Table of contents

Sample characteristics

Characteristics of the sample are summarized in the table below (Table S1).

Table S1. Sample characteristics.

	N(%) or M(SD)
Country	
Brazil	302 (20.2%)
France	147 (9.8%)
Holland	210 (14.0%)
Italy	280 (18.7%)
Spain	222 (14.8%)
United Kingdom	336 (22.4%)
Sex	
Males	706 (52.8%)
Females	791 (47.2%)
Age	36.1 (12.9)
Migrant status	
Non-migrant	1,205 (80.5%)
Migrant	292 (19.5%)
Ethnicity	
Asian	33 (2.2%)
Black	121 (8.1%)
Mixed	116 (7.7%)
North-African	24 (1.6%)
Other	24 (1.6%)
White	1,179 (78.8%)
Education	
Higher	554 (37.0%)
School/college/vocational	871 (58.2%)
No qualification	72 (4.8%)
Relational status	
Other	1,014 (67.7%)
Single	483 (32.3%)
Employment	
Other	1,285 (85.8%)
Unemployed	212 (14.2%)
Current use of cannabis	
No	1,337 (89.3%)
Yes	160 (10.7%)
Childhood trauma	6.9 (2.2)
Bullying	
Never	1,079 (72.1%)
Ever	418 (27.9%)

Missing data

The proportion of participants with missing data was generally low, ranging from none on sex to 173 (11.6%) on one SIS-R item. Complete data were available for 1,143 participants (76.4%). Missing data patterns were inspected visually using the R packages "dlookr" and "visdat"[1,2]. Missing data were assumed to be missing at random. Thus, we conducted an imputation of the missing values the R package "missRanger"[3], which is largely based on the algorithm of "missForest"[4], an iterative imputation approach based on random forest (RF). This approach has been widely used and has been recognized to produce low imputation errors and to outperform other popular imputation methods, such as MICE[5]. The parameter num.trees, which defaults at 100, was set at 5,000. Following imputation, out-of-bag errors were computed for each variable as a measure of imputation accuracy. Out-of-bag errors range from 0 to 1, with values closer to 0 indicating a better performance. Errors were generally low (<0.10), except for the CTQ subscale of sexual abuse (NMRSE=0.26).

Multidimensional Item Response Theory models

Data from SIS-R and CAPE were analysed using Multidimensional Item Response Theory (MIRT). To ensure enough covariance coverage for item response modelling, only variables with at least 10% of valid frequency were used.

SIS-R. Controls recruited in Verona were not administered SIS-R due to a divergent study protocol; thus, this site was excluded from the analyses involving this instrument. Based on previous literature and on a valid theoretical background, we fitted five different MIRT models: a) a unidimensional model (one unique general factor); b) a bidimensional model including positive (magic ideation, illusions, psychotic phenomena, and depersonalization/derealization, referential thinking, and suspiciousness) and negative (social isolation, introversion, blunted affect, hypersensitivity) schizotypy; c) a multidimensional model with three uncorrelated factors, i.e. Cognitive-Perceptual (magic ideation, illusions, psychotic phenomena, and depersonalization/derealization), Negative (social isolation, introversion, blunted affect, hypersensitivity), and Paranoid (referential thinking and suspiciousness); d) a multidimensional model with the three aforementioned factors but correlate; and e) a bifactor model with one general latent factor along with three specific uncorrelated factors. The model with the best fit was identified comparing the fit indices of the different models (LL, AIC, BIC, SABIC, RMSEA, SRMSR, TLI, CFI). As shown in Table S1, the bifactor model was the one with the best fit.

CAPE. Based on previous research[6], we fitted a bifactor MIRT model to estimate the general factor ($CAPE_{GEN}$) and the depressive ($CAPE_{DEP}$), negative ($CAPE_{NEG}$), and positive ($CAPE_{POS}$) domains. Fit indices are shown in Table S1.

Table S2. Comparison of SIS-R MIRT models fit statistics.

