Association Between Population Density and Genetic Risk for Schizophrenia

Lucía Colodro-Conde, PhD; Baptiste Couvy-Duchesne, PhD; John B. Whitfield, PhD; Fabian Streit, PhD; Scott Gordon, PhD; Kathryn E. Kemper, PhD; Loic Yengo, PhD; Zhili Zheng, MD, PhD; Maciej Trzaskowski, PhD; Eveline L. de Zeeuw, PhD; Michel G. Nivard, PhD; Marjolijn Das, PhD; Rachel E. Neale, PhD; Stuart MacGregor, PhD; Catherine M. Olsen, PhD; David C. Whiteman, MBBS, PhD; Dorret I. Boomsma, PhD; Jian Yang, PhD; Marcella Rietschel, PhD; John J. McGrath, MD, PhD; Sarah E. Medland, PhD; Nicholas G. Martin, PhD

IMPORTANCE Urban life has been proposed as an environmental risk factor accounting for the increased prevalence of schizophrenia in urban areas. An alternative hypothesis is that individuals with increased genetic risk tend to live in urban/dense areas.

OBJECTIVE To assess whether adults with higher genetic risk for schizophrenia have an increased probability to live in more populated areas than those with lower risk.

DESIGN, SETTING, AND PARTICIPANTS Four large, cross-sectional samples of genotyped individuals of European ancestry older than 18 years with known addresses in Australia, the United Kingdom, and the Netherlands were included in the analysis. Data were based on the postcode of residence at the time of last contact with the participants. Community-based samples who took part in studies conducted by the Queensland Institute for Medical Research Berghofer Medical Research Institute (QIMR), UK Biobank (UKB), Netherlands Twin Register (NTR), or QSkin Sun and Health Study (QSKIN) were included. Genome-wide association analysis and mendelian randomization (MR) were included. The study was conducted between 2016 and 2018.

EXPOSURES Polygenic risk scores for schizophrenia derived from genetic data (genetic risk is independently measured from the occurrence of the disease). Socioeconomic status of the area was included as a moderator in some of the models.

MAIN OUTCOMES AND MEASURES Population density of the place of residence of the participants determined from census data. Remoteness and socioeconomic status of the area were also tested.

RESULTS The QIMR participants (15 544; 10 197 [65.6%] women; mean [SD] age, 54.4 [13.2] years) living in more densely populated areas (people per square kilometer) had a higher genetic loading for schizophrenia ($r^2 = 0.12\%$; $P = 5.69 \times 10^{-5}$), a result that was replicated across all 3 other cohorts (UKB: 345 246; 187 469 [54.3%] women; age, 65.7 [8.0] years; NTR: 11 212; 6727 [60.0%] women; age, 48.6 [17.5] years; and QSKIN: 15 726; 8602 [54.7%] women; age, 57.0 [7.9] years). This genetic association could account for 1.7% (95% CI, 0.8%-3.2%) of the schizophrenia risk. Estimates from MR analyses performed in the UKB sample were significant ($b = 0.049; P = 3.7 \times 10^{-7}$ using GSMR), suggesting that the genetic liability to schizophrenia may have a causal association with the tendency to live in urbanized locations.

CONCLUSIONS AND RELEVANCE The results of this study appear to support the hypothesis that individuals with increased genetic risk tend to live in urban/dense areas and suggest the need to refine the social stress model for schizophrenia by including genetics as well as possible gene-environment interactions.

Published online June 23, 2018.

© 2018 American Medical Association. All rights reserved.
In 2011, Lederbogen and colleagues published a functional magnetic resonance imaging study that showed greater brain activation of the stress-processing pathways in participants living in urban vs rural areas and suggested there and in a later study that the greater social stress of urban living could explain the well-documented higher prevalence of schizophrenia observed in urban than rural environments (odds ratio, 1.72; 95% CI, 1.53-1.92). Herein, we investigate an alternative, but not incompatible, explanation: people with higher genetic risk for schizophrenia tend to live in more urbanized areas owing to selective migration in either past or current generations.

