Phenome-wide associations with polygenic scores for schizophrenia and major depression in 100,000 Chinese adults

Baihan Wang[1], Sam Morris[1], Hannah Fry[1], Andri Iona[1], Jonathan Clarke[1], Kuang Lin[1], Christiana Kartsonaki[1], Derrick A Bennett[1], Yiping Chen[1], Huaidong Du[1], Ling Yang[1], Daniel Avery[1], Dan Schmidt-Valle[1], Shixian Feng[2], Dianjianyi Sun[3,4,5], Canqing Yu[3,4,5], Jun Lv[3,4,5], Pei Pei[4], Junshi Chen[6], Karoline Kuchenbaecker[7], Liming Li[3,4,5], Robin G Walters[1], Zhengming Chen[1], Iona Y Millwood[1], on behalf of China Kadoorie Biobank Collaborative Group

1. Clinical Trial Service Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
2. NCDs Prevention and Control Department, Henan CDC, Zhengzhou, China
3. Department of Epidemiology & Biostatistics, School of Public Health, Peking University, Beijing, China
4. Peking University Center for Public Health and Epidemic Preparedness and Response, Beijing, China
5. Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China
6. China National Center for Food Safety Risk Assessment, Beijing, China
7. Division of Psychiatry, University College London, London, UK

Address for correspondence

Dr Iona Y Millwood
NDPH
University of Oxford
Old Road Campus
Roosevelt Drive
Oxford OX3 7LF
iona.millwood@ndph.ox.ac.uk

Word count: abstract 150; main text 5,152

1 table, 4 figures, 5 supplementary tables, and 11 supplementary figures

03 March 2025

Abstract (150 words)

Phenome-wide association studies of polygenic scores (PGS) can reveal shared genetic influences between mental disorders and other traits. Based on summary statistics in East Asian (EAS) and European (EUR) ancestries, we tested associations of PGS for schizophrenia and major depression with 254 phenotypes in 100,640 Chinese adults. PGS predicted schizophrenia (R^2=1.96%-3.49%) and major depression (R^2=0.19%-0.55%) and were associated with various socio-demographic, lifestyle, and physical measures, including poor subjective health and lower household size. The EAS-schizophrenia-PGS was inversely associated with smoking initiation, and the EAS-depression-PGS was inversely associated with BMI. EAS and EUR populations had opposing genetic correlations for smoking-schizophrenia (inverse-EAS, positive-EUR) and BMI-depression (inverse-EAS, positive-EUR). Causality of these relationships was supported by Mendelian Randomisation in EUR, but multivariable analyses suggested pleiotropic effects on other related traits influenced the findings. Our study suggests the context specificity of relationships between mental disorders and other traits, highlighting a potential role of sociocultural factors.

Keywords: *PheWAS, schizophrenia, major depression, polygenic score, mental health*

Introduction

Mental disorders impose a significant burden on the global population, accounting for 80 million disability-adjusted life years (DALYs) in the past three decades[1,2]. In Asia, mental disorders ranked as the 8th leading cause of disease burden in 2019, increasing from 13th in 1990[3]. China accounts for 16% of the global burden of mental disorders[2,4,] creating challenges for China's national health care system. The number of DALYs due to mental disorders in China has been rising steadily, with depressive disorders, anxiety disorders, and schizophrenia (SCZ) contributing to the majority of the burden[3,5]. However, there is still limited recognition of mental disorders by the general public in China, leading to barriers to diagnosis and treatment.[6] The resources available for mental health care in China are also considerably lower than those in other high-income countries, with uneven distributions across urban and rural areas[7,8].

Many mental disorders show significant heritability, with SCZ and major depression (MD) having around 80% and 30% heritability, respectively[9,10]. Breakthroughs in understanding the genetic basis of SCZ and MD have been made by genome-wide association studies (GWAS) in diverse populations[11–14]. The genetic correlation between East Asian ancestry (EAS) and European ancestry (EUR) populations has been estimated to be substantial (> 0.9) for both SCZ and MD[12–14], indicating largely overlapping genetic bases across populations for these disorders. Nevertheless, important differences still exist between the two ancestries. For example, Chen et al. reported larger genetic differences between EAS and EUR for variants associated with SCZ at genome-wide significance, compared to variants randomly selected from the whole genome [15]. Moreover, unique associations between mental disorders and other traits have been reported in EAS that were not previously discovered in EUR. For example, using Mendelian Randomisation (MR), O'Loughlin et al. reported that lower body mass index (BMI) was causally associated with a higher risk of MD in the Chinese population, which was opposite in direction to the causal association in EUR populations[16]. This suggests that sociocultural factors potentially influence the relationships between mental disorders and other health-related traits[16].

Polygenic scores (PGS) are the sum of trait-associated variants weighted by their effect sizes and broadly represent an individual's genetic predisposition for a given

trait[17,18]. They have been found to be predictive of mental disorders in clinical and population-based studies.[19,20] PGS can also be employed in phenome-wide association studies (PheWAS), to reveal associations between the genetic risk for mental disorders and a range of other traits across the human phenome.[21,22] For example, previous PheWAS conducted in Western and EUR populations reported associations of the PGS for SCZ or MD with other mental disorders, physical illnesses, behavioural traits, and brain structure measures[23–26], suggesting shared genetic mechanisms and new modifiable risk factors.

Nevertheless, few PheWAS using the PGS for SCZ or MD have been conducted in the Chinese population. Such studies will help discover setting-specific associations, clarify disease aetiology, and inform public health policies. Therefore, the aims of the current study are: 1) to identify associations of the genetic risk for SCZ and MD with a wide range of phenotypes measured in 100,640 Chinese adults; 2) to examine the consistency of these associations in both EAS and EUR populations with genetic correlation analyses; 3) to assess the causal relevance of these associations using univariable and multivariable MR.

Methods

Study population

The China Kadoorie Biobank (CKB) is a prospective cohort study with > 512,000 adults aged 30–79 years, recruited during 2004-08 from 10 regions across urban and rural areas in China[27]. At baseline, participants were assessed by laptop-based questionnaires on socio-demographic characteristics, medical history, and lifestyle habits. Physical measurements (e.g. anthropometry, blood pressure, and heart rate) were also collected at baseline. After the baseline survey, the long-term health of the participants was monitored via linkage with local death or disease registries, as well as the national health insurance systems recording all hospitalisation episodes. The present study is based on follow-up to 1st January 2019 (median = 12.36 [Q1 – Q3: 11.20 – 13.27] years follow-up).

CKB complies with all the required ethical standards for medical research on human subjects. Ethical approvals were granted and have been maintained by the relevant

institutional ethical research committees in the UK and China. Written informed consent was obtained from all participants.

Genotyping

A total of 100,640 CKB participants were genotyped using a custom Affymetrix array and included in the current study[28]. Of these participants, 23,518 were cases of cardiovascular disease (CVD) or chronic obstructive pulmonary disease (COPD) selected for a nested case-control study, while the remaining 77,122 were randomly selected and population-representative. Details of genotyping and quality control procedures in CKB have been previously described[28]. Genotyped variants were pre-phased using SHAPEIT v4.2 (SHAPEIT v2.904 for chromosome X)[29] and uploaded to the TOPMed[30] or Westlake Biobank for Chinese[31] server for imputation. Two sets of the imputed data were merged by selecting the imputed genotype with a higher imputation INFO score for each variant.

Polygenic score computation

We computed PGS for SCZ and MD based on two types of GWAS summary statistics: one from GWAS conducted in EAS; and the other from both GWAS in EAS and GWAS in EUR. This is because variants discovered by GWAS in EAS may include population- and setting-specific genetic associations, while incorporating GWAS in multi-ancestries has been found to improve the overall predictive power of PGS[32]. As a result, four different PGS were computed for each participant in CKB: PGS-SCZ-EAS, PGS-SCZ-multi (EAS+EUR), PGS-MD-EAS, and PGS-MD-multi (EAS+EUR).

GWAS summary statistics for SCZ[11,13] and MD[12,14] in EAS and EUR were downloaded from the Psychiatric Genomics Consortium (PGC) website (https://pgc.unc.edu/). Since CKB contributed to the GWAS on MD in EAS, the GWAS summary statistics excluding CKB data were used here to ensure no sample overlap. We used PRS-CSx[32,33] to infer the posterior SNP effect sizes of variants via a continuous shrinkage prior within EAS, or via a shared prior across EAS and EUR. We used samples from the EAS and EUR superpopulations in the 1000 Genomes Project[34] as linkage disequilibrium (LD) references. Only HapMap3 variants with a minor allele frequency (MAF) > 0.01 in the relevant ancestry were included in the analysis, as recommended by PRS-CSx[32,33,35].