	LL	AIC	BIC	SABIC	RMSEA	SRMSR	TLI	CFI
SIS-R								
Unidimensional	-11,364	22,809	23,021	22,894	0.121 (0.110-0.133)	0.095	0.780	0.868
Bidimensional	-11,312	22,704	22,916	22,789	0.110 (0.099-0.121)	0.161	0.819	0.892
Tridimensional (uncorrelated factors)	-12,773	25,637	25,876	25,733	0.141 (0.132-0.151)	0.197	0.715	0.817
Tridimensional (three correlated factors)	-12,457	25,010	25,265	25,113	0.067 (0.056-0.077)	0.060	0.937	0.965
Bifactor	**-12,379**	**24,865**	**25,146**	**24,978**	**0.040 (0.027-0.053)**	**0.034**	**0.978**	**0.991**
CAPE								
Bifactor	-23,367	46,915	47,393	47,107	0.039 (0.037-0.042)	0.038	0.954	0.961

Table S3. Loadings of factors in the SIS-R bifactor model.

SIS-R item	Factor	General factor loading	Specific factor loading	Communality (h2)
Social isolation	*Negative*	0.588	0.528	0.625
Introversion	*Negative*	0.556	0.755	0.879
Blunted affect	*Negative*	0.383	0.423	0.326
Referential: being watched	*Paranoid*	0.672	0.520	0.721
Referential: remarks	*Paranoid*	0.752	0.529	0.845
Hypersensitivity	*General*	0.602	-	0.362
Suspiciousness	*General*	0.755	-	0.570
Magic ideation	*Cognitive-Perceptual*	0.564	0.407	0.483
Illusions	*Cognitive-Perceptual*	0.562	0.603	0.680
Psychotic phenomena	*Cognitive-Perceptual*	0.639	0.497	0.656
Derealisation/Depersonalisation	*Cognitive-Perceptual*	0.435	0.555	0.497

Table S4. Loadings of factors in the bifactor model of CAPE.

CAPE item	Factor	General factor loading	Specific factor loading	Communality (h2)
CAPE03	*Negative*	0.533	0.234	0.365
CAPE04	*Negative*	0.412	0.272	0.244
CAPE08	*Negative*	0.403	0.559	0.475
CAPE16	*Negative*	0.575	0.142	0.351
CAPE18	*Negative*	0.703	0.235	0.550
CAPE23	*Negative*	0.577	0.121	0.347
CAPE25	*Negative*	0.624	0.088	0.398
CAPE27	*Negative*	0.483	0.736	0.775
CAPE29	*Negative*	0.544	0.333	0.407
CAPE32	*Negative*	0.546	0.599	0.656
CAPE35	*Negative*	0.648	0.135	0.439
CAPE36	*Negative*	0.716	0.105	0.523
CAPE37	*Negative*	0.582	0.093	0.347
CAPE02	*Positive*	0.539	0.274	0.365
CAPE05	*Positive*	0.380	0.354	0.270
CAPE06	*Positive*	0.439	0.289	0.276
CAPE07	*Positive*	0.509	0.250	0.322
CAPE10	*Positive*	0.617	0.306	0.475
CAPE11	*Positive*	0.217	0.812	0.706
CAPE13	*Positive*	0.345	0.626	0.511
CAPE15	*Positive*	0.312	0.387	0.247
CAPE17	*Positive*	0.286	0.527	0.360
CAPE20	*Positive*	0.313	0.393	0.252
CAPE22	*Positive*	0.566	0.203	0.362
CAPE28	*Positive*	0.458	0.398	0.368
CAPE09	*Depressive*	0.633	0.249	0.463
CAPE12	*Depressive*	0.641	0.510	0.670
CAPE14	*Depressive*	0.628	0.418	0.569
CAPE19	*Depressive*	0.455	0.129	0.223
CAPE39	*Depressive*	0.665	0.321	0.546

Figure S1. Mean SIS-R factor scores across study sites.

