**SUPPLEMENTARY INFORMATION (SI)**

**Linking prolonged Childhood and Adolescent Loneliness to Schizophrenia Spectrum Disorders: Results from the EUGEI study**

**SM1. Participants description**

We used data from the Work-package 6 (WP6) of the European Network of National Networks studying Gene-Environment Interactions in Schizophrenia (EU-GEI). Data collection took place between 2010 and 2015 in Turkey, Spain, and Serbia, comprising a total of 1,261 patients, 1,282 unaffected siblings and 1,525 controls.

Patients were diagnosed with schizophrenia spectrum disorders (SSD) according to the DSM-IV-TR. The diagnosis was subsequently confirmed using the Operational Criteria Checklist for Psychotic and Affective Illness. Siblings and unrelated controls with no lifetime history of psychotic disorders were recruited from the same population as the cases. Exclusion criteria for all participants included a diagnosis of psychotic disorder due to another medical condition, a history of head injury with loss of consciousness, and an intelligence quotient (IQ) less than 70. Ethical approval for the projects was obtained from the medical ethics committees of all participating sites, and the research was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all respondents, and participants under the age of 18 provided assent, with parental informed consent also obtained.

To maintain high quality and homogeneity in clinical, experimental, and environmental assessments, standardized instruments were administered by psychiatrists, psychologists, or trained research assistants. These professionals underwent mandatory on-site training sessions and online training modules, which included interactive interview videos and self-assessment tools. Both on-site and online training sessions were repeated annually to ensure high inter-rater reliability throughout the study enrollment period.

**SM2. Statistical Analyses:**

We used mixed regression models with a random intercept for the country—to account for clustering of participants within countries—to assess the association between CAL and i)psychosis risk and ii)psychotic symptoms. For each set of analyses, we ran separate models for LNL12 and LNL12-16. We assessed the relationship between LNL12/LNL12-16 with case-control and sibling-control status with multilevel logistic models. We included age, sex, and LNL12/LNL12-16 as fixed effects and country as a random effect. To explore sex-specific effects we ran additional models including a loneliness\*sex interaction and repeated the models after stratifying by sex.

To explore the effect of potentially confounding variables on our findings, we conducted sensitivity analyses after adjusting for i) social isolation and ii)CAPE-dep, and iii)after removing participants with a younger age at onset using different cut-offs (age at onset <18, <19, and <20 years).

We assessed the association of LNL12/LNL12-16 with log-transformed CAPE mean scores (CAPE-pos, CAPE-neg, CAPE-dep, and CAPE total) using mixed linear regression models. We included the same set of fixed and random effects, with clinical group(case/sibling/control) as an additional fixed effect. We additionally ran stratified analyses by clinical group.

In a second set of mixed logistic regression analyses, we assessed the effect of the interaction between LNL12/LNL12-16 and dichotomized PGSSZ(CAL\*PGSSZ) on i)psychosis risk and ii) psychotic symptoms. For each set of analyses, we ran separate models for LNL12 and LNL12-16. We initially assessed the existence of a correlation structure between PGSSZ and CAL in the general population by regressing the effect of PGSSZ on LNL12/LNL12-16 (outcome) in a multilevel logistic regression model in the control sample only.

We assessed the effect of the interaction between CAL and PGSSZ on case-control and sibling-control status with multilevel logistic models. We included age, sex, and the 10 first PCA ancestry components as fixed effects and country as a random effect. Following previous recommendations from the EUGEI consortium for gene-environment interaction models (GxE)[1](https://www.zotero.org/google-docs/?vLzNxn), we assessed the existence of a CAL\*PGSSZ interaction on an additive rather than a multiplicative scale. Departure from additivity was evaluated as follows: first, 1) four states resulting from the combination of loneliness(LNL12 or LNL12-16) and binary PGSSZ risk state variables were included as independent dummy variables. Then, the relative excess risk due to interaction (RERI=ORLNL&SZ-PGS−ORLNL−ORSZ-PGS+1) was estimated. We calculated 95% confidence intervals(95%CI) with both the Delta method and the bootstrap percentile method(1,000 permutations). RERI values with a 95%CI not containing zero were considered a positive deviation from additivity.