Appendix I in the Supplement presents a short review of the literature in this area. In summary, living in an urban environment, which is itself partially heritable, is associated with increased risk of developing schizophrenia after controlling for potential confounders (age, sex, ethnicity, drug use, social class, family history, and season of birth) and using different measures of urbanicity (population size or density, window of exposure (birth, upbringing, or illness onset), and disease definition (schizophrenia or broad psychosis)). Although the association is established, its putative (familial) environmental or genetic components are unclear. It has been suggested that approximately 30% of all schizophrenia cases could be potentially prevented if the exposure to urban environments was removed, assuming urban environment is a causal factor. However, it is not clear whether urban residence has a causal effect on mental health or whether urban residence is a consequence of the disease (e.g., migration to the city of people in the prodromal stages of the disorder).

In the present study, we sought to examine the nature of this association by testing whether adults older than 18 years with higher genetic risk for schizophrenia are more likely to live in urbanized and populated areas (measured as population density) than those with lower genetic risk. If so, this finding would suggest that the higher prevalence of schizophrenia in cities is not only a consequence of the urban environment. This determination has been made possible by the advances in the identification of common genetic variants associated with schizophrenia in large discovery samples and the development of polygenic risk scores (PRS) in independent samples. In addition, we checked that the association could not be explained by differences in socioeconomic status (SES) of the residential areas. We also investigated the direction of causation between schizophrenia and population density using multi-instrument mendelian randomization (MR) study. For completeness, we present the estimates of the twin heritability and genome-wide association analyses (GWAS) of our main phenotypes.

Methods

Cohorts and Variables

We performed the analyses using a discovery cohort of 15,544 participants genotyped as part of a series of studies of general health conditions conducted by the Genetic Epidemiology Unit at Queensland Institute for Medical Research Berghofer Medical Research Institute (QIMR), Australia (Table 1; eAppendix 2, eTable 1, and eFigure 1 in the Supplement provide cohort and variable descriptions). We used the UK Biobank (UKB) (n = 456,426), the Netherlands Twin Register (NTR) (n = 16,434), and the Australian QSkin Sun and Health Study (QSKIN) sample (n = 17,267) to replicate and extend our analyses (Table 1; eAppendixes 3-5, eTable 2, eTable 3, eFigures 2-5 in the Supplement provide cohort and variable descriptions).

The QIMR Berghofer Medical Research Institute-Human Research Ethics Committee approved the study. The study was conducted between 2016 and 2018.

Statistical Analysis

Variance Component Analysis

We analyzed data from 5894 twins from the QIMR sample to estimate the contribution of additive genetic influences (narrow sense heritability), shared/familial environment, and unique environment to the interpersonal differences in population density, remoteness, and SES of the residential area (Appendix 6 in the Supplement provides more information on twin and family studies). We used the OpenMx package in R to estimate the parameters of the mixed models. Significance of the variance components was tested using likelihood ratio tests on nested models. We used the same approach to replicate the results on population density in the NTR cohort.

In the QIMR data, we fit a gene-environment moderator effect model that allows the variance components (heritability, shared environment, and unique environment) to vary across age (eAppendix 7 in the Supplement). This approach reflected previous results from Whitfield et al that suggested that heritability of rural/urban living in Australia increases with age. Finally, we estimated the genetic and environmental correlations between population density, remoteness, and SES using bivariate twin models. All models included age, age × sex, age2 × sex, GWAS array, and 4 genetic principal components as covariates.