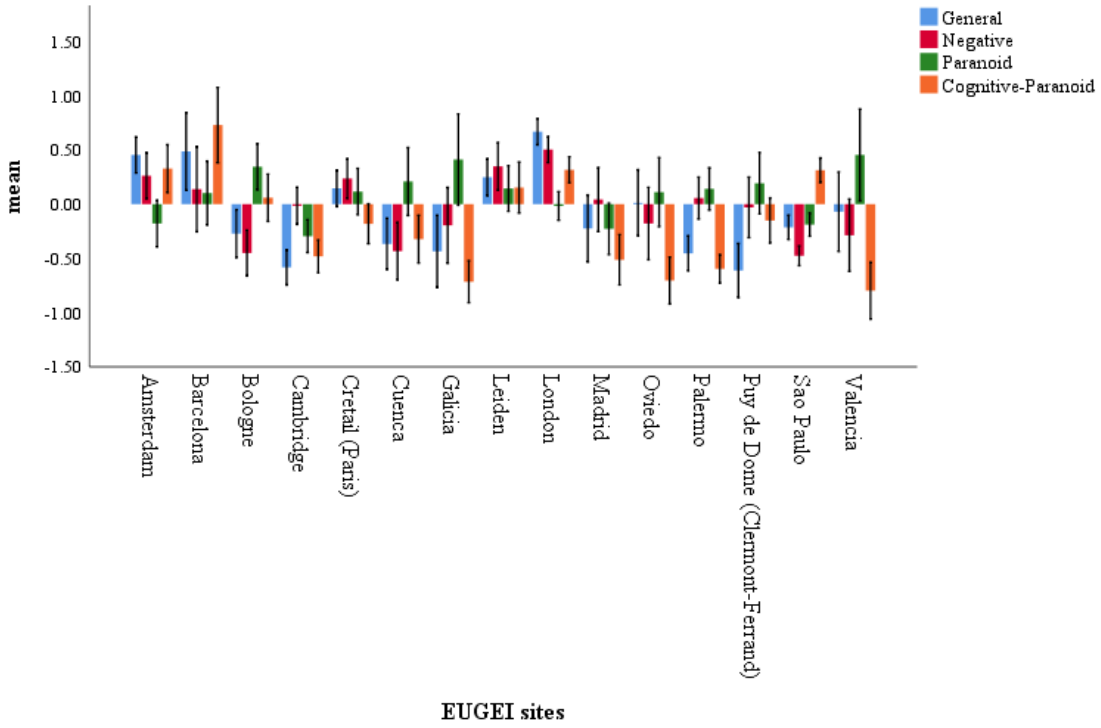
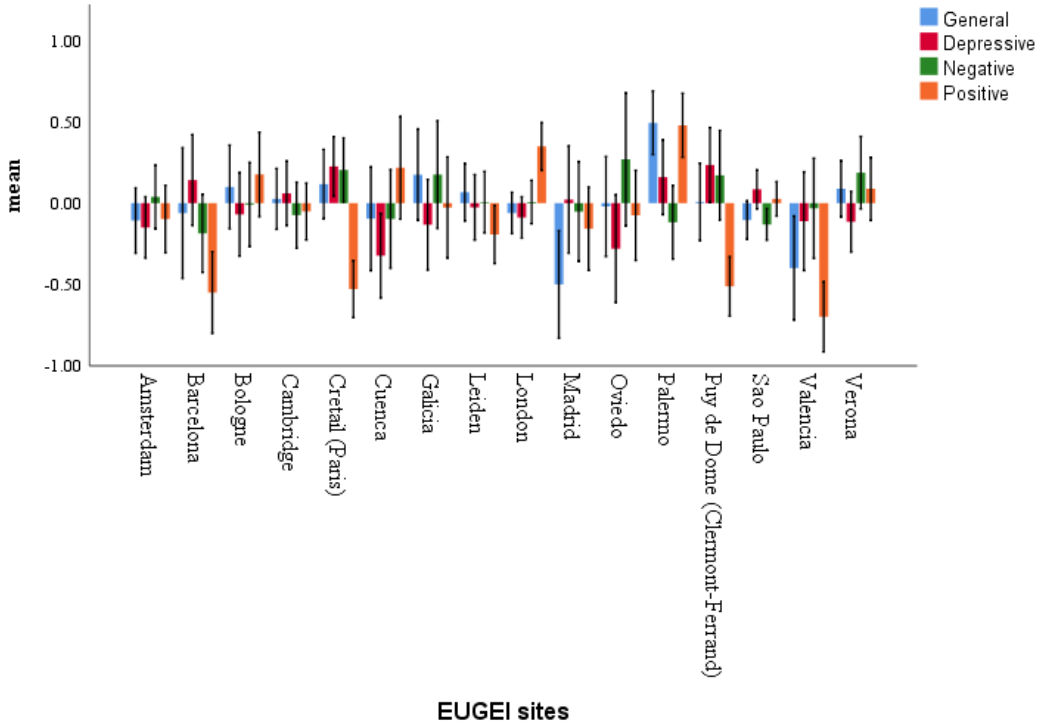


Figure S2. Mean CAPE factor scores across study sites.



