GENETIC AND GEOGRAPHICAL ASSOCIATIONS WITH SIX DIMENSIONS OF PSYCHOTIC EXPERIENCES IN ADOLESCENCE

Jessye Maxwell[*,1,2], *Angelica Ronald*[3], *Alastair G. Cardno*[4], *Gerome Breen*[1,2], *Kaili Rimfeld*[1,5], *Evangelos Vassos*[1,2]

[1]Social Genetic and Developmental Psychiatric Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [2]NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Trust, London, UK; [3]Department of Psychological Sciences, Birkbeck, University of London, London, UK; [4]Division of Psychological and Social Medicine, University of Leeds, Leeds, UK; [5]Department of Psychology, Royal Holloway, University of London, Surrey, UK.

[*]To whom correspondence should be addresses; Social Genetic and Developmental Psychiatry Centre, Memory Lane, London, SE5 8AF; tel: +44 (0)20 7848 0873; e-mail: jessye.maxwell@kcl.ac.uk.

Abstract word count: 242
Manuscript word count: 3974

Abstract

Background and hypothesis: Large-scale epidemiological and genetic research have shown that psychotic experiences in the community are risk factors for adverse physical and psychiatric outcomes. We investigated the associations of 6 types of specific psychotic experiences and negative symptoms assessed in mid-adolescence with well-established environmental and genetic risk factors for psychosis.

Study design: Fourteen polygenic risk scores (PRS) and nine geographical environmental variables from 3,590 participants of the Twins Early Development Study (mean age 16) were associated with paranoia, hallucinations, cognitive disorganisation, grandiosity, anhedonia, and negative symptoms scales. The predictors were modelled using LASSO regularisation separately (Genetic and Environmental models) and jointly (GE model).

Study results: In joint GE models, we found significant genetic associations of negative symptoms with educational attainment PRS (β = -0.07; 95%CI = -0.12 to -0.04); cognitive disorganisation with neuroticism PRS (β = 0.05; 95%CI = 0.03 to 0.08); paranoia with MDD (β = 0.07; 95%CI = 0.04 to 0.1), BMI (β = 0.05; 95%CI = 0.02 to 0.08), and neuroticism PRS (β = 0.05; 95%CI = 0.02 to 0.08). From the environmental measures only family SES (β = -0.07, 95%CI = -0.10 to -0.03) and regional education levels (β = -0.06; 95%CI = -0.09 to -0.02) were associated with negative symptoms.

Conclusions: Our findings advance understanding of how genetic propensity for psychiatric, cognitive, and anthropometric traits, as well as environmental factors, together play a role in creating vulnerability for specific psychotic experiences and negative symptoms in mid-adolescence.

Key words: environmental risk, geographical variables, polygenic risk scores, psychosis scales, schizophrenia, prediction models

Introduction

It has been hypothesized that there is a psychosis continuum in the general population, with clinical psychotic disorders such as schizophrenia and bipolar disorder at the extreme end of continuously distributed phenotypes.[1,2] Symptoms of psychotic disorders include positive (delusions, hallucinations and disorganised symptoms), and negative (lack of volition, reduced speech output, and flattening of affect) domains.[3–7] 'Psychotic experiences' refers to subclinical psychotic-like features measured in the general population, representing the full range of severity.[8–10] Consistent with clinical psychotic disorders, psychotic experiences have been shown to have a multidimensional structure, although there is variation in the number and definition of the dimensions across studies.[8,11–15]

Large-scale epidemiological research has shown that psychotic experiences in the community are risk factors for a range of later physical and mental health disorders and adverse outcomes.[16–18] Consequently, psychotic experiences may be an important target for providing insights into the causes of psychosis and for preventative strategies.[19,20] Studies have demonstrated a shared aetiology between psychotic experiences and clinical psychotic disorders, including both genetic and environmental risk factors.[21–23] Environmental measures in common with clinical psychotic disorders include urbanicity, migrant status and socio-economic status.[3,21,24] Many studies also suggest that psychotic experiences in childhood may be associated with a broader set of psychiatric disorders and behaviours with onset in early adulthood, such as affective, anxiety, and substance use disorders and suicidality.[16,25–30]

Twin studies have estimated the heritability of psychotic experiences around 30-50%[15,31–36] and molecular genetic studies have attributed 3-17% of the variance in psychotic experiences to common genetic variation.[22,23,37] Schizophrenia polygenic risk score (PRS) has been associated with psychotic experiences in adolescence,[23,38] however, there is some inconsistency across studies,[39,40] which may be attributable to differences in sample characteristics or definitions of psychotic experiences.[22,41] Psychotic experiences are also known to be associated with genetic predisposition to depression and neurodevelopmental disorders.[23,42,43] As such, in the current study, we employ a multi-PRS approach,[44] which allows us to investigate the association of a variety of psychiatric, cognitive and anthropometric trait PRS with psychotic experiences.

Psychotic experiences have been associated with urban upbringing.[21,34,45–47] Studies aimed at identifying what underlies these findings have discovered that psychotic experiences are associated with certain characteristics of the urban environment including neighbourhood adversity and deprivation[48,49] or air pollution.[50] However, uncertainty over the interpretation of these associations comes from the recent discovery that aspects of the environment, including urban living, are themselves partially heritable, and the associations between urbanicity and psychotic experiences may not be independent of a genetic predisposition to psychiatric disorders.[51,52] In the sample used in our study, psychotic experiences have been shown to share some genetic influences with stressful life events and bullying,[53,54] but more distal environmental factors have not so far been investigated.