PRS Analysis

The PRS in the QIMR sample were calculated from the imputed genotype dosage scores using GWAS summary statistics from the GWAS meta-analysis from the 2014 Psychiatric Genetics Conference.
# Table 1. Description of the Cohorts Used for the Analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Discovery Cohort</th>
<th>Replication Cohorts</th>
<th>NTR&lt;sup&gt;a&lt;/sup&gt;</th>
<th>QSKIN&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QIMR&lt;sup&gt;c&lt;/sup&gt;, (Australia)</td>
<td>UKB&lt;sup&gt;d&lt;/sup&gt;, (United Kingdom)</td>
<td>(the Netherlands)</td>
<td>(Australia (limited to Queensland))</td>
</tr>
<tr>
<td>Sample size, No.</td>
<td>15 544</td>
<td>456 426 (345 246 with PRS)</td>
<td>16 434 (11 212 with PRS)</td>
<td>15 726</td>
</tr>
<tr>
<td>Inclusion criteria&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Adult genotyped participants of QIMR studies</td>
<td>Genotyped participants of the UKB</td>
<td>Adult genotyped participants of the NTR</td>
<td>Unrelated (GRM &lt;0.1) adult genotyped participants of the QSKIN who had not previously participated in QIMR studies</td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, mean (SD):</td>
<td>54.4 (11.2)</td>
<td>65.7 (8.0)</td>
<td>48.6 (17.5)</td>
<td>57.0 (7.9)</td>
</tr>
<tr>
<td>From 7015 families</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample comprises 1119 complete MZ pairs, 1104 complete DZ pairs, and 1448 singleton twins</td>
<td>Sample comprises 345 258 unrelated individuals (GRM &lt;0.05)</td>
<td>Sample comprises 1740 complete MZ pairs, 1114 DZ complete twin pairs, and 812 singleton twins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetic data</td>
<td>Participants genotyped using commercial arrays</td>
<td>Participants genotyped using 2 closely related arrays (the UK BILEVE and the UK Biobank Axiom Arrays)</td>
<td>Participants genotyped using commercial arrays</td>
<td>Same as QIMR</td>
</tr>
<tr>
<td>Genotype data cleaned (by batch) for call rate (≥95%)</td>
<td>We used HRC imputed data provided by the biobank</td>
<td>Genotype data cleaned for call rate (≥95%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAF (≥1%)</td>
<td>Variants with MAF &lt;0.005%, missingness &lt;0.05, pHWE &lt;10&lt;sup&gt;-4&lt;/sup&gt;, and MAC &gt;5 excluded</td>
<td>MAF ≥0.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardy-Weinberg equilibrium</td>
<td>Data were checked for non-European ancestry</td>
<td>Hardy-Weinberg equilibrium (P ≥ 10&lt;sup&gt;-7&lt;/sup&gt;), allele frequency difference with GONL &lt;0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data checked for pedigree, sex, and mendelian errors and for non-European ancestry</td>
<td>Data checked for pedigree, sex, and mendelian errors and for non-European ancestry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imputation to the 1000 genomes (phase 3, release 5) performed on the Michigan Imputation Server&lt;sup&gt;22&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measures of population density and SES</td>
<td>Population density, remoteness, and SES variables generated from the postcode provided by participants at the time of last contact (1990-2015)</td>
<td>From the Easting and Northing coordinates rounded to the kilometer, we performed reverse geocoding to identify the postcode district in which the participants likely lived</td>
<td>Population density and SES derived from the most recent participants' postcodes</td>
<td>Population density and SES variables were generated from the postcode provided by the participants at the time of last contact (2010-2012)</td>
</tr>
<tr>
<td>Population density and SES</td>
<td>Numbers corresponded to the neighborhood data published in 2015-2016 by the Netherlands’s national statistical agency (CBS), which defines a neighborhood as the part of a municipality that is homogeneously demarcated from either a demographic or socioeconomic structure</td>
<td>We matched the postcodes as described in the QIMR sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population density expressed in number of residents per squared kilometer</td>
<td>Population density expressed in number of residents per hectare</td>
<td>Population density expressed in number of residents per squared kilometer</td>
<td>Population density expressed in number of residents per squared kilometer</td>
<td></td>
</tr>
<tr>
<td>SES based on the IRSAD&lt;sup&gt;c&lt;/sup&gt;,&lt;sup&gt;d&lt;/sup&gt; which can be used to measure socioeconomic well-being in a continuum, from the most disadvantaged areas (low values) to the most advantaged areas (high values)</td>
<td>We used the Townsend deprivation index&lt;sup&gt;c&lt;/sup&gt;,&lt;sup&gt;e&lt;/sup&gt; as a measure of SES</td>
<td>SES measured using the average personal income per person and the average market value of residential properties</td>
<td>SES based on the IRSAD&lt;sup&gt;c&lt;/sup&gt;,&lt;sup&gt;d&lt;/sup&gt; as described in the QIMR sample</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) population density was 1169 (1350) people/km&lt;sup&gt;2&lt;/sup&gt; (range, 0.01-5506)</td>
<td>Mean (SD) population density was 24.10 (25.1) inhabitants per hectare (ie, 2410 per km&lt;sup&gt;2&lt;/sup&gt;) (range, 0.1-222.5)</td>
<td>Mean (SD) population density in the total sample was 4669 (3793) (range, 1-17 797)</td>
<td>Mean (SD) population density was 748 people/km&lt;sup&gt;2&lt;/sup&gt; (945) (range, 0.01-3771)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: DZ, dizygotic; GONL, Genome of the Netherlands; GRM, genetic relationship matrices; HRC, Haplotype Reference Consortium; IRSAD, Index of Relative Socioeconomic Advantage and Disadvantage; MAC, minor allele count; MAF, minor allele frequency; MZ, monozygotic; NTR, Netherlands Twin Register; pHWE, P value for the Hardy-Weinberg test statistic; PRS, polygenic risk scores; QIMR, QIMR Berghof Medical Research Institute; QSKIN, QSkin Sun and Health Study; SES, socioeconomic status; UKB, United Kingdom Biobank.

