Examining the association between genetic risk for depression, wellbeing and schizophrenia, and proximity to greenspace

Zoe E. Reed[1,2*], Tim T. Morris[2,3], Oliver S.P. Davis[2,3,4,5], George Davey Smith[2,3], Marcus R. Munafò[1,2,4], Gareth J. Griffith[2,3]

*Corresponding author

1 School of Psychological Science, University of Bristol, Bristol, BS8 1TU, UK

2 Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, BS8 2BN, UK

3 Population Health Sciences, Bristol Medical School, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK

4 National Institute for Health Research Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, BS8 2BN, UK

5 Alan Turing Institute, British Library, London, NW1 2DB, UK

Abstract

It is unclear whether associations between between greenspace and mental health reflects those with better mental health seeking out greener environments.

To examine this we 1) test associations between depression, wellbeing and schizophrenia polygenic scores (PGS) with two greenspace measures in UK Biobank (N=238,306 and 293,922), 2) estimate multilevel-models (MLM), clustering individuals by local geography and, 3) conduct one-sample Mendelian randomisation (MR) to estimate causal effects.

Depression and schizophrenia PGS were associated with lower greenspace, whilst wellbeing PGS was associated with higher greenspace. Locally-clustered MLM demonstrated attenuation for the individual wellbeing PGS association and a reversal of effect for the schizophrenia PGS association. MR revealed evidence of a causal effect of increased depression liability on decreased greenspace.

Therefore, greater liability to depression may cause selection into less green neighbourhoods. Our results also highlight how individual-level PGS associations can be biased by contextual, between-area differences in outcome.

Introduction

Living in areas with greater greenspace has been shown to associate with both decreased prevalence (and symptoms) of mental health conditions and increased wellbeing[1,2]. Evidence has been found across many indicators of greenspace, measured in terms of both land use[3,4] and vegetation[5–7], and across mental health outcomes, for instance, decreased psychological distress and depression[5,8–10] and increased wellbeing[11]. In this research we consider two of the most commonly used and widely validated indicators. The first uses the natural reflective properties of photosynthesising plants to infer vegetation density from satellite imagery[12], and the second uses administrative data from the land use database to estimate greenspace availability[13].

Similarly, residing in more urban areas, which have less greenspace, is associated with increased prevalence of mental health conditions[14], including depression and schizophrenia[15–19]. Recent meta-analyses of urbanicity and mental health studies found that risk for a range of mental health conditions was higher in urban areas[15,18], with odds ratios (ORs) between 1.13 and 1.33. A recent Danish population-based study found incidence rate ratios between 1.15 and 2.17 for the most urban areas compared to rural areas across multiple mental health conditions[20]. Greenspace has been suggested to drive a component of urban-rural differences in health outcomes[21].

These associations are commonly ascribed to the “social drift” hypothesis[22]. Social drift proposes that those with greater risk of current or future mental health conditions may migrate, or “drift”, to more deprived, urban areas, driving the

observed concentration of such conditions in urban centres[14], and in turn be less likely to have access to greenspace[19]. There is, however, a long history of research contesting the causal directionality of this relationship[23,24]. In the contrary hypothesis, socioeconomic position (SEP) or greenspace are considered causal risk factors for developing a mental health condition, rather than simply confounding the relationship between poor mental health and selective migration. Such debates are further complicated by spatial confounding, as urbanicity and greenspace are intimately related to many socio-demographic variables known to associate with mental health[1,21,25]. Therefore, identifying a causal effect of greenspace over that which might be caused by correlated, contextual confounds is difficult. The resulting identification issues contribute to the ongoing uncertainty around the causal nature of the mental health and greenspace, or urbanicity, relationship.

Previous studies exploring causal relationships between mental health and urbanicity found evidence for a causal effect of increased genetic risk for schizophrenia on living in more densely populated and deprived areas[22,26]. Recent systematic reviews found evidence for mental health effects of environmental interventions to increase greenspace was absent for adults[27]. However, risk-of-bias assessments in this study indicated that only four of the 14 studies provided robust data, so studies with more robust designs are needed to draw reliable conclusions.

Further evidence is therefore required to establish whether a causal relationship between greenspace and mental health liability does exist. Furthermore, such analyses must determine whether it is greenspace specifically and not urbanicity or SEP confounds driving these associations.

77 In this study, we assessed the association between genetic risk for several mental
78 health outcomes and current residential greenspace for participants in UK Biobank.
79 We also tested whether observed associations may be causal using one-sample
80 Mendelian randomisation (MR)[28–30]. MR exploits the randomisation of genotypes at
81 conception to estimate causal effects of modifiable exposures on outcomes, here
82 mental health on greenspace. We used polygenic scores (PGS) for depression,
83 wellbeing and schizophrenia. We suggest that an association would support that
84 those with a genetic predisposition for more positive mental health outcomes may
85 preferentially select into environments with more greenspace.

Results

Sample characteristics

Sample characteristics for our UK Biobank sample are presented in Table 1.

Table 1. Sample characteristics

	N with data available	Mean (SD) or percentage (%)
Age	336,997	56.87 (8.00)
Sex	336,997	53.79% females
NDVI	238,306	0.16 (0.17)
Percentage greenspace	293,922	36.86 (23.64)
Depression diagnosis	336,997	12.53%
MHQ depression	79,946	26.82%
Wellbeing	109,943	4.59 (0.78)
Schizophrenia diagnosis	293,922	0.17%

SD=Standard Deviation, NDVI=Normalised Difference Vegetation Index, MHQ=Mental Health Questionnaire

Phenotypic associations

Results from linear regression models between depression, wellbeing and schizophrenia and two greenspace outcomes, adjusting for age and sex are presented in Supplementary Table S1. The two measures of greenspace we used were the mean Normalised Difference Vegetation Index (NDVI) and the percentage of an area within a 300m radius of the participants home classified as greenspace (see Methods for further details). We found negative associations between the depression and schizophrenia diagnosis phenotypes and greenspace measures, although there was weaker evidence of this for the Mental Health Questionnaire

(MHQ) depression measure. We found positive associations between wellbeing and the greenspace measures, although evidence for this was weaker with the NDVI outcome.

Linear regression model associations between polygenic scores and NDVI and percentage greenspace

Results from the linear regression model adjusting for age, sex and the first 25 principal components (PCs) are presented in Table 2 for NDVI and Table 3 for percentage greenspace. Table 2 and 3 present results for PGS calculated at a p-value threshold of 0.05 (see Supplementary Tables S2 and S3 for other p-value thresholds). Polygenic risk for depression was negatively associated with residing in green areas for both NDVI (b=-0.008; 95% CI: -0.01 to -0.004; p=0.0001) and percentage greenspace (b=-0.005; 95% CI: -0.01 to -0.001; p=0.01). This was in a broadly consistent direction across most PGS p-value thresholds (Supplementary Tables S2 and S3); however, we only found strong evidence of this association at less stringent thresholds ($1x10^{-03}$ for the NDVI measure and 0.01 for percentage greenspace).

The wellbeing PGS was positively associated with residing in greener areas for percentage greenspace only (b=0.01; 95% CI: 0.003 to 0.01; p=0.0005). This was in a consistent direction across all PGS p-value thresholds. However, we again only found strong evidence of this using a p-value threshold above 0.01 for percentage greenspace. We did not find strong evidence of this association with the NDVI measure (b=-0.0004; 95% CI: -0.004 to 0.003; p=0.83), although the direction of effect was positive for most other p-value thresholds.

