FIRST-EPISODE PSYCHOSIS FIRST-EPISODE PSYCHOSIS PATIENTS WHO DETERIORATED IN THE PREMORBID PERIOD DO NOT HAVE HIGHER POLYGENIC RISK SCORES THAN OTHERS. A CLUSTER ANALYSIS OF EU-GEI DATA.

Supplementary material

Supplementary Methods

1. Participants' recruitment

Centrally trained researchers screened all potential FEP patients, aged 18-64 years between 2010 and 2015, presenting to mental health services with a clinical diagnosis for an untreated non-affective or affective FEP, based on ICD-10 (ICD-10 F20-33). Patients excluded were previously treated for psychosis or met criteria for organic psychosis (ICD-10: F06) or transient psychotic symptoms resulting from acute intoxication (ICD-10: F1X.5) [1]. The 17 catchment areas were distributed across six countries: United Kingdom, Italy, Spain, Holland, France, and Brazil[2,3].

Population controls were aged 18-64 years and residents in the same catchment areas as the patients and recruited using a combination of random and quota sampling to ensure representativeness of each local population in terms of age, sex, and self-ascribed ethnicity. They did not report any lifetime treatment for psychosis.

2. Instruments

A modified version of the Medical Research Council (MRC) sociodemographic schedule [4] collected such characteristics. Each subject received a research-based diagnosis based on the OPCRIT algorithm[5]. The Cannabis Experience Questionnaire (CEQ) collected cannabis and other substance use information [6] categorized as "no use," "occasional," and "daily" use. Based on available information on potency of cannabis used, a category included "no use," "daily use of high potency cannabis" [tetrahydrocannabinol (THC) 10%], and "any other pattern of use" [6]. The Nottingham Onset Scale [7] listed the age of onset and weeks of untreated psychosis (DUP). A medication list recorded antipsychotic (AP) and other treatments. An abbreviated version of the WAIS, including information (verbal comprehension), block design (perceptual reasoning), digit symbol (processing speed), and arithmetic (working memory) subtest, estimated IQ in patients and controls [8]. Given the multisite design, we could not use the same psychometric test (i.e., a reading test) among countries to assess premorbid IQ. Instead, nine scales from the PAS [9,10] examined premorbid social (PSF) and academic functioning (PAF), from childhood to age 11 (<12 years) and in early adolescence (12-16 years). IQ calculation and the PAS dimensions included in these analyses, adjusted by age, are fully described in our previous study [11].

3. Cluster analysis

The Two-Step Cluster Analysis procedure in SPSS, version 24 is able to analyze large data files efficiently by using a likelihood distance measure assuming that the variables included are independent and normally distributed. However, this analysis is robust enough to the violations of both assumptions. The first step constructs a Cluster Features (CF) Tree, and each case is placed into an existing or a new leaf node containing variable information about the case. Thus, the CF tree contains a summary of the data file. The second step uses an agglomerative clustering algorithm and produces a range of solutions, which are then compared using Schwarz's Bayesian Criterion (BIC) or Akaike Information Criteria (AIC) [12–14] Thus, this analysis gives different results depending on how the observations are ordered. To prevent collinearity between childhood and early adolescence measures in PAS and PSF (r=0.7 in both scales) and focus on potential deterioration, the input variables were: $PSF_{<12}$; $PAF_{<12}$; change scores between <12 years and 12-16 years for both social (PSF_{change}) and academic domain (PAF_{change}), and current IQ z-scores.

Firstly, unsupervised clustering was run 1000 times by changing the random order of the subjects to determine the optimal number of clusters (maximum solution expected=6). We used the stepwise decrease in log-likelihood as the distance measure for identifying clusters and change in the Bayesian Information Criterion to determine the number of clusters to retain (best ratio change of cluster distance at least >1.15) [15]. Secondly, we ran a 50 subjects' assignment solution by pre-determining the chosen number of clusters with random reordering. Then, subjects were allocated to the category they were assigned most of the time. ANOVAs for repeated measures were adjusted by age, sex, country, and self-ascribed ethnicity, having clusters of patients and controls as the between-group factor. This analysis also allowed us to see whether each patients' group differed from controls in PSF and PAF changes. ANOVAs, chi-square tests, ordinal and logistic regression were used to compare groups on demographic, cannabis use, and clinical characteristics. In addition, between-group comparisons of IQ and PAS, by ANOVA, were adjusted by age, sex, country, and self-ascribed ethnicity. Based on the demographic data of each site, a weighted score was generated to reduce the probability for biased estimation of the prevalence of exposures among controls.

4. Genotyping and PRS calculation

Subjects were genotyped at the MRC Centre for Neuropsychiatric Genetics and Genomics in Cardiff (UK) by a custom Illumina HumanCoreExome-24 BeadChip genotyping array, including 570,038 genetic variants. Quality control was conducted in PLINK v1.07 or with custom Perl scripts. Variants with call rate <98% or with Hardy-Weinberg Equilibrium p-value < 1e-6 and samples with call rate <98% were excluded from the dataset. The genotypic sex of samples was calculated from X

chromosome data and samples with different genotypic and recorded sex were excluded. A Principal Component Analysis was performed on pruned linkage disequilibrium variants and the first 10 principal components (PCs) were retained and included as covariates in the genetic analyses to control for the effects of population stratification. Genotypic ancestry was defined using the first two PCs and only individuals of European ancestry were included in the genetic analyses.

Supplementary Tables

S-Table 1. Representativeness of the sample. Included and not included subjects' comparisons.

	FEP				CONTROLS			
	EU-GEI Consented (N=1130)	Included (N=802)	Not Included (N=328)	p-value	EU-GEI (N=1497)	Included (N=1263)	Not included (N=234)	p-value
Sex, N (%)				0.637				0.199
Male	697 (61.7)	491 (61.2)	206 (62.8)		706 (49.0)	605 (47.9)	101 (43.2)	
Female	433 (38.3)	311 (38.8)	122 (37.2)		791 (51.)	658 (52.1)	133 (56.8)	
Age, Mean (SD)	30.8 (10.6)	30.5 (10.3)	32.6 (11.1)	**<0.001**	36.1 (12.8) [a]	36.1 (13.0)	35.5 (12.1)	0.469
Self-ascribed ethnicity, N (%)				0.055	[b]			0.273
White	715 (63.3)	523 (65.2)	192 (58.6)		1178 (73.3)	982 (77.8)	196 (83.8)	
Black	183 (16.2)	119 (14.8)	64 (19.5)		121 (11.5)	109 (8.8)	12 (5.1)	
Other ethnicities	232 (20.5)	160 (20)	72 (21.9)		197 (15.2)	172 (13.6)	25 (10.7)	

Missing data on: [a] 2 controls; [b] 1 control;

S-Table 2. Sociodemographic and Clinical Characteristics of the Sample by Cases and Controls.

Variables	Cases	N	Controls	N	test	df	*p-value*
Sex, N (%)		802		1263	34.9	1	**<0.005**
Male	491 (61.2)		605 (47.9)				
Female	311 (38.8)		658 (52.1)				
Age, Mean (SD)	30.5 (10.3)	802	36.1 (13.0)	1261	-10.3	2061	**<0.005**
Self-ascribed ethnicity, N (%)		802		1263	35.6	1	**<0.005**
White	523 (65.2)		982 (77.8)				
Black	119 (14.8)		109 (8.8)				
Other ethnicities	160 (20)		172 (13.6)				
Genetic ancestry							
European	**491 (78.8)**	**623**	**836 (83.6)**	**1000**	**5.9**	**1**	**0.015**
Education*, N (%)		798		1258	191.5	1	**<0.005**
No qualification	113 (14.2)		46 (3.7)				
Compulsory edu.	207 (25.9)		171 (13.6)				
1st level/job related edu.	345 (43.2)		549 (43.6)				
University/Post-graduate	133 (16.7)		492 (39.1)				
Occupation*, N (%)		787		1253	205.9	1	**<0.005**
Unemployed	450 (57.2)		320 (25.5)				

Student/Employed	337 (42.8)		933 (74.5)				
Relationship*, N (%)		799		1260	255.8	1	**<0.005**
Married/Steady rel.	233 (29.2)		810 (64.3)				
Separated/widowed	49 (6.1)		67 (5.3)				
Single	517 (64.7)		383 (30.4)				
Living Status*, N (%)		794		1253	232.7	1	**<0.005**
Partner/Friend/Child	218 (27.5)		751 (59.9)				
Alone	113 (14.2)		161 (12.8)				
Parents/Other	463 (58.3)		341 (27.2)				
Cannabis Use, N (%)		802		1263	60.5	1	**<0.005**
Yes	530 (66.1)		614 (51.4)				
No	272 (48.6)		649 (48.6)				
Current Use, N (%)		529		613	13.9	1	**<0.005**
Yes	175 (33.1)		142 (23.2)				
No	354 (66.9)		471 (76.8)				
Age at First Use**, (mean, SD)	16.7 (4.1)	527	17.8 (4.4)	614	-4.3	1139	**<0.005**
Frequency of use**, N (%)		796		1262	138.1	1	**<0.005**
Everyday	246 (30.9)		88 (7.0)				
Occasionally	278 (34.9)		525 (41.6)				
Never	272 (34.2)		649 (51.4)				
Type used**, N (%)		486		548	14.5	1	**<0.005**
Low-THC cannabis	211 (43.4)		303 (43.4)				
High-THC cannabis	275 (56.6)		245 (56.6)				

Legend: SD=standard deviation; df=degree of freedom; edu.=education; rel.=relationship. * aggregation of categories was made through linear regression.