<sup>a</sup> The Supplement contains documents on genetic and phenotypic data for all cohorts.

<sup>b</sup> European ancestry was an inclusion criterion across all cohorts.
Genome-wide Association Analyses
We performed GWAS of population density or SES in our largest sample, the UKB, using BOLT-LMM29 including age, age², sex, age x sex, age² x sex, and 4 genetic principal components30 on 12,722,635 SNPs with MAF greater than 0.005%. We ran additional GWAS of population density controlling for SES and of SES controlling for population density.

We also conducted GWAS of population density in the QIMR, NTR, and QSKIN cohorts. In the QIMR sample, we used RAREMETALWORKER31 and controlled for age, age², sex, age x sex, age² x sex, SES, GWAS array, and 4 genetic principal components as covariates. We explicitly corrected for relatedness using the kin pedigree option. Single-nucleotide polymorphisms with MAF less than 0.5% or imputation r²<0.6 were excluded, leaving 8,495,074 SNPs for analyses. In the NTR, the GWAS, corrected for age and sex, was performed using GCTA-MLMA32,33 to account for relatedness and population stratification using 2 genetic relationship matrices (one corrects for the related individuals; the second corrects for the population stratification in the distantly related individuals) and 5 genetic principal components. The number of markers included after quality control (r²>0.8 and MAF<0.01%) was 7,636,917. Finally, we performed GWAS in QSKIN using PLINK, version 1.90b4.128 with age, sex, and 4 genetic principal components as covariates, using a total of 7,672,045 markers after selecting those with r²>0.6 and MAF less than 0.01%.

We used LD score regressions34 to confirm the SNP heritability and the genetic correlations between the measures of population density across samples. eAppendices 2-5 in the Supplement provide details on genetic data of the 4 cohorts.

Mendelian Randomization
Mendelian randomization methods allow us to generate hypotheses concerning the direction of causation between 2 heritable variables.13 Herein, we relied on the 2014 GWAS meta-analysis summary results from the Psychiatric Genomics Consortium for schizophrenia and the GWAS results calculated from our samples for population density and SES. We used MR-Base35 (TwoSampleMR R package36) and GEMRM37 to conduct MR using known schizophrenia SNPs as instruments, thus testing the selection hypothesis that having a higher propensity to schizophrenia (ie, higher PRS) may have a possible causal association with the tendency to live in a denser and less remote area (eAppendix 9 in the Supplement) gives more details on MR and our analysis. We also investigated the reverse hypothesis (population density or SES inducing onset of schizophrenia) using GWAS results from our largest sample, the UKB.

Results
Variance Component Analysis
Population density, remoteness, and SES were all significantly correlated at a phenotypic level in all samples considered (eAppendix 10 and eTable 4 in the Supplement). Population density and remoteness were heritable (heritability [h²]) in the QIMR sample (Figure 1) (h² for population density = 16.9%; 95% CI, 3.4-30.4; P = .01; h² for remoteness = 16.3%; 95% CI, 3.5-29.0; P = .01); the heritability of SES was not significant (h² = 11.0%; 95% CI, 0.00-24.6; P = .12).