We ran additional case-control models after stratifying by sex (males and females separately). We compared the RERI distributions in males and females using Wilcoxon-Mann-Whitney tests after bootstrap resampling(5,000 permutations) of 500 schizophrenia and 500 HC subjects in each sex separately.

We assessed the effect of the interaction between LNL12/LNL12-16 and PGSSZ on log-transformed CAPE mean scores (CAPE-pos, CAPE-neg and CAPE-dep, CAPE-total) using mixed linear regression models. We included the same set of fixed and random effects, with clinical group(case/control) as an additional fixed effect. We additionally ran stratified analyses by clinical group and sex.

We evaluated the performance of case-control models by assessing the Area Under the Curve (AUC) and the amount of variance explained using Nakagawa R2. We used the likelihood ratio test(LRT) to assess whether the addition of the LNL12/LNL12-16 by PGSSZ term in the model statistically improved the explained variance. We also assessed the sensitivity, specificity, and positive predictive value(PPV) for LNL12 or LNL12-16, PGSSZ and the interaction(CAL\*PGSSZ) terms based on model predictions. Participantswere classified as patients by the logistic model when the predicted probability was greater than 0.5. The PPV was calculated as PPV = (sensitivity \* prevalence) / ((sensitivity \* prevalence) + ((1 - specificity) \* (1 - prevalence))). For these calculations, a prevalence of 1% for schizophrenia was used. 95%CI for PPV were calculated after Monte Carlo simulations based on sensitivity, specificity, and prevalence.

P-values were corrected for multiple testing using the Benjamini and Hochberg false discovery rate(FDR)method[2](https://www.zotero.org/google-docs/?tOdfRB). All statistical analyses were conducted using R employing lmer and glmer functions from the lme4 package.

**SM3. Stratified and sensitivity analyses**

We conducted a comprehensive analysis to evaluate the relationship between CAL (LNL12, LNL12-16) with schizophrenia in both case-control and sibling-control groups, as well as with associated symptomatology. In all instances, mixed regression models were utilized to account for the clustering of subjects within participant countries.

Initially, we examined the relationship between LNL12 and LNL12-16 and their associations with case-control and sibling-control status using random intercept multilevel logistic models. All models also included age and sex as covariates.

Additionally, to control for the potential influence of social isolation during childhood and adolescence we included as a covariate in the case-control random intercept multilevel logistic models. Social isolation was assessed using the Premorbid Adjustment Scale (PAS) for psychosis, focusing on three dimensions during two developmental periods (before age 12 and between ages 12-16): peer relationships, sociability and withdrawal, and Adaption to school.

The PAS variables included the following:

1. **Peer Relationships (before age 12 and during adolescence 12-16):**
	* **0:** Many friends (more than 5), close relationships ("best friends" or people you could confide in) with several.
	* **1:**
	* **2:** Close relationships with a few friends (1 or 2), casual friendships with others.
	* **3:** Only casual friends.
	* **4:** Deviant (unusual) friendship patterns: friendly with children younger or older only, or relatives only, or casual relationships only.
	* **6:** Social isolate, no friends, not even superficial relationships.
2. **Sociability and Withdrawal (before age 12 and during adolescence 12-16):**
	* **0:** No introversion, actively and frequently seeks social contact.
	* **1:**
	* **2:** Mild introversion, enjoys socializing when involved, occasionally seeks opportunities to socialize.
	* **3:**
	* **4:** Moderate introversion, prone to daydreaming and being excessively imaginative, passively accepts being included in social interactions but does not seek them.
	* **5:**
	* **6:** Does not interact with others, shows introversion and isolation, actively avoids contact.
3. **Adaption to school (before age 12 and during adolescence 12-16):**
	* **0:** Good adaptation, likes school, minimal or no behavioral problems, has friends at school, and likes most teachers.
	* **1:**
	* **2:** Reasonable adaptation, occasional behavioral problems, not very interested in school but rarely skips. Has friends but rarely participates in extracurricular activities.
	* **3:**
	* **4:** Poor adaptation, dislikes school, frequently skips, and exhibits frequent behavioral problems (may have been expelled).
	* **5:**
	* **6:** Total rejection of school, engages in delinquency or vandalism directed at the school.