Patients comprised more men, were younger, and were more frequently self-ascribed to a minority ethnic group than controls. Compared to controls, they were also less educated, more likely to be unemployed, single, and live with their parents or other families. They were also more likely to have used cannabis in their lifetime, with more THC, daily, to have started cannabis earlier in their life, and to currently use cannabis, compared to controls.

S-Table 3. Mean IQ, PSF, and PAF Scores in the Four Patients' Clusters and Controls – descriptives, unadjusted

Variable	GROUP	Mean (SD)	SE	95% CI Lower	95% CI Upper
IQ	CONTROLS	102.6 (17.6)	0.45	101.7	103.5
	HIGH	106.1 (14.2)	0.90	104.3	107.9
	INTERMEDIATE	80.8 (11.9)	0.86	79.1	82.5
	DETERIORATING	80.6 (12.9)	1.06	78.5	82.7
	LOW	73.9 (12.7)	0.86	72.2	75.7
PSF<12	CONTROLS	0.16 (0.89)	0.02	0.11	0.21
	HIGH	-0.48 (0.96)	0.07	-0.62	-0.33
	INTERMEDIATE	-0.44 (1.3)	0.07	-0.58	-0.30
	DETERIORATING	0.19 (0.77)	0.08	0.02	0.36
	LOW	-0.003 (1.03)	0.07	-0.14	0.13
PAF<12	CONTROLS	0.23 (0.85)	0.02	0.19	0.28
	HIGH	0.52 (0.64)	0.05	0.42	0.62
	INTERMEDIATE	-0.16 (0.68)	0.04	-0.25	-0.07
	DETERIORATING	0.23 (0.63)	0.05	0.12	0.35
	LOW	-1.51 (0.82)	0.04	-1.60	-1.41
PSF 12-16	CONTROLS	0.18 (0.85)	0.02	0.13	0.24
	HIGH	-0.23 (0.95)	0.06	-0.36	-0.10
	INTERMEDIATE	-0.19 (1.11)	0.06	-0.32	-0.07
	DETERIORATING	-0.52 (1.28)	0.07	-0.68	-0.37
	LOW	-0.13 (1.12)	0.06	-0.25	-0.004
PAF 12-16	CONTROLS	0.25 (0.88)	0.02	0.20	0.30
	HIGH	0.39 (0.72)	0.05	0.29	0.49
	INTERMEDIATE	-0.03 (0.85)	0.05	-0.14	0.07
	DETERIORATING	-0.78 (0.98)	0.08	-0.93	-0.62
	LOW	-1.01 (0.82)	0.05	-1.1	-0.90

Legend: HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*

S-Table 4. Within-groups pairwise comparisons to test significance of the change in PAS over time

	Mean Difference	SE	Sig.[c]	95% CI for Difference[c] Lower	Upper
CONTROLS					
(PSF 12-16) – (PSF<12)	0.032	0.020	0.109	-0.007	0.071
(PAF 12-16) – (PAF<12)	0.005	0.020	0.791	-0.034	0.044
HIGH					
(PSF 12-16) – (PSF<12)	0.240	0.049	**<0.001**	0.144	0.336
(PAF 12-16) – (PAF<12)	-0.108	0.049	**0.028**	-0.204	-0.012
INTERMEDIATE					
(PSF 12-16) – (PSF<12)	0.244	0.047	**<0.001**	0.152	0.336
(PAF 12-16) – (PAF<12)	0.130	0.047	**0.005**	0.039	0.222
DETERIORATING					
(PSF 12-16) – (PSF<12)	-0.741	0.058	**<0.001**	-0.854	-0.627
(PAF 12-16) – (PAF<12)	-0.968	0.058	**<0.001**	-1.081	-0.854
LOW					
(PSF 12-16) – (PSF<12)	-0.134	0.048	**0.005**	-0.228	-0.041
(PAF 12-16) – (PAF<12)	0.510	0.047	**<0.001**	0.417	0.603

Legend: Based on estimated marginal means. c. Adjustment for multiple comparisons: Bonferroni. Analysis accounting for sex, age, country, and self-ascribed ethnicity. HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*

S-Table 5. *Sociodemographic Characteristics by Patients' Clusters*

	HIGH	INTERMEDIATE	DETERIORATING	LOW	F or χ^2 (df)	p
Country[a], N (%)					71.7 (15)	**<0.001**
UK	45 (22)	31 (13.8)	39 (26)	34 (15.2)		
Holland	52 (25.4)	49 (21.9)	47 (31.3)	20 (9)		
Spain	40 (19.5)	52 (23.2)	21 (14)	64 (28.7)		
France	7 (3.4)	16 (7.1)	12 (8)	30 (13.5)		
Italy	26 (12.7)	22 (9.8)	18 (12)	27 (12.1)		
Brazil	35 (17.1)	54 (24.1)	13 (8.7)	48 (21.5)		
Sex, N (%)					6.6 (3)	0.084
Male	119 (58)	129 (57.6)	91 (60.7)	152 (68.2)		
Female	86 (42)	95 (42.4)	59 (39.3)	71 (3.8)		
Age, mean (sd)	31.3 (9.8)	32.4 (10.5)	28.2 (9.6)	29.3 (10.5)	6.5 (3)	**<0.001**
Self-ascribed ethnicity, N (%)					28.3 (6)	**<0.001**
White	162 (79)	144 (64.3)	85 (56.7)	132 (59.2)		
Black	14 (6.8)	32 (14.3)	30 (20)	43 (19.3)		
Other	29 (14.1)	48 (21.4)	35 (23.3)	48 (21.5)		
Genetic ancestry*, N (%)						
European	139 (85.3)	142 (81.1)	77 (70%)	133 (76)	10.5 (3)	**0.014**
Education, N (%)					109.7 (9)	**<0.001**
No qualification	8 (3.9)	32 (14.4)	24 (16)	49 (22)		
Compulsory edu.	31 (15.3)	62 (27.9)	44 (29.3)	70 (31.4)		
1st level/job related edu.	90 (44.3)	93 (41.9)	70 (46.7)	92 (41.3)		
University/Post-graduate	74 (36.5)	35 (15.8)	12 (8)	12 (5.4)		
Occupation, N (%)					18.9 (3)	**<0.001**
Unemployed	92 (45.8)	122 (55.5)	95 (64.2)	141 (64.7)		
Student/Employed	109 (54.2)	98 (45.5)	53 (35.8)	77 (35.3)		
Relationship, N (%)					7.2 (6)	0.302
Married/Steady rel.	67 (32.8)	67 (29.9)	42 (28)	57 (25.8)		
Separated/widowed	13 (6.4)	17 (7.6)	4 (2.7)	15 (6.8)		
Single	124 (60.8)	140 (62.5)	104 (69.3)	149 (67.4)		
Living Status, N (%)					9.4 (6)	0.147
Partner/Friend/Child	64 (31.4)	67 (30.2)	33 (22.1)	54 (24.7)		
Alone	34 (16.7)	32 (14.4)	22 (14.8)	25 (11.4)		
Parents/Other	106 (52)	123 (55.4)	94 (63.1)	140 (63.9)		

*among those who were genotyped. HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*

There was a difference in the country distribution of the subgroup of patients (χ^2(15)=71.7, p<0.001). UK and HOLLAND presented higher proportion of *high-functioning* (UK, N=45, 30.2%; Holland, N=52, 31%) and *deteriorating* patients (UK, N=34, 26.2%; Holand, N=47, 28%), while Italy had higher proportions of *high-functioning* (N=26, 28%) and *low-functioning* (N=27, 29%) patients.

Spain mostly presented *intermediate* (N=52, 29.4%) and *low-functioning* patients (N=64, 36.2%). Brazil had the highest proportion of *intermediate* patients (N=54, 36%). France presented the highest proportion of *low-functioning* patients (N=30, 46.2%).