Based on the SNP effect sizes generated by PRS-CSx, we then used PLINK[36] to compute PGS for all 100,640 genotyped participants in CKB. The PGS were then standardised based on the means and SDs among the 77,122 population-representative CKB participants.

Associations with schizophrenia and major depression

We first tested the associations between the newly computed PGS and their corresponding phenotypes (SCZ and MD) in CKB. Cases of SCZ were identified based on hospitalisation records at follow-up (ICD-10 code: F25). Cases of MD were identified based on both the Composite International Diagnostic Inventory-short form (CIDI-SF) completed at baseline, and hospitalisation records at follow-up (ICD-10 codes: F32, F33, F34.1, and F38.1). With SCZ/MD as the outcome variable, we ran logistic regression analyses with the standardised PGS for SCZ/MD as the main explanatory variable, including age, age^2, sex, study region, and the first 11 genomic principal components as covariates. To assess associations in different risk groups, the PGS were further split into quartiles to compute odds ratios with the first quartile as the reference.

Phenome-wide association studies

We included 254 phenotypes measured at baseline or during follow-up in the PheWAS. These included 67 baseline characteristics capturing socio-demographics, physical health, mental health and sleep, clinical measurements, lifestyle, and female reproductive history. Categorical variables with more than two levels were converted to binary variables, as indicated by their names in Supplementary Data 1 and 2. Hospitalisation episodes for diseases and conditions at follow-up were originally recorded as ICD-10 codes. We then used phecodes (version 1.2)[37] to curate the ICD-10 codes and generate 187 clinically defined phecodes in CKB. To maximise power, we analysed the three-digit parent codes in the PheWAS and only kept phecodes with > 100 cases. This resulted in 80% power to detect an OR of 1.5 per SD higher PGS at $p < 0.001$.

We applied logistic and linear regression models to assess the associations between PGS and categorical and continuous phenotypes, respectively, in CKB. Age, age^2, sex, study region, and the first 11 genomic principal components were included as covariates. To maximise power and avoid potential biases, we identified cases of all

disease-related phenotypes from the overall dataset (including both the selected CVD/COPD cases and the population-representative participants), while controls were only identified from the population-representative subset (Supplementary Figure 1). For each phecode, we excluded controls with related conditions, as defined by its phecode exclude range[37]. Analyses of all other phenotypes were restricted to the population-representative participants. Multiple testing was corrected with a false discovery rate = 0.05. The same procedure was repeated in sex-specific analyses, for all phenotypes were included in the overall analysis. For the associations identified in the PheWAS, we also performed sensitivity analyses by including household size, household income, ownership index, education level, BMI, smoking, and alcohol drinking as additional covariates.

Genetic correlations

We also estimated the genetic correlations, based on GWAS summary statistics, for mental disorder PGS-phenotype pairs that showed associations in the PheWAS. This was to examine if the findings in CKB were generalizable to the larger EAS population, and if the associations were consistent between EAS and EUR. We downloaded the GWAS summary statistics for those phenotypes from published meta-analyses, genomics consortia, or other cohorts/biobanks, if they were publicly available in both EAS and EUR. For GWAS in EAS, we selected studies that were not conducted in CKB, or meta-analyses that included CKB as only one of the many contributing cohorts. These GWAS summary statistics are summarised in Supplementary Table 1. All GWAS summary statistics were processed and harmonised with the GWASLab package[38].

We used LDSC[39] to compute the genetic correlations (r_g) for all mental disorder-phenotype pairs within each ancestry. We also performed additional analyses using Popcorn[40] to compute the genetic-effect correlations (ρ_{ge}) across EAS and EUR. All genetic correlation analyses were conducted with variants with MAF > 0.01 and present in HapMap3[35], based on EAS and EUR LD references from the 1000 Genomes Project[34].

Mendelian Randomisation

For mental disorder-phenotype pairs that showed associations in both the PheWAS and genetic correlation analyses, we used MR to assess their causal relevance.

There are three core assumptions of MR: 1) the genetic variants are associated with the exposure (relevance assumption); 2) there are no unmeasured confounders between the genetic variants and the outcome (independence assumption); 3) the genetic variants do not affect the outcome expect via the exposure (exclusion restriction assumption).

We conducted bi-directional two-sample MR[41] using the TwoSampleMR[42] package. To maximise power and distinguish MR from the PheWAS, we performed summary-level MR based on publicly available GWAS summary statistics instead of individual-level MR. To assess MR assumption 1), we calculated the mean F-statistics for each exposure to assess instrument strength. To meet MR assumption 2), we performed MR with exposure and outcome GWAS from the same ancestry, separately for EAS and EUR, to avoid confounding due to ancestry. To remove correlated variants, we performed LD clumping with window = 10000 kb and $r^2 = 0.001$ based on EAS/EUR LD references from the 1000 Genomes Project[34]. Due to the relatively small sample sizes of GWAS in EAS, we selected variants that reached suggestive genome-wide significance ($p = 1 \times 10^{-5}$) as genetic instruments in the main analyses[41]. We also performed sensitivity analyses with variants that reached genome-wide significance ($p = 5 \times 10^{-8}$) as genetic instruments in EAS. Genetic instruments in EUR were selected only if they reached genome-wide significance. We used four different methods to perform the two-sample MR: inverse-variance weighting (IVW; main analyses)[41]; MR-Egger[43]; weighted median[44]; weighted mode[45].

To assess MR assumption 3), we further performed summary-level multivariable MR[46] using the TwoSampleMR[42] and MVMR[47] packages to account for potential pleiotropic effects on other phenotypes (i.e. BMI, smoking, educational attainment, and poverty) in EUR. This was done by adding each of the extra phenotypes as an additional exposure, and then comparing the multivariable MR results with the univariable MR results. To conduct multivariable MR, we first performed LD clumping for all pairs of exposures with window = 10000 kb, $r^2 = 0.001$, and $p = 5 \times 10^{-8}$, based on the EUR LD reference from the 1000 Genomes Project[34]. We then estimated the pairwise covariances between each instrument and its corresponding pairs of exposures, based on the phenotypic correlations between exposures in the UK Biobank (Supplementary Note 2; Supplementary Table 5). We also estimated conditional F-statistics to test instrument strength and Q-statistics to test horizontal

pleiotropy[46]. The main multivariable MR analysis was conducted using an IVW model, taking into account the covariance matrix between exposures. Additionally, we used Q-statistic minimisation to re-estimate beta, which is robust to weak instruments and pleiotropy[46]. All univariable and multivariable MR analyses were conducted in R-4.4.1[48].

Results

PGS associations with schizophrenia and major depression

For SCZ, there were 158 cases recorded during follow-up and 77,148 population-representative controls in CKB (Table 1). Both PGS-SCZ-EAS and PGS-SCZ-multi were positively associated with the risk of SCZ. PGS-SCZ-multi explained more variance in SCZ on the liability scale (R^2 = 3.49%) than PGS-SCZ-EAS (R^2 = 1.96%). For MD, there were 906 cases (744 recorded at baseline and 169 at follow-up; 7 overlapping) and 77,321 population-representative controls. Similarly, both PGS-MD-EAS and PGS-MD-multi were positively associated with the risk of MD, but PGS-MD-multi explained more variance on the liability scale than PGS-MD-EAS (R^2 = 0.55% vs 0.19%). The distribution of PGS between cases and controls, as well as OR per PGS quartile are shown in Supplementary Figures 2 and 3. All PGS were associated with their corresponding phenotypes in the sex-specific analyses (Supplementary Table 2; Supplementary Figures 4-7).