For the schizophrenia PGS, we found strong evidence of an association with percentage greenspace only (*b*=-0.004; 95% CI: -0.01 to -0.001; *p*=0.02). This was in a consistent direction across most PGS p-value thresholds, although there is only strong evidence of this using a p-value threshold above 0.05 (Supplementary Table S3). We did not find evidence of an association with the NDVI measure (*b*=0.0005; 95% CI: -0.004 to 0.005; *p*=0.80).

Table 2. Associations between polygenic scores for depression, wellbeing and schizophrenia and NDVI for p-value threshold of 0.05 (N=238,306)

Phenotype	**Beta***	**2.5% CI**	**97.5% CI**	**P**
Depression	-0.008	-0.01	-0.004	0.0001
Wellbeing	-0.0004	-0.004	0.003	0.83
Schizophrenia	0.0005	-0.004	0.005	0.80

*CI=Confidence Interval. *Beta reflects the NDVI SD change predicted by an SD increase in the PGS*

Table 3. Associations between polygenic scores for depression, wellbeing and schizophrenia and percentage greenspace for p-value threshold of 0.05 (N=293,922)

Phenotype	**Beta***	**2.5% CI**	**97.5% CI**	**P**
Depression	-0.005	-0.009	-0.001	0.007
Wellbeing	0.006	0.003	0.01	0.0005
Schizophrenia	-0.004	-0.008	-0.0006	0.02

*CI=Confidence Interval. *Beta reflects the percentage greenspace SD change predicted by an SD increase in the PGS*

Multilevel model associations between polygenic scores and NDVI and percentage greenspace, accounting for census area

The random-intercept models allowing neighbourhood or district differentials both fit better than the linear model (Supplementary Table S4 shows comparisons of fit statistics for all three models). The level-2 VPC, which assesses the proportion of total, unexplained variation in outcome accounted for by a specific level in the model (see Supplementary Material Section 1 for further details), for the neighbourhood and district models respectively were 0.83 and 0.52 for the NDVI measure, and 0.48 and 0.24 for percentage greenspace. This implies that after adjusting for mental health PGS (across any measure), and 25 principal components, 83% of the remaining unexplained variation in individual-level NDVI exposure can be explained by the neighbourhood of the individual, or equally (from the ICC definition) we would expect a correlation of 0.83 between NDVI measures taken from pairs of individuals within the same neighbourhood.

Results from these models are presented in Table 4 for NDVI (and Supplementary Table S5) and Table 5 for percentage greenspace (and Supplementary Table S6). Results from these models are comparable to those from the linear model for depression liability with NDVI under both the neighbourhood specification (b=-0.002; 95% CI: -0.003 to -0.0004; p=0.01) and the district specification (b=-0.004; 95% CI: -0.01 to -0.002; $p=2.60\times10^{-04}$). They were also comparable for the percentage greenspace measure for neighbourhood (b=-0.004; 95% CI: -0.01 to -0.001; $p=5.31\times10^{-03}$) and district specifications (b=-0.01; 95% CI: -0.01 to -0.004; $p=1.61\times10^{-05}$). For these variables, within-area effect estimates were consistent direction with global effect estimates.

The coefficient for wellbeing and percentage greenspace attenuated in both the neighbourhood specification (b=0.002; 95% CI: -0.001 to 0.005; p=0.12) and district

specification (*b*=0.004; 95% CI: 0.001 to 0.01; *p*=0.01), indicating that the initial association may reflect aggregate differences in the wellbeing PGS between neighbourhoods and/or districts. However, there was still no clear evidence of an association with the NDVI measure in either specification.

Finally, for schizophrenia liability we still did not find any evidence of an effect on NDVI in either neighbourhood or district specification. However, for the percentage greenspace measure we observed an interesting effect in that the direction of the association was reversed in both the neighbourhood (*b*=0.003; 95% CI: 0.0002 to 0.01; *p*=0.03) and district specifications (*b*=0.001; 95% CI: -0.002 to 0.005; *p*=0.50). There was weaker evidence of this for the district specification; however, for the neighbourhood specification this effect was observed for all p-value thresholds. In specifying a random intercept model, we allow for between-neighbourhood (and between-district) differences in NDVI to be estimated as normally distributed residuals from a global mean. This means that the beta estimates functionally represent a within-area estimate, for instance, the effect of PGS on NDVI, given that you are in a neighbourhood with a higher-than-average NDVI. The change in direction of this sign implies that although the global relationship between schizophrenia PGS and greenspace was negative, this is mostly driven by neighbourhoods with high NDVI tending to have lower than average PGS for schizophrenia. When we account for between-area differences in our intercept, we see that within-neighbourhoods the relationship was positive. This can be considered an example of Simpson's paradox[52,53] where an individual model would fail to address the important context of neighbourhood in determining an individual's NDVI.

Table 4. Associations between polygenic scores for depression, wellbeing and schizophrenia and NDVI measure, with neighbourhood or district level random intercepts, for p-value threshold of 0.05 (N=238,306).

	Neighbourhood				**District**			
Phenotype	**Beta***	**2.5% CI**	**97.5% CI**	**P**	**Beta***	**2.5% CI**	**97.5% CI**	**P**
Depression	-0.002	-0.003	-0.0003	0.02	-0.004	-0.006	-0.002	2.13×10^{-04}
Wellbeing	0.001	-0.0006	0.003	0.22	0.002	-0.0004	0.004	0.11
Schizophrenia	-0.0006	-0.002	0.001	0.45	-0.0007	-0.003	0.001	0.52

*Neighbourhood= Middle Layer Super Output Area or intermediate zone, District= Local Authority District or Local Authority, CI=Confidence Interval. **Beta reflects the NDVI SD change predicted by an SD increase in the PGS, for the average individual in a typical neighbourhood or district.*

Table 5. Associations between polygenic scores for depression, wellbeing and schizophrenia and percentage greenspace with neighbourhood or district level random intercepts, for p-value threshold of 0.05 (N=293,922).

	Neighbourhood				**District**			
Phenotype	**Beta***	**2.5% CI**	**97.5% CI**	**P**	**Beta***	**2.5% CI**	**97.5% CI**	**P**
Depression	-0.004	-0.006	-0.001	6.06×10^{-03}	-0.007	-0.01	-0.004	1.10×10^{-05}
Wellbeing	0.002	-0.0006	0.005	0.14	0.004	0.0009	0.007	0.01
Schizophrenia	0.003	0.0003	0.006	0.03	0.001	-0.002	0.004	0.51

*Neighbourhood= Middle Layer Super Output Area or intermediate zone, District= Local Authority District or Local Authority, CI=Confidence Interval. *Beta reflects the percentage greenspace SD change predicted by an SD increase in the PGS, for the average individual in a typical neighbourhood or district.*

199 *One-sample Mendelian randomisation for depression and wellbeing on greenspace*
200 *measures*

201 We used one-sample MR analyses to assess whether there was evidence of a
202 causal effect of depression liability on the NDVI measure and percentage
203 greenspace, of wellbeing on percentage greenspace, and of schizophrenia liability
204 on percentage greenspace. Results from the one-sample MR analyses are shown in
205 Table 6 and results from sensitivity analyses are presented in Supplementary Table
206 S7. We found strong evidence for a causal effect of the liability of having a diagnosis
207 of depression on lower percentage greenspace (*b*=-0.51; 95% CI: -0.93 to -0.10;
208 *p*=0.01). This was not apparent for the MHQ depression measure, although the
209 direction of effect was the same. We did not find strong evidence of a causal effect of
210 liability for the depression measures on the NDVI outcome; however, the direction of
211 effect was the same for both measures. We did not find strong evidence of a causal
212 effect of wellbeing on percentage greenspace for the PGS constructed for genome-
213 wide significant SNPs or at the p-value threshold of $1x10^{-05}$, although the direction of
214 effect was positive. Finally, we did not find strong evidence of a causal effect of
215 schizophrenia liability on percentage greenspace.