Sensitivity analyses

To strengthen robustness of our findings we performed sensitivity analyses on the complete-case sample. Sensitivity analyses were further adjusted for inverse probability weights accounting for any over- or under-sampling of controls relative to the population at-risk. Each control was given a weight inversely proportional to the probability of selection on key demographics (age, sex, binary majority/minority ethnicity status) using census data on relevant populations. Results of sensitivity analyses are shown in the tables below.

Table S5. Multilevel regression analysis of SIS-R General factor score from 15 catchment areas in Europe and Brazil (complete-case analysis).

	Null model (N=1,143)	Model 1 (N=1,143)	Model 2 (N=1,143)	Model 3 (N=1,143)
Fixed effects				
Individual level				
Age				
		-0.006 (-0.011- -0.003)	**-0.005 (-0.009- -0.001)**	**-0.005 (-0.009- -0.001)**
Sex				
Female		Ref.	Ref.	Ref.
Male		**-0.087 (-0.173-0.001)**	-0.061 (-0.144-0.023)	-0.061 (-0.145-0.022)
Education				
Higher			Ref.	Ref.
School, college, vocational			**0.196 (0.105-0.286)**	**0.200 (0.104-0.506)**
No qualification			**0.306 (0.105-0.506)**	**0.305 (0.104-0.506)**
Relational status				
Other			Ref.	Ref.
Single			0.078 (-0.015-0.170)	0.077 (-0.016-0.169)
Employment				
Other			Ref.	Ref.
Unemployed			**0.139 (0.023-0.254)**	**0.141 (0.025-0.256)**
Current cannabis use				
No			Ref.	Ref.
Yes			0.074 (-0.060-0.208)	0.075 (-0.059-0.209)
Migrant status				
Non-migrant			Ref.	Ref.
Migrant			**0.127 (0.021-0.233)**	**0.128 (0.022-0.234)**
CTQ				
			0.063 (0.044-0.082)	**0.062 (0.043-0.081)**
Bullying				
Never			Ref.	Ref.
Ever			**0.316 (0.222-0.410)**	**0.308 (0.214-0.402)**
Site level				
Incidence of FEP				**0.256 (0.117-0.395)**
Random effects				
Individual variance	0.55 (0.51-0.60)	0.54 (0.50-0.59)	0.48 (0.46-0.52)	0.48 (0.44-0.52)
Site variance	0.12 (0.05-0.26)	0.12 (0.05-0.27)	0.10 (0.04-0.23)	0.04 (0.02-0.13)
PCV				
PCV between individuals	Ref..	0.0%	12.7%	12.7%
PCV between sites	Ref..	0.0%	16.7%	66.7%
ICC	0.18 (0.09-0.32)	0.18 (0.09-0.33)	0.18 (0.09-0.33)	0.08 (0.03-0.21)
Log likelihood	-1,301	-1,295	-1,219	-1,215
AIC	2,607	2,560	2,465	2,459
BIC	2,623	2,625	2,536	2,535

CTQ: Childhood Trauma Questionnaire; SCZ-PRS: Schizophrenia Polygenic Risk Score; FEP: First-Episode Psychosis; PCV: Proportional change in variance; ICC: Intraclass correlation coefficitent; AIC: Akaike Information Criterion.

Table S6. Multilevel regression analysis of CAPE General factor score from 16 catchment areas in Europe and Brazil (complete case analysis).