Phenome-wide associations

PheWAS results for the four PGS are shown in Figure 1 and Supplementary Data 1. Both PGS-SCZ-EAS and PGS-SCZ-multi were associated with higher risks of psychotic disorders (phecode: 295) recorded during follow-up, and with self-reported psychiatric disorder, depressive symptoms, worse subjective health, and lower household size, at baseline ($p < 1 \times 10^{-4}$). Higher PGS-SCZ-EAS was associated with lower BMI (beta per SD higher PGS = -0.06; 95% CI = -0.08 - -0.03; $p = 4.14 \times 10^{-6}$), while the association between PGS-SCZ-multi and BMI was directionally consistent but did not survive correction for multiple testing (beta = -0.03, 95% CI = -0.06 - -0.01; $p = 0.004$). PGS-SCZ-EAS was inversely associated with ever regular smoking (OR = 0.95; 95% CI = 0.93 - 0.98; $p = 1.53 \times 10^{-4}$) and current regular

smoking (OR = 0.96; 95% CI = 0.94 - 0.99; $p = 1.06 \times 10^{-3}$) at baseline, and these associations with PGS-SCZ-multi were directionally consistent (ever regular smoking: OR = 0.98; 95% CI = 0.96 - 1.00; $p = 0.069$; current regular smoking: OR = 0.97; 95% CI = 0.95 - 0.99; $p = 0.011$). In contrast, higher PGS-SCZ-multi was associated with more cigarettes smoked per day (beta = 0.27, 95% CI = 0.14 - 0.40; $p = 7.93 \times 10^{-5}$), and was directionally consistent with PGS-SCZ-EAS (beta = 0.16, 95% CI = 0.03 - 0.30; $p = 0.018$). Higher PGS-SCZ-multi was also associated with higher levels of education, lower standing/sitting height, and lower physical activity ($p < 1 \times 10^{-3}$), with similar associations for PGS-SCZ-EAS reaching nominal significance ($p < 0.05$). Sex-specific PheWAS showed generally consistent results with the sex-combined analyses (Supplementary Figures 8 and 9; Supplementary Data 1), except that the association between PGS-SCZ-EAS and smoking was only present in males, as few women in CKB smoke. We also observed an inverse association between PGS-SCZ-EAS and polyp of female genital organs in the female-specific analysis (OR = 0.79; 95% CI = 0.69 - 0.90; $p = 4.28 \times 10^{-4}$). Adjusting for additional covariates did not substantially change the associations between PGS-SCZ and these phenotypes (Supplementary Data 1).

Higher PGS-MD-EAS and PGS-MD-multi were associated with higher risks of self-reported depressive symptoms and MD measured by the CIDI questionnaire at baseline, as well as lower household size ($p < 2 \times 10^{-3}$). However, PGS-MD-multi yielded more associations in general than PGS-MD-EAS. Higher PGS-MD-multi was associated with higher risks of generalised anxiety disorder, anxiety symptoms, neurasthenia, and gallstone/gallbladder disease, as well as worse subjective health, and shorter/worse sleep at baseline ($p < 1 \times 10^{-3}$). PGS-MD-multi was also positively associated with several diseases at follow-up ($p < 2 \times 10^{-3}$), including ischemic heart disease (phecode: 411), cerebrovascular disease (phecode: 433), chronic airway obstruction (phecode: 496), bonchritis (phecode: 497), and abdominal pain (phecode: 785), but no such associations were found for PGS-MD-EAS even at nominal significance. Conversely, PGS-MD-EAS had an inverse association with BMI (beta = -0.04, 95% CI = -0.06 - -0.01, $p = 1.77 \times 10^{-3}$) that was not observed for PGS-MD-multi (beta = -0.01, 95% CI = -0.03 - -0.01, $p = 0.320$). Sex-specific findings were generally consistent with the sex-combined analyses (Supplementary Figures 7 and 8; Supplementary Data 1), including a positive association between PGS-MD-multi

and number of pregnancies (beta = 0.03, 95% CI = 0.02 - 0.04, $p = 2.16 \times 10^{-6}$). Adjusting for additional covariates attenuated the associations between PGS-MDD and urban residency and household size, but other associations did not change substantially (Supplementary Data 1).

Genetic correlations

Among the PheWAS associations, GWAS summary statistics were available in both EAS and EUR for 5 of the phenotypes related to PGS-SCZ and 8 related to PGS-MD (Supplementary Table 1). We also tested the genetic correlations between SCZ and MD. As shown in Figure 2 and Supplementary Table 3, there were positive genetic correlations between SCZ and MD in both EAS and EUR. There was also an inverse genetic correlation between SCZ and BMI in EUR (r_g = -0.11; 95% CI = -0.14 - -0.08; $p = 4.20 \times 10^{-13}$) but not in EAS, although it was directionally consistent (r_g = -0.05; 95% CI = -0.12 - 0.02; p = 0.143). Moreover, we found positive genetic correlations between SCZ and educational attainment in both EAS and EUR ($p < 0.05$). Consistent with the PheWAS results, an inverse genetic correlation was found between SCZ and smoking initiation in EAS (r_g = -0.10; 95% CI = -0.18 - -0.03; $p = 4.70 \times 10^{-3}$). In contrast, in EUR, SCZ had a positive genetic correlation with smoking initiation (r_g = 0.17; 95% CI = 0.13 - 0.21; $p = 6.86 \times 10^{-19}$) and cigarettes per day (r_g = 0.10; 95% CI = 0.06 - 0.15; $p = 1.28 \times 10^{-5}$). Cross-ancestry analyses showed positive genetic correlations between SCZ and MD, as well as inverse genetic correlations between SCZ and BMI (Supplementary Figure 10 and Supplementary Table 4).

In EAS, there was an inverse genetic correlation between MD and BMI (r_g = -0.19; 95% CI = -0.33 - -0.05; $p = 8.60 \times 10^{-3}$). In contrast, in EUR, MD showed a positive genetic correlation with BMI (r_g = 0.11; 95% CI = 0.08 - 0.15; $p = 1.04 \times 10^{-9}$), as well as with smoking initiation, cigarettes per day, and several physical disease outcomes. Cross-ancestry analyses showed an inverse genetic correlation between MD-EAS and BMI-EUR, as well as positive genetic correlations between MD-EUR and smoking-EAS (Supplementary Figure 9 and Supplementary Table 4).

Mendelian Randomisation

As BMI, smoking initiation, and cigarettes per day showed associations with SCZ and MD in both the PheWAS and genetic correlation analyses, we conducted two-

sample bi-directional MR to assess their causal relevance in EAS and EUR. All exposures' genetic instruments had a mean F-statistic > 20, suggesting a low likelihood of weak instrument bias (Supplementary Data 2). In EAS, there was an inverse association between genetically-predicted BMI and MD (beta = -0.11; 95% CI = -0.20 - -0.02; p = 0.021; Figure 3), but this association was no longer significant when the weighted median and weighted mode methods were used (Supplementary Data 2). There were no other causal associations between SCZ/MD and these phenotypes in either direction in EAS. Sensitivity analyses with genetic instruments selected at a more stringent threshold also showed no evidence of associations in EAS (Supplementary Figure 11 and Supplementary Data 2). In EUR, genetically predicted BMI was inversely associated with SCZ (beta = -0.19; 95% CI = -0.29 - -0.08; p < 0.001) but positively associated with MD (beta = 0.10; 95% CI = 0.05 - 0.14; p < 0.001). There were also strong positive bi-directional causal associations between smoking initiation and both SCZ and MD in EUR. Additionally, genetically-predicted SCZ and MD were both positively associated with cigarettes per day in EUR. Sensitivity analyses using other MR methods showed consistent results with the IVW method in EUR (Supplementary Data 2).

We performed multivariable MR to assess if the causal associations observed in EUR could be influenced by pleiotropic effects on other phenotypes (i.e. BMI, smoking, educational attainment, and poverty). Compared to univariable MR results, adding educational attainment as an additional exposure in multivariable MR attenuated the causal relationship of BMI on SCZ and MD (Figure 4, Supplementary Data 3). The causal relevance of BMI to MD was also attenuated when including smoking initiation or poverty as an additional exposure. Furthermore, we found that the bi-directional association between smoking initiation and SCZ was attenuated by adding cannabis use disorder as an additional exposure. The conditional F-statistics were > 10 for most exposures, except for smoking initiation, cigarettes per day, and cannabis use in certain models, indicating potential weak instrument bias. As all models also had a p < 0.05 for their Q-statistics, indicating potential pleiotropy, we used Q-statistic minimisation to re-estimate the beta for each exposure. This shifted most betas closer to 0 in our sensitivity analyses (Supplementary Data 3).