Shared environment (common environment [c²]) effects explained a more substantial and highly significant proportion of the trait variance (Figure 1) (c² for population density = 24.3%; 95% CI, 13.1-35.1; c² for remoteness = 29.1%; 95% CI, 19.0-40.0; and c² for socioeconomic status = 26.8%; 95% CI, 15.6-37.1; all P < .001), which highlights that people tend to live with or close to their parents or other relatives. Population density was also heritable in the NTR (h² = 12.1%; 95%
CI, 1.3-23.2; \( P = .28 \) and showed shared environmental sources of variance (\( \hat{c}^2 = 36.5; 95\% CI, 26.8-45.7; \hat{P} = 7 \times 10^{-13} \)).

In the QIMR cohort, population density was more heritable and less influenced by shared environmental sources as participants became older (eAppendix 11 in the Supplement), with the heritability increasing from 9.0% to 25.6% between ages 20 and 80 years. Over the same lifespan, \( c^2 \) decreased from 44.5% to less than 10.0% (eAppendix 11 in the Supplement), but variance explained by unique environmental sources (including measurement error) remained constant (eAppendix 11 in the Supplement). Similar results were obtained using standardized estimates, suggesting constant phenotypic variance across age. In addition, population density, remoteness, and SES shared environmental influences as indicated by significant environmental correlations from the twin models (eTable 4 in the Supplement).

Polygenic Risk Scores Analysis

Figure 2 shows the percentage of variance of the population density of the place of residence explained by the PRS for schizophrenia, with and without controlling for SES. In the QIMR sample, PRS calculated from all semi-independent SNPs across the genome (Figure 2) explained the greatest amount of variance in population density (\( \hat{r}^2 = 0.12\% ; \hat{P} = 5.69 \times 10^{-5} \)), and still explained (\( \hat{r}^2 = 0.074\% ; \hat{P} < .001 \)) when accounting for SES. Schizophrenia PRS also were significantly associated with remoteness (\( \hat{r}^2 = 0.06\% ; \hat{P} = .003 \)) when including all of the independent SNPs, although the association disappeared when correcting for SES. eFigure 6 and eAppendix 12 in the Supplement provide information on the association between genetic risk for schizophrenia, remoteness, and SES. We did not find evidence of interactions between sex or age and PRS for schizophrenia (\( \hat{P} > .05 \)) contributing to population density or remoteness in the QIMR sample.

The association between schizophrenia risk score and population density was replicated in the NTR (\( \hat{r}^2 = 0.14\% ; \hat{P} = 8.3 \times 10^{-4} \) and \( \hat{r}^2 = 0.073\% ; \hat{P} = .002 \) when correcting for SES), in the UKB (\( \hat{r}^2 = 0.088\% ; \hat{P} = 7.7 \times 10^{-59} \) \( ; \hat{r}^2 = 0.012\% ; \hat{P} = 1.2 \times 10^{-11} \) when accounting for SES) and in QSKIN (\( \hat{r}^2 = 0.027\% ; \hat{P} = .02 \) and \( \hat{r}^2 = 0.015\% ; \hat{P} = .047 \) when correcting for SES) (Figure 2). All correlations were in the same direction, pointing to increased PRS for participants living in more densely populated areas. Results for remoteness and SES in QSKIN were consistent with those in QIMR and are presented in eFigure 7 in the Supplement.

In addition, we tested the association between SES and schizophrenia PRS. The association did not reach statistical significance in the QIMR or QSKIN sample when taking into account multiple testing (\( \hat{P} > .01 \)) (eFigures 6 and 7 in the Supplement) but was significant in the UKB (\( \hat{r}^2 = 0.084\% ; \hat{P} = 8.6 \times 10^{-64} \), correcting for population density) (eFigure 8 in the Supplement) likely because of the gain of power owing to its very large sample size.

Genome-wide Association Analyses

Six genomic regions reached genome-wide significance for population density in the UKB, and this number increased to 12 when correcting for SES. Similarly, we identified 13 loci associated with SES in the UKB when correcting for population density. We observed fewer significant associations with population density or remoteness in the smaller samples, but they did not correspond to the SNP associations found in the UKB. The SNP heritability ranged from 0.6% (QIMR: SE, 3.2%) to 9.3% (QSKIN: SE, 4.0%) as estimated by LD score regression (eTable 5 in the Supplement). The genetic correlation between population density across samples ranged from 0.30 (SE, 0.44; \( \hat{P} = .49 \), QSKIN-NTR) to 0.61 (SE, 0.29; \( \hat{P} = .04 \), UKB-NTR). eAppendix 13, eFigures 9-16, and eTable 5 in the Supplement provide all GWAS and LD score regression results (including GWAS meta-analysis).