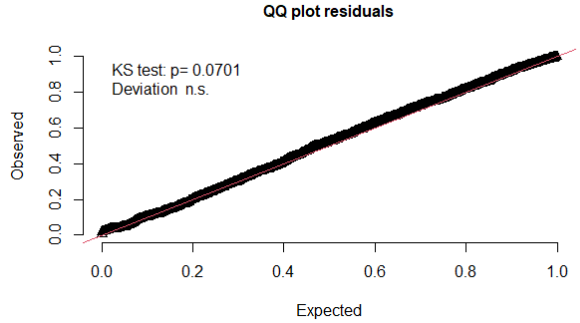
	Null model (N=1,497)	Model 1 (N=1,497)	Model 2 (N=1,493)	Model 3 (N=1,493)
Fixed effects				
Individual level				
Age				
		-0.004 (-0.009-0.001)	-0.001 (-0.006-0.003)	-0.001 (-0.075-0.003)
Sex				
Female		Ref.	Ref.	Ref.
Male		**-0.145 (-0.250- -0.041)**	**-0.122 (-0.223- -0.020)**	**-0.120 (-0.222- -0.019)**
Education				
Higher			Ref.	Ref.
School, college, vocational			0.065 (-0.044-0.173)	-0.058 (-0.051-0.166)
No qualification			-0.068 (-0.312-0.176)	-0.079 (-0.322-0.164)
Relational status				
Other			Ref.	Ref.
Single			**0.138 (0.026-0.250)**	**0.138 (0.026-0.250)**
Employment				
Other			Ref.	Ref.
Unemployed			**0.141 (0.000-0.251)**	**0.142 (0.001-0.282)**
Current cannabis use				
No			Ref.	Ref.
Yes			0.152 (-0.011-0.315)	0.155 (-0.008-0.318)
Migrant status				
Non-migrant			Ref.	Ref.
Migrant			-0.016 (-0.144-0.113)	-0.008 (-0.137-0.120)
CTQ				
			0.100 (0.077-0.123)	**0.101 (0.078-0.125)**
Bullying				
Never			Ref.	Ref.
Ever			**0.333 (0.220-0.447)**	**0.344 (0.230-0.458)**
Site level				
Incidence of FEP				**-0.099 (-0.181- -0.017)**
Random effects				
Individual variance	0.81 (0.75-0.89)	0.81 (0.74-0.88)	0.71 (0.66-0.78)	0.71 (0.66-0.78)
Site variance	0.02 (0.01-0.08)	0.02 (0.01-0.08)	0.02 (0.01-0.06)	0.01 (0.00-0.06)
PCV				
PCV between individuals	Ref.	0.0%	12.3%	12.3%
ICC	0.03 (0.01-0.09)	0.03 (0.01-0.09)	0.03 (0.01-0.08)	0.01 (0.00-0.08)
Log likelihood	-1,512	-1,507	-1,437	-1,435
AIC	3,031	3,027	2,902	2,901
BIC	3,046	3,057	2,973	2,973

CTQ: Childhood Trauma Questionnaire; SCZ-PRS: Schizophrenia Polygenic Risk Score; FEP: First-Episode Psychosis; PCV: Proportional change in variance; ICC: Intraclass correlation coefficitent; AIC: Akaike Information Criterion; Bayesian Information Criterion.

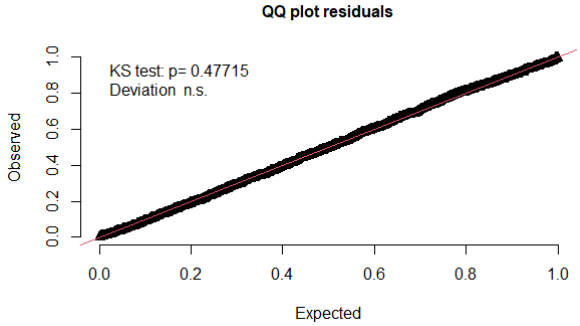
Residuals diagnostics

The R package “DHARMa” was used to test the normality of the residuals of all the mixed-effects linear regression models using a simulation-based approach. We visually inspected the distribution of residuals on a QQ-plot (Figure S3) and performed a formal KS test, which was non-significant for all the fitted SIS-R and CAPE models. Thus, we can conclude that the homoskedasticity assumption was met.