Discussion

This is the first study to assess the associations of PGS for SCZ and MD with ~250 phenotypes in over 100,000 Chinese adults. Both PGS were associated with their corresponding mental disorders, as well as with a range of socio-demographic, lifestyle, and physical measurements in CKB. In particular, we identified novel population-specific inverse associations between genetic susceptibility for SCZ and smoking initiation, and between genetic susceptibility for MD and BMI. Genetic correlation analyses revealed opposing findings in EAS and EUR, where smoking initiation-SCZ, and BMI-MD, were inversely correlated in EAS and positively correlated in EUR, although MR only supported causal relationships in EUR. Further multivariable MR analyses suggested that such causality in EUR might be accounted for by pleiotropic effects on other phenotypes, such as cannabis use, in the relationship between smoking and SCZ.

The PGS for SCZ and MD were associated with various mental-health-related phenotypes in the current study. As expected, the PGS for SCZ and MD were both associated with their corresponding phenotypes in CKB, but PGS-SCZ explained more variance on the liability scale than PGS-MD. This is consistent with previous findings in EUR, as the estimated heritability of SCZ (~ 80%) was much higher than that of MD (~ 30%)[9,10]. In the PheWAS, we also observed associations between PGS-SCZ/MD and many other measures related to mental health, including depressive symptoms, anxiety symptoms, generalised anxiety disorder, continuous pain, self-reported psychiatric disorder, sleep, and neurasthenia. This agrees with the considerable shared heritability observed between mental disorders in EUR[49,50].It is worth highlighting that both PGS-SCZ and PGS-MD were strongly associated with self-reported subjective health measures, in line with the high burden of mental disorders in China [3,5]. The association with subjective health may also reflect the high prevalence of somatic symptoms in patients with mental disorders in China, where one study reported that over 78% of 2,191 patients with MD had moderate to severe somatic symptoms[51].

The PGS for SCZ and MD were also associated with various non-psychiatric phenotypes in the PheWAS. For example, we found an inverse genetic correlation between PGS-SCZ-EAS and BMI, which also reached nominal significance for PGS-

SCZ-multi. This was supported by the inverse genetic correlations between SCZ and BMI in both EAS and EUR, replicating previous research findings[52,53]. Moreover, we observed a positive association between PGS-SCZ-multi and education level, which reached nominal significance for PGS-SCZ-EAS and was supported by genetic correlation analyses. The finding is also consistent with the positive genetic correlation between SCZ and educational attainment in EUR[54]. Although counterintuitive, the positive association might be explained by the higher language ability and creativity in people with SCZ[55,56]. Additionally, PGS-SCZ-multi was inversely associated with sitting height in our study, which also reached nominal significance for PGS-SCZ-EAS. This is consistent with previous observational findings on standing height and SCZ in Western populations[57], and sitting height might be more related to SCZ since it has been associated with early-life growth[58]. However, we did not find genetic correlations between standing height and SCZ, and we were unable to perform this analysis for sitting height due to the lack of publicly available GWAS. Finally, we found positive associations between PGS-MD-multi and number of pregnancies in our study, consistent with previous observational findings[59] and the relatively high heritability of perinatal depression (54%)[60]. PGS-MD-multi was also associated with the risks of several physical illnesses at follow-up in CKB, supporting previous observational associations of MD with ischaemic heart disease[61], stroke[62], and COPD[63] in diverse populations. However, those associations with pregnancy and physical illnesses were not found for PGS-MD-EAS or in the following genetic correlation analyses within EAS. This might be due to the limited power of GWAS on MD in EAS, or could suggest potentially different genetic architecture of MD between EAS and EUR.

In the PheWAS, we observed a unique, novel inverse association between PGS-MD-EAS and BMI, supported by their inverse genetic correlation in EAS. Conversely, the present and previous studies showed a positive genetic correlation between BMI and MD in EUR[52,64]. Further MR analyses revealed a positive association between genetically predicted BMI and MD in EUR but an inverse association in EAS, although the latter was not significant in our sensitivity analyses. Our MR results generally agreed with previous MR studies on BMI and MD[65], including the inverse association between genetically predicted BMI and MD in EAS reported by O'Loughlin et al[16]. The weak association between genetically predicted BMI and MD

in EAS in our study might be explained by the limited power of GWAS on MD and BMI in EAS. However, considering their inverse association in the PheWAS and genetic correlation analyses, it is unlikely that an increase in GWAS sample size would change its direction in MR. Differences in sociocultural factors may offer an alternative explanation: It has been hypothesised that higher BMI is often associated with higher socio-economic status among middle-aged and old adults in China, while people with higher BMI often face more weight-based discrimination in Western societies[66]. Indeed, in our multivariable MR analysis in EUR, adding smoking initiation, educational attainment, or poverty as an additional exposure attenuated the causality of BMI on MD. This suggests that pleiotropic effects on BMI and these phenotypes might have influenced the association between BMI and MD.

In the PheWAS, PGS-SCZ-EAS was inversely associated with smoking initiation, which was supported by the inverse genetic correlation between SCZ and smoking initiation in EAS. Conversely, both smoking initiation and cigarettes per day showed positive genetic correlations with SCZ in EUR. Many previous studies in Western populations reported a positive relationship between smoking and SCZ[67], with MR studies suggesting a bi-directional causal relationship in EUR[68]. The self-medication hypothesis postulates that smoking may act as a way to relieve symptoms among people with schizophrenia[69]. However, studies in China reported a similar prevalence of smoking in people with schizophrenia to the prevalence in the general population[70,71]. Our MR analysis in EAS also revealed a lack of causal associations between SCZ and smoking initiation or cigarettes per day. Such difference between populations may be explained by differences in smoking behaviour: while there were fewer current smokers (11.9%) than ex-smokers (24.9%) in the UK[72], the proportion of current smokers (24.3%) was much higher than that of ex-smokers (3.1%) in China[73]. This is also related to the fact that smoking in China was often accepted as a social activity, especially for men[74]. Therefore, it is possible that people with higher PGS-SCZ participate less in social activities and thus are less likely to smoke. Cannabis use is another possible explanation for this difference, as it is a strong risk factor for SCZ[75,76] and often co-occurs with cigarette smoking in Western populations[77]. In contrast, very few people use cannabis in China due to strict regulations, so the relationship between smoking and SCZ would not be influenced by cannabis use. Indeed, in our multivariable MR analysis, adding cannabis use

disorder as an additional exposure attenuated the causal associations between smoking initiation and SCZ in EUR. This suggests that the causality of smoking initiation and SCZ might be, at least partially, explained by pleiotropic effects on smoking initiation and cannabis use disorder.

The current study has limitations. First, since the CKB follow-up data came from hospitalisation and death records, certain medical conditions might have been under-represented or missed. These include episodes of mental disorders that might not have required in-patient treatment, which might explain the relatively low prevalence of SCZ and MD in CKB. Moreover, although we computed PGS based on the largest GWAS available in EAS and EUR at the time of the study, GWAS on SCZ and MD in EAS still had much smaller sample sizes than those in EUR. This limited the statistical power of the PGS, as reflected by the lower R^2 explained by PGS-EAS than PGS-multi. Nevertheless, all PGS were still strongly associated with SCZ and MD in our current study, supporting the validity of their usage in the PheWAS. Additionally, it is worth highlighting that the associations between PGS and phenotypes do not indicate causality. Although we employed MR for causal inference, we were unable to do this for certain phenotypes (e.g. sleep duration) due to the lack of publicly available GWAS in EAS. Furthermore, although all exposures' genetic instruments had a mean F-statistic > 20 in the univariable MR analyses, some of them (e.g. cannabis use disorder) had a conditional F-statistic < 10 in the multivariable MR analyses. Nevertheless, since Q-statistic minimisation in the sensitivity analyses generally shifted betas towards the null, we believe that the conclusions drawn from the multivariable MR are still valid despite potential weak instrument bias. Finally, some of our findings in the PheWAS still require further explanation, as it is unclear why PGS-SCZ-EAS was inversely associated with female genital polyps. Although oestrogen deficiency has been associated with SCZ[78], more research is needed to replicate this association in other cohorts.

To conclude, the current study found that the PGS for SCZ and MD were associated with a range of phenotypes among ~100,000 Chinese adults, suggesting their shared genetic bases. The distinct associations of smoking-SCZ and BMI-MD across different populations highlight the important role sociocultural factors may play in those relationships. More research in diverse populations is still needed to clarify the context specificity of mental disorders and their links with other health-related traits.

Policymakers should also be aware of the influence of sociocultural factors to develop better mental health strategies tailored for specific populations.