216 We did not find any strong evidence for causal effects in the sensitivity analyses
217 (Supplementary Table S7), although the direction of effect was consistent with our
218 main results for depression liability and percentage greenspace.

219 Table 6. One-sample Mendelian randomisation results

Exposure	**Outcome**	**N**	**Beta***	**2.5% CI**	**97.5% CI**	**P**
Depression	NDVI	238,306	-0.26	-0.74	0.23	0.30

diagnosis						
MHQ depression	NDVI	64,740	-0.01	-0.51	0.50	0.97
Depression diagnosis	Percentage greenspace	293,922	-0.51	-0.93	-0.10	0.01
MHQ depression	Percentage greenspace	77,358	-0.54	-1.14	0.06	0.08
Wellbeing (GW PGS)	Percentage greenspace	97,099	0.16	-0.49	0.81	0.63
Wellbeing (PGS at $1x10^{-05}$)	Percentage greenspace	97,099	0.06	-0.75	0.87	0.88
Schizophrenia	Percentage greenspace	293,922	6.75	-7.95	21.45	0.37

*CI=Confidence Interval, GW=Genome-wide, PGS=Polygenic score, NDVI=Normalised Difference Vegetation Index, MHQ=Mental Health Questionnaire. *Beta reflects the NDVI or percentage greenspace SD change predicted by an increased genetic liability for depression and schizophrenia or every one unit increase in the wellbeing score*

Discussion

Our results indicate that genetic risk for depression and wellbeing are associated with the greenspace of the areas which participants reside in. We found the strongest evidence for depression, with increased genetic risk associated with decreased greenspace for two different measures of greenspace. We also found evidence that this effect might be causal for percentage greenspace – in other words, that those at a higher genetic risk for depression may select into less green areas. Our MR sensitivity analyses did not provide strong evidence of a causal effect; therefore, our results should be interpreted with caution. However, the sensitivity analyses used individual SNPs and our results may therefore have been impacted by low statistical power. The positive association with wellbeing was only observed for the percentage greenspace measure, but the direction of effect was the same with the NDVI measure. Our results support previous observational studies which found an association between greenspace and depression and wellbeing[2,5,9,17,18] and extend these findings by suggesting a causal effect of liability to depression diagnosis on greenspace. As greenspace cannot cause genetic risk for these mental health outcomes, this overcomes issues around reverse causation.

In addition, models specifying a random intercept specification for local context indicated that the relationship for depression and greenspace was negative both within- and between-neighbourhoods and districts. Our concern that global effect estimates may produce misleading PGS relationships was borne out for all phenotypes, as the inclusion of an area-level random intercept attenuated results for depression liability and wellbeing, and inverted the sign of the schizophrenia liability relationship. For this study (and, we argue, for other genetically informed study

248 designs) the effect estimate of interest is the relationship between PGS and outcome
249 over and above that which is due to area-level compositional differences and other
250 residual area-level confounding.

251 This suggests that at least part of the PGS-greenspace relationships we see were
252 due to aggregate between-area differences, and not due to the impact of those PGS
253 within an area. For instance, attenuation for random effects model estimates for
254 wellbeing with percentage greenspace suggests that the initial association was likely
255 due to individuals' wellbeing PGS being consistent within an area. Thus, when we
256 accounted for between-area differences via the random intercept specification, the
257 observed relationship within-area became far less strong, as most of this initial
258 relationship was driven by certain areas having both high wellbeing PGS and high
259 greenspace levels.

260 This is evidenced more explicitly by the extremely high level-2 VPC values. We note
261 that these were not identical for all mental health PGS but the amount of variance
262 explained by each PGS was very small, so the VPC was almost the same for each
263 model. A VPC of 0.83 implies that simply knowing individuals' neighbourhoods would
264 allow you to account for 83% of the unexplained, observed variation in phenotypic
265 NDVI. This is important as, coupled with the reversal of sign of the effect estimate in
266 our schizophrenia liability model, it implies our initial global effect estimate was
267 mostly driven by individuals living within the same area (both neighbourhood or
268 district areas) being more phenotypically and genetically similar to their immediate
269 neighbours than to those living in other areas. This is particularly concerning as this
270 is what adjustment for genetic principal components is often cited as insulating
271 against, with the implication that adjustment for genetic ancestry should deal with

spatial confounding between UK Biobank participants. This appears to not be the case, and as such we propose that local neighbourhood random intercept specifications be included in PGS analyses as a routine robustness check to ensure that global PGS estimates are not being biased by such spatial confounding. That said, our study design is likely an extreme case of such confounding, as the greenspace value for any given individual can be reasonably inferred to be nearly identical of that of their immediate neighbours, simply because the outcome is defined within a spatial boundary.

Our results for depression and wellbeing PGS are consistent with the social drift hypothesis, which supports a role of the movement of genetically similar individuals into more urban/less green areas. Here, this suggests those with greater genetic risk of poor mental health may select into areas with less greenspace. We did not test a bidirectional effect for mental health with greenspace as we were unable to construct strong genetic instruments for greenspace. We did run GWAS for both greenspace measures in UK Biobank, but only found one genome-wide significant SNP for NDVI, and none for percentage greenspace so analyses would be impossible, or underpowered to detect any effects. If larger GWAS of contextual phenotypes were to be conducted, this may help investigate causation in this direction.

The relationship we observe for schizophrenia liability is particularly interesting. We found that for both greenspace measures when we modelled the association with schizophrenia liability the direction of the association reversed when specifying an area-level random-intercept parameter, an example of Simpson's paradox[54]. There are multiple reasons why this may occur, but one which is consistent with known UK Biobank selection issues is that this is caused by non-random selection within areas.

296 UK Biobank participation is known to be patterned by both phenotypic schizophrenia
297 and schizophrenia PGS[36]. We suggest that the results we see are consistent with
298 collider bias where, within-areas, those with low schizophrenia PGS and in greener
299 areas are more likely to have participated in UK Biobank. Assuming no complex
300 interactions, this would induce a positive relationship between schizophrenia PGS
301 and greenness as a result of collider bias, where individuals with a high
302 schizophrenia PGS are more likely to have high greenspace values. What is striking,
303 however, is that schizophrenia PGS and greenspace are still clustered strongly
304 enough that between areas we see the anticipated relationship, implying that areas
305 with higher greenspace tend to be populated with individuals with lower
306 schizophrenia PGS to a strong enough degree that the global effect estimate is
307 negative in the full UK Biobank population.