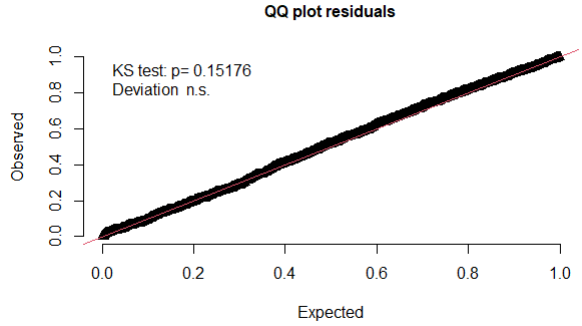
Figure S3. QQ-plots of mixed linear regressions standardized residuals.
Age and gender adjusted SIS-R model.



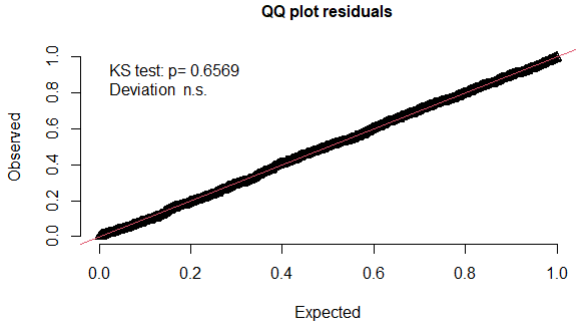
SIS-R model adjusted for all covariates without incidence.



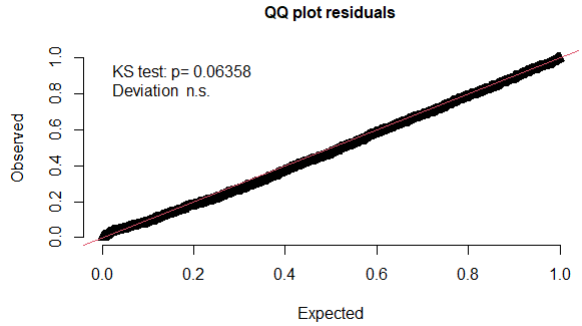
CAPE model adjusted for all covariates without incidence.

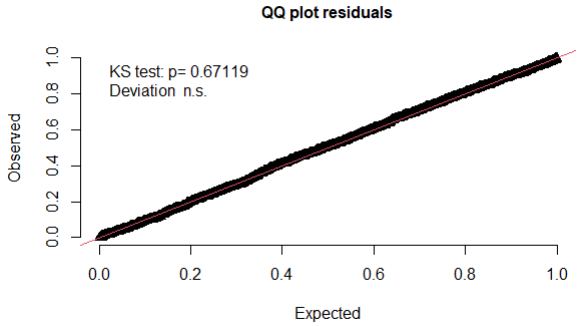


SIS-R model adjusted for all covariates plus incidence at 2-level.



CAPE model adjusted for all covariates plus incidence at 2-level.





Multicollinearity checks

We checked for multicollinearity examining correlations between independent variables and estimating individual Variance Inflation Factors (VIF). There was no evidence of a strong correlation between our predictors (coefficient of at least ±0.6). A correlation plot is shown below (Table S7). Finally all VIFs were around 1 and not exceeding the value of 1.15. We can conclude that multicollinearity did not represent an issue in our analyses.

Table S8. Correlation matrix of predictors included in the fully-adjusted models.

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1.	1										
2.	-.07	1									
3.	-.11	.01	1								
4.	-.02	-.03	.31	1							
5.	.27	-.11	-.05	-.04	1						
6.	.01	-.02	-.06	-.11	-.04	1					
7.	.16	-.14	-.02	-.06	-.05	-.04	1				
8.	.01	.02	-.06	.01	.06	-.02	.03	1			
9.	-.07	.09	-.17	-.12	-.05	-.04	-.02	-.10	1		
10.	.10	.09	.04	.07	.04	.01	.03	-.05	.02	1	
11.	-.00	-.01	.02	.05	-.01	.01	-.00	-.05	.02	.06	1

1. Age
2. Gender
3. Educational attainment: no qualification
4. Educational attainment: school, college or vocational
5. Relationship status
6. Employment
7. Cannabis use
8. Migrant status
9. Child maltreatment
10. Bullying
11. Local incidence of first-episode psychosis