S-Table 6. Comparisons in Patterns of Cannabis Use and other Clinical Characteristics by Patients' Clusters

	HIGH	INTERMEDIATE	DETERIORATING	LOW	F or χ^2 (df)	p
Cannabis Use, N (%)					10.2 (2)	**0.006**[a,#]
Yes	143 (69.8)	133 (59.4)	111 (74)	143 (64.1)		
No	62 (30.2)	91 (40.6)	39 (26)	80 (35.9)		
Current Use, N (%)						
Yes	38 (26.8)	46 (34.6)	35 (31.5)	56 (39.2)	3.9 (1)	**0.046**[a]
No	104 (73.2)	87 (65.4)	76 (68.5)	87 (60.8)		
Age First Use**, M (SD)	16.7 (3.1)	17.3 (4.2)	16.3 (3.8)	16.3 (5.2)	1.9 (3)	0.124[a]
Median (IQR)	16 (13-19)	16 (13-19)	15 (11-19)	16 (12-20)		
Frequency of use**, N (%)					10.7 (3)	**0.013**[a,#]
Everyday	60 (29.7)	54 (24.1)	55 (36.9)	77 (34.8)		
Occasionally	80 (39.6)	79 (35.3)	55 (36.9)	64 (29)		
No use	62 (30.7)	91 (40.6)	39 (26.2)	80 (36.2)		
Type used**, N (%)					0.7 (1)	0.401[a]
Low-THC cannabis	55 (43)	53 (43.4)	35 (34.7)	68 (50.4)		
High-THC cannabis	73 (57)	69 (56.5)	66 (65.3)	67 (49.6)		
Frequency*potency					12.1 (6)	0.058[a]
No use	62 (30.7)	91 (40.6)	39 (26.2)	80 (36.2)		
Daily use of high potency	41 (20.3)	33 (14.7)	37 (24.8)	42 (19)		
Any other pattern of use	99 (49)	100 (44.6)	73 (49)	99 (44.8)		
DUP in weeks	52.9 (142.2)	52.7 (115.9)	78.5 (262.4)	65.9 (188.4)	0.76 (3)	0.516
AP treatment						
AP free	119 (59.2)	149 (68.7)	78 (52.7)	139 (63.5)	13.1 (6)	**0.041**
1 AP	53 (26.4)	37 (17.1)	37 (25)	46 (21)		
more than 1 A	29 (14.4)	31 (14.3)	33 (22.3)	34 (14.5)		
ICD-10 Diagnoses					38 (39)	0.515
Bipolar Affective disorder	6 (3)	3 (1.4)	5 (3.4)	4 (1.8)		
Delusional disorder	9 (4.5)	7 (3.2)	7 (4.7)	10 (4.6)		
Hypomanic disorder	4 (2)	2 (0.9)	3 (2)	3 (1.4)		
Mania with psychosis	23 (11.4)	11 (5.1)	10 (6.8)	16 (7.3)		
Manic disorder	6 (3)	4 (1.8)	0 (0)	4 (1.8)		
Mild depression disorder	7 (3.5)	6 (2.8)	3 (2)	7 (3.2)		
Moderate depression disorder	4 (2)	6 (2.8)	2 (1.4)	3 (1.4)		
Moderate depression with somatic syndrome	4 (2)	8 (3.7)	4 (2.7)	9 (4.1)		
Other non-organic	64 (31.8)	93 (42.9)	55 (37.2)	71 (32.4)		

psychotic syndrome						
Schizoaffective disorder, bipolar type	2 (1)	0 (0)	0 (0)	0 (0)		
Schizoaffective disorder, depressed type	3 (1.5)	3 (1.4)	4 (2.7)	2 (0.9)		
Schizoaffective disorder, manic type	3 (1.5)	1 (0.5)	3 (2)	3 (1.4)		
Schizophrenia	62 (30.8)	64 (29.5)	45 (30.4)	82 (37.4)		
Severe depression with psychotic symptoms	4 (2)	9 (4.1)	7 (4.7)	5 (2.3)		
ICD-10 Categories						
Non-affective	143 (71.1)	168 (77.4)	114 (77)	168 (76.7)	2.8 (3)	0.414
Affective	58 (28.9)	49 (22.6)	34 (23)	51 (23.3)		

Legend: HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*.

a Bonferroni-adjusted and further corrected using the Benjamini-Hochberg (B-H) procedure, pre-determining a 5% chance of a false discovery rate.

Significant after Benjamini-Hochberg (B-H) procedure.

High-functioning (N=143, 69.8%) and *deteriorating* (N=111, 74%) groups had a higher proportion of cannabis lifetime users within their groups compared with the other two groups ($\chi^2(2)$=10.2, p=0.006). The *high-functioning* group had a higher proportion of occasional users (N=80, 39.6%) compared to the other groups ($\chi^2(3)$=10.7, p=0.013). The *deteriorating* group included the highest proportion of daily users of high potency cannabis (N=37, 24.8%); though, no significant difference between groups was detected (12.1(6), p=0.058). There were no differences in terms of DUP (p=0.516).

S-Table 7. IQ Subtests Differences Between Patients's Clusters and Controls

Dependent Variable	Statistic	*p*	Contrast
Processing Speed	F (4, 2047)=195.9	<0.001	C > H > I = D = L
Perceptual Reasoning	F (4, 2047)=113.1	<0.001	H = C > I = D = L
Working Memory	F (4, 2047)=106.7	<0.001	H = C > I = D > L
Verbal Comprehension	F (4, 2047)=87.7	<0.001	H > C > I = D > L

C=CONTROLS; H=HIGH COGNITIVE-FUNCTIONING; I=INTERMEDIATE; D=DETERIORATING; L=LOW COGNITIVE-FUNCTIONING. Adjusted by age, sex, country, and self-ascribed ethnicity.

Processing speed was impaired in all subgroups of patients, as compared to controls (F(4,2047)=195.9, p<0.001). *High-functioning* patients performed better than controls in verbal comprehension (M_{diff}=1.2, 95% CI 0.49, 1.9, p<0.001). The *deteriorating* and *intermediate* groups performed identically in all subtests. The Low group was the worst-performing in working memory (F(4,2047)=113.1, p<0.001) and verbal comprehension (F(4,2047)=87.7, p<0.001) (S-Table 6).

S-Table 8. Comparisons by Standardized Symptom Dimensions

	High	*Intermediate*	*Deteriorating*	*Low*	F (3,778)	*p*	Comparisons
General	0.358 (0.06)	0.3 (0.06)	0.157 (0.07)	0.323 (0.05)	1.875	0.132	H>D*
Positive	-0.006 (0.08)	0.206 (0.07)	0.084 (0.09)	0.237 (0.07)	2.163	0.091	L=I>H*
Negative	-0.087 (0.08)	0.147 (0.07)	0.277 (0.08)	0.105 (0.07)	4.128	**0.006**	L=I=D>H**
Disorganization	-0.111 (0.07)	0.063 (0.07)	-0.051 (0.08)	-0.009 (0.06)	1.199	0.309	I=H=D=L
Mania	0.323 (0.08)	0.128 (0.07)	0.043 (0.08)	0.251 (0.07)	2.848	**0.037**	H=L>D=I[#]
Depressive	0.109 (0.07)	0.259 (0.06)	0.431 (0.08)	0.027 (0.06)	6.186	**0.0003**	D=I>L=H[##]

Legend: estimated marginal means and standard errors from Multivariate GLM by clusters are provided for Z-Scores of symptom dimensions. Adjusted by age, sex, country, and self-ascribed ethnicity. Bonferroni corrected. *differences between these groups are significant. Nonetheless, they cannot be considered because the general F-test for that dimension is not significant. ** L>H p=0.053. [#]L>D=0.052. [##] I=H (p=0.195). HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*.

High-functioning patients had the highest scores of manic symptoms (F(3,778)=2.848, p=0.037), with a greater differentiation from *deteriorating* (M_{diff}=0.3, 95% CI 0.06, 0.49, p=0.010) and *intermediate* patients (M_{diff}=0.2, 95% CI 0.004, 0.38, p=0.046). They had also the fewest positive and negative symptoms, these last lower than *deteriorating* patients (M_{diff}=-0.36, 95% CI -0.64, -0.07, p=0.004). *Deteriorating* patients had also more depressive symptoms than both *high-functioning* (M_{diff}=0.32, 95% CI 0.05, 0.59, p=0.011) and *low-functioning* patients (M_{diff}=0.4, 95% CI 0.13, 0.67, p=0.001). The *low-functioning* cluster presented less depressive symptoms than the *intermediate* (M_{diff}=-0.2, 95% CI -0.05, -0.14, p=0.011) (S-Tables 8, 9).