In light of the evidence that psychotic experiences show a range of transdiagnostic associations in mental health as well as associations with other types of adverse outcomes, we investigated the association of genetic predisposition to a range of psychiatric, cognitive, and anthropometric traits with a six-dimensional representation of psychotic experiences (i.e., paranoia, hallucinations, cognitive disorganization, grandiosity, anhedonia, and negative symptoms).[8] Furthermore, we investigated the association of psychotic experiences with factors associated with urban living, a well-established environmental risk factor for psychotic disorders. Lastly, we modelled genetic and environmental measures in one model to assess whether the two predictors are independent of each other.[55] The study was preregistered at https://osf.io/pts7m/.

Methods

Sample

Participants included in the current study are part of the Longitudinal Experiences and Perceptions (LEAP) study, which is drawn from the Twins Early Development Study (TEDS).[56] TEDS is a community sample, which constitutes around 10,000 twin pairs who were born in England and Wales between 1994 and 1996. The recruitment of these participants was designed to obtain a sample of families that are representative of the population in England and Wales.[57] Of the 10,874 TEDS families that were contacted for inclusion in the LEAP study, 5,059 (47%) twin pairs provided psychotic experiences data (*mean age* = 16.32 years; *s.d.* = 0.68). Exclusions leading to the removal of 316 families included individuals who did not provide consent, had a severe medical disorder, perinatal complication, or had unknown zygosity. Further details on the LEAP study are found elsewhere.[8] All twins provided written consent to participate in the study at age 16.

Genotyping

Genotyping was done either on the Affymetrix GeneChip 6.0 or Illumina HumanOmniExpressExome (61% of genotyped sample) DNA microarrays.[57] Genotypes from the two platforms were separately phased using EAGLE2,[58] and imputed into the Haplotype Reference Consortium (release 1.1). Further details are found elsewhere.[59] After merging, there were 7,363,646 genotyped or well-imputed SNPs (information score [INFO] > 0.75) available for analysis. After randomly selecting one twin from each pair, we obtained a sample of 4040 (56% female) individuals with genotype data at age 16.

The first 10 ancestry informed principal components were calculated using 39,353 autosomal SNPs with minor allele frequency > 5% and imputation INFO score of 1, selected after pruning to remove SNPs in linkage disequilibrium (r^2) > 0.1 and excluding regions with known high linkage disequilibrium.[59]

After the exclusion of individuals without relevant genotype and phenotype data at age 16, we were left with a sample size of 3590 (complete cases only).

Polygenic risk scores

We calculated polygenic risk scores for 14 psychiatric, cognitive and anthropomorphic traits (chosen based on their selection in a previous study[55]) to estimate their association with psychotic experience and negative symptoms at age 16. GWAS summary statistics were downloaded for attention deficit hyperactivity disorder (ADHD),[60] anorexia nervosa,[61], anxiety,[62], autism spectrum disorders (ASD),[63] bipolar disorder,[64] major depressive disorder (MDD),[65] education years,[66] extraversion,[67] intelligence (excluding the 3,414 TEDS participants used in the reported GWAS),[68] schizophrenia,[69] subjective well-being (excluding the 2,148 TEDS participants used in the reported GWAS),[70] neuroticism,[70] height[72] and BMI[72] (***Supplementary Table 1***). We used PRScs software with default parameter settings to calculate posterior SNP effect sizes under continuous shrinkage priors for each of the GWAS listed above.[73] Overlapping SNPs between the selected GWAS summary statistics and 7,363,646 genotyped or imputed SNPs available for TEDS data were used to generate the PRS. PRS were then calculated using plink 1.9 as the sum of risk alleles weighted by SNP effect sizes and standardized.

Environmental measures

Geographical variables from pollution, census, and landcover data were cross-referenced with the participants' postcodes provided in 2008 (mean age of participants = 13, s.d. = 0.58). We obtained 2008 pollution data for routinely collected particles (PM_{10} and $PM_{2.5}$), nitrogen dioxide (NO_2) and oxides of nitrogen (NO_X) from resources on annual pollution statistics based on 1x1 km grid squares in the UK (https://uk-air.defra.gov.uk/data/pcm-data). These specific pollutants were selected based on relevant literature.[50,74–79]

We included measures of population density, urban/rural binary classification, Townsend deprivation index, regional education levels, and regional levels of low social class based on the 2011 census. Census statistics for output areas (OA, i.e., geographical regions created specifically for collecting census data) from the 2011 census were downloaded from Nomisweb (https://www.nomisweb.co.uk/; ***Supplementary Methods 1***).

Measures of greenspace were based on land cover maps from the Centre of Ecology and Hydrology (CEH). The percentage land cover types are derived from satellite images and digital cartography. There are 23 classes of land cover that encompasses the entire range of UK habitats. The greenspace variable was created by combining percentages for all 21 of the rural land cover classes. Land cover maps used in the current study were generated in 2007 and based on 1x1 km grid squares (https://www.ceh.ac.uk/services/land-cover-map-2007).