Acknowledgments

EU-GEI WP2 Group: The European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI) WP2 Group members include : Kathryn Hubbard, MSc, Department of Health Service and Population Research, Institute of Psychiatry, King's College London, De Crespigny Park, Denmark Hill, London, England; Stephanie Beards, PhD, Department of Health Service and Population Research, Institute of Psychiatry, King's College London, De Crespigny Park, Denmark Hill, London, England; Pedro Cuadrado, MD, Villa de Vallecas Mental Health Department, Villa de Vallecas Mental Health Centre, Hospital Universitario Infanta Leonor/Hospital Virgen de la Torre, Madrid, Spain; José Juan Rodríguez Solano, 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; Angel Carracedo, MD, PhD, Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario, Santiago de Compostela, Spain; Gonzalo López, PhD, Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, Investigación Sanitaria del Hospital Gregorio Marañón (Centro de Investigación Biomédica en Red de Salud Mental), Madrid, Spain; Silvia

Amoretti, PhD, Department of Psychology and Psychiatry, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Centro de Investigación Biomédica en Red de Salud Mental, University of Barcelona, Barcelona, Spain; Eduardo J. Aguilar, MD, PhD, Department of Medicine, University of Valencia, INCLIVA, CIBERSAM, Valencia, Spain.; Paz Garcia-Portilla, MD, PhD, Department of Medicine, Psychiatry Area, School of Medicine, Universidad de Oviedo, Centro de Investigación Biomédica en Red de Salud Mental, Oviedo, Spain; Javier Costas, PhD, Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario, Santiago de Compostela, Spain; Estela Jiménez-López, MSc, Department of Psychiatry, Servicio de Psiquiatría Hospital "Virgen de la Luz," Cuenca, Spain; Mario Matteis, MD, Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, Investigación Sanitaria del Hospital Gregorio Marañón (Centro de Investigación Biomédica en Red de Salud Mental), Madrid, Spain; Emiliano González, PhD, Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, Investigación Sanitaria del Hospital Gregorio Marañón (Centro de Investigación Biomédica en Red de Salud Mental), Madrid, Spain; Emilio Sánchez, MD, Department of Psychiatry, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, Investigación Sanitaria del Hospital Gregorio Marañón (Centro de Investigación Biomédica en Red de Salud Mental), Madrid, Spain; Nathalie Franke, MSc, Department of Psychiatry, Early Psychosis Section, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands; Jean-Paul Selten, MD, PhD, Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, the Netherlands, and Rivierduinen Institute for Mental Health Care, Leiden, the Netherlands; Fabian Termorshuizen, 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 Rivierduinen Centre for Mental Health, Leiden, the Netherlands; Daniella van Dam, PhD, Department of Psychiatry, Early Psychosis Section, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands; Elles Messchaart, MSc, Rivierduinen Centre for Mental Health, Leiden, the Netherlands; Marion Leboyer, MD, PhD, Univ Paris Est Creteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, Fondation FondaMental, F-94010 Creteil, France.; Franck Schürhoff, MD, PhD, AP-HP, Univ Paris Est Creteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, Fondation FondaMental, F-94010 Creteil, France.; Stéphane Jamain, PhD, Univ Paris Est Creteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, Fondation FondaMental, F-94010 Creteil, France.; Flora Frijda, MSc, Etablissement Public de Santé Maison Blanche, Paris, France; Grégoire Baudin, MSc, Univ Paris Est Creteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, Fondation FondaMental, F-94010 Creteil, France.; Aziz Ferchiou, MD, AP-HP, Univ Paris Est Creteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, Fondation FondaMental, F-94010 Creteil, France.; Baptiste Pignon, MD, PhD, Univ Paris Est Creteil (UPEC), AP-HP, Hôpitaux Universitaires « H. Mondor », DMU IMPACT, INSERM, IMRB, Fondation FondaMental, F-94010 Creteil, France.; Jean-