S-Table 9. Pairwise Comparisons of Estimated Marginal Means From Multivariate GLM of Symptom Dimensions by Clusters

Dependent variable	(I) CLUSTERS	(J) CLUSTERS	Mean DIfference (I-J)	SE.	*p*	95% CI for Mean Differences	
GENERAL	HIGH	INTERMEDIATE	0.058	0.080	0.471	-0.100	0.216
		DETERIORATING	0.202*	0.089	**0.024**	0.027	0.377
		LOW	0.035	0.082	0.670	-0.127	0.197
	INTERMEDIATE	DETERIORATING	0.144	0.088	0.104	-0.030	0.317
		LOW	-0.023	0.079	0.773	-0.178	0.133
	DETERIORATING	LOW	-0.167	0.089	0.061	-0.341	0.008
POSITIVE	HIGH	INTERMEDIATE	-0.212*	0.104	**0.043**	-0.417	-0.007
		DETERIORATING	-0.090	0.116	0.438	-0.318	0.138
		LOW	-0.243*	0.107	**0.024**	-0.453	-0.032
	INTERMEDIATE	DETERIORATING	0.122	0.115	0.290	-0.104	0.347
		LOW	-0.031	0.103	0.765	-0.233	0.171
	DETERIORATING	LOW	-0.152	0.115	0.187	-0.379	0.074
NEGATIVE	HIGH	INTERMEDIATE	-0.234*	0.097	**0.016**	-0.424	-0.044
		DETERIORATING	-0.363*	0.107	**0.001**	-0.574	-0.153
		LOW	-0.192	0.099	0.053	-0.387	0.002
	INTERMEDIATE	DETERIORATING	-0.130	0.106	0.223	-0.338	0.079
		LOW	0.042	0.095	0.663	-0.146	0.229
	DETERIORATING	LOW	0.171	0.107	0.110	-0.039	0.381
DISORGANISATION	HIGH	INTERMEDIATE	-0.174	0.094	0.063	-0.357	0.010
		DETERIORATING	-0.060	0.104	0.566	-0.263	0.144
		LOW	-0.101	0.096	0.291	-0.289	0.087
	INTERMEDIATE	DETERIORATING	0.114	0.103	0.267	-0.087	0.316
		LOW	0.073	0.092	0.431	-0.108	0.254
	DETERIORATING	LOW	-0.042	0.103	0.688	-0.245	0.161
MANIA	HIGH	INTERMEDIATE	0.194*	0.097	**0.046**	0.004	0.385
		DETERIORATING	0.280*	0.108	**0.010**	0.068	0.491
		LOW	0.071	0.099	0.474	-0.124	0.266
	INTERMEDIATE	DETERIORATING	0.086	0.107	0.423	-0.124	0.295
		LOW	-0.123	0.096	0.199	-0.311	0.065
	DETERIORATING	LOW	-0.208	0.107	0.052	-0.419	0.002
DEPRESSION	HIGH	INTERMEDIATE	-0.150	0.093	0.105	-0.332	0.032
		DETERIORATING	-0.322*	0.103	**0.002**	-0.524	-0.121
		LOW	0.082	0.095	0.389	-0.104	0.268
	INTERMEDIATE	DETERIORATING	-0.172	0.102	0.091	-0.372	0.027
		LOW	0.232*	0.091	**0.011**	0.053	0.411
	DETERIORATING	LOW	0.404*	0.102	**0.000**	0.203	0.605

Legend: HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*

S-Table 10. Bivariate Correlations Between PRSs

		SCZ_PRS	BD_PRS	MDD_PRS	CUD_PRS	IQ_PRS
SCZ_PRS	*r*	1	0.521**	0.289**	0.101**	-0.104**
	p-value		<0.001	<0.001	<0.001	<0.001
BD_PRS	*r*	0.521**	1	0.273**	0.066*	-0.011
	p-value	<0.001		<0.001	0.017	0.697
MDD_PRS	*r*	0.289**	0.273**	1	0.066*	-0.101**
	p-value	<0.001	<0.001		0.016	<0.001
CUD_PRS	*r*	0.101**	0.066*	0.066*	1	-0.037
	p-value	<0.001	0.017	0.016		0.181
IQ_PRS	*r*	-0.104**	-0.011	-0.101**	-0.037	1
	p-value	<0.001	0.697	<0.001	0.181	

** Correlation is significant at the 0.01 level (2-tailed). Unajusted.

* Correlation is significant at the 0.05 level (2-tailed). Unajusted.

S-Table 11. Bivariate Correlations Between PRSs Tested and Measures Used in the Clustering Solution

		PSF<12	PSF 12-16	PAF<12	PAF 12-16	IQ
SCZ_PRS	*r*	0.035	0.059*	-0.139**	-0.074**	-0.228**
	p-value	0.2	0.031	<0.001	0.007	<0.001
BD_PRS	*r*	0.034	0.049	-0.089**	-0.057*	-0.212**
	p-value	0.219	0.075	0,001	0.039	<0.001
MDD_PRS	*r*	-0.028	-0.025	-0.129**	-0.086**	-0.148**
	p-value	0.304	0.36	<0.001	0.002	<0.001
IQ_PRS	*r*	-0.059*	-0.026	0.150**	0.093**	0.161**
	p-value	0.033	0.35	<0.001	0.001	<0.001
CUD_PRS	*r*	-0.019	0.006	-0.093**	-0.093**	-0.110**
	p-value	0.488	0.837	0.001	0.001	<0.001

** Correlation is significant at the 0.01 level (2-tailed). Unajusted.

S-Table 12. Pairwise Comparisons of PRSs between groups[a]

Dependent variable	(I) CLUSTERS	(J) CLUSTERS	Mean DIfference (I-J)	SE.	*p*	95% CI for Mean Difference[ac]	
SCZ	CONTROLS	HIGH	-0.534*	0.088	**0.000**	-0.781	-0.287
		INTERMEDIATE	-0.478*	0.087	**0.000**	-0.723	-0.233
		DETERIORATING	-0.364*	0.114	**0.015**	-0.685	-0.042
		LOW	-0.537*	0.091	**0.000**	-0.792	-0.283
	HIGH	INTERMEDIATE	0.056	0.114	1.000	-0.264	0.377
		DETERIORATING	0.171	0.136	1.000	-0.211	0.552
		LOW	-0.003	0.117	1.000	-0.333	0.326
	INTERMEDIATE	DETERIORATING	0.114	0.135	1.000	-0.265	0.494
		LOW	-0.060	0.115	1.000	-0.384	0.265
	DETERIORATING	LOW	-0.174	0.137	1.000	-0.560	0.212

BD	CONTROLS	HIGH	-0.328*	0.088	**0.002**	-0.575	-0.081
		INTERMEDIATE	-0.439*	0.087	**0.000**	-0.684	-0.193
		DETERIORATING	-0.315	0.115	0.060	-0.637	0.007
		LOW	-0.206	0.091	0.237	-0.461	0.050
	HIGH	INTERMEDIATE	-0.111	0.114	1.000	-0.432	0.210
		DETERIORATING	0.013	0.136	1.000	-0.369	0.395
		LOW	0.123	0.117	1.000	-0.207	0.452
	INTERMEDIATE	DETERIORATING	0.124	0.135	1.000	-0.256	0.504
		LOW	0.233	0.116	0.435	-0.091	0.558
	DETERIORATING	LOW	0.110	0.138	1.000	-0.277	0.497
MD	CONTROLS	HIGH	-0.111	0.088	1.000	-0.359	0.138
		INTERMEDIATE	-0.205	0.088	0.197	-0.452	0.042
		DETERIORATING	-0.175	0.115	1.000	-0.498	0.149
		LOW	-0.380*	0.091	**0.000**	-0.637	-0.124
	HIGH	INTERMEDIATE	-0.094	0.115	1.000	-0.416	0.228
		DETERIORATING	-0.064	0.136	1.000	-0.447	0.320
		LOW	-0.270	0.118	0.223	-0.601	0.062
	INTERMEDIATE	DETERIORATING	0.030	0.136	1.000	-0.351	0.412
		LOW	-0.175	0.116	1.000	-0.502	0.151
	DETERIORATING	LOW	-0.206	0.138	1.000	-0.594	0.183
IQ	CONTROLS	HIGH	0.055	0.091	1.000	-0.200	0.309
		INTERMEDIATE	0.210	0.090	0.195	-0.043	0.463
		DETERIORATING	0.133	0.118	1.000	-.199	0.465
		LOW	0.347*	0.093	**0.002**	0.084	0.610
	HIGH	INTERMEDIATE	0.156	0.118	1.000	-0.175	0.486
		DETERIORATING	0.078	0.140	1.000	-0.315	0.472
		LOW	0.292	0.121	0.157	-0.047	0.632
	INTERMEDIATE	DETERIORATING	-0.077	0.139	1.000	-0.469	0.314
		LOW	0.136	0.119	1.000	-0.198	0.471
	DETERIORATING	LOW	0.214	0.142	1.000	-0.185	0.613

Legend: HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*.
Based on estimated marginal means. Dependent variables are adjusted by PCA, age, gendersex, and country *Mean difference is sig. at 0.05. a. Regression of minimum weighted squares (wt1). c. Multiple comparisons adjusted by Bonferroni.

The model, including covariates, the four PRSs and the ten PCs, predicted group membership with a moderate effect size ($\Delta\chi^2$=392.3, df=84, p=2.4^{-41}, R^2=0.284).

The model, including covariates and the four PRSs, predicted patients' cluster membership with a modest effect size (($\Delta\chi^2$=129, df=63; p=0.000002, effect size=0.247). Compared to the covariates-only model (R^2 =0.220), the introduction of PRSs determined a meaningless increase in significance (R^2=0.027), consisting in a 2.7% of the within-clusters variance explained by the PSRs.