308 Our results also differed between the NDVI and greenspace measure used. Our
309 results are consistent with previous observational studies showing increased NDVI,
310 which indicates visible greenness of an area was associated with fewer depressive
311 symptoms, decreased job-related stress and psychological distress[5–7]. However, in
312 other studies using a similar measure to ours based on data from the Generalized
313 Land Use Database (i.e., percentage of greenspace in an area), associations with
314 mental health outcomes were not generally found[3,55,56].

315 Our study has several limitations. First, PGS analyses have well known limitations;
316 for example, different p-value thresholds to construct scores affects the predictive
317 accuracy of the score. More SNPs may increase the variance explained but may
318 also suffer from overfitting. In our main analyses we focused on one p-value
319 threshold but included other thresholds as sensitivity analyses (see Supplementary

320 Tables S2 and S3) and found consistent results across these. Second, our large
321 sample size means that the estimated associations will have increased precision and
322 we may overstate confidence in associations which may not generalise to a wider
323 population. Similarly, these results were obtained from a European ancestry sample
324 and may not be generalisable to other ancestry groups. Third, it is possible that the
325 GWAS for our mental health phenotypes also include some SNPs associated with
326 "upstream" traits like greenspace (i.e., the primary phenotype is mis-specified)[57]. As
327 we do not know what SNPs influence greenspace, we are unable to account for this.
328 However, the PRS analyses are more likely to be impacted by this than the MR
329 analyses given the less stringent p-values used and therefore the greater number of
330 SNPs included that may capture these "upstream" traits. Fourth, whilst we may
331 increase the likelihood of obtaining generalisable estimates by including a random
332 intercept for area-context, participation in UK Biobank at baseline was extremely
333 non-random as we outline above. Fifth, we were not able to be prescriptive with our
334 operationalisation of greenspace. A recent study illustrated the sensitivity of results
335 to greenspace specification; for instance, availability, accessibility, and visibility of
336 greenspace produced very different results when included in analyses[58].The NDVI
337 measure likely offers a more bespoke measure of greenspace in rural settings as it is
338 calculated at a household level, rather than at OA level (rural OAs are sparser and
339 larger). However, NDVI simply captures surface reflectivity from satellite imagery, so
340 has no sensitivity to the accessibility of the greenspace it is capturing, unlike the
341 percentage greenspace measure. Future research into this must carefully consider
342 which components of greenspace are of interest for each research question. Sixth, in
343 our MR analyses we were possibly underpowered to detect an effect for wellbeing
344 due to the small number of genome-wide significant SNPs identified in the GWAS.

Seventh, a key limitation of MLM approaches is that endogenous, omitted variables associated with the higher-level contexts can reduce generalisability. If the aim of an analysis is solely to adjust for higher level context, then a fixed effects approach might be more appropriate (see Abdellaoui et al[59]). Finally, we also need to consider that geographical location will also be, to some extent, accounted for in our PGS[60], as well as other types of structure, which may include population stratification and assortative mating[61].

In conclusion, we found evidence to suggest that those at risk of poorer mental health outcomes, specifically depression and wellbeing, may select into less green areas. Further studies investigating bidirectionality are called for as if lack of greenspace further exacerbates poor mental health then interventions targeting greenspace may be important. Finally, our study highlights the potential for spatial confounding to bias genetic analyses of geographically clustered phenomena. We suggest therefore, that increasing availability of genetic datasets with geographical location will allow for future research to understand and quantify these potential biases.

361 **Methods**

362 **Cohort description**

363 The UK Biobank is a large population-based prospective health research resource of
364 503,317 participants, aged between 38 and 73 years at recruitment, recruited
365 between 2006 and 2010, from across the UK[31]. The study aimed to identify
366 determinants of human disease in middle and old age. A range of data have been
367 collected across sociodemographic, health status, lifestyle, cognitive function, self-
368 reported measures and physical and mental health measures. Data were collected
369 via a number of methods, including paper and web-based questionnaires, computer
370 assisted interviews, clinic visits and data linkage. Baseline assessment took place at
371 22 assessment centres in the UK to enable recruitment from a range of locations.
372 Further information can be found on the UK Biobank website
373 (www.ukbiobank.ac.uk). UK Biobank received ethical approval from the Research
374 Ethics Committee (REC reference for UK Biobank is 11/NW/0382). We excluded
375 participants based on the latest withdrawal lists at the time of data extraction
376 (04/02/2020) for our UK Biobank project (project number: 21829).

377 **Phenotypic measures**

378 The proposed mechanism of greenspace impact is inconsistently defined in the
379 literature[32]. Therefore, we used two measures to capture participants' residential
380 access to greenspace. These measures were made available to UK Biobank by
381 Sarkar et al., and were derived using exact postcodes[33]. Both measures were
382 standardised prior to inclusion in models.

383 *Normalised Difference Vegetation Index*

Our first outcome was the mean Normalised Difference Vegetation Index (NDVI), calculated for a 500m radius of the participant's home location. NDVI is a commonly used parameterisation of green-space access as it can be calculated from satellite imagery, rather than requiring administrative land use information[6,7], and closely approximates subjective exposure estimates[12]. The NDVI ranges between -1 and +1, with positive values indicating greener areas based on reflectance measures in Colour InfraRed (CIR) satellite data[33]. The distribution in UK Biobank was -0.50 to 0.70. After exclusions based on withdrawn consent and quality control filtering for genetic data (described below) and including only those with genetic data, the final sample size for NDVI was 238,306 (see Figure 1).