Further details on the geographical variables included in the current study are shown in ***Table 1***, including descriptive statistics for the TEDs sample and the means and standard deviations of the age 16 population in England and Wales. T-tests for significant differences between TEDs sample means and age 16 English and Welsh population means were all found to be highly significant with P-values < 0.001 (details in ***Supplementary Methods 2***).

We also included a family socio-economic status (SES) measure, which were collected at first contact (mean age of participants = 18 months). This was calculated based on the mother's

and father's qualification levels and employment status and the mother's age at birth of the first child. A higher score corresponds to higher SES.

Continuous measures were transformed using the Yeo-Johnson power transformation method (***Supplementary Methods 3*** & ***Supplementary Figure 1***). Standardisation was applied after transformation of the data by subtracting the mean and dividing by the standard deviation. Details on all predictor variables included analysed are found in ***Supplementary Table 2***. Correlations between variables are shown in ***Supplementary Figure 2***. We removed variables if they had a correlation coefficient of > 0.8, i.e., PM_{10} and NO_2.

Table 1: Descriptive statistics for geographical environmental variables in TEDs, compared with the age 16 population in England and Wales.

Geographical variable	**Sample mean [SD] or %**	**Population mean [SD] or %**
$PM_{2.5}$	10 [1.8]	11 [2]
Oxides of nitrogen (NO_X)	24 [10.5]	28 [14]
Population density (people per hectare)	37 [32]	54 [62]
Urban classification	69%	81%
Townsend deprivation index (TDI)	-1.8 [2.3]	-0.07 [3.5]
Regional % persons aged 16+ with level-4 qualifications (university degree)	30 [13]	26 [13]
Regional % persons ages 16-74 in the lowest social class	19 [12]	26 [15]
Greenspace (% natural land cover)	48 [31]	41 [32]

Outcome measures

Dependent measures included six quantitative subscales of the Specific Psychotic Experiences Questionnaire (SPEQ) measures at mean age 16.3 (standard deviation = 0.68). These comprised paranoia, hallucinations, cognitive disorganisation, grandiosity, anhedonia, all self-reported, and parent-rated negative symptoms. Questionnaire items for each of the subscales are found in ***Supplementary Table 3*** Further details on how these scales were derived and validated are found elsewhere.[8] Distributions of the psychotic experiences and negative symptom scales are shown in ***Supplementary Figure 3*** and correlations between these six scales in ***Supplementary Figure 4***. Each of the psychotic experiences and parent-rated negative symptom scales were standardized by subtracting the mean and dividing by the standard deviation.

Statistical analyses

First, we performed partial correlations between each of the six psychotic experiences and negative symptom scales and each of the 23 genetic and environmental factors adjusted for

age and sex using the "ppcor" package in R. PRS included in partial correlations and in the subsequent analyses were adjusted for genetic covariates by regressing each PRS on 10 ancestry informed principal components and genotyping chip and using the residuals in analyses.

Next, we built three models for each of the six psychotic experiences and negative symptom scales. These included separately and jointly modelled genetic and environmental models, one with all the PRS (G), one with all the environmental factors (E) and a joint model of both genetic and environmental factors (GE). Modelling was performed in R using "glmnet" and "caret" packages.

We used LASSO (least absolute shrinkage and selection operator) regression,[80] as it reduces overfitting and the sensitivity of the regression coefficients to multi-collinearity. LASSO includes a penalty function that eliminates correlated coefficients, which improves the model in case of collinearity but also conducts automated feature selection. This is especially warranted when modelling multiple geographical measures due to their correlation. To ensure that the age and sex covariates were not removed during the model selection procedure, each of the psychotic experience scales was regressed on age and sex and the residuals were used in the analysis.

We compared the variance explained R-squared (R^2) of the three models to assess whether environmental measures had independent effects on psychotic experience and negative symptoms scales when adjusting for the genetic effects, and vice versa.[55] We used a nested cross-validation procedure with an inner and outer loop for model selection and measuring model performance, respectively. The performance of each model was assessed by computing the average R^2 in the hold-out set for each fold of the outer loop (details in ***Supplementary Methods 4***). Comparisons between the models were made using the William's test to calculate the significant difference in the correlation between the observed and predicted values for each of these models ("paired.r" function from the "psych" package in R).[81,82] Predicted values were the average predictive values across each fold of the outer loop in our nested cross-validation. A Bonferroni adjusted P-value threshold of 0.002 was used (0.05 *P*-value adjusted for 20 tests, 5 model comparisons for each of the six outcomes).

For inspection of the model coefficients, we built final models using the whole dataset and model selection was performed using the inner loop cross-validation procedure. Model coefficients were estimated using post-selection inference methods to adjust for the variable selection procedure. [83] Variables were reported as significant if they had p-values less than a Bonferroni adjusted threshold of 0.008 (0.05 threshold adjusted for each model for the six psychotic experiences scales). Further details in ***Supplementary Methods 4***.

Results

Partial correlations between each of the six psychotic experiences and negative symptom scales and the predictor variables, adjusted for age and sex, are shown in ***Figure 1***. Cognitive disorganisation and paranoia were both significantly positively correlated with neuroticism (partial correlations of 0.08 and 0.06, respectively) and MDD PRS (0.07 and 0.09, respectively).