Mendelian Randomization

Finally, we selected between 88 and 94 genome-wide significant SNPs for schizophrenia as instruments to perform MR analyses with population density as the outcome variable after excluding SNPs showing evidence of pleiotropic effects by the heterogeneity in dependent instruments outlier analysis (implemented in the GSMR software). These numbers are consistent with the 108 independent associations reported by the Psychiatric Genomics Consortium; the difference arose from SNPs not being present or not passing quality control in GWAS of population density and SES.

Estimates from MR analyses performed in the UKB sample were significant (\( \hat{b} = 0.049; \hat{P} = 3.7 \times 10^{-7} \)) using GSMR (Table 2, Figure 3), suggesting that the genetic liability to schizophrenia has a causal association with the tendency to live in urbanized locations. We observed similar effect sizes in all other samples, although the MR results were not significant. We found no evidence of confounding heterogeneity of effect sizes (\( \hat{P} > .30 \)) or from pleiotropy (\( \hat{P} > .05 \)) using the tests implemented in MR-Base. Reverse MR testing (propensity to live in a more dense or low socioeconomic status area as a cause of schizophrenia) was only suggestive of a (larger) association with population density or SES on schizophrenia (\( \hat{b} = 0.20; \hat{P} = .01 \) using GSMR) as it did not survive multiple testing correction. This lack of evidence could have been the result of reduced statistical power, as both genetic instruments had a smaller number of SNPs (n = 12) (Table 2; eAppendix 14 and eTables 6-10 in the Supplement provide detailed MR results).

Discussion

The present study investigated the association between genetic risk for schizophrenia and characteristics of a person’s place of residence (population density, remoteness, and SES) to test the genetic nature of the association between schizophrenia and population density and infer the direction of causation. We used data on where people live collected as part of 4 studies from 3 countries (Australia, United Kingdom, and the Netherlands) for a total of 504,130 participants.

In all 4 nonclinical cohorts, genetic risk for schizophrenia was associated with greater population density of the postcode of residence beyond what could be explained by SES of the area (Figure 2). Results were consistent for remoteness. Our results show that the geographic distribution of the genetic risk for schizophrenia can be explained by genetic factors, with a significant genetic correlation between population density and schizophrenia estimated to be around 0.12%.
Association Between Population Density and Genetic Risk for Schizophrenia

A-D, Genetic risk for schizophrenia showed positive effects for population density in all significant associations. Results in bold highlight significant results after correction for multiple testing. In the discovery Queensland Institute of Medical Research sample, we accounted for the number of tests performed. In the Netherlands Twin Register, UK Biobank, and QSkin Sun and Health Study, we used a significance threshold of .05 as we aimed to replicate the results found with the polygenic risk scores (PRS) calculated over all independent genomic regions. However, when multiple PRS were available, we present all of the results for completeness. Python-based software (LDpred) was used to conduct the analyses in the NTR cohort.

Schizophrenia is not uniform and that participants with higher genetic risk levels live in areas with higher population density over what is expected by chance. We also found a significant association between genetic risk for schizophrenia and SES of the place of residence.
place of residence across the United Kingdom, although this association was not replicated in the 2 Australian samples. Our MR results in the UKB suggest that schizophrenia risk could be a causal factor in the choice to live in more densely populated and low socioeconomic status areas, although more powered analyses would be required to confirm this result in the other cohorts (Figure 3). These results are consistent with the selective migration hypothesis that individuals with genetic liability for schizophrenia tend to move to or remain in urban areas. Larger GWAS for the environmental variables are required to confirm a reverse causation (eTable 9 and eTable 10 in the Supplement) and clarify how psychopathologic traits and residential location relate to each other. More data are needed to clarify the influence of comorbid psychiatric risks and associated traits (eg, educational attainment, creativity, or risk taking) on the reported association.

