulary test score	-0.22	0.011
Flanker test score	-0.12	0.178
List sorting test score	0.01	0.896
Card sort test score	0.12	0.178
Pattern comparison test score	-0.02	0.879
Picture sequence memory test score	0.01	0.896
Reading recognition test score	-0.11	0.233
Penn Progressive Matrices - Fluid intelligence	-0.09	0.343
DDisc - AUC in small reward trails ($200)	-0.34	6.36E-05
DDisc - AUC in large reward trails ($400)	-0.27	0.002
NIH Toolbox - Sadness	0.15	0.092
NIH Toolbox - Life satisfaction	-0.19	0.032
NIH Toolbox - Meaning and Purpose	-0.31	2.95E-04
NIH Toolbox - Positive Affect	-0.06	0.494
NIH Toolbox - Emotional Support	-0.08	0.424
NIH Toolbox - Instrumental Support	-0.12	0.178
NIH Toolbox - Perceived Stress	0.19	0.029

Two-sided t-test was used for comparison, adjusting for age, sex and handedness. BH-FDR was used for multiple testing. ASR, Achenbach Adult Self Report; SSAGA, Semi-Structured Assessment for the Genetics of Alcoholism; DDisc, Delay discounting. A smaller AUC in DDisc indicates steep delay discounting. All the emotional responses were measured according to NIH Toolbox.

Supplementary Table 5. Comparisons in neurobehavioral performances between SUDs and HCs in UKB.

Measurements	Cohen's d	P_{adj}
Risk taking	0.21	1.30E-27
Anhedonia	0.07	0.001
Mood swings	0.10	5.58E-07
Irritability	0.19	9.01E-22
Sensitivity / hurt feelings	0.05	0.029
Fed-up feelings	0.07	4.90E-04
Worry too long after embarrassment	0.02	0.374
Suffer from 'nerves'	-0.02	0.302
Pairs matching - incorrect matches	0.01	0.683
Fluid intelligence	0.002	0.932
Matrix pattern completion - Correct number	-0.04	0.183
Reaction time	-0.03	0.211
Trail making - mean completion time	0.05	0.087
Numeric memory	-0.06	0.052
Tower rearranging test - correct rate	-0.06	0.021
Symbol digit substitution - correct rate	-0.03	0.029

Two-sided t-test was used for comparison, adjusting for age, sex and handedness. BH-FDR was used for multiple testing. Fluid intelligence was calculated as the correct rate for participants who have answered at least half of the 13 questions.

Supplementary Table 6. Cognitive terms selected from Neurosynth platform.

affect	affective	anxiety	anxiety disorders	appraisal	attentional control	awareness
choice	choices	choose	cognition	cognitive	cognitive control	cognitive deficits
cognitive emotional	cognitive function	cognitive functions	cognitive impairment	cognitive performance	cognitive processes	cognitive tasks
cognitively	conscious	consciousness	control network	control processes	control task	controlled
controlling	daily life	decision	decision making	decision task	decisions	effort
effortful	emotion regulation	emotional	emotional faces	emotional information	emotional responses	emotional stimuli
emotional valence	emotionally	emotions	everyday life	executive control	executive function	executive functions
expectancy	expectation	expectations	fear	fearful	fearful faces	gambling
goal	goal directed	goals	happy	happy faces	high risk	impulsive
impulsivity	inhibit	inhibition	inhibitory	inhibitory control	judgment	judgment task
judgments	life	memory processes	memory tasks	memory wm	monetary	monetary incentive
monetary reward	monitoring	mood	motivation	motivational	negative affect	negative emotional
negative emotions	nogo	prediction	prediction error	predictions	predictive	preference
preferences	probabilistic	probability	reappraisal	reference	referential	referred
response inhibition	reward	reward anticipation	rewarding	rewards	risk	risk taking
risky	self referential	self report	self reported	shifting	shifts	social
social cognition	social cognitive	social interaction	social interactions	socially	stop signal	stress
stress disorder	stroop	stroop task	subjective	suppressed	suppression	switch
switching	unexpected	value	values	wm	wm task	working memory

Supplementary Table 7. Correlations between longitudinal GMV changes and the severity of SUD.

ROI	Correlation (P_{adj})		
	Alcohol	Marijuana	Other drugs
Left_Accumbens_area	-0.05 (0.78)	0.03 (0.91)	-0.18 (0.78)
Right_Accumbens_area	-0.03 (0.87)	-0.12 (0.68)	0.07 (0.87)
Left_Amygdala	0.03 (0.87)	-0.03 (0.91)	0.04 (0.87)
Right_Amygdala	-0.01 (0.87)	-0.1 (0.72)	0.34 (0.47)
Left_Caudate	0.02 (0.87)	0.08 (0.78)	0.06 (0.87)
Right_Caudate	-0.01 (0.87)	0.22 (0.45)	0.13 (0.78)
Left_Hippocampus	0.004 (0.93)	0.14 (0.61)	0.12 (0.78)
Right_Hippocampus	0.03 (0.87)	-0.06 (0.83)	0.17 (0.78)
Left_Putamen	0.04 (0.87)	0.19 (0.47)	0.03 (0.87)
Right_Putamen	0.07 (0.44)	0.02 (0.91)	0.14 (0.78)
Left_lateralorbitofrontal	-0.09 (0.42)	-0.21 (0.46)	-0.13 (0.78)
Right_lateralorbitofrontal	-0.08 (0.42)	-0.12 (0.68)	-0.33 (0.47)
Left_medialorbitofrontal	-0.01 (0.93)	-0.19 (0.47)	0.03 (0.87)
Right_medialorbitofrontal	-0.11 (0.31)	-0.14 (0.61)	-0.03 (0.87)
Left_frontalpole	0.02 (0.87)	0.08 (0.78)	0.07 (0.87)
Right_frontalpole	-0.02 (0.87)	-0.01 (0.93)	-0.16 (0.78)
Left_superiorfrontal	-0.06 (0.49)	-0.26 (0.45)	-0.11 (0.78)
Right_superiorfrontal	-0.08 (0.42)	-0.26 (0.45)	-0.16 (0.78)
Left_rostralanteriorcingulate	-0.1 (0.31)	-0.24 (0.45)	0.22 (0.78)
Right_rostralanteriorcingulate	-0.02 (0.87)	-0.15 (0.61)	-0.07 (0.87)
Left_caudalanteriorcingulate	-0.03 (0.87)	-0.07 (0.78)	0.17 (0.78)
Right_caudalanteriorcingulate	-0.07 (0.44)	-0.3 (0.45)	-0.15 (0.78)