Acknowledgement

The chief acknowledgment is to the participants, the project staff, and the China National Centre for Disease Control and Prevention (CDC) and its regional offices for assisting with the fieldwork. We thank Judith Mackay in Hong Kong; Yu Wang, Gonghuan Yang, Zhengfu Qiang, Lin Feng, Maigeng Zhou, Wenhua Zhao, Yan Zhang in China CDC; Lingzhi Kong, Xiucheng Yu, and Kun Li in the Chinese Ministry of Health; and Garry Lancaster, Sarah Clark, Martin Radley, Mike Hill, Hongchao Pan, and Jill Boreham in the CTSU, Oxford, for assisting with the design, planning, organization, and conduct of the study.

The CKB baseline survey and the first re-survey were supported by the Kadoorie Charitable Foundation in Hong Kong. The long-term follow-up has been supported by Wellcome grants to Oxford University (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z) and grants from the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0510100), and the National Natural Science Foundation of China (82192900, 82192901, 82192904, 82388102). The UK Medical Research Council (MC_UU_00017/1, MC_UU_12026/2, MC_U137686851), Cancer Research UK (C16077/A29186; C500/A16896) and the British Heart Foundation (CH/1996001/9454), provide core funding to the Clinical Trial Service Unit and Epidemiological Studies Unit at Oxford University for the project. DNA extraction and genotyping were supported by GlaxoSmithKline and the UK Medical Research Council (MC-PC-13049, MC-PC-14135). Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Analyses in the UK Biobank have been conducted using the UK Biobank Resource under Application Number 50474. Baihan Wang is funded by the Nuffield Department of Population Health Early Career Research Fellowship.

Data availability

In CKB, non-genetic data (e.g. baseline, resurveys, biomarkers, and disease endpoints) are available and updated periodically for access by bona fide researchers. Details of the CKB Data Sharing Policy, data release schedules and data request application procedures are available at www.ckbiobank.org. All queries about data access can be made to ckbaccess@ndph.ox.ac.uk. Access to individual participant genetic data (e.g. genotyping, whole genome sequence) is currently constrained by China's Administrative Regulations on Human Genetic Resources, for which collaboration with CKB researchers is generally required, which may be subject to separate regulatory approvals in China if it involves substantial sharing of unpublished data. Further information is available from the corresponding authors upon request. Summary statistics of genome-wide association studies used in this study are publicly available, as summarised in Supplementary Table 1.

References

1. Ferrari A. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry*. 2022;9(2):137-150. doi:10.1016/S2215-0366(21)00395-3

2. Abbafati C, Abbas KM, Abbasi-Kangevari M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020;396(10258):1204-1222. doi:10.1016/S0140-6736(20)30925-9

3. Chen Q, Huang S, Xu H, et al. The burden of mental disorders in Asian countries, 1990-2019: an analysis for the global burden of disease study 2019. *Transl Psychiatry*. 2024;14(1). doi:10.1038/S41398-024-02864-5

4. Deng Y, Sun S, Wu S, et al. Burden and trends of mental disorders in China from 1990 to 2019: findings from the Global Burden of Disease Study 2019.

Soc Psychiatry Psychiatr Epidemiol. 2023;1:1-14. doi:10.1007/S00127-023-02594-X/FIGURES/1

5. Ma C, Yu S, Huang Y, et al. Burden of Mental and Substance Use Disorders — China, 1990−2019. *China CDC Wkly*. 2020;2(41):804. doi:10.46234/CCDCW2020.219

6. Huang D, Yang LH, Pescosolido BA. Understanding the public's profile of mental health literacy in China: A nationwide study. *BMC Psychiatry*. 2019;19(1):1-12. doi:10.1186/S12888-018-1980-8/FIGURES/2

7. Que J, Lu L, Shi L. Development and challenges of mental health in China. *Gen Psychiatr*. 2019;32(1). doi:10.1136/GPSYCH-2019-100053

8. Xiang YT, Ng CH, Yu X, Wang G. Rethinking progress and challenges of mental health care in China. *World Psychiatry*. 2018;17(2):231. doi:10.1002/WPS.20500

9. Hilker R, Helenius D, Fagerlund B, et al. Heritability of schizophrenia and schizophrenia spectrum based on the nationwide danish twin register. *Biol Psychiatry*. 2018;83(6):492-498. doi:10.1016/j.biopsych.2017.08.017

10. Flint J, Kendler KS. The Genetics of Major Depression. *Neuron*. 2014;81(3):484-503. doi:10.1016/j.neuron.2014.01.027

11. Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature 2022 604:7906*. 2022;604(7906):502-508. doi:10.1038/s41586-022-04434-5

12. Meng X, Navoly G, Giannakopoulou O, et al. Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. *Nature Genetics 2024 56:2*. 2024;56(2):222-233. doi:10.1038/s41588-023-01596-4

13. Lam M, Chen CY, Li Z, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet*. 2019;51(12):1670-1678. doi:10.1038/s41588-019-0512-x

14. Giannakopoulou O, Lin K, Meng X, et al. The Genetic Architecture of Depression in Individuals of East Asian Ancestry: A Genome-Wide Association

Study. *JAMA Psychiatry*. 2021;78(11):1258-1269. doi:10.1001/JAMAPSYCHIATRY.2021.2099

15. Chen S, Tang D, Deng L, Xu S. Asian-European differentiation of schizophrenia-associated genes driven by admixture and natural selection. *iScience*. 2024;27(5). doi:10.1016/J.ISCI.2024.109560/ASSET/CF9A7140-DEBB-494F-95CC-1495492B1710/MAIN.ASSETS/GR3.JPG

16. O'Loughlin J, Casanova F, Fairhurst-Hunter Z, et al. Mendelian randomisation study of body composition and depression in people of East Asian ancestry highlights potential setting-specific causality. *BMC Med*. 2023;21(1). doi:10.1186/S12916-023-02735-8

17. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics*. 2015;31(9):1466-1468. doi:10.1093/BIOINFORMATICS/BTU848

18. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols 2020 15:9*. 2020;15(9):2759-2772. doi:10.1038/s41596-020-0353-1

19. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics 2018 19:9*. 2018;19(9):581-590. doi:10.1038/s41576-018-0018-x

20. Lewis CM, Vassos E. Polygenic risk scores: From research tools to clinical instruments. *Genome Med*. 2020;12(1):1-11. doi:10.1186/S13073-020-00742-5/TABLES/2

21. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome–phenome relationship using phenome-wide association studies. *Nature Reviews Genetics 2016 17:3*. 2016;17(3):129-145. doi:10.1038/nrg.2015.36

22. Bastarache L, Denny JC, Roden DM. Phenome-Wide Association Studies. *JAMA*. 2022;327(1):75-76. doi:10.1001/JAMA.2021.20356

23. Zheutlin AB, Dennis J, Linnér RK, et al. Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 106,160 patients across four health care systems. *American Journal of Psychiatry*. 2019;176(10):846-855.

doi:10.1176/APPI.AJP.2019.18091085/ASSET/IMAGES/LARGE/APPI.AJP.2019.18091085F3.JPEG

24. Shen X, Howard DM, Adams MJ, et al. A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. *Nature Communications 2020 11:1*. 2020;11(1):1-16. doi:10.1038/s41467-020-16022-0

25. Fang Y, Fritsche LG, Mukherjee B, Sen S, Richmond-Rakerd LS. Polygenic Liability to Depression Is Associated With Multiple Medical Conditions in the Electronic Health Record: Phenome-wide Association Study of 46,782 Individuals. *Biol Psychiatry*. 2022;92(12):923-931. doi:10.1016/J.BIOPSYCH.2022.06.004

26. Socrates A, Maxwell J, Glanville KP, et al. Investigating the effects of genetic risk of schizophrenia on behavioural traits. *npj Schizophrenia 2021 7:1*. 2021;7(1):1-9. doi:10.1038/s41537-020-00131-2

27. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;40(6):1652-1666.