Percentage greenspace

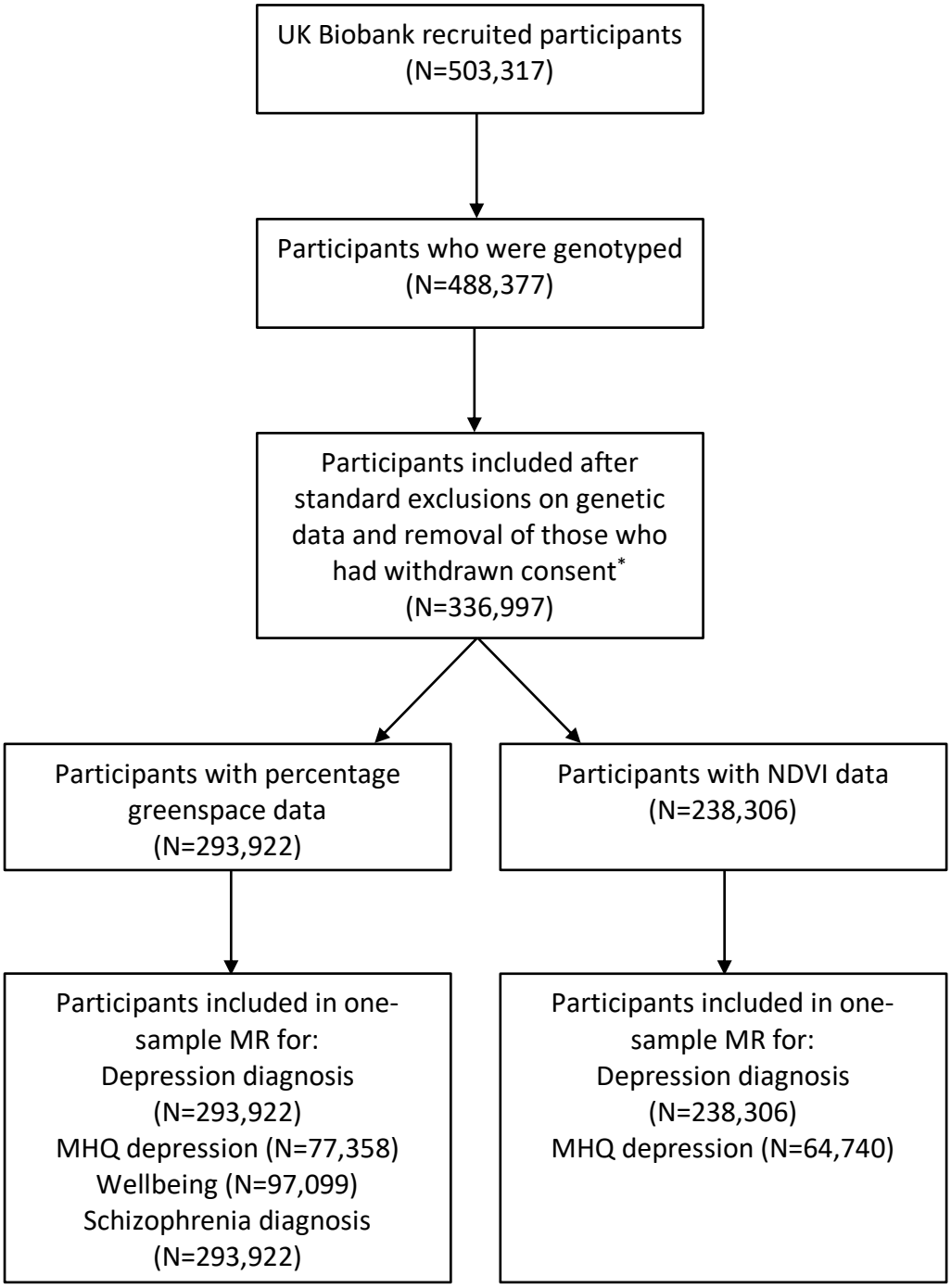
Our second outcome refers to the percentage of an area within a 300m radius of the participants home classified as greenspace. This is based on land use data from the Generalized Land Use Database (GLUD)[13] for England 2005 data at the 2001 census output area (OA) level. Census output areas were designed to have a consistent population, thus their area varies considerably with population density. Each participant's home location was allocated an area weighted mean of these data for output areas intersecting the 300m radius of a home. After exclusions and including only those with genetic data, the final sample size used in analyses with this outcome was 293,922 (see Figure 1).

The percentage greenspace measure was only weakly correlated with the NDVI measure (r=0.16). This is likely because they capture different aspects of greenspace; the NDVI measure captures greenness via the reflectance of the land surface, whereas the percentage greenspace classification captures functional

408 greenspace, including areas of natural agricultural land, recreational areas and other

409 areas of grassland, and excluding domestic gardens[13].

410 Figure 1. Flowchart of number of participants included in analyses



411

412 **Participants who withdrew consent by 04/02/2020*

413 *NDVI= Normalised Difference Vegetation Index*

414 *Depression*

415 For one-sample MR analyses, we also require phenotypic data on mental health
416 outcomes. We used data from a combination of both ICD-10 diagnoses and
417 diagnoses reported in verbal interviews with a nurse at the assessment centre.
418 Anyone who had a diagnosis of depression from any of these sources was coded as
419 having depression. Anyone without a diagnosis indicated in these data sources was
420 coded as not having depression. We refer to this combined measure as diagnosis of
421 depression. Data were also available for a sub-sample of UK Biobank participants
422 who completed questions related to depressive and manic symptoms during the final
423 two years of recruitment. From these data, a variable for having bipolar disorder and
424 major depression was derived, where a participant reporting both was coded only as
425 having the most severe condition[34]. We excluded participants who had bipolar
426 disorder type 1 or 2 and identified those as having depression if they had probable
427 recurrent major depression (severe or moderate) or a single probable major
428 depression episode. We refer to this as the mental health questionnaire (MHQ)
429 depression measure.

430 *Wellbeing*

431 As part of an on-line mental health self-assessment questionnaire issued in 2016,
432 participants were asked the question 'In general how happy are you?'. Participants
433 could select from the following options: 'Prefer not to answer', 'Do not know',
434 'Extremely happy', 'Very happy', 'Moderately happy', 'Moderately unhappy, 'Very
435 unhappy' and 'Extremely unhappy'. We excluded participants who selected either of
436 the first two options and coded the other options so that a higher number indicated a
437 more positive response.

438 *Schizophrenia*

439 Similar to the depression diagnosis variable, we used data from a combination of
440 both ICD-10 diagnoses and diagnoses reported in verbal interviews with a nurse at
441 the assessment centre. Participants with a diagnosis of schizophrenia from any of
442 these sources were coded as having schizophrenia. Participants without a diagnosis
443 indicated in these data sources were coded as not having schizophrenia. In UK
444 Biobank 0.17% of participants were coded as having schizophrenia, compared to
445 0.7% reported for psychotic disorders in England[35]. This is consistent with known
446 selection pressures in UK Biobank , as demonstrated previously with reduced
447 participation in individuals with greater liability to schizophrenia[36].

448 *Potential confounders*

449 We included age at time of initial assessment and sex in all of our analyses. As we
450 examined phenotypes related to geographical location, we also included the first 25
451 principal components of population structure (PCs) as confounders in all of our
452 models including PGS. The number of PCs was chosen to reduce geographical
453 confounding. A recent study found that adjusting for 18 PCs was a sensible
454 correction for population structure[37], but in this study we corrected for 25 PCs as this
455 was where we saw a flattening in the variance explained by the PCs in our outcomes
456 (Supplementary Figure S1).

457 *Structural Variables*

458 In our multilevel models we specified geographical clustering using the
459 administrative area that each participant resided in at the time of recruitment.
460 Participant locations were matched on easting and northing coordinates[38] to the

most recently defined local administrative units at two spatial scales, one to represent local neighbourhood context, and one the broader city and commuting context. We consider Middle Layer Super Output Areas (MSOAs, average population = 7,200) for England and Wales analogous to Intermediate Zones (IZ, average population = 4,200) for Scotland – hereafter, “neighbourhoods”. Similarly, we consider Local Authority Districts (LADs, average population 140,000) analogous to Local Authorities (LAs, average population 170,000) for Scotland, hereafter “districts”. Supplementary Figures S2 and S3 show the number of UK Biobank participants per neighbourhood and district respectively. Data were rounded to the nearest kilometre to preserve anonymity, meaning some participants’ recorded locations fell outside of census boundaries (0.6% of participants). These participants were assigned to the nearest neighbourhood or district based on Euclidean distance to the census output area centre.