As only 9 participants were left with neuroimaging measurements available at both BL and FU2, and an FTND score greater than 0, correlation results between FTND score, which represented the severity of tobacco use, and GMV changes were not provided. BH-FDR method was used for multiple testing.

Supplementary Table 8. Comparisons in neurobehavioral performances at baseline between follow-up SUDs and follow-up HCs in ABCD and IMAGEN.

Measurements	Cohen's d	P_{adj}
ABCD		
CBCL - rule breaking	0.03	0.256
CBCL - conduct score	0.02	0.435
CBCL - stress score	3.40E-4	0.986
BIS/BAS - fun seeking score	0.09	5.79E-06
UPPS-P total impulsivity score	0.18	1.02E-19
SST - stop signal reaction time (integration estimation)	0.00	0.940
MID - Positive rate in small reward trails	0.04	0.435
MID - Positive rate in large reward trails	0.09	0.031
IMAGEN		
Conduct problem	0.20	0.005
TCI - disorderliness vs. regimentation	0.37	8.50E-08
TCI - novelty seeking total score	0.46	1.79E-11
CGT - delay aversion	-0.02	0.874
CGT - deliberation time	-8.58E-4	0.990
CGT - overall proportion bet	0.06	0.724
CGT - quality of decision making	-0.07	0.679
CGT - risk adjustment	-0.05	0.724
CGT - risk taking	0.05	0.724
SST - stop signal reaction time	0.11	0.268
SURPS - sensation seeking	0.34	5.33E-07
SURPS - impulsivity	0.20	0.005
KIRBY - k	-0.05	0.724
KIRBY - k for small LDRs	-0.03	0.874
KIRBY - k for Medium LDRs	-0.02	0.874
KIRBY - k for Large LDRs	-0.02	0.874

Two-sided t-test was used for comparison, adjusting for age, sex and handedness. BH-FDR was used for multiple testing within studies. CBCL, Child Behavior Check List; BIS/BAS, Behavioral inhibition and Behavioral Activations scales; SST, Stop Signal Test; MID, Monetary Incentive Delay task; TCI, Temperament and Character Inventory; CGT, Cambridge Guessing Task; KIRBY, Monetary-Choice Questionnaire. All the rate calculations in reward trails were adjusted for the corresponding rate in neural trails.

Supplementary Table 9. Demographic characteristics of the samples in validation samples.

NKI-RS (*n*=1112)			
	HC (*n*=862)	SU (*n*=250)	*P*
Age	42.96 (20.06)	41.42 (17.67)	0.272
Gender			<0.001
Male	296 (34.3%)	122 (48.8%)	
Female	566 (65.7%)	128 (51.2%)	
Handedness			0.596
Right	771 (89.4%)	220 (88.0%)	
Left	91 (10.6%)	30 (12.0%)	
Ethnics			0.017
White	623 (72.4%)	195 (78.9%)	
Black	150 (17.4%)	41 (16.6%)	
SES	6.82 (2.37)	6.28 (2.74)	0.006
IQ	101.57 (13.43)	100.52 (13.20)	0.277

Cam-CAN (*n*=445)			
	HC (*n*=140)	SU (*n*=250)	*P*
Age	54.97 (18.83)	55.80 (17.86)	0.654
Gender			
Male	47 (33.6%)	171 (56.1%)	<0.001
Female	93 (66.4%)	134 (43.9%)	
Handedness			
Right	129 (92.1%)	272 (89.2%)	0.478
Left	7 (5.0%)	25 (8.2%)	
Mixed	4 (2.9%)	8 (2.6)	
Ethnics			
White	128 (92.1%)	300 (98.4%)	0.003
Other	11 (7.9%)	5 (1.6%)	
SES			0.025
A	24 (17.1%)	75 (24.6%)	
B	34 (24.3%)	61 (20.0%)	
C	32 (22.9%)	78 (25.6%)	
D	32 (22.9%)	44 (14.4%)	
E	10 (7.1%)	12 (3.9%)	
F	8 (5.7%)	35 (11.5%)	
IQ	30.94 (7.14)	32.37 (6.48%)	0.039

BRAINMINT (*n*=570)			
	HC (*n*=414)	SU (*n*=156)	*P*
Age	16.6 (3)	19 (2.1)	<0.001
Gender			0.55
Male	121 (29,2%)	41 (26,3%)	
Female	293 (70,8%)	115 (73,7%)	
Handedness			1
Right	375 (90,5%)	141 (90,4%)	

Left	35 (8,5%)	15 (9.6%)	
Mixed	4 (1%)	0 (0%)	

Continuous variables were provided as mean(s.d.) and discrete variables were provided as n(%). In CamCAN, handedness is determined by the handedness score: homeint_handedness > 50 – Right-handed; homeint_handedness < -50 – Left-handed; -50 ≤ homeint_handedness ≤ 50 – Mixed-handed.