28. Walters RG, Millwood IY, Lin K, et al. Genotyping and population characteristics of the China Kadoorie Biobank. *Cell Genomics*. 2023;0(0):100361. doi:10.1016/J.XGEN.2023.100361

29. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012;9(2):179-181. doi:10.1038/nmeth.1785

30. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287. doi:10.1038/ng.3656

31. Cong PK, Bai WY, Li JC, et al. Genomic analyses of 10,376 individuals in the Westlake BioBank for Chinese (WBBC) pilot project. *Nature Communications 2022 13:1*. 2022;13(1):1-15. doi:10.1038/s41467-022-30526-x

32. Ruan Y, Lin YF, Feng YCA, et al. Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics 2022 54:5*. 2022;54(5):573-580. doi:10.1038/s41588-022-01054-7

33. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications 2019 10:1*. 2019;10(1):1-10. doi:10.1038/s41467-019-09718-5

34. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

35. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature 2010 467:7311*. 2010;467(7311):52-58. doi:10.1038/nature09298

36. Purcell SM, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795

37. Bastarache L. Using Phecodes for Research with the Electronic Health Record: From PheWAS to PheRS. *Annu Rev Biomed Data Sci*. 2021;4:1. doi:10.1146/ANNUREV-BIODATASCI-122320-112352

38. He Y, Koido M, Shimmori Y, Kamatani Y. GWASLab: a Python package for processing and visualizing GWAS summary statistics. doi:10.51094/JXIV.370

39. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics 2015 47:3*. 2015;47(3):291-295. doi:10.1038/ng.3211

40. Brown BC, Ye CJ, Price AL, Zaitlen N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am J Hum Genet*. 2016;99(1):76-88. doi:10.1016/j.ajhg.2016.05.001

41. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658-665. doi:10.1002/GEPI.21758

42. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet*. 2017;13(11):e1007081. doi:10.1371/JOURNAL.PGEN.1007081

43. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-525. doi:10.1093/IJE/DYV080

44. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;40(4):304. doi:10.1002/GEPI.21965

45. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46(6):1985-1998. doi:10.1093/IJE/DYX102

46. Burgess S, Thompson SG. Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *Am J Epidemiol*. 2015;181(4):251. doi:10.1093/AJE/KWU283

47. Sanderson E, Spiller W, Bowden J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. *Stat Med*. 2021;40(25):5434-5452. doi:10.1002/SIM.9133

48. R Core Team. R: A Language and Environment for Statistical Computing. Published online 2024. http://www.r-project.org/index.html

49. Anttila V, Bulik-Sullivan B, Finucane HK, et al. Analysis of shared heritability in common disorders of the brain. *Science (1979)*. 2018;360(6395). doi:10.1126/SCIENCE.AAP8757/SUPPL_FILE/AAP8757_TABLE_S7.XLSX

50. Bourque VR, Poulain C, Proulx C, et al. Genetic and phenotypic similarity across major psychiatric disorders: a systematic review and quantitative assessment. *Translational Psychiatry 2024 14:1*. 2024;14(1):1-10. doi:10.1038/s41398-024-02866-3

51. Vaccarino AL, Sills TL, Evans KR, Kalali AH. Prevalence and association of somatic symptoms in patients with Major Depressive Disorder. *J Affect Disord*. 2008;110(3):270-276. doi:10.1016/J.JAD.2008.01.009

52. Perry BI, Bowker N, Burgess S, et al. Evidence for Shared Genetic Aetiology Between Schizophrenia, Cardiometabolic, and Inflammation-Related Traits: Genetic Correlation and Colocalization Analyses. *Schizophr Bull Open*. 2022;3(1):sgac001. doi:10.1093/SCHIZBULLOPEN/SGAC001

53. Chen W, Feng J, Jiang S, et al. Mendelian randomization analyses identify bidirectional causal relationships of obesity with psychiatric disorders. *J Affect Disord*. 2023;339:807-814. doi:10.1016/J.JAD.2023.07.044

54. Bansal V, Mitjans M, Burik CAP, et al. Genome-wide association study results for educational attainment aid in identifying genetic heterogeneity of schizophrenia. *Nature Communications 2018 9:1*. 2018;9(1):1-12. doi:10.1038/s41467-018-05510-z

55. Power RA, Steinberg S, Bjornsdottir G, et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci*. 2015;18(7):953-955. doi:10.1038/NN.4040

56. Rajagopal VM, Ganna A, Coleman JRI, et al. Genome-wide association study of school grades identifies genetic overlap between language ability, psychopathology and creativity. *Scientific Reports 2023 13:1*. 2023;13(1):1-16. doi:10.1038/s41598-022-26845-0

57. Zammit S, Rasmussen F, Farahmand B, et al. Height and body mass index in young adulthood and risk of schizophrenia: a longitudinal study of 1 347 520 Swedish men. *Acta Psychiatr Scand*. 2007;116(5):378-385. doi:10.1111/J.1600-0447.2007.01063.X

58. Wadsworth M, Hardy R, Paul A, Marshall S, Cole T. Leg and trunk length at 43 years in relation to childhood health, diet and family circumstances; evidence from the 1946 national birth cohort. *Int J Epidemiol*. 2002;31(2):383-390. doi:10.1093/IJE/31.2.383

59. Wang Y, Wei R, Chen Z, et al. The association between the number of pregnancies and depressive symptoms: A population-based study. *J Affect Disord*. 2024;350:411-419. doi:10.1016/J.JAD.2024.01.161

60. Viktorin A, Meltzer-Brody S, Kuja-Halkola R, et al. Heritability of Perinatal Depression and Genetic Overlap With Nonperinatal Depression. *Am J Psychiatry*. 2016;173(2):158-165. doi:10.1176/APPI.AJP.2015.15010085

61. Gan Y, Gong Y, Tong X, et al. Depression and the risk of coronary heart disease: a meta-analysis of prospective cohort studies. *BMC Psychiatry*. 2014;14(1):371. doi:10.1186/S12888-014-0371-Z

62. Pan A, Sun Q, Okereke OI, Rexrode KM, Hu FB. Depression and the Risk of Stroke Morbidity and Mortality: A Meta-analysis and Systematic Review. *JAMA□: the journal of the American Medical Association*. 2011;306(11):1241. doi:10.1001/JAMA.2011.1282

63. Atlantis E, Fahey P, Cochrane B, Smith S. Bidirectional Associations Between Clinically Relevant Depression or Anxiety and COPD: A Systematic Review and Meta-analysis. *Chest*. 2013;144(3):766-777. doi:10.1378/CHEST.12-1911

64. Zhang H, Zheng R, Yu B, et al. Dissecting shared genetic architecture between depression and body mass index. *BMC Med*. 2024;22(1):455. doi:10.1186/S12916-024-03681-9/FIGURES/5

65. Tyrrell J, Mulugeta A, Wood AR, et al. Using genetics to understand the causal influence of higher BMI on depression. *Int J Epidemiol*. 2019;48(3):834-848. doi:10.1093/IJE/DYY223

66. Zhang L, Liu K, Li H, et al. Relationship between body mass index and depressive symptoms: the “fat and jolly” hypothesis for the middle-aged and elderly in China. *BMC Public Health*. 2016;16(1):1-15. doi:10.1186/S12889-016-3864-5/TABLES/6

67. De Leon J, Diaz FJ. A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. *Schizophr Res*. 2005;76(2-3):135-157. doi:10.1016/J.SCHRES.2005.02.010

68. Wootton RE, Richmond RC, Stuijfzand BG, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med*. 2020;50(14):2435-2443. doi:10.1017/S0033291719002678

69. Kumari V, Postma P. Nicotine use in schizophrenia: The self medication hypotheses. *Neurosci Biobehav Rev*. 2005;29(6):1021-1034. doi:10.1016/J.NEUBIOREV.2005.02.006

70. Cao XL, Li Y, Zhong BL, et al. Current cigarette smoking in Chinese female patients with schizophrenia: A meta-analysis. *Psychiatry Res*. 2016;235:203-205. doi:10.1016/J.PSYCHRES.2015.11.050

71. Li Y, Cao XL, Zhong BL, et al. Smoking in male patients with schizophrenia in China: A meta-analysis. *Drug Alcohol Depend*. 2016;162:146-153. doi:10.1016/J.DRUGALCDEP.2016.02.040

72. Office for National Statistics (ONS). *Adult Smoking Habits in the UK: 2023*.; 2024. Accessed January 7, 2025. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2023

73. Liu Z, Li Y hua, Cui Z yang, et al. Prevalence of tobacco dependence and associated factors in China: Findings from nationwide China Health Literacy Survey during 2018–19. *Lancet Reg Health West Pac*. 2022;24:100464. doi:10.1016/j.lanwpc.2022.100464