S-Table 13. IQ Boundaries and Subgroups

	CONTROLS	HIGH	INTERMEDIATE	DETERIORATING	LOW
IQ (min, max)	102.68 (47, 150)	106.17 (70, 149)	80.89 (45, 128)	80.65 (45, 107)	74 (45, 110)
IQ <70 N (%)	35 (2.7)	0 (0)	33 (14.7)	27 (18)	78 (34.9)

IQ 70-85 N (%)	173 (13.6)	13 (6.3)	116 (51.7)	62 (41.3)	105 (47.2)
IQ 86-115	758 (60)	143 (69.7)	73 (32.5)	62 (41.3)	40 (17.9)
IQ>115	297 (23.5)	49 (23.9)	2 (0.8)	0 (0)	0 (0)

Legend: HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*

S-Table 14. Model Summary of Group Comparisons in Terms of PRSs. Estimated Marginal Means and 95% CIs

PRS	Controls (N=836)	High (N=139)	Intermediate (N=142)	Deteriorating (N=77)	Low (N=133)	η^2	Contrasts	Within-patient statistics* F	p
SZ	-0.172	0.336	0.238	0.182	0.275	0.045	H=L=I=D > C	0.58	0.622
	-0.23, -0.11	0.18, 0.48	0.09, 0.38	-0.02, 0.36	0.12, 0.42				
BD	-0.126	0.210	0.238	0.236	0.051	0.024	I =D=H>C=L	1.27	0.282
	-0.18, -0.06	0.05, 0.36	0.09, 0.38	-0.04, 0.42	-0.09, 0.19				
MD	-0.059	0.021	0.109	0.036	0.230	0.009	L > C=I=D=H	1.51	0.210
	-0.12, -0.03	-0.13, 0.17	-0.03, 0.25	-0.15, 0.22	0.08, 0.37				
IQ	0.066	0.033	-0.138	-0.099	-0.246	0.012	L < C=H=D=I	2.27	0.079
	0.04, 0.12	-0.12, 0.18	-0.28, 0.01	-0.28, 0.08	-0.39, -0.09				

The model is corrected by age, sex, country, and the ten ancestry PC as covariates and weighted by case/control.
High=*high cognitive-functioning*; Low=*low cognitive-functioning*
*statistics are relative to within-patient only comparisons by assuming the *high cognitive-functioning* group as the baseline category. Between-patient contrasts are represented in the "contrasts" column.

S-Table 15. Pairwise Comparisons between controls and Clusters of FEP patients in CUD_PRS (pt 0.005)

(I) CLUSTERS	(J) CLUSTERS	Mean DIfference (I-J)	SE.	p	95% CI for Mean Difference Lower Bound	Upper Bound
CONTROLS	HIGH	$-7.73E^{-05}$	0.000034	0.251	-0.000174	$1.96E^{-05}$
	INTERMEDIATE	$-2.99E^{-05}$	0.000034	1.000	-0.000126	$6.61E^{-05}$
	DETERIORATING	$-6.10E^{-05}$	0.000044	1.000	-0.000186	$6.40E^{-05}$
	LOW	-0.000113*	0.000036	0.017	-0.000215	$-1.20E^{-05}$

Based on estimated marginal means. * The mean difference is significant at the 0,05 level. c Adjustment for multiple comparisons: Bonferroni.

Supplementary Figures

S-Figure 1: Cluster solutions after 1000 unsupervised Two-step cluster analysis with random reordering
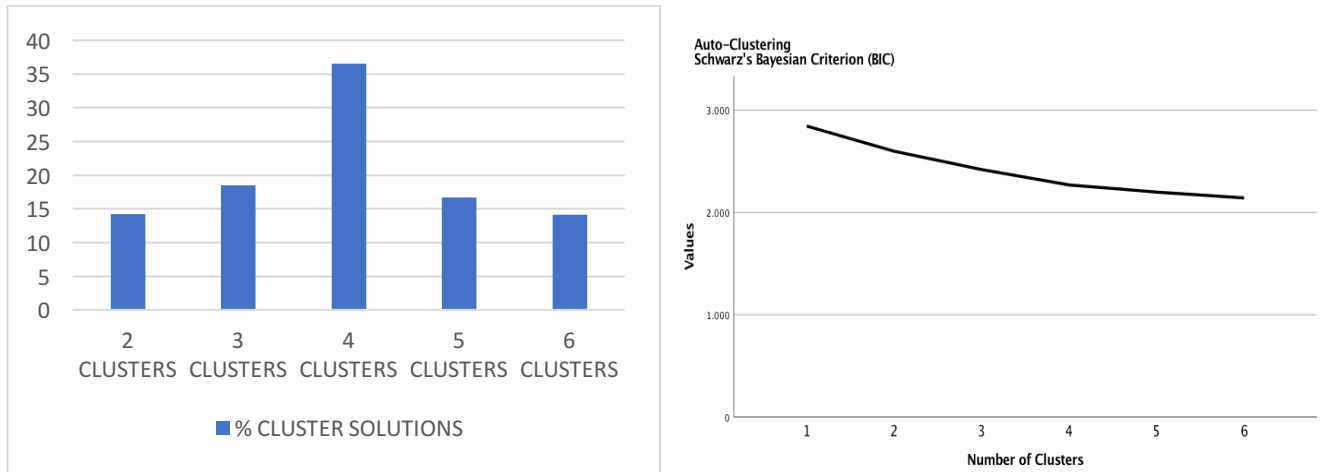
Auto-Clustering Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change[a]	Ratio of BIC Changes[b]	Ratio of Distance Measures[c]
1	2843.891			
2	2598.160	-245.731	1.000	1.270
3	2418.946	-179.213	0.729	1.133
4	**2268.545**	**-150.401**	**0.612**	**1.591**
5	2198.855	-69.690	0.284	1.101
6	2141.638	-57.217	0.233	1.291

a. The changes are from the previous number of clusters in the table.

b. The ratios of changes are relative to the change for the two cluster solution.

c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

S-Figure 2. Three-dimensional Representation of FEP Clusters in terms of premorbid change and IQ
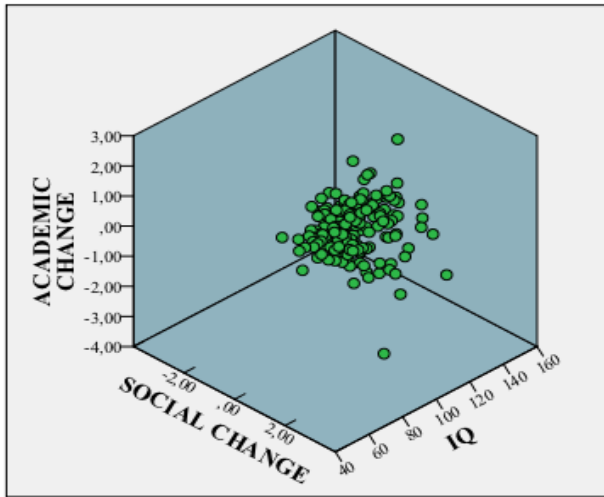
HIGH
n=205
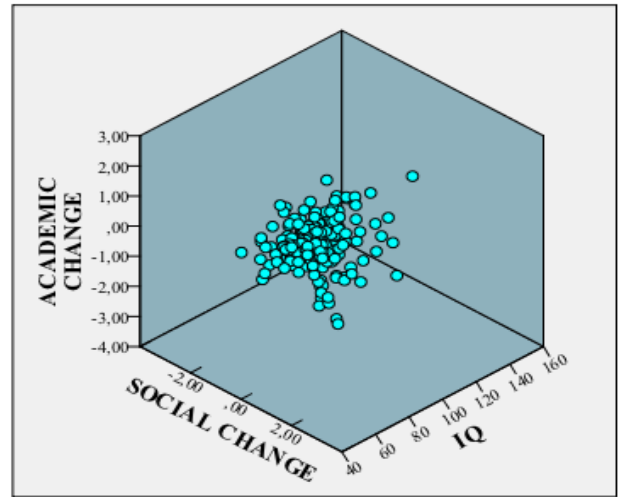
INTERMEDIATE
n=224
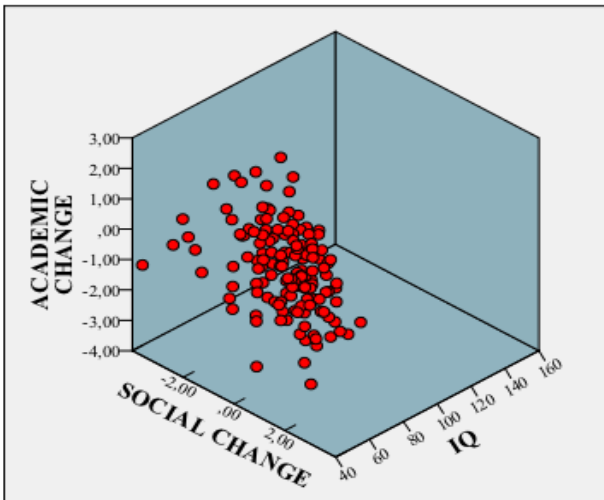
DETERIORATING
n=150
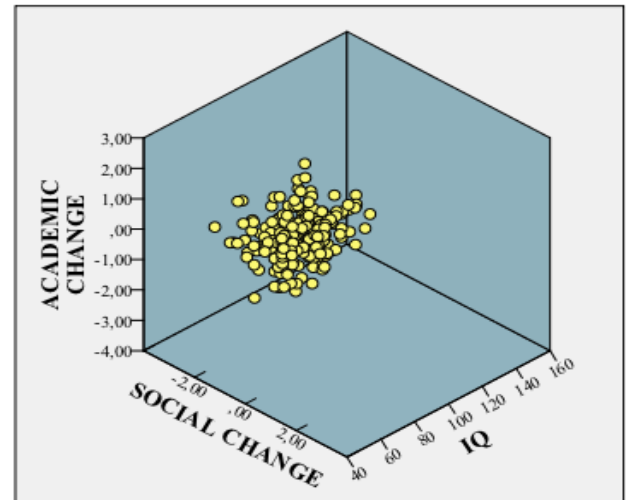
LOW
n=223