Cognitive disorganisation and parent-rated negative symptoms were both significantly correlated with genetic and environmental factors relating to education and socio-economic status (***Supplementary Table 5***).

Parent-rated negative symptoms achieved the highest overall variance explained in the nested cross-validation. The joint modelling of genetic and environmental factors (GE model) explained 2.3% of the variance (***Figure 2***; ***Supplementary Table 6***). Nested comparisons between the GE model and the more parsimonious models showed that the E but not the G model contributed significantly to the variance explained in parent-rated negative symptoms (William's test p = 0.002 and 0.05, respectively; ***Supplementary Table 7***). In post-selection inference on the best performing GE model, parent-rated negative symptoms were significantly negatively associated with both a genetic predisposition to education (β = -0.07; 95%CI = -0.12 to -0.04; p = $3x10^{-4}$; ***Figure 3; Supplementary Table 8***) and regional educational attainment (β = -0.06; 95%CI = -0.09 to -0.02; p = 0.004; ***Figure 3; Supplementary Table 8***). Additionally, parent-rated negative symptoms were associated with lower SES (β = -0.07, 95%CI = -0.10 to -0.03; p = 0.002; ***Figure 3; Supplementary Table 8***).

The joint modelling of genetic and environmental factors (GE model) explained 1.3% of the variance in cognitive disorganisation (***Figure 2***; ***Supplementary Table 6***). However, only the G model contributed significantly to the variance explained (William's test p = 0.006). In the best performing GE model, the only significant association for cognitive disorganisation was the neuroticism PRS (β = 0.05; 95%CI = 0.03 to 0.08; p = 0.002; ***Figure 3; Supplementary Tables 9***).

For the paranoia scale, it was found that the best performing E model shrunk all the coefficients to zero, meaning no linear combination of any subset of the included environmental factors were useful for predicting paranoia. The GE model explained 1 % of the variance in paranoia (***Figure 2; Supplementary Table 6***). The paranoia scale was associated with higher PRS for MDD (β = 0.07; 95%CI = 0.04 to 0.1; p = $1x10^{-4}$; ***Figure 3; Supplementary Table 10***, BMI (β = 0.05; 95%CI = 0.02 to 0.08; p = 0.003; ***Figure 3; Supplementary Table 10***), and neuroticism (β = 0.05; 95%CI = 0.02 to 0.08; p = 0.008; ***Figure 3; Supplementary Table 10***).

For the hallucinations scale, the GE model achieved the highest variance explained (median R^2 = 0.4%; ***Figure 2***; ***Supplementary Table 6***). Model comparisons indicated that neither the E nor G model contributed significantly to the variance explained (William's test p = 0.09 and 0.05, respectively; ***Supplementary Table 7***). Furthermore, we found no specific significant associations with the hallucinations scale (***Figure 3***; ***Supplementary Table 11***). The grandiosity and anhedonia scales had no significant partial correlations for any of the included genetic and environmental factors. Furthermore, in the LASSO models it was found that the best performing model shrunk all the coefficients to zero, meaning no linear combination of any subset of the included predictors were useful for predicting these two outcomes.

Discussion

In this study, we investigated the association of sets of polygenic scores and geographical environmental risk factors with six psychotic experiences and negative symptom scales in adolescents. Our findings strengthen previous literature demonstrating psychotic experiences' transdiagnostic associations in mental health through highlighting differential associations between PRS and the psychotic experience subscales, i.e., parent-rated negative symptoms with educational attainment and subjective well-being PRS; cognitive disorganisation with neuroticism PRS; paranoia with BMI, MDD and neuroticism PRS. We also show that family SES and environmental measures of regional educational attainment are associated with parent-rated negative symptoms. These results emphasize the value of studying psychotic experience subscales separately. Furthermore, the results are from models including both genetic and environmental factors, demonstrating that the effects remain significant when adjusting for potential gene-environment correlation.

The highest variance explained was seen for parent-rated negative symptom scale, which was associated with lower regional educational attainment and lower family socio-economic (SES) measures. This is supported by previous research showing an association of negative symptoms with lower SES and educational attainment.[86] The PRS for educational attainment was also significantly associated with parent-rated negative symptoms,[66] suggesting that the PRS influences negative symptoms independently of the educational attainment phenotype. However, nested model comparisons demonstrated that only environmental factors had a significant impact on the model's variance explained.

The second highest variance explained was seen for cognitive disorganisation, with only genetic factors impacting on the prediction model. Despite the obvious difference of this scale with the disorganisation dimension derived from factor analysis of cases with psychosis, previous evidence suggests that disorganisation has the highest heritability and includes influences that are independent of psychosis liability.[84] The predominant association for cognitive disorganisation with neuroticism PRS is supported by a recent study demonstrating a shared genetic aetiology between cognitive disorganisation and childhood emotional and behavioural problems.[85]

Comparing the hallucinations and paranoia scales, within the context of the current study, we notice that hallucinations appear to be affected by both the genetic and environmental factors, while paranoia only by the genetic factors examined. This is consistent with previous findings in the same sample showing that paranoia has the highest heritability.[87] Furthermore, in the same study, shared environment only had a significant influence on hallucinations and negative symptom scales,[87] which is supported by the association of these scales with regional education levels and SES seen here.