Genetic data

There were 488,377 participants with genotyped samples, of which 49,979 were genotyped using the UK BiLEVE array and 438,398 using the UK Biobank axiom array. Pre-imputation quality control, phasing and imputation have been described elsewhere[39]. In summary, multiallelic SNPs and those with a MAF≤1% were removed. Phasing of genotype data was performed using a modified version of the SHAPEIT2 algorithm. The SNPs used were imputed to the Haplotype Reference Consortium (HRC) reference panel, using IMPUTE2 algorithms. A graded filtering with different imputation qualities for different MAF ranges was used (Info>0.3 for MAF>3%, Info>0.6 for MAF 1-3%, Info>0.8 for MAF 0.5-1% and Info>0.9 for MAF 0.1-0.5%), where MAF and info scores were recalculated on an in-house derived

485 ‘European’ subset. Individuals with sex-mismatch or sex-chromosome aneuploidy
486 were excluded (N=814). In-house quality control filtering of the UK Biobank data is
487 described in a published protocol[40]. We restricted the sample to individuals of white
488 British ancestry who self-report as “White British” and who have very similar
489 ancestral backgrounds according to principal components analysis (N=409,703), as
490 described by Bycroft and colleagues[39]. Estimated kinship coefficients using the KING
491 toolset[41] identified 107,162 pairs of related individuals. An in-house algorithm was
492 then applied to this list and preferentially removed the individuals related to the
493 greatest number of other individuals until no related pairs remain. These individuals
494 were excluded (N=79,450).

495 **Polygenic scores**

496 We constructed 12 different PGS using Plink (version 2)[42] for each phenotype using
497 summary statistics from genome wide association studies (GWAS) of major
498 depressive disorder (MDD)[43], wellbeing[44] and schizophrenia[45]. The summary
499 statistics for the MDD and wellbeing GWAS excluded UK Biobank to avoid sample
500 overlap, and also excluded 23andMe participants. The PGS we constructed were
501 based on different p-value thresholds of the trait GWAS (0.5, 0.4, 0.3, 0.2, 0.1, 0.05,
502 0.01, $1x10^{-3}$, $1x10^{-4}$, $1x10^{-5}$, $1x10^{-6}$, $5x10^{-8}$). We used the PGS constructed at a p-
503 value threshold of 0.05 for our main analyses due to this explaining the largest
504 proportion of variance in depression[43] and schizophrenia[45] in the original GWAS
505 studies. This was not reported for wellbeing; therefore, we used a p-value threshold
506 of 0.05 for these analyses for consistency. We present results from other p-value
507 thresholds as sensitivity analyses to assess the consistency of the observed effects
508 across different thresholds. For the PGS constructed at a p-value threshold of 0.05,

509 the depression PGS explained 0.4% of the variance in having a diagnosis of
510 depression in the UK Biobank, the wellbeing PGS explained 0.3% of the variance in
511 wellbeing, and the schizophrenia PGS explained 2.7% of the variance in having a
512 diagnosis of schizophrenia. SNPs from the GWAS and UK Biobank were
513 harmonised, aligning the effect estimates and alleles. SNPs were clumped used the
514 European subsample of the 1000 genomes project, with $R^2 < 0.25$ and a window of
515 500 kb. The PGS were created by multiplying the number of effect alleles for each
516 participant in UK Biobank by the effect estimate of the SNP from the relevant GWAS,
517 then summing across all SNPs associated with each trait. PGS were z-standardised;
518 results should be interpreted as per standard deviation (SD) increase in score.

519 **Statistical analysis**

520 *Linear regression models*

521 Initial analyses were conducted in R version 3.5.1 and 3.6.2[46] using linear regression
522 models. We first conducted analyses to examine phenotypic associations in UK
523 Biobank where NDVI or percentage greenspace were the outcomes and the mental
524 health phenotypes were the exposures. We then conducted analyses where NDVI or
525 percentage greenspace were the outcomes and the PGS for MDD, wellbeing and
526 schizophrenia were the exposures in separate analyses. These global regression
527 models cannot meaningfully distinguish within- and between-area components of a
528 given relationship but are useful to demonstrate aggregate relationships across a
529 dataset. It also allows us to observe whether there are differences when accounting
530 for geographical context which may indicate contextually varying relationships.

531 *Multilevel models (MLM)*

532 Our interest for this study is in modelling the impact of the mental health PGS on
533 greenspace, but also in quantifying the degree to which the ostensibly individual
534 relationship can be distorted by higher level spatial confounding. As such, we fit
535 multilevel (or mixed or hierarchical) models (MLM) to efficiently estimate the within-
536 area effect of PGS exposure, whilst we account for between-area differences in
537 greenspace and PGS, retaining the capacity to quantify the impact of contextual
538 clustering that is sacrificed in traditional fixed effect approaches[47]. We present this as
539 a Variance Partitioning Coefficient (VPC), for the area level, which gives the
540 proportion of unexplained variation in a coefficient, that is explained by the area level
541 rather than solely being assumed a result of individual differences. In the two level
542 case, this is exactly equivalent to the intra-class correlation (ICC), giving the
543 expected correlation between randomly selected pairs of individuals within the same
544 higher level unit[48]. This ensures we do not misattribute between-area aggregate
545 variation in mental-health PGS exposures, to independent between-individual
546 variation. The MLM's are further described in the Supplementary Material (Section
547 1).

548 For MLMs we used the R package *lme4*[49]. We fitted two MLMs, adjusting for sex,
549 age and 25 PCs as fixed effects. The first model included a normally distributed
550 random effect around the modelled intercept, which varies at neighbourhood level.
551 By allowing each area to have its own modelled intercept, we account for the fact
552 that different geographical areas have different mean values for exposure and
553 outcome. This means our PGS effect estimate represents the effect of a unit
554 increase in PGS *within* a given neighbourhood. The second MLM extends the
555 formulation above but with random effects specified for district rather than
556 neighbourhood. We evaluated the model fit using Akaike's Information Criteria (AIC)

557 and compared across all three models, taking the best fitting model as the one with
558 lowest AIC. We also present the proportion of the total residual variance attributable
559 to variation between neighbourhoods or districts, rather than between-individual,
560 within-area variation, by calculating ICCs.

561 *One-sample Mendelian randomisation analyses*

562 Where we found evidence of an association between the PGS and the greenspace
563 outcomes, we ran one-sample MR analyses, with externally derived weights, to
564 assess whether these relationships may be causal. MR is a causal inference method
565 which uses genetic variants (typically SNPs), which are randomly assigned at
566 conception, as instrumental variables (IVs) for an exposure of interest. MR is more
567 robust to assumptions of confounding and reverse causation than conventional
568 observational analyses. There are three underlying core assumptions for MR, which
569 are i) that the genetic IV is robustly associated with the exposure (relevance), ii) the
570 genetic IV is only associated with the outcome via the exposure i.e., no direct effects
571 on the outcome (exclusion restriction), and iii) there is no confounding of the genetic
572 IV and outcome (independence).

573 In these analyses, for the genetic IV for depression, wellbeing and schizophrenia, we
574 used weighted PGS created using SNPs that reached genome-wide significance
575 ($p<5x10^{-8}$) in the GWAS including all samples. For schizophrenia we used the
576 corresponding weights from the GWAS including all samples, however for
577 depression and wellbeing we used weights from the GWAS excluding UK Biobank
578 and 23andMe samples[43–45]. For the genetic IV for wellbeing, we also used a
579 weighted PGS created at $p<1x10^{-05}$, due to the low number of SNPs (n=3) reaching
580 genome-wide significance.