Legend: HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*

S-Figure 3. Diagnoses distribution among clusters.
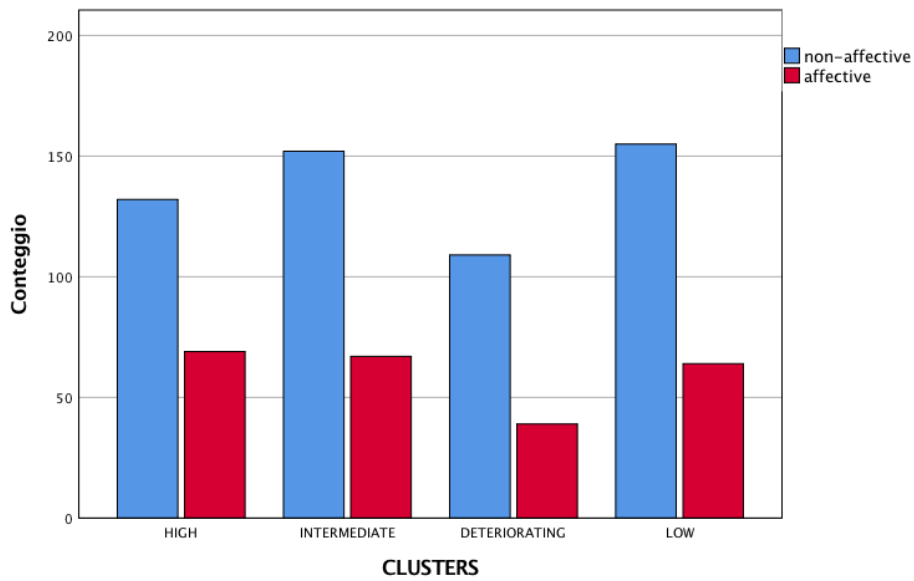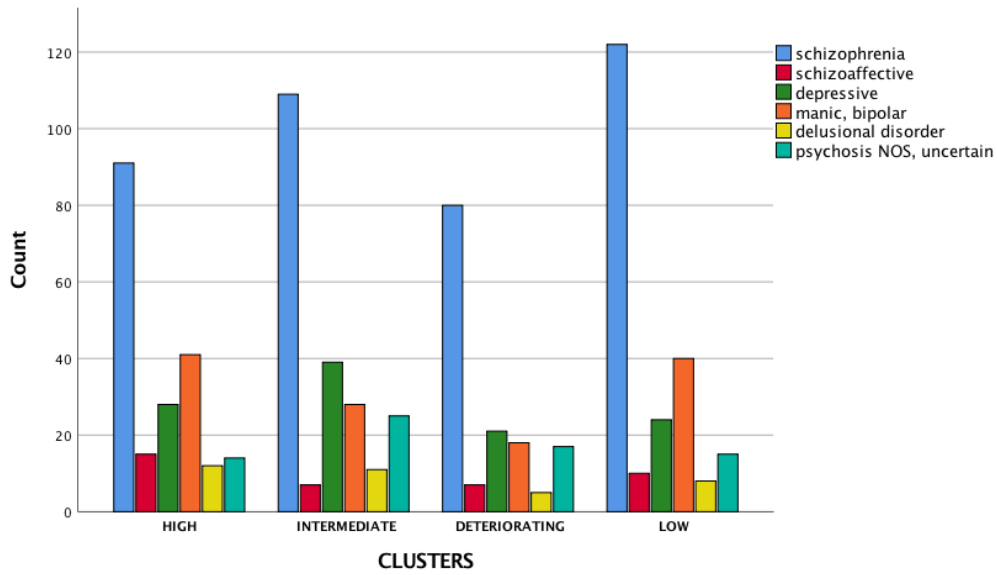
Legend: HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*

S-Figure 4 Polygenic Risk Scores Across Different Clusters of Patients and Controls (all Genetic Ancestries)
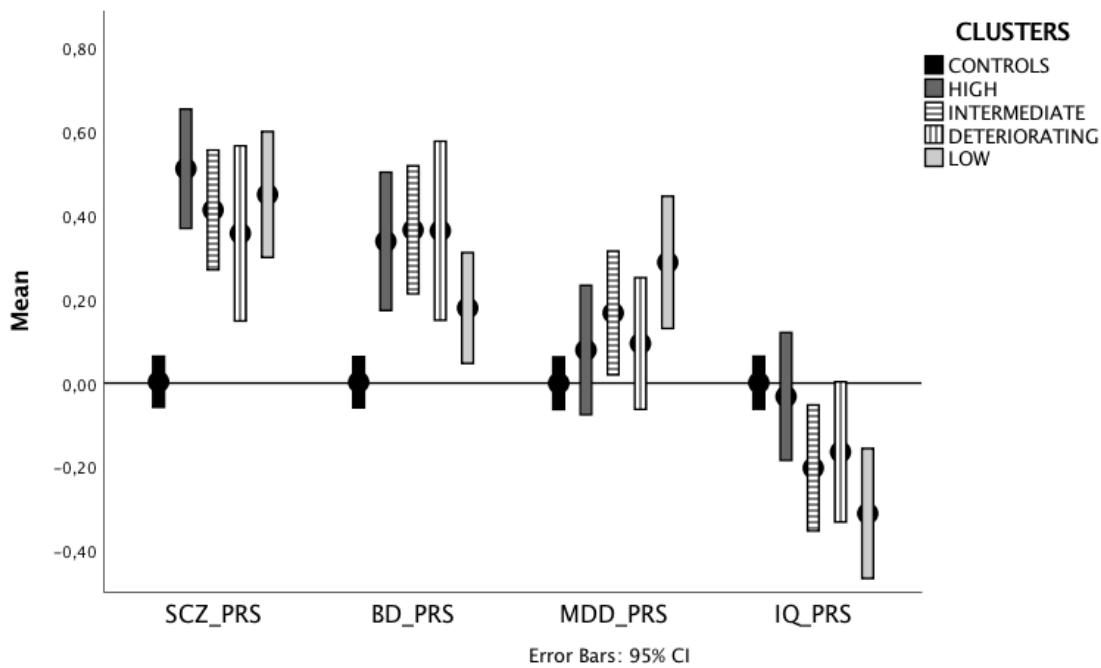
Legend: The Y-axis indicates z-scores for PRSs. Higher standardized scores indicate higher genetic risk for the SCZ_PRS, BD_PRS, and MD_PRS. For IQ_PRS, lower standardized scores indicate a predisposition to worse IQ. AlPRSs were adjusted to account for age, sex, country, and ten PCs and standardized around controls' means and standard deviations. Each controls' PRS represents a mean of 0 and a standard deviation of 1. Error bars represent 95% CIs. HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*.

S-Figure 5 Polygenic Risk Scores Across Different Clusters of Patients and Controls (SCZ_PRS_PT_1.0)
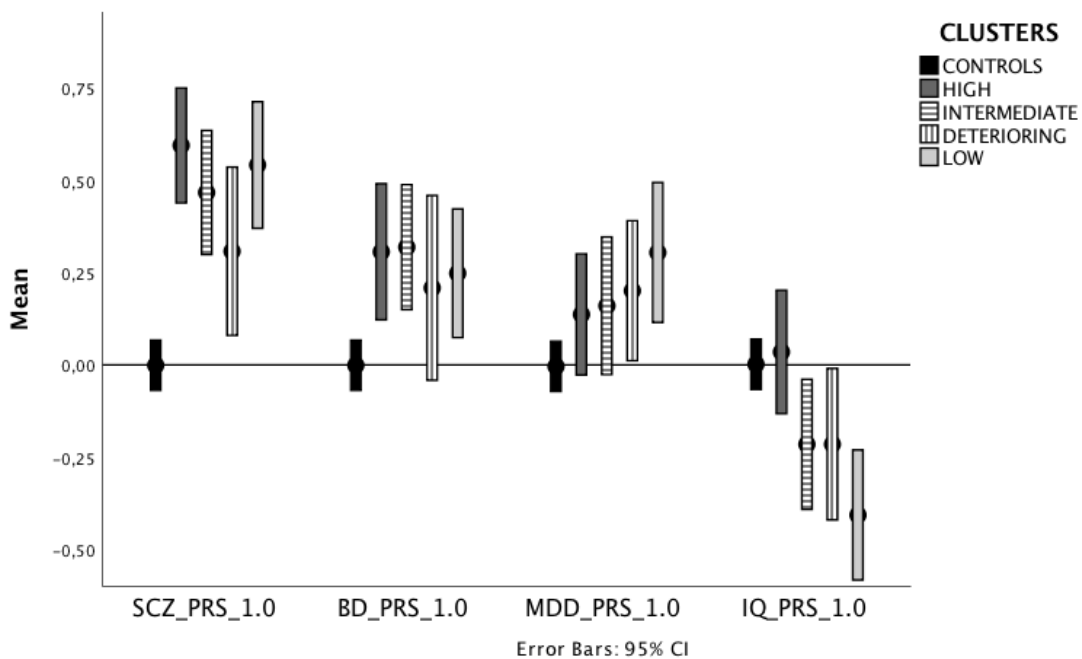
Legend: The Y-axis indicates z-scores for PRSs. Higher standardized scores indicate higher genetic risk for the SCZ_PRS, BD_PRS, and MD_PRS. For IQ_PRS, lower standardized scores indicate a predisposition to worse IQ. All PRSs were adjusted to account for age, sex, country, and ten PCs and standardized around controls' means and standard deviations. Each controls' PRS represents a mean of 0 and a standard deviation of 1. Error bars represent 95% CIs. HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*.