Contrary to the hypothesis that psychotic experiences are part of a 'psychosis continuum',[10] schizophrenia PRS was not associated significantly with any of the psychotic experiences or negative symptom scales, even though it is the most powerful mental health PRS in predicting case-control status. This adds to previous studies on this topic which provided mixed evidence to date on the degree of association between schizophrenia PRS and psychotic experiences or showed stronger association with PRS for other psychiatric disorders.[23,42,88] However, most of these studies examined older individuals, which may have different psychotic experience profile. Using multiple PRS and the LASSO shrinkage method, our study provides evidence

that if psychotic experiences in adolescence are associated with the schizophrenia PRS, the strength of the association is weaker than other disorder-trait genetic associations.[89]

Here, we did not replicate associations between psychotic experiences and air pollution that have previously been reported.[50] However, in the previous study, more fine-scale pollution measures were used for multiple frequently visited addresses, which may account for the differences in the reported associations. Another potential explanation for this disparity is the inconsistent definition of psychotic experiences across studies, which include single composite scores and, three- to six- factor dimensions.[11–15,90] It has also been shown that some co-occurring psychotic experiences, such as cognitive disorganisation and negative symptoms, are associated with more schizophrenia-relevant variables.[86] Prevalence of psychotic experiences in adolescence, depending on its definition, range from 7% to 95%,[14,91] which is an indicator of the challenge in defining these outcomes and may be the reason for the inconsistency in study findings.[41] A systematised definition of psychotic experiences in non-clinical populations is necessary to achieve replicability and generalizability of research findings.

There are several limitations to this study. Geographical variables were linked to the participants' addresses in 2008, around three years prior to the collection of the psychotic experience questionnaires, so did not include the effect of early life or cumulative exposure to these factors. Furthermore, air pollution data was linked to the participants' home addresses, so there was no information on exposures at school or other frequently visited locations, which may lead to exposure misclassification. We also did not include any information on indoor air pollution or smoking status or, more importantly, cannabis use, a known risk factor for psychosis.[92] Generally, our results are limited by the included genetic and environmental variables as many other unmeasured variables may influence psychotic experiences over and above those that were included. The predictive ability of the PRS included are limited by the power of the training GWAS and we may see different results when PRS explain more of the total genetic variation. Furthermore, the psychotic experience and negative symptom questionnaire measures current symptoms rather than lifetime symptoms, which may limit power to detect effects. TEDS is a twin cohort and results may not generalize, although it has been shown that there is a similar prevalence of psychotic experiences in twins vs non-twins cohorts.[87] Selection bias may also be present as our cohort had better education, higher SES, and lower exposure to pollutants than the average population of the same age. Finally, with regards to the inherent assumptions of the linear regression method used in the current study, some may be violated, including 1) regression coefficients reflect unconditional relationships only (i.e., no interaction effects between predictors) and 2) normality and homoscedasticity of the residuals was not directly tested. Furthermore, the models developed in the current study were not externally validated, thus it is unknown how well the model coefficients and R^2 will generalize to other samples.

In conclusion, we find that both genetic risk and geographical environmental factors contribute significantly to the reporting of psychotic experiences in adolescence. The differential predictive ability of specific genetic and environmental risk factors with the six specific psychotic experiences highlights the value of studying these domains separately. We can conclude that cognitive disorganisation and negative symptoms during mid-adolescence were most predicted from the genetic and environmental risk factors examined in our study.

Adolescence is known to be a critically vulnerable stage for mental health when most mental health conditions begin. Identifying the genetic and environmental risk factors associated with red flags for poor mental health in adolescence will help to identify suitable targets for early intervention programmes.

Conflict of interest

The Authors have declared that there are no conflicts of interest in relation to the subject of this study.

Acknowledgements

TEDS is supported by a programme grant to Professor Robert Plomin from the UK Medical Research Council (MR/V012878/1 and previously MR/M021475/1), with additional support from the US National Institutes of Health (AG046938). We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ grant agreement n° 602768. K.R. is supported by a Sir Henry Wellcome Postdoctoral Fellowship. This work contains National Statistics data. Crown copyright and database right (2021). Source: Office for National Statistics licensed under the Open Government Licence v.3.0. LCM2007 and database right NERC 2011; all rights reserved. This work contains data from Defra, UK. This report represents independent research partially funded by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

References

1. Van Os, J., Hanssen, M., Bijl, R. V. & Ravelli, A. Strauss (1969) revisited: a psychosis continuum in the general population? *Schizophr. Res.* **45**, 11–20 (2000).
2. Van Os, J. & Linscott, R. J. Introduction: The Extended Psychosis Phenotype—Relationship With Schizophrenia and With Ultrahigh Risk Status for Psychosis. *Schizophr. Bull.* **38**, 227 (2012).
3. Jauhar, S., Johnstone, M. & McKenna, P. J. Schizophrenia. *Lancet* **399**, 473–486 (2022).
4. Liddle, P. F. The symptoms of chronic schizophrenia. A re-examination of the positive-negative dichotomy. *Br. J. Psychiatry* **151**, 145–151 (1987).
5. Andreasen, N. C., Arndt, S., Alliger, R., Miller, D. & Flaum, M. Symptoms of schizophrenia. Methods, meanings, and mechanisms. *Arch. Gen. Psychiatry* **52**, 341–351 (1995).
6. Grube, B. S., Bilder, R. M. & Goldman, R. S. Meta-analysis of symptom factors in schizophrenia. *Schizophr. Res.* **31**, 113–120 (1998).
7. Shafer, A. & Dazzi, F. Meta-analysis of the positive and Negative Syndrome Scale (PANSS) factor structure. *J. Psychiatr. Res.* **115**, 113–120 (2019).