These measures allowed us to assess the effects of loneliness while controlling for variables that approximate social isolation retrospectively, as well as behavioral and social functioning during the same developmental periods.

Furthermore, to control for the potential influence of depressive symptomatology at the time of completing the questionnaire, we included the mean score of depressive symptoms as measured by the CAPE-NEG scale as a covariate in the sensitivity analysis. This allowed us to account for the possibility that current depressive symptoms could confound the observed relationships.

For the final sensitivity analyses, we stratified the analyses by different time points for the onset of psychosis (18 to 20 years). We calculated the year of psychosis onset and conducted subset analyses exclusively with onsets occurring after 18, 19, and 20 years.

**SM4. Genetic data processing and polygenic score estimation.**

Genotyping was performed at the Cardiff University Institute of Psychological Medicine and Clinical Neurology using the custom Illumina HumanCoreExome-24 BeadChip arrays, which probed for 570,038 genetic variants. The genotypic data underwent processing through the GenomeStudio package and were converted to PLINK format for subsequent analyses. QC measures implemented have been described in previous works using the same genetic datasets [1,3,4](https://www.zotero.org/google-docs/?xURwES). Briefly, variants with less than 98% call rate and diverging from the Hardy-Weinberg equilibrium (p-value < 1 x 10-6) across Turkish, Northern European, or Southern European cohorts were removed. A total of 559,505 variants were retained. Samples below a 98% call rate were removed, and a linkage disequilibrium pruned set of variants was prepared for further analysis. Further steps involved removing outlier samples based on homozygosity F values and samples with incorrect sibling relationships based on identity-by-descent values and excluding chips with high error rates. After QC, genotypes were imputed using the Haplotype Reference Consortium reference panel through Eagle and Minimac3 for haplotype phasing and imputation, respectively. Following stringent criteria (imputation r2 > 0.6, minor allele frequency > 0.1%, and call rate > 99%), 8,277,535 variants were finally selected, and PLINK was used to generate best-guess genotypes from genotype probabilities.

Polygenic scores (PGS) for each participant with available genetic data passing QC were estimated using PRS-CS [5](https://www.zotero.org/google-docs/?izon9E) (<https://github.com/getian107/PRScs>) following standard parameters. With a Bayesian architecture, PRS-CS infers posterior SNP effect sizes under continuous shrinkage priors using GWAS summary statistics and an external LD reference panel. This method was selected as it has been observed to be one of the best-performing polygenic scoring methods in the context of psychiatric disorders [6](https://www.zotero.org/google-docs/?VWllcd). European LD reference from 1000 Genomes project provided by PRS‐CS was used for all computations. The global shrinkage parameter was set to φ=1x10‐2 to reflect the polygenic nature of schizophrenia, as recommended by the developers. To avoid sample overlapping, leave-one-out summary statistics from the latest schizophrenia PGC3 GWAS [7](https://www.zotero.org/google-docs/?hMreqF) was used (NSZ = 67,122, NHC = 92,855). A logistic regression model was applied to test the association of SZ-PGS with case‐control and case-sibling status (adjusted for ancestry using the first ten principal components, age, sex and country), and liability scale R2 was calculated using Lee et al [8](https://www.zotero.org/google-docs/?poJTUa), assuming schizophrenia prevalence of 1%. In order to understand the direction of effect of the PGS across the different partitions, SZ-PGS comparisons across ranked deciles were also performed. The target sample was first separated into 10 deciles of increasing PGS. The P-threshold with the lowest p-value was selected for each partition. The phenotype values of each decile were compared to those of the reference decile (the median decile (5th) was used as reference) one-by-one, with decile status as predictor of target phenotype (5th decile was coded 0 and tested decile 1) in a logistic regression model. OR values for each comparison were estimated from regression coefficients of these decile-status predictors. Sex, age, country and 10 first PCA ancestry components were used as covariates. See supplementary figures below for AUC estimation and comparison across risk deciles.

**SF1. Predicted case-control status based on childhood (A) and adolescence (B) loneliness, stratified by sex. Logistic regression models incorporating interaction terms between loneliness and sex were used, adjusted for age and country. Points represent model predictions, while lines indicate trends for males (blue) and females (pink).**

A) Predicted case-control status as a function of childhood loneliness (before age 12), illustrating how early loneliness impacts schizophrenia risk differently for males and females.