We used a two-stage least squares regression approach, with the Applied Econometrics with R (AER) package (v1.2-9), to estimate the causal effects of depression liability, wellbeing and schizophrenia liability on the greenspace outcomes. Two-stage least squares regression allows the IV-exposure effect to be estimated in the first step and then in the second step the effect on the outcome is estimated by regressing the outcome on the estimates from step one. For depression, we used the variable for diagnosis of depression to estimate the genetic IV and exposure relationship in our main analyses, and we used the MHQ depression measure in sensitivity analyses. For schizophrenia we used the variable for diagnosis of schizophrenia to estimate the genetic IV and exposure relationship.

We also ran sensitivity analyses for all one-sample MR analyses using individual SNP data (as these were not possible to apply to the PGS), except for the genome-wide p-value threshold wellbeing analysis as there were too few SNPs. We used the collider-correction approach[50], which applies two-sample MR methods to one-sample MR data. This approach artificially induces collider bias between the exposure and outcome and then corrects for it, in order to facilitate simple adjustment for weak instrument bias and pleiotropy. We used collider-corrected inverse-variance weighted (IVW), Simulation Extrapolation (SIMEX) adjusted MR-Egger and least absolute deviation approaches. The IVW approach constrains the regression intercept to pass through zero and assumes there is no unbalanced pleiotropic effects of SNPs. The IVW estimate ignores uncertainty in the SNP-exposure estimate by assuming that its variance is negligible (known as the NO Measurement Error [NOME] assumption). MR Egger assesses and counteracts unbalanced pleiotropy by not constraining the intercept to pass through zero. SIMEX allows for adjustment due to regression dilution bias from weak instruments. Finally, IVW and

606 MR Egger rely on the InSIDE assumption (that the SNP-exposure association is not
607 correlated with a path from the SNP to the outcome, independent of the exposure)
608 and the least absolute deviation approach does not. The least absolute deviation
609 approach is also SIMEX adjusted. We also calculated the I-squared statistic
610 (indicative of strength of NOME violation for MR Egger) and the mean F-statistic
611 (indicative of instrument strength)[51]. Finally, we ran MR Robust Adjusted Profile
612 Score (RAPS) for the wellbeing MR using a p-value threshold of $p<1x10^{-05}$. As it
613 allows large numbers of weak instruments below the conventional GWAS threshold
614 ($5x10^{-08}$) to be included, accounting for weak instrument bias.

615 **Data availability**

616 UK Biobank data are available through a procedure described at
617 https://www.ukbiobank.ac.uk/enable-your-research.

618 **Code availability**

619 The analysis code that forms the basis of the results presented here is available from
620 https://github.com/ZoeReed/Greenspace_MH_PGS_analysis.

621 **Acknowledgements:** This research has been conducted using data from UK
622 Biobank, a major biomedical database (www.ukbiobank.ac.uk, application No
623 21829). We thank the participants of UK Biobank. Quality Control filtering of the UK
624 Biobank data was conducted by R.Mitchell, G.Hemani, T.Dudding, L.Corbin,
625 S.Harrison, L.Paternoster as described in the published protocol (doi:
626 10.5523/bris.1ovaau5sxunp2cv8rcy88688v).

627 **Funding:** This work was supported in part by the UK Medical Research Council
628 Integrative Epidemiology Unit at the University of Bristol (Grant ref:
629 MC_UU_00011/1, MC_UU_00011/7). GG was supported by an ESRC Postdoctoral
630 Fellowship (Grant ref: ES/T009101/1). MRM is supported by the National Institute for
631 Health Research Bristol Biomedical Research Centre. OSPD is funded by the Alan
632 Turing Institute under the EPSRC grant EP/N510129/1. MRM and OSPD were also
633 supported by the National Institute for Health Research (NIHR) Biomedical Research
634 Centre at the University Hospitals Bristol NHS Foundation Trust and the University of
635 Bristol.

636 **Conflicts of interest:** There were no conflicts of interest

637 **Author contributions:** Conceptualization: ZER, MRM, GJG; Methodology: ZER,
638 GJG, TTM; Formal Analysis: ZER; Resources: MRM, OSPD, GDS; Data Curation:
639 ZER; Writing—Original Draft: ZER, GJG; Writing—Review and Editing: ZER, GJG,
640 TTM, OSPD, GDS, MRM; Supervision: MRM; Project Administration: MRM; Funding
641 Acquisition: MRM, GDS, OSPD.

642 **References**

643 1. Gong, Y., Palmer, S., Gallacher, J., Marsden, T. & Fone, D. A systematic review of the
644 relationship between objective measurements of the urban environment and
645 psychological distress. *Environment International* **96**, 48–57 (2016).
646 2. Zhang, Y., Mavoa, S., Zhao, J., Raphael, D. & Smith, M. The association between green
647 space and adolescents mental well-being: A systematic review. *International Journal*
648 *of Environmental Research and Public Health* **17**, 1–26 (2020).
649 3. Feng, X. & Astell-Burt, T. The relationship between neighbourhood green space and
650 child mental wellbeing depends upon whom you ask: Multilevel evidence from 3083
651 children aged 12–13 years. *International Journal of Environmental Research and*
652 *Public Health* vol. 14 235 (2017).
653 4. White, M. P., Alcock, I., Wheeler, B. W. & Depledge, M. H. Would You Be Happier
654 Living in a Greener Urban Area? A Fixed-Effects Analysis of Panel Data. *Psychological*
655 *Science* **24**, 920–928 (2013).
656 5. Bezold, C. P. *et al.* The Association Between Natural Environments and Depressive
657 Symptoms in Adolescents Living in the United States. *Journal of Adolescent Health* **62**,
658 488–495 (2018).
659 6. Herrera, R. *et al.* Greenness and job-related chronic stress in young adults: A
660 prospective cohort study in Germany. *BMJ Open* **8**, e021599 (2018).
661 7. Wang, P., Meng, Y. Y., Lam, V. & Ponce, N. Green space and serious psychological
662 distress among adults and teens: A population-based study in California. *Health and*
663 *Place* **56**, 184–190 (2019).
664 8. Sarkar, C., Webster, C. & Gallacher, J. Residential greenness and prevalence of major
665 depressive disorders: a cross-sectional, observational, associational study of 94 879
666 adult UK Biobank participants. *The Lancet Planetary Health* **2**, e162–e173 (2018).
667 9. Bezold, C. P. *et al.* The relationship between surrounding greenness in childhood and
668 adolescence and depressive symptoms in adolescence and early adulthood. *Annals of*
669 *Epidemiology* **28**, 213–219 (2018).
670 10. Li, D., Deal, B., Zhou, X., Slavenas, M. & Sullivan, W. C. Moving beyond the
671 neighborhood: Daily exposure to nature and adolescents' mood. *Landscape and*
672 *Urban Planning* **173**, 33–43 (2018).
673 11. Ward, J. S., Duncan, J. S., Jarden, A. & Stewart, T. The impact of children's exposure to
674 greenspace on physical activity, cognitive development, emotional wellbeing, and
675 ability to appraise risk. *Health and Place* **40**, 44–50 (2016).
676 12. Rhew, I. C., Vander Stoep, A., Kearney, A., Smith, N. L. & Dunbar, M. D. Validation of
677 the Normalized Difference Vegetation Index as a Measure of Neighborhood
678 Greenness. *Annals of Epidemiology* **21**, 946–952 (2011).
679 13. Department for Communities and Local Government. Generalised Land Use Database
680 Statistics for England 2005 (Enhanced Basemap). (2007).
681 14. Faris, R. E. L. & Dunham, H. W. *Mental disorders in urban areas: an ecological study of*
682 *schizophrenia and other psychoses. Mental disorders in urban areas: an ecological*
683 *study of schizophrenia and other psychoses.* (Univ. Chicago Press, 1939).
684 15. Purtle, J. *et al.* Urban–Rural Differences in Older Adult Depression: A Systematic
685 Review and Meta-analysis of Comparative Studies. *American Journal of Preventive*
686 *Medicine* **56**, 603–613 (2019).