Our work builds on previous research that reported that the density of population of where one lives is significantly heritable and on nonmolecular studies showing evidence of a familial effect (ie, owing to genetics and/or family environment) in the association between schizophrenia and urban dwelling. Our results complement 2 recent publications on the interplay between schizophrenia risk and neighborhood. The first, from the Swedish registries (N = 759,536), reported an association between schizophrenia PRS and low socioeconomic status neighborhood, which we replicated in the UKB. In addition, we found that SES could not completely explain the association between schizophrenia PRS and population density.

We found a large environmental correlation between SES and population density in Australia (bivariate model including additive genetic, common environmental, and unique environmental factors, $r_c = 0.86$, $P < .0001$; $r_e = 0.33$, $P < .0001$) compared with the estimated genetic correlation ($r_c = 0.35$, $P = .34$) (eTable 4 in the Supplement). Thus, SES is a potential confounder of genetic analyses of population density, and conversely, population density is a confounder of SES genetic analyses. Composition of SES measures studied here differed in each country and also differ from the Swedish study.

A study from the Danish registries found an association between schizophrenia PRS and urban living only for individuals aged 15 years, but not at birth, and did not study the neighborhood during adult age. Herein, we showed that the population density of where a person lives is mostly explained by shared and unique environment, with the heritability increasing with age (eAppendix II in the Supplement). Thus, we are uncertain whether shared environment confounded the results observed in the Danish population. However, we cannot rule out a genetic association between upbringing environment and the disease risk (ie, a passive gene-environment correlation) in which the association is driven by the genotype that a child inherits and the environment in which they are raised. Herein, we rather focused on an active gene-environment correlation, presumably driven by selective migration, by including only older participants who have a higher degree of independence in choosing where they live. More work is needed to confirm and examine these results over age groups, which will likely require large, longitudinal cohorts, such as national registries.

We highlighted the importance of age in our analysis by replicating and expanding previous results: that place of residence is heritable and the heritability increases over time, while the influence of family environment declines (eAppendix II in the Supplement). This age effect, together with sex differences in prevalence and age of onset of schizophrenia, justified the study of interactions between PRS and age and sex that may contribute to the choice of neighborhood; however, these interactions were not significant in our analyses.

### Limitations

A limitation of our study arises from the low power of the GWAS to detect all variants associated with schizophrenia. This lack

---

**Table 2. Summary of the MR-Based Analysis of Schizophrenia and Population Density Across the 4 Cohorts**

<table>
<thead>
<tr>
<th>Method</th>
<th>QIMR No. of SNPs</th>
<th>QIMR b (SE)</th>
<th>QIMR P</th>
<th>UKB No. of SNPs</th>
<th>UKB b (SE)</th>
<th>UKB P</th>
<th>NTR No. of SNPs</th>
<th>NTR b (SE)</th>
<th>NTR P</th>
<th>QSKIN No. of SNPs</th>
<th>QSKIN b (SE)</th>
<th>QSKIN P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed-effects meta-analysis</td>
<td>93</td>
<td>0.045</td>
<td>0.050</td>
<td>.36</td>
<td>92*</td>
<td>0.054</td>
<td>0.009</td>
<td>1.6 × 10^-3*</td>
<td>.37</td>
<td>88</td>
<td>0.043</td>
<td>0.050</td>
</tr>
<tr>
<td>Fixed-effects meta-analysis</td>
<td>93</td>
<td>0.047</td>
<td>0.051</td>
<td>.35</td>
<td>92*</td>
<td>0.049</td>
<td>0.013</td>
<td>3.4 × 10^-7*</td>
<td>.40</td>
<td>88</td>
<td>0.042</td>
<td>0.050</td>
</tr>
<tr>
<td>Random-effects meta-analysis</td>
<td>93</td>
<td>0.043</td>
<td>0.054</td>
<td>.43</td>
<td>92*</td>
<td>0.046</td>
<td>0.013</td>
<td>5.8 × 10^-4*</td>
<td>.40</td>
<td>88</td>
<td>0.042</td>
<td>0.050</td>
</tr>
<tr>
<td>Maximum likelihood</td>
<td>93</td>
<td>0.045</td>
<td>0.051</td>
<td>.37</td>
<td>92*</td>
<td>0.056</td>
<td>0.011</td>
<td>9.6 × 10^-3*</td>
<td>.37</td>
<td>88</td>
<td>0.044</td>
<td>0.050</td>
</tr>
<tr>
<td>MR Egger</td>
<td>93</td>
<td>0.49</td>
<td>0.026</td>
<td>.06</td>
<td>92</td>
<td>0.16</td>
<td>0.06</td>
<td>.013</td>
<td>88</td>
<td>.006</td>
<td>0.240</td>
<td>.97</td>
</tr>
<tr>
<td>Weighted median</td>
<td>93</td>
<td>0.11</td>
<td>0.075</td>
<td>.13</td>
<td>92*</td>
<td>0.053</td>
<td>0.015</td>
<td>7.7 × 10^-4*</td>
<td>.31</td>
<td>88</td>
<td>0.074</td>
<td>0.074</td>
</tr>
<tr>
<td>Inverse variance weighted</td>
<td>93</td>
<td>0.045</td>
<td>0.056</td>
<td>.41</td>
<td>92*</td>
<td>0.053</td>
<td>0.013</td>
<td>1.0 × 10^-4*</td>
<td>.37</td>
<td>88</td>
<td>0.043</td>
<td>0.050</td>
</tr>
<tr>
<td>GSmR</td>
<td>93</td>
<td>0.019</td>
<td>0.051</td>
<td>.41</td>
<td>92*</td>
<td>0.049</td>
<td>0.011</td>
<td>3.7 × 10^-7*</td>
<td>.33</td>
<td>88</td>
<td>0.049</td>
<td>0.051</td>
</tr>
</tbody>
</table>