74. Ma GX, Shive SE, Ma XS, et al. Social influences on Cigarette Smoking Among Mainland Chinese and Chinese Americans: A Comparative Study. *Am J Health Stud*. 2013;28(1):12. Accessed January 6, 2025. https://pmc.ncbi.nlm.nih.gov/articles/PMC3914219/

75. Marconi A, Di Forti M, Lewis CM, Murray RM, Vassos E. Meta-Analysis of the association between the level of cannabis use and risk of psychosis. *Schizophr Bull*. 2016;42(5):1262-1269. doi:10.1093/schbul/sbw003

76. Vaucher J, Keating BJ, Lasserre AM, et al. Cannabis use and risk of schizophrenia: a Mendelian randomization study. *Molecular Psychiatry 2018 23:5*. 2017;23(5):1287-1292. doi:10.1038/mp.2016.252

77. Agrawal A, Budney AJ, Lynskey MT. The Co-occurring Use and Misuse of Cannabis and Tobacco: A Review. *Addiction (Abingdon, England)*. 2012;107(7):1221. doi:10.1111/J.1360-0443.2012.03837.X

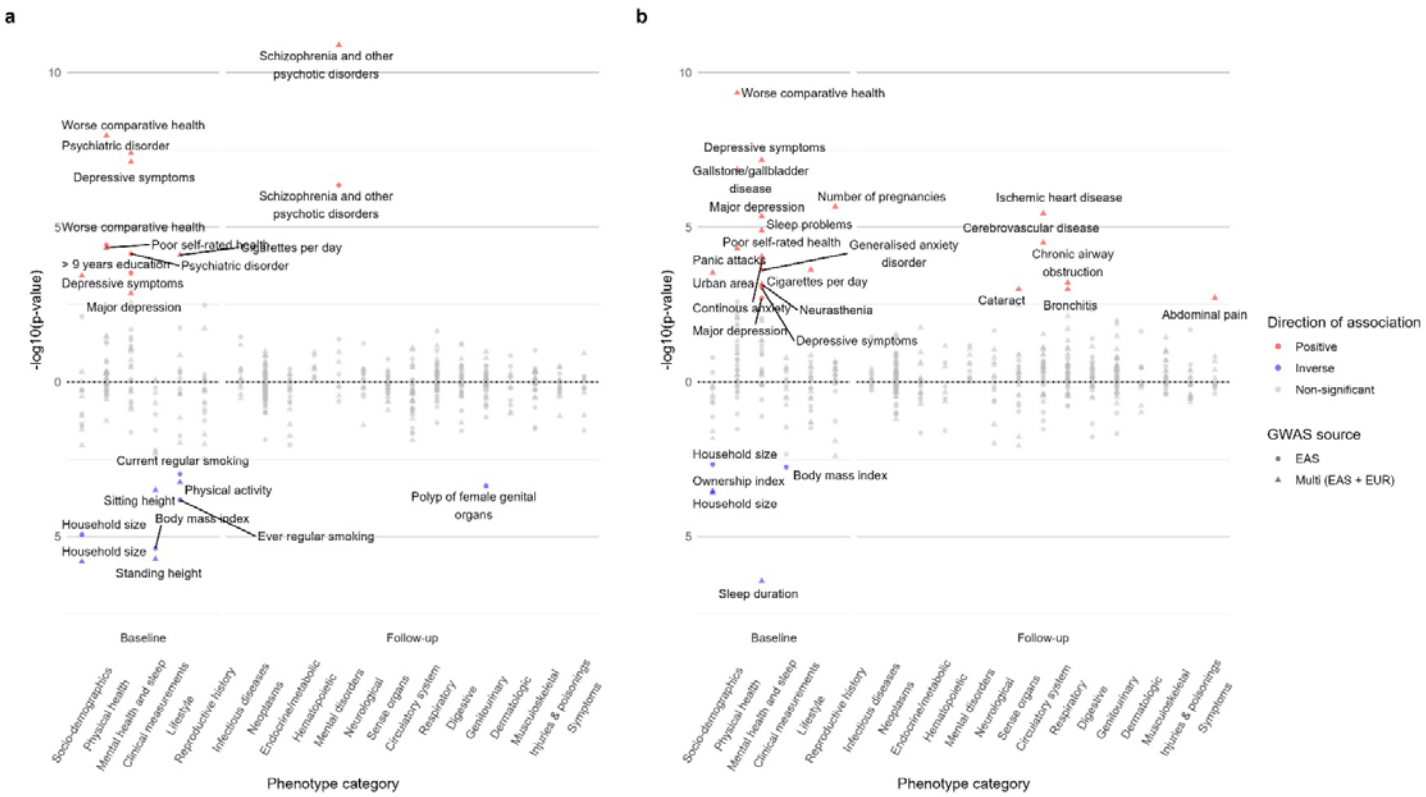
78. Gogos A, Sbisa AM, Sun J, Gibbons A, Udawela M, Dean B. A Role for Estrogen in Schizophrenia: Clinical and Preclinical Findings. *Int J Endocrinol*. 2015;2015(1):615356. doi:10.1155/2015/615356

Table 1. Associations between polygenic scores for schizophrenia/major depression and their corresponding phenotypes in China Kadoorie Biobank.

Mental disorder	n (cases)	n (controls)	GWAS source	OR per SD higher PGS (95%CI)	*p*	R^2 on liability scale
SCZ	158	77,148	EAS	1.52 (1.29 - 1.79)	6.25×10^{-7}	1.96%
			Multi (EAS + EUR)	1.73 (1.47 - 2.03)	3.11×10^{-11}	3.49%
MD	906	77,321	EAS	1.13 (1.06 - 1.21)	2.05×10^{-4}	0.19%
			Multi (EAS + EUR)	1.21 (1.14 - 1.30)	9.04×10^{-9}	0.55%

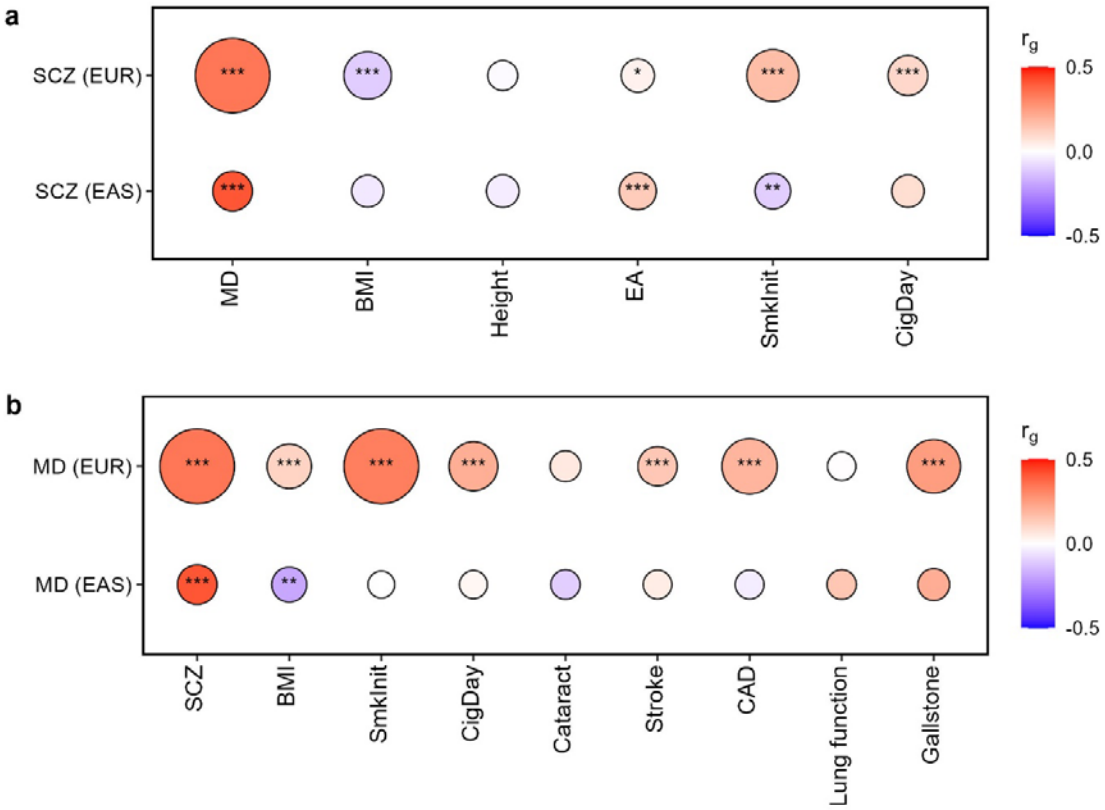
Note. Cases of mental disorders were identified from the overall dataset, while controls were identified from the population-representative subset. Two types of polygenic scores were assessed for each mental disorder, one based on GWAS in EAS and one based on GWAS in both EAS and EUR. SCZ: Schizophrenia. MD: Major depression. GWAS: Genome-wide association studies. EAS: East Asian ancestry. EUR: European ancestry. OR: Odds ratio per one standard deviation higher polygenic scores. PGS: Polygenic score. CI: Confidence interval.