S-Figure 6. CUD Polygenic Risk Score Across Different Clusters of Patients and Controls
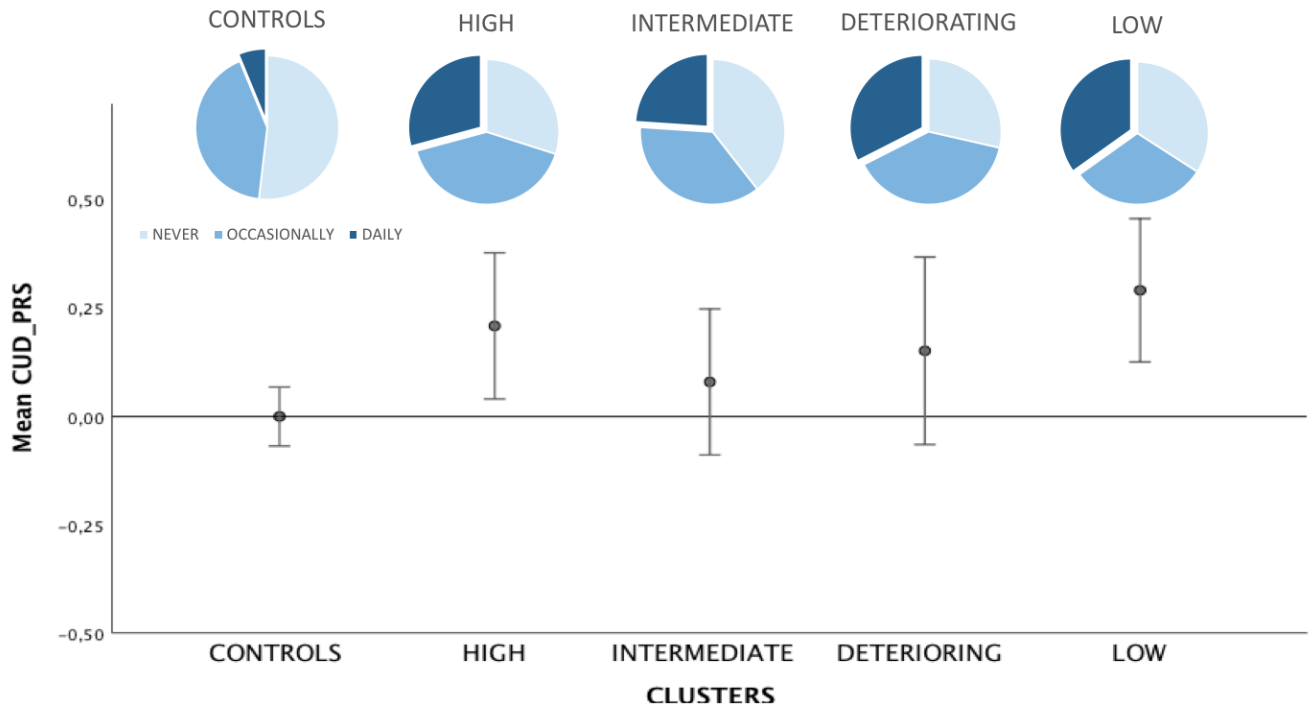
Legend: The upper part of the graph reports the distribution of different patterns of cannabis use in the clusters and in controls. The Y-axis indicates z-scores for CUD PRS. Higher standardized scores indicate higher genetic risk for CUD. PRS was adjusted to account for age, sex, country, and ten PCs and standardized around controls' mean and standard deviation. Controls' PRS represents a mean of 0 and a standard deviation of 1. Error bars represent 95% CIs. HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*.

S-Figure 7. Height Polygenic Risk Score[16] Across Different Clusters of Patients and Controls – CONTROL PRS
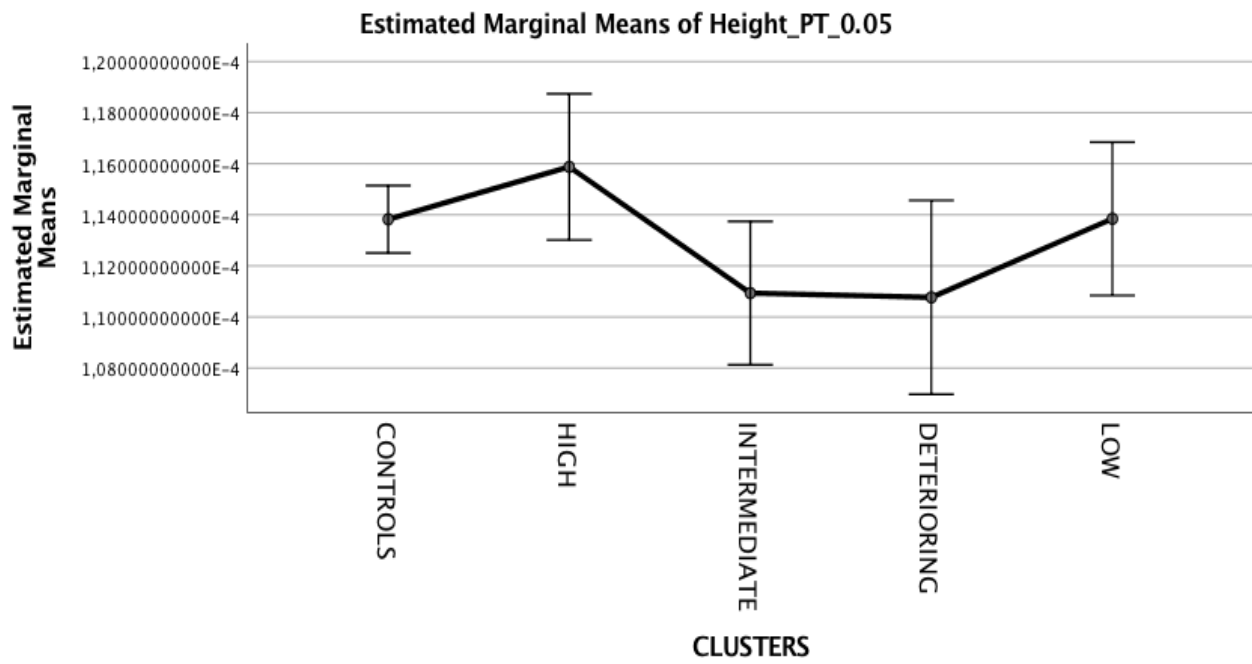
Error bars: 95% CI

Legend: The Y-axis indicates Height_PRS. Higher scores predict higher Height. PRS was adjusted to account for age, sex, country, and ten PCs. Error bars represent 95% CIs. HIGH=*high cognitive-functioning*; LOW=*low cognitive-functioning*. The Heigh_PRS was not related to group membership (F(4,1274)=2.122, p=0.079).