Abbreviations: GSMR, generalized summary data-based mendelian randomization; MR, mendelian randomization; NTR, Netherlands Twin Register; QIMR, QIMR Berghofer Medical Research Institute; QSKIN, QSkin Sun and Health Study; SNPs, single-nucleotide polymorphisms; UKB, United Kingdom Biobank.

*Statistically significant results.
of power results in a limited PRS instrument that explains only 11.6% of the total trait heritability in the population.\(^\text{11}\) Thus, the variance explained by the current PRS, which is based on common variants, may be only one-tenth of what one would observe with a PRS capturing the whole genetic signal. As a consequence, the small association with population density reported herein may account for 1.7% (95% CI, 0.8%-3.2%) of the schizophrenia risk (based on population density explaining 0.2% of schizophrenia PRS in the QIMR data and 0.002 of 0.116 = 0.017). Ancestry may also confound PRS analyses,\(^\text{47}\) especially for the variables studied herein (eAppendix 15 and eFigures 17-19 in the Supplement\(^\text{48}\)). We tried to overcome this issue using mixed models that are equivalent to fitting all genetic principal components as covariates (eAppendices 8 and 15 in the Supplement).\(^\text{48}\)

Another limitation is the possible sample overlap between the UKB sample and data used in the schizophrenia GWAS, which may inflate results from PRS and MR analyses (eAppendix 16 in the Supplement). However, we estimated the overlap to be negligible (64 participants or 0.01%, eAppendix 16 in the Supplement), and we did not observe larger effect sizes in the UKB compared with the 3 other cohorts.

Conclusions

Our findings support the notion that the increased schizophrenia prevalence in urbanized areas is not only owing to the environmental stressors of the city or other putative risk factors associated with urbanicity (eg, increased risk of infection, low vitamin D levels, and substance abuse)\(^\text{49}\) but also on the genetic risk for the disease. The associated PRS prediction was replicated across 3 different countries that likely differ in availability of space, social mobility/ opportunities, associations between population density and SES of the area, and historical constraints on living environ-
Association Between Population Density and Genetic Risk for Schizophrenia

Future disease models will need to include both genetic selection and environmental factors of urban stress on schizophrenia to inform implications for intervention. In addition, there is a need to address the potential gene-by-environment interactions that would arise if genetic variants influencing schizophrenia also influence the choice of a stressful neighborhood, which would contribute to the interaction between urbanicity and family history of schizophrenia that has been reported in the Danish population.30 Such diathesis-stress interaction studies using PRS have been published for depression,51,52 but given the lower prevalence of schizophrenia, national registries will likely be required for investigation.

REFERENCES