8. Ronald, A. *et al.* Characterization of psychotic experiences in adolescence using the specific psychotic experiences questionnaire: Findings from a study of 5000 16-Year-Old Twins. *Schizophr. Bull.* **40**, 868–877 (2014).
9. Wigman, J. T. W. *et al.* Replication of the five-dimensional structure of positive psychotic experiences in young adulthood. *Psychiatry Res.* **197**, 353–355 (2012).
10. Van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P. & Krabbendam, L. A systematic review and meta-analysis of the psychosis continuum: Evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychol. Med.* **39**, 179–195 (2009).
11. Vollema, M. G. & Hoijtink, H. The multidimensionality of self-report schizotypy in a psychiatric population: an analysis using multidimensional Rasch models. *Schizophr. Bull.* **26**, 565–575 (2000).
12. Cyhlarova, E. & Claridge, G. Development of a version of the Schizotypy Traits Questionnaire (STA) for screening children. *Schizophr. Res.* **80**, 253–261 (2005).
13. Fonseca-Pedrero, E., Paíno-Piñeiro, M., Lemos-Giráldez, S., Villazón-García, Ú. & Muñiz, J. Validation of the Schizotypal Personality Questionnaire-Brief Form in adolescents. *Schizophr. Res.* **111**, 53–60 (2009).
14. Wigman, J. T. W. *et al.* The Structure of The Extended Psychosis Phenotype in Early Adolescence—A Cross-sample Replication. *Schizophr. Bull.* **37**, 850 (2011).
15. Ericson, M., Tuvblad, C., Raine, A., Young-Wolff, K. & Baker, L. A. Heritability and longitudinal stability of schizotypal traits during adolescence. *Behav. Genet.* **41**, 499–511 (2011).
16. McGrath, J. J. *et al.* The bidirectional associations between psychotic experiences and DSM-IV mental disorders. *Am. J. Psychiatry* **173**, 997–1006 (2016).
17. Bromet, E. J. *et al.* Association Between Psychotic Experiences and Subsequent Suicidal Thoughts and Behaviors: A Cross-National Analysis From the World Health Organization World Mental Health Surveys. *JAMA psychiatry* **74**, 1136–1144 (2017).
18. Navarro-Mateu, F. *et al.* The association between psychotic experiences and disability: results from the WHO World Mental Health Surveys. *Acta Psychiatr. Scand.* **136**, 74–84 (2017).
19. Kelleher, I. & Cannon, M. Psychotic-like experiences in the general population: characterizing a high-risk group for psychosis. *Psychol. Med.* **41**, 1–6 (2011).
20. Morrison, A. P. *et al.* Cognitive therapy for the prevention of psychosis in people at ultra-high risk: Randomised controlled trial. *Br. J. Psychiatry* **185**, 291–297 (2004).
21. Linscott, R. J. & Van Os, J. An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: On the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychological Medicine* **43**, 1133–1149 (2013).
22. Ronald, A. & Pain, O. A systematic review of genome-wide research on psychotic experiences and negative symptom traits: new revelations and implications for psychiatry. *Hum. Mol. Genet.* **27**, R136–R152 (2018).
23. Pain, O. *et al.* Genome-wide analysis of adolescent psychotic-like experiences shows genetic overlap with psychiatric disorders. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **177**, 416–425 (2018).
24. McGrath, J., Saha, S., Chant, D. & Welham, J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol. Rev.* **30**, 67–76 (2008).
25. Cederlöf, M. *et al.* A longitudinal study of adolescent psychotic experiences and later