The EU-GEI WP2 GROUP – Acknowledged authors

Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IiSGM (CIBERSAM), Madrid, Spain; **Hubbard, Kathryn** Institute of Psychiatry, Psychology and Neuroscience, King's College London, London; **Jamain, Stéphane** PhD, Univ Paris-Est-Créteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, translational Neuropsychiatry, Fondation FondaMental F-94010 Creteil, France; **Jiménez-López, Estela** MSc, PhD, Universidad de Castilla-La Mancha. Health and Social Research Center, Cuenca, Spain, Department of Psychiatry, Hospital Virgen de la Luz, Cuenca, Spain, CIBERSAM, Madrid, Spain. **Leboyer, Marion** Univ Paris-Est-Créteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, translational Neuropsychiatry, Fondation FondaMental F-94010 Creteil, France; **Llorca, Pierre Michel** EA 7280 Npsydo, Université Clermont Auvergne, Clermont-Ferrand, France **López Montoya, Gonzalo** PhD, Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IiSGM (CIBERSAM), Madrid, Spain; **Lorente-Rovira, Esther** Department of Psychiatry, Hospital Clinico de Valencia, Valencia, Spain and CIBERSAM, Madrid, Spain; **Marcelino Loureiro, Camila** PhD, Departamento de Neurociências e Ciencias do Comportamento, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brasil, and Núcleo de Pesquina em Saúde Mental Populacional, Universidade de São Paulo, São Paulo, Brasil; **Marrazzo, Giovanna** MD, PhD, Unit of Psychiatry, "P. Giaccone" General Hospital, Palermo, Italy; **Martínez Díaz-Caneja, Covadonga MD**, Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, IiSGM, CIBERSAM, School of Medicine, Universidad Complutense, Madrid, Spain. **Messchaert, Elles** MSc, Rivierduinen Centre for Mental Health, Leiden, the Netherlands; **Mezquida, Gisela** Mezquida, Gisela PhD, Centre for Biomedical Research in the Mental Health Network (CIBERSAM), Madrid, Spain; Barcelona Clinic Schizophrenia Unit, Neuroscience Institute, Hospital Clinic of Barcelona, Barcelona, Spain; Department of Clinical Foundations, Pharmacology Unit, University of Barcelona, Barcelona, Spain. **Moltó, Ma Dolores,** Department of Genetics, University of Valencia, Campus of Burjassot, Biomedical Research Institute INCLIVA, Valencia, Spain, Centro de Investigacion Biomedica en Red de Salud Mental (CIBERSAM), Madrid, Spain **Moreno, Carmen** MD, Department of Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IiSGM (CIBERSAM), Madrid, Spain; **Nacher Juan,** Neurobiology Unit, Department of Cell Biology, Interdisciplinary Research Structure for Biotechnology and Biomedicine (BIOTECMED), Universitat de València, Valencia, Spain, Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM): Spanish National Network for Research in Mental Health, Madrid, Spain, Fundación Investigación Hospital Clínico de Valencia, INCLIVA, Valencia, Spain. **Olmeda, Ma Soledad** Department of Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, Madrid, Spain; **Parellada, Mara** MD, PhD, Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IiSGM (CIBERSAM), Madrid, Spain; **Pignon, Baptiste** MD, Univ Paris-Est-Créteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, translational Neuropsychiatry, Fondation FondaMental F-94010 Creteil, France; **Rapado-Castro, Marta,** PhD, Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IiSGM (CIBERSAM), Madrid, Spain; **Richard, Jean-Romain** MSc, Univ Paris-Est-Créteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, translational Neuropsychiatry, Fondation FondaMental F-94010 Creteil, France; **Rodríguez Solano, José Juan** MD, Puente de Vallecas Mental Health Department, Hospital Universitario Infanta Leonor/Hospital Virgen de la Torre, Centro de Salud Mental Puente de Vallecas, Madrid, Spain; **Roldán, Laura** PhD, Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IiSGM (CIBERSAM), Madrid, Spain; Department of Personality, Assessment and Clinical Psychology. School of Psychology, Universidad Complutense de Madrid, Spain. **Rossi Menezes, Paulo** Population Mental Health Center, University

of S.o Paulo, Brazil, Department of Preventive Medicine, Faculty of Medicine, University of S.o Paulo, Brazil; **Ruggeri, Mirella** Professor of Psychiatry, University of Verona – Italy, Director, Institute of Psychiatry, Verona Hospital Trust - AOUI; **Sáiz, Pilar A,** MD, PhD, Department of Psychiatry, School of Medicine, University of Oviedo; Centro de Investigación en Red Salud Mental, CIBERSAM; Instituto de Investigación Sanitaria del Principado de Asturias, ISPA; Instituto de Neurociencias del Principado de Asturias, INEUROPA; Servicio de Salud del Principado de Asturias (SESPA), Oviedo, Spain; **Sánchez-Gutierrez, Teresa** PhD, Faculty of Health Science, Universidad Internacional de La Rioja (UNIR), Madrid, Spain. **Sánchez, Emilio** MD, Department of Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IiSGM (CIBERSAM), Madrid, Spain; **Sanjuàn, Julio** Department of Psychiatry, School of Medicine, Universidad de Valencia, CIBERSAM, Valencia, Spain; **Santos, Jose Luis** Department of Psychiatry, Servicio de Psiquiatr.a Hospital "Virgen de la Luz," Cuenca, Spain; São Paulo, Brasil; **Schürhoff, Franck** MD, PhD, AP-HP, Univ Paris-Est-Créteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, translational Neuropsychiatry, Fondation FondaMental F-94010 Creteil, France; **Seri, Marco** PhD, Department of Medical and Surgical Science, Psychiatry Unit, Alma Mater Studiorum Università di Bologna, Via Massarenti 9, 40138 Bologna, Italy; **Shuhama, Rosana**, PhD, Departamento de Neurociências e Ciencias do Comportamento, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brasil, and Núcleo de Pesquina em **Termorshuizen, Fabian** PhD, Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, South Limburg Mental Health Research and Teaching Network, Maastricht University Medical Centre, Maastricht, the Netherlands, and Riverduinen Centre for Mental Health, Leiden, the Netherlands; **Tronche, Anne-Marie** MD, Fondation Fondamental, Créteil, France, CMP B CHU, Clermont Ferrand; **van Dam, Daniella** PhD, Department of Psychiatry, Early Psychosis Section, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands; **van der Ven, Elsje** PhD, Department of of Clinical, Neuro- & Developmental Psychology, Faculty of Behavioural and Movement Sciences, University of Amsterdam, Amsterdam, the Netherlands;

Supplementary References

1. World Health Organization. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. https://www.who.int/classifications/icd/en/bluebook.pdf. Published 1992. Accessed June 8, 2016.

2. Gayer-Anderson C, Jongsma HE, Di Forti M, et al. The EUropean Network of National Schizophrenia Networks Studying Gene–Environment Interactions (EU-GEI): Incidence and First-Episode Case–Control Programme. *Soc Psychiatry Psychiatr Epidemiol*. 2020;55(5):645-657. doi:10.1007/s00127-020-01831-x

3. Jongsma HE, Gayer-Anderson C, Lasalvia A, et al. Treated Incidence of Psychotic Disorders in the Multinational EU-GEI Study. *JAMA Psychiatry*. 2018;75(1):36. doi:10.1001/jamapsychiatry.2017.3554

4. Mallett R, Leff J, Bhugra D, Pang D, Zhao JH. Social environment, ethnicity and schizophrenia. A case-control study. *Soc Psychiatry Psychiatr Epidemiol*. 2002;37(7):329-335.

5. McGuffin P, Farmer A, Harvey I. A Polydiagnostic Application of Operational Criteria in Studies of Psychotic Illness: Development and Reliability of the OPCRIT System. *Arch Gen Psychiatry*. 1991;48(8):764-770.

6. Di Forti M, Quattrone D, Freeman TP, et al. The contribution of cannabis use to variation in the incidence of psychotic disorder across Europe (EU-GEI): a multicentre case-control study. *The Lancet Psychiatry*. 2019;6(5):427-436. doi:10.1016/S2215-0366(19)30048-3

7. Singh SP, Cooper JE, Fisher HL, et al. Determining the chronology and components of psychosis onset: The Nottingham Onset Schedule (NOS). *Schizophr Res*. 2005;80(1):117-130. doi:10.1016/j.schres.2005.04.018

8. Velthorst E, Levine SZ, Henquet C, et al. To cut a short test even shorter: Reliability and validity of a brief assessment of intellectual ability in Schizophrenia - A control-case family study. *Cogn Neuropsychiatry*. 2013;18(6):574-593. doi:10.1080/13546805.2012.731390

9. Cannon-Spoor HE, Potkin SG, Wyatt RJ. Measurement of Premorbid Adjustment in Chronic Schizophrenia. *Schizophr Bull*. 1982;8(3):470-484.

10. Rabinowitz J, Levine SZ, Brill N, Bromet EJ. The Premorbid Adjustment Scale Structured Interview (PAS-SI): Preliminary findings. *Schizophr Res*. 2007;90(1):255-257. doi:https://doi.org/10.1016/j.schres.2006.10.008

11. Ferraro L, La Cascia C, Quattrone D, et al. Premorbid Adjustment and IQ in Patients with First-Episode Psychosis: A Multisite Case-Control Study of Their Relationship with

Cannabis Use. *Schizophr Bull*. 2020;46(3):517-529. doi:10.1093/schbul/sbz077

12. IBM Corporation. IBM SPSS Statistics for Macintosh. In: New York, NY: Armonk; 2017.

13. Norusis MJ. *IBM SPSS Statistics 19 Statistical Procedures Companion*. NJ: Prentice Hall: Upper Saddle River; 2011.

14. Bacher J, Wenzig K, Vogles M. SPSS Two-Step Cluster: A first evaluation. In: *The RC33 Sixth International Conference on Social Science Methodology*. Amsterdam; 2004.

15. Liu Y, Li QL, Dong LY, Wen BC. Combination clustering analysis method and its application. *J Appl Sci*. 2013;13(8):1251-1255. doi:10.3923/JAS.2013.1251.1255

16. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27(20):3641-3649. doi:10.1093/hmg/ddy271