Figure 1. Phenome-wide associations with polygenic scores for schizophrenia and major depression in China Kadoorie Biobank.



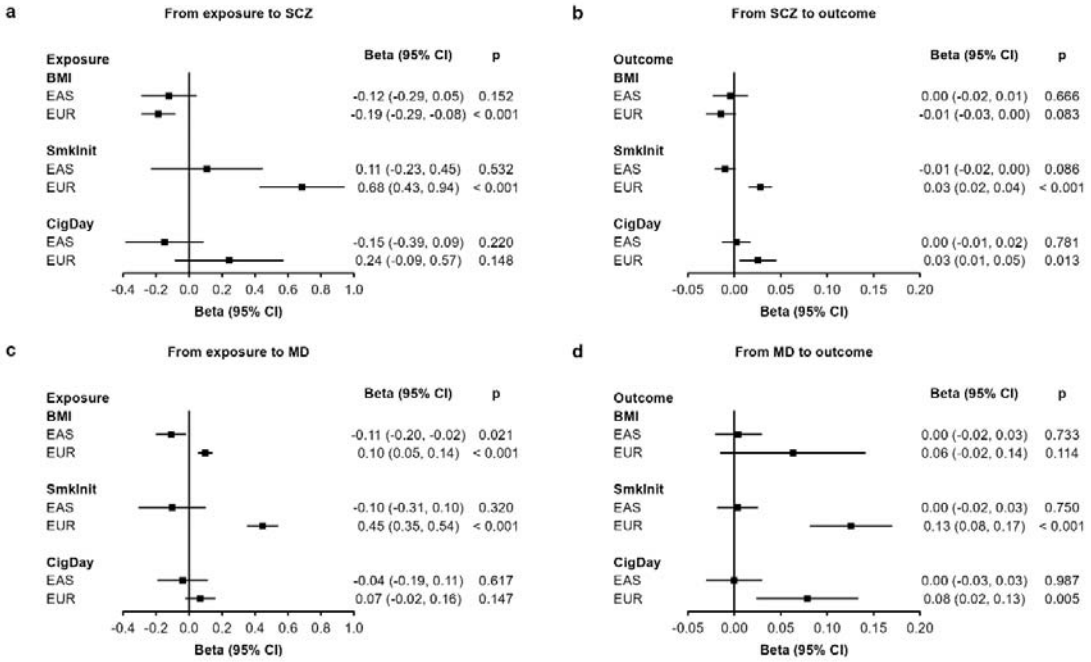
Note. a, Results of polygenic scores for schizophrenia. b, Results of polygenic scores for major depression. A total of 254 phenotypes (67 baseline measures and 187 phecodes recorded during follow-up) were tested. Phenotypes only present in one sex are included (sex-specific results shown). The shape of dots indicates the GWAS source, while the colour of dots indicates the direction of association. Results were corrected for multiple testing with a false discovery rate = 0.05. GWAS: Genome-wide association studies. EAS: East Asian ancestry. EUR: European ancestry.

Figure 2. Genetic correlations between schizophrenia/major depression and associated phenotypes.



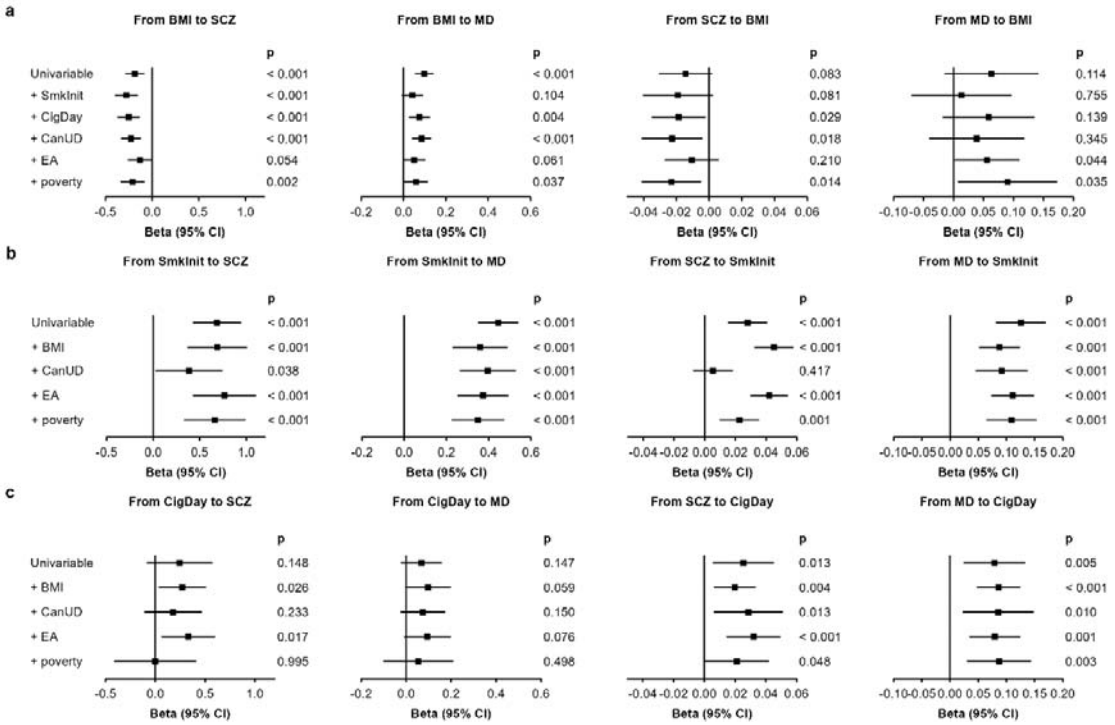
Note. a, Genetic correlations with SCZ. b, Genetic correlations with MD. Genetic correlations (r_g) were tested by LDSC, using GWAS on SCZ/MD and phenotypes within the same ancestry. Mental disorder-phenotype pairs that were significant in the phenome-wide association analysis were tested here. Only phenotypes with publicly available GWAS in both EAS and EUR were included. The colour of the circles indicates the correlation coefficient. The size of the circles is scaled to -log10(p-value). *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. SCZ: Schizophrenia. MD: Major depression. EAS: East Asian ancestry. EUR: European ancestry. BMI: Body mass index. EA: Educational attainment. SmkInit: Smoking initiation. CigDay: Cigarettes per day. CAD: Coronary artery disease.

Figure 3. Bi-directional Mendelian Randomisation between schizophrenia/major depression and associated phenotypes.



Note. a, From exposure to SCZ. b, From SCZ to outcome. c, From exposure to MD. d, From MD to outcome. Results shown here are based on estimates from the inverse variance weighted method. Genetic instruments were selected based on $p < 1 \times 10^{-5}$ in EAS and $p < 5 \times 10^{-8}$ in EUR after clumping. SCZ: Schizophrenia. MD: Major depression. EAS: East Asian ancestry. EUR: European ancestry. BMI: Body mass index. SmkInit: Smoking initiation. CigDay: Cigarettes per day. CI: Confidence interval.

Figure 4. Multivariable Mendelian Randomisation between schizophrenia/major depression and other phenotypes in the European ancestry population.



Note. a, Results for BMI. b, Results for SmkInit. c, Results for CigDay. All multivariable Mendelian Randomisation analyses were based on GWAS summary statistics in the European ancestry population with the inverse variance weighted method. Genetic instruments were selected based on $p < 5 \times 10^{-8}$ after clumping. An additional phenotype was added to the univariable model to assess its influence on the association each time. SCZ: Schizophrenia. MD: Major depression. BMI: Body mass index. SmkInit: Smoking initiation. CigDay: Cigarettes per day. CanUD: Cannabis use disorder. EA: Educational attainment. CI: Confidence interval.