Romain Richard, MSc, Institut National de la Santé et de la Recherche Médicale, U955, Créteil, France, and Fondation Fondamental, Créteil, France; Thomas Charpeaud, MD, Fondation Fondamental, Créteil, France, CMP B CHU, Clermont Ferrand, France, and Université Clermont Auvergne, Clermont-Ferrand, France; Anne-Marie Tronche, MD, Fondation Fondamental, Créteil, France, CMP B CHU, Clermont Ferrand, France, and Université Clermont Auvergne, Clermont-Ferrand, France; Giovanna Marrazzo, MD, PhD, Unit of Psychiatry, "P. Giaccone" General Hospital, Palermo, Italy; Lucia Sideli, PhD, Department of Biomedicine, neurosciences, and advanced diagnostics, University of Palermo, , Via G. La Loggia 1, 90129 Palermo, Italy.; Crocettarachele Sartorio, PhD, Unit of Psychiatry, "P. Giaccone" General Hospital, Palermo, Italy;; Fabio Seminerio, PhD, Unit of Psychiatry, "P. Giaccone" General Hospital, Palermo, Italy; Camila Marcelino Loureiro, MD, 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; Rosana Shuhama, 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; Mirella Ruggeri, MD, PhD, Section of Psychiatry, Department of Neuroscience, Biomedicine and Movement, University of Verona, Verona, Italy; Chiara Bonetto, PhD, Section of Psychiatry, Department of Neuroscience, Biomedicine and Movement, University of Verona, Verona, Italy; Doriana Cristofalo, MA, Section of Psychiatry, Department of Neuroscience, Biomedicine and Movement, University of Verona, Verona, Italy; Marco Seri , Department of Medical and Surgical Sciences, Bologna University; Elena Bonora, Department of Medical and Surgical Sciences, Bologna University

Author contributions

All the authors in the EU-GEI group collected or supervised the data collection. RMM, DQ, and GD were responsible for the conception and design of the study. GD, KM, DQ, and CG-A cleaned and prepared the data for this paper analysis. GD, KM and DQ did the data analysis and wrote the findings in the initial manuscript. GD and KM contributed to the creation of the figures and tables. RMM, MDF, DQ, and IT provided a careful statistical and methodological revision of the manuscript and contributed to the final draft. GD, DQ, MDF, and RMM contributed to the interpretation of the results. All authors had full access to all data (including statistical reports and tables) in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Declarations of interest

MDF reports personal fees from Janssen, outside the submitted work. RMM reports personal fees from Janssen, Lundbeck, Sunovion, and Otsuka, outside of the submitted work. PML reports personal fees from Janssen, Lundbeck, and Otsuka, outside of the submitted work. CA has been a consultant to or has received honoraria or grants from Acadia, Angelini, Gedeon Richter, Janssen Cilag, Lundbeck, Minerva, Otsuka,

Roche, Sage, Servier, Shire, Schering Plough, Sumitomo Dainippon Pharma, Sunovion and Takeda. All authors declare no competing interests. JBK is supported by the National Institute for Health Research University College London Hospital Biomedical Research Centre and has received consultancy fees from Roche and the Health Services Executive, Ireland. MB has been a consultant for, received grant/research support and honoraria from, and been on the speakers/advisory board of ABBiotics, Adamed, Angelini, Casen Recordati, Janssen-Cilag, Menarini, Rovi and Takeda.

References

1. Tierney N. visdat: Visualising Whole Data Frames. *J Open Source Softw*. 2017;2(16):355. doi:10.21105/JOSS.00355
2. Ryu C. Tools for Data Diagnosis, Exploration, Transformation [R package dlookr version 0.5.2]. Published online November 22, 2021.
3. Stekhoven DJ, Bühlmann P. Fast Imputation of Missing Values [R package missRanger version 2.1.3]. *Bioinformatics*. 2021;28(1):112-118. doi:10.1093/BIOINFORMATICS/BTR597
4. Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-118. doi:10.1093/bioinformatics/btr597
5. Waljee AK, Mukherjee A, Singal AG, et al. Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*. 2013;3(8). doi:10.1136/bmjopen-2013-002847
6. Quattrone D, Ferraro L, Tripoli G, et al. Daily use of high-potency cannabis is associated with more positive symptoms in first-episode psychosis patients: the EU-GEI case–control study. *Psychol Med*. 2021;51(8):1329-1337. doi:10.1017/S0033291720000082