16. Solmi, F., Dykxhoorn, J. & Kirkbride, J. B. Urban-Rural Differences in Major Mental Health Conditions. in 27–132 (Springer, Singapore, 2017). doi:10.1007/978-981-10-2327-9_7.
17. Vassos, E., Agerbo, E., Mors, O. & Bøcker Pedersen, C. Urban-rural differences in incidence rates of psychiatric disorders in Denmark. *British Journal of Psychiatry* **208**, 435–440 (2016).
18. Peen, J., Schoevers, R. A., Beekman, A. T. & Dekker, J. The current status of urban-rural differences in psychiatric disorders. *Acta Psychiatrica Scandinavica* **121**, 84–93 (2010).
19. Pedersen, C. B. Persons with schizophrenia migrate towards urban areas due to the development of their disorder or its prodromata. *Schizophrenia Research* **168**, 204–208 (2015).
20. Vassos, E., Agerbo, E., Mors, O. & Bøcker Pedersen, C. Urban-rural differences in incidence rates of psychiatric disorders in Denmark. *British Journal of Psychiatry* **208**, 435–440 (2016).
21. Verheij, R. A., Maas, J. & Groenewegen, P. P. Urban-rural health differences and the availability of green space. *European Urban and Regional Studies* **15**, 307–316 (2008).
22. Sariaslan, A. *et al.* Schizophrenia and subsequent neighborhood deprivation: Revisiting the social drift hypothesis using population, twin and molecular genetic data. *Translational Psychiatry* **6**, e796–e796 (2016).
23. Giggs, J. A. The Distribution of Schizophrenics in Nottingham. *Transactions of the Institute of British Geographers* **59**, 55–76 (1973).
24. Giggs, J. A. The Distribution of Schizophrenics in Nottingham: A Reply. *Transactions of the Institute of British Geographers* **64**, 150–156 (1975).
25. March, D. *et al.* Psychosis and place. *Epidemiologic Reviews* vol. 30 84–100 (2008).
26. Colodro-Conde, L. *et al.* Association between population density and genetic risk for schizophrenia. *JAMA Psychiatry* **75**, 901–910 (2018).
27. Moore, T. H. M. *et al.* The effects of changes to the built environment on the mental health and well-being of adults: Systematic review. *Health and Place* vol. 53 237–257 (2018).
28. Davey Smith, G. & Ebrahim, S. “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* **32**, 1–22 (2003).
29. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics* **23**, R89-98 (2014).
30. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* **27**, 1133–1163 (2008).
31. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Medicine* **12**, e1001779 (2015).
32. Cusack, L., Larkin, A., Carozza, S. E. & Hystad, P. Associations between multiple green space measures and birth weight across two US cities. *Health and Place* **47**, 36–43 (2017).
33. Sarkar, C., Webster, C. & Gallacher, J. UK Biobank Urban Morphometric Platform (UKBUMP) – a nationwide resource for evidence-based healthy city planning and public health interventions. *Annals of GIS* **21**, 135–148 (2015).

34. Smith, D. J. *et al.* Prevalence and characteristics of probable major depression and bipolar disorder within UK Biobank: Cross-sectional study of 172,751 participants. *PLoS ONE* **8**, e75362 (2013).
35. McManus, S. *et al. Mental health and wellbeing in England : Adult Psychiatric Morbidity Survey 2014*. (2016).
36. Tyrrell, J. *et al.* Genetic predictors of participation in optional components of UK Biobank. *Nature Communications* **12**, 886 (2021).
37. Sarmanova, A., Morris, T. & Lawson, D. J. Population stratification in GWAS meta-analysis should be standardized to the best available reference datasets. *bioRxiv* (2020) doi:10.1101/2020.09.03.281568.
38. Office for National Statistics. 2011 Census: Aggregate Data. (2020) doi:http://doi.org/10.5257/census/aggregate-2011-2.
39. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
40. Mitchell, R. *et al.* UK Biobank Genetic Data: MRC-IEU Quality Control, Version 2. (2019) doi:https://doi.org/10.5523/bris.1ovaau5sxunp2cv8rcy88688v.
41. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
42. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
43. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668–681 (2018).
44. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics* **48**, 624–633 (2016).
45. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
46. R Core Team. R: A language and environment for statistical computing. (2016).
47. Bell, A. & Jones, K. Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods* **3**, 133–153 (2015).
48. Merlo, J., Chaix, B., Yang, M., Lynch, J. & Råstam, L. A brief conceptual tutorial of multilevel analysis in social epidemiology: Linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of Epidemiology and Community Health* **59**, 443–449 (2005).
49. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, (2015).
50. Barry, C. *et al.* Exploiting collider bias to apply two-sample summary data Mendelian randomization methods to one-sample individual level data. *PLOS Genetics* **17**, e1009703 (2021).
51. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-Egger regression: The role of the I 2 statistic. *International Journal of Epidemiology* **45**, 1961–1974 (2016).
52. Hernán, M. A., Clayton, D. & Keiding, N. The simpson's paradox unraveled. *International Journal of Epidemiology* **40**, 780–785 (2011).

53. Julious, S. A. & Mullee, M. A. Confounding and Simpson's paradox. *BMJ (Clinical research ed.)* **309**, 1480–1 (1994).
54. Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J. & Borsboom, D. Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology* **4**, 513 (2013).
55. Weeland, J., Laceulle, O. M., Nederhof, E., Overbeek, G. & Reijneveld, S. A. The greener the better? Does neighborhood greenness buffer the effects of stressful life events on externalizing behavior in late adolescence? *Health and Place* **58**, 102163 (2019).
56. Mueller, M. A. E., Flouri, E. & Kokosi, T. The role of the physical environment in adolescent mental health. *Health and Place* **58**, 102153 (2019).
57. Richmond, R. C. & Davey Smith, G. Mendelian Randomization: Concepts and Scope. *Cold Spring Harbor Perspectives in Medicine* a040501 (2021) doi:10.1101/cshperspect.a040501.
58. Labib, S. M., Lindley, S. & Huck, J. J. Estimating multiple greenspace exposure types and their associations with neighbourhood premature mortality: A socioecological study. *Science of the Total Environment* **789**, 147919 (2021).
59. Abdellaoui, A., Verweij, K. J. H. & Nivard, M. G. Geographic Confounding in Genome-Wide Association Studies. *bioRxiv* 2021.03.18.435971 (2021) doi:10.1101/2021.03.18.435971.
60. Haworth, S. *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nature Communications* **10**, 333 (2019).
61. Morris, T. T., Davies, N. M., Hemani, G. & Smith, G. D. Population phenomena inflate genetic associations of complex social traits. *Science Advances* **6**, eaay0328 (2020).