development of substance use disorder and suicidal behavior. *Schizophr. Res.* **181**, 13–16 (2017).
26. Fisher, H. L. *et al.* Specificity of childhood psychotic symptoms for predicting schizophrenia by 38 years of age: A birth cohort study. *Psychol. Med.* **43**, 2077–2086 (2013).
27. Kelleher, I., Cederlöf, M. & Lichtenstein, P. Psychotic experiences as a predictor of the natural course of suicidal ideation: A Swedish cohort study. *World Psychiatry* **13**, 184–188 (2014).
28. Kelleher, I. *et al.* Psychotic symptoms in adolescence index risk for suicidal behavior: Findings from 2 population-based case-control clinical interview studies. *Arch. Gen. Psychiatry* **69**, 1277–1283 (2012).
29. Werbeloff, N. *et al.* Self-reported attenuated psychotic symptoms as forerunners of severe mental disorders later in life. *Arch. Gen. Psychiatry* **69**, 467–475 (2012).
30. Healy, C. *et al.* Childhood and adolescent psychotic experiences and risk of mental disorder: a systematic review and meta-analysis. *Psychol. Med.* **49**, 1589–1599 (2019).
31. Hay, D. A. *et al.* Phenotypic and Genetic Analyses of a Short Measure of Psychosis-proneness in a Large-scale Australian Twin Study. *Twin Res.* **4**, 30–40 (2003).
32. Hur, Y. M., Cherny, S. S. & Sham, P. C. Heritability of hallucinations in adolescent twins. *Psychiatry Res.* **199**, 98–101 (2012).
33. Linney, Y. M. *et al.* A quantitative genetic analysis of schizotypal personality traits. *Psychol. Med.* **33**, 803–816 (2003).
34. Polanczyk, G. *et al.* Etiological and clinical features of childhood psychotic symptoms results from a birth cohort. *Arch. Gen. Psychiatry* **67**, 328–338 (2010).
35. Wigman, J. T. W. *et al.* A twin study of genetic and environmental determinants of abnormal persistence of psychotic experiences in young adulthood. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **156**, 546–552 (2011).
36. Zavos, H. M. S. *et al.* Consistent etiology of severe, frequent psychotic experiences and milder, less frequent manifestations: A twin study of specific psychotic experiences in adolescence. *JAMA Psychiatry* **71**, 1049–1057 (2014).
37. Sieradzka, D. *et al.* Heritability of Individual Psychotic Experiences Captured by Common Genetic Variants in a Community Sample of Adolescents. *Behav. Genet.* **45**, 493–502 (2015).
38. Taylor, M. J. *et al.* Association of Genetic Risk Factors for Psychiatric Disorders and Traits of These Disorders in a Swedish Population Twin Sample. *JAMA Psychiatry* **76**, 280–289 (2019).
39. Jones, H. J. *et al.* Phenotypic manifestation of genetic risk for schizophrenia during adolescence in the general population. *JAMA Psychiatry* **73**, 221–228 (2016).
40. Sieradzka, D. *et al.* Are Genetic Risk Factors for Psychosis Also Associated with Dimension-Specific Psychotic Experiences in Adolescence? *PLoS One* **9**, e94398 (2014).
41. Hinterbuchinger, B. & Mossaheb, N. Psychotic-Like Experiences: A Challenge in Definition and Assessment. *Front. Psychiatry* **12**, 270 (2021).
42. Legge, S. E. *et al.* Association of Genetic Liability to Psychotic Experiences with Neuropsychotic Disorders and Traits. *JAMA Psychiatry* **76**, 1256–1265 (2019).
43. Barkhuizen, W., Pain, O., Dudbridge, F. & Ronald, A. Genetic overlap between psychotic experiences in the community across age and with psychiatric disorders.

Transl. Psychiatry **10**, (2020).
44. Krapohl, E. *et al.* Multi-polygenic score approach to trait prediction. *Mol. Psychiatry* **23**, 1368–1374 (2018).
45. Vassos, E., Pedersen, C. B., Murray, R. M., Collier, D. A. & Lewis, C. M. Meta-Analysis of the Association of Urbanicity With Schizophrenia. *Schizophr. Bull.* **38**, 1118–1123 (2012).
46. Newbury, J. *et al.* Why Are Children in Urban Neighborhoods at Increased Risk for Psychotic Symptoms? Findings from a UK Longitudinal Cohort Study. *Schizophr. Bull.* **42**, 1372–1383 (2016).
47. Singh, S. P., Winsper, C., Wolke, D. & Bryson, A. School mobility and prospective pathways to psychotic-like symptoms in early adolescence: A prospective birth cohort study. *J. Am. Acad. Child Adolesc. Psychiatry* **53**, 518-527.e1 (2014).
48. Newbury, J. B. *et al.* In the eye of the beholder: Perceptions of neighborhood adversity and psychotic experiences in adolescence. *Dev. Psychopathol.* **29**, 1823–1837 (2017).
49. Newbury, J. *et al.* Cumulative Effects of Neighborhood Social Adversity and Personal Crime Victimization on Adolescent Psychotic Experiences. *Schizophr. Bull.* **44**, 348–358 (2018).
50. Newbury, J. B. *et al.* Association of air pollution exposure with psychotic experiences during adolescence. *JAMA Psychiatry* **76**, 614–623 (2019).
51. Colodro-Conde, L. *et al.* Association Between Population Density and Genetic Risk for Schizophrenia. *JAMA Psychiatry* **75**, 901 (2018).
52. Maxwell, J. M., Coleman, J. R. I., Breen, G. & Vassos, E. Association Between Genetic Risk for Psychiatric Disorders and the Probability of Living in Urban Settings. *JAMA Psychiatry* (2021). doi:10.1001/JAMAPSYCHIATRY.2021.2983
53. Shakoor, S. *et al.* Association between stressful life events and psychotic experiences in adolescence: evidence for gene-environment correlations. *Br. J. Psychiatry* **208**, 532–538 (2016).
54. Shakoor, S. *et al.* A shared genetic propensity underlies experiences of bullying victimization in late childhood and self-rated paranoid thinking in adolescence. *Schizophr. Bull.* **41**, 754–763 (2015).
55. Allegrini, A. G. *et al.* Multivariable G-E interplay in the prediction of educational achievement. *PLOS Genet.* **16**, e1009153 (2020).
56. Haworth, C. M. A., Davis, O. S. P. & Plomin, R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Res. Hum. Genet.* **16**, 117–125 (2013).
57. Rimfeld, K. *et al.* Twins Early Development Study: A Genetically Sensitive Investigation into Behavioral and Cognitive Development from Infancy to Emerging Adulthood. *Twin Res. Hum. Genet.* **22**, 508–513 (2019).
58. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
59. Selzam, S. *et al.* Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. *PLoS Genet.* **14**, e1007757 (2018).
60. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
61. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and

implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).
62. Purves, K. L. *et al.* A major role for common genetic variation in anxiety disorders. *Mol. Psychiatry* (2019). doi:10.1038/s41380-019-0559-1
63. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
64. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
65. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
66. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
67. van den Berg, S. M. *et al.* Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behav. Genet.* **46**, 170–182 (2016).
68. Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).
69. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nat. 2022 6047906* **604**, 502–508 (2022).
70. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
71. Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
72. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
73. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1–10 (2019).
74. Pun, V. C., Manjourides, J. & Suh, H. Association of Ambient Air Pollution with Depressive and Anxiety Symptoms in Older Adults: Results from the NSHAP Study. *Environ. Health Perspect.* **125**, 342–348 (2017).
75. Szyszkowicz, M., Kousha, T., Kingsbury, M. & Colman, I. Air Pollution and Emergency Department Visits for Depression: A Multicity Case-Crossover Study. *Environ. Health Insights* **10**, (2016).
76. Klompmaker, J. O. *et al.* Associations of combined exposures to surrounding green, air pollution and traffic noise on mental health. *Environ. Int.* **129**, 525–537 (2019).
77. Antonsen, S. *et al.* Exposure to air pollution during childhood and risk of developing schizophrenia: a national cohort study. *Lancet Planet. Heal.* **4**, e64–e73 (2020).
78. Latham, R. M. *et al.* Childhood exposure to ambient air pollution and predicting individual risk of depression onset in UK adolescents. *J. Psychiatr. Res.* **138**, 60–67 (2021).
79. Tzivian, L. *et al.* Effect of long-term outdoor air pollution and noise on cognitive and psychological functions in adults. *International Journal of Hygiene and Environmental Health* **218**, 1–11 (2015).

80. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
81. Williams, E. J. The Comparison of Regression Variables. *J. R. Stat. Soc. Ser. B* **21**, 396–399 (1959).
82. Steiger, J. H. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* **87**, 245–251 (1980).
83. Liu, K., Markovic, J. & Tibshirani, R. More powerful post-selection inference, with application to the Lasso. *arXiv* (2018).
84. Rijsdijk, F. V., Gottesman, I. I., Mcguffin, P. & Cardno, A. G. Heritability estimates for psychotic symptom dimensions in twins with psychotic disorders. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **156B**, 89–98 (2011).
85. Shakoor, S., McGuire, P., Cardno, A. G., Freeman, D. & Ronald, A. A twin study exploring the association between childhood emotional and behaviour problems and specific psychotic experiences in a community sample of adolescents. *J. Child Psychol. Psychiatry.* **59**, 565 (2018).
86. Cardno, A. G., Selzam, S., Freeman, D. & Ronald, A. Psychotic-Like Experiences in Adolescence Occurring in Combination or Isolation: Associations with Schizophrenia Risk Factors. *Psychiatr. Res. Clin. Pract.* **3**, 67–75 (2021).
87. Zavos, H. M. S. *et al.* Consistent etiology of severe, frequent psychotic experiences and milder, less frequent manifestations: A twin study of specific psychotic experiences in adolescence. *JAMA Psychiatry* **71**, 1049–1057 (2014).
88. García-González, J. *et al.* The effects of polygenic risk for psychiatric disorders and smoking behaviour on psychotic experiences in UK Biobank. *Transl. Psychiatry* **10**, 1–10 (2020).
89. Ronald, A. Common (Genetic) Links Between Clinics and the Community: New Evidence From a Tourette Syndrome Polygenic Score. *Biol. Psychiatry* **85**, 281–282 (2019).
90. Ronald, A. *et al.* Characterization of psychotic experiences in adolescence using the specific psychotic experiences questionnaire: Findings from a study of 5000 16-Year-Old Twins. *Schizophr. Bull.* **40**, 868–877 (2014).
91. Kelleher, I. *et al.* Clinicopathological significance of psychotic experiences in non-psychotic young people: Evidence from four population-based studies. *Br. J. Psychiatry* **201**, 26–32 (2012).
92. Marconi, A., Di Forti, M., Lewis, C. M., Murray, R. M. & Vassos, E. Meta-analysis of the Association Between the Level of Cannabis Use and Risk of Psychosis. *Schizophr. Bull.* **42**, 1262–1269 (2016).

Figure legends

***Figure 1**: Partial correlations estimated using Pearson's correlation and marked as statistically significant with an asterix (P-value < 0.0004).*

***Figure 2**: Distribution of hold-out set variance explained (R^2) by the genetic (G), environmental (E), and joint genetic and environmental (GE) models estimated in a nested cross-validation procedure. The median and interquartile range of the R^2 from the 500-fold outer loop of the nested cross-validation are plotted in boxplots with the top and bottom whiskers set at the 97.5^{th} and 2.5^{th} percentile.*

***Figure 3**: Coefficient estimation for genetic and environmental associations with six psychotic experience and negative symptom scales. Best GE model selected via 10-fold cross validation repeated 100 times. Coefficients were estimated using post-selection inference analysis.*