B) Predicted case-control status as a function of adolescence loneliness (ages 12-16), highlighting the significant interaction between loneliness during adolescence and gender, with a more pronounced effect on schizophrenia risk observed in females compared to males



**SF2. Quantile plot of SZ-PGS predictions in case-control (A) and sibling-control (B) comparisons.** To avoid sample overlapping, leave-one-out summary statistics from the latest schizophrenia PGC3 GWAS was used (NSZ = 67,122, NHC = 92,855). The EU-GEi WP6 target sample is separated into deciles of increasing SZ-PGS. The case-control / sibling-control status of each decile is compared to the median (5th decile), one-by-one, using a logistic regression model with covariates (sex, age, country and 10 PCA ancestry components). OR values for each comparison were estimated from regression coefficients of these decile-status predictors.

**A)**



**B)**



**SF3. Sex differences on the** i**nteraction between PGSSCZ and CAL on SSD risk.** Differences in the effects of the interaction between LNL12 (**A**) and LNL12-16 (**B**) and dichotomized PGS-SZ (above 75% of control population) on case-control status were evaluated after bootstrap resampling (5,000 permutations with replacement) of 500 schizophrenia and 500 HC subjects in each sex separately. Differences between the RERI distributions in males and females were statistically assessed using Wilcoxon-Mann-Whitney tests.



**SF4. Area Under the Curve (AUC) estimation for case-control (A) and sibling-control (B) predictions.**

**A)**



**B)**



**SF5.Positive predictive value (PPV) of developing schizophrenia for subjects from the general population in comparison to those having high genetic risk, those who suffered childhood loneliness (LNL12) and those having both genetic and loneliness factors.** Multilevel logistic regression models accounting for participant country, including SZ-PGS, LNL12 and interaction SZ-PGS\*LNL12 terms and adjusted for sex, age, and the first 10 ancestry principal components were performed. The probability of a subject classified by the regression model as schizophrenia patient is colored in red in the following contexts: A) All subjects, B) Only males, C) Only females and D) considering PGS-SZ cutoff of 90% of the control population.



**References**

[1. Guloksuz S, Pries LK, Delespaul P, et al. Examining the independent and joint effects of molecular genetic liability and environmental exposures in schizophrenia: results from the EUGEI study. *World Psychiatry*. 2019;18(2):173-182. doi:10.1002/wps.20629](https://www.zotero.org/google-docs/?e67gRO)

[2. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x](https://www.zotero.org/google-docs/?e67gRO)

[3. Os J van, Pries LK, Have M ten, et al. Evidence, and replication thereof, that molecular-genetic and environmental risks for psychosis impact through an affective pathway. *Psychological Medicine*. 2022;52(10):1910-1922. doi:10.1017/S0033291720003748](https://www.zotero.org/google-docs/?e67gRO)

[4. Pries LK, Ferro GAD, Os J van, et al. Examining the independent and joint effects of genomic and exposomic liabilities for schizophrenia across the psychosis spectrum. *Epidemiology and Psychiatric Sciences*. 2020;29:e182. doi:10.1017/S2045796020000943](https://www.zotero.org/google-docs/?e67gRO)

[5. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10(1):1776. doi:10.1038/s41467-019-09718-5](https://www.zotero.org/google-docs/?e67gRO)

[6. Ni G, Zeng J, Revez JA, et al. A comparison of ten polygenic score methods for psychiatric disorders applied across multiple cohorts. *Biological Psychiatry*. Published online May 4, 2021. doi:10.1016/j.biopsych.2021.04.018](https://www.zotero.org/google-docs/?e67gRO)

[7. Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604(7906):502-508. doi:10.1038/s41586-022-04434-5](https://www.zotero.org/google-docs/?e67gRO)

[8. Lee SH, Goddard ME, Wray NR, Visscher PM. A Better Coefficient of Determination for Genetic Profile Analysis. *Genetic Epidemiology*. 2012;36(3):214-224. doi:https://doi.org/10.1002/gepi.21614](https://www.zotero.org/google-docs/?e67gRO)