

Stratifying depression by schizophrenia risk

Word count
Abstract 250
Main text: 3771

Figures 2
Tables 3

**Dissection of Major Depressive Disorder using polygenic risk scores for
Schizophrenia in two independent cohort.**

¹Whalley HC*, ¹Adams MJ, ¹ Hall LS, ¹ Clarke T-K, ¹Fernandez-Pujals AM,
¹Gibson J, ¹Wigmore E, ¹Jonathan Hafferty, ²Hagenaars SP, ^{2,3}Davies G,
⁴Campbell A, ⁵Hayward C, ¹Lawrie SM, ⁴Porteous DJ, ^{2,3}Deary IJ, ¹McIntosh
AM.

¹Division of Psychiatry, University of Edinburgh, Edinburgh, UK.

²Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh,
Edinburgh, UK.

³Department of Psychology, University of Edinburgh, Edinburgh, UK

⁴Centre for Genetics and Molecular Medicine, IGMM, Western General Hospital, University of
Edinburgh, Edinburgh, UK.

⁵MRC Human Genetics Unit, IGMM, University of Edinburgh

* Correspondence: Dr Heather C Whalley

University of Edinburgh

Royal Edinburgh Hospital

Morningside Park

Edinburgh EH10 5HF

UK

Telephone: +44 131 537 6502

Email: heather.whalley@ed.ac.uk

Stratifying depression by schizophrenia risk

Abstract

Major depressive disorder (MDD) is known for its substantial clinical and suspected causal heterogeneity. It is characterised by low mood, psychomotor slowing, and increased levels of the personality trait neuroticism; factors which are also associated with schizophrenia (SCZ). It is possible that some cases of MDD may have a substantial genetic loading for SCZ. A sign of the presence of SCZ-like MDD sub-groups would be indicated by an interaction between MDD status and polygenic risk of SCZ on cognitive, personality and mood measures. In the current study, we hypothesised that higher SCZ-polygenic risk would define larger MDD case-control differences in cognitive ability, and smaller differences in distress and neuroticism. Polygenic risk scores (PGRS) for SCZ and their association with cognitive variables, neuroticism, mood, and psychological distress were estimated in a large population-based cohort (Generation Scotland: Scottish Family Health Study, GS:SFHS). Individuals were divided into those with, and without, depression ($n=2587$ & $n=16,764$ respectively) to test whether there was an interaction between MDD status and schizophrenia risk. Replication was sought in UK Biobank ($n=33,525$). In both GS:SFHS and UK Biobank we found significant interactions between SCZ-PGRS and MDD status for measures of psychological distress and neuroticism. In both cohorts there was a reduction of case-control differences on a background of higher genetic risk of SCZ. These findings suggest that depression on a background of high genetic risk for SCZ may show attenuated associations with distress and neuroticism. This may represent a causally distinct form of MDD more closely related to SCZ.

Stratifying depression by schizophrenia risk

Introduction

Major depressive disorder (MDD) has a lifetime prevalence of approximately 16% and is a heritable and severely disabling psychiatric condition (1). It is highly clinically heterogeneous, relying on diagnostic criteria that are fulfilled through the presence of an arbitrary threshold of symptoms that lack empirical validation. Further, individuals passing this threshold can have a very different combination of symptoms. As a likely consequence of causal heterogeneity, progress in our understanding of depression's underlying risk factors and biological mechanisms has been limited.

MDD is characterised by sustained negative affect, psychological distress and by differences in personality and behavioural traits, such as increased levels of the personality trait neuroticism (2). People with MDD also manifest, on average, reduced executive function, attention, processing speed and working memory (3), but relatively preserved general cognitive ability (4). These behavioural and personality features are partly heritable, enduring features of illness that are often genetically correlated with liability to MDD (5-8). Notably however, these clinical and personality associations are not specific to MDD and also occur in other psychiatric disorders including schizophrenia (SCZ) (9-11).

Genetic factors account for ~37% of liability to MDD and depressive symptoms (12, 13). Genome-wide association studies (GWAS) have indicated that ~30% of the heritability for MDD is accounted for by an additive polygenic genetic architecture, where risk is conferred by an accumulation of many alleles of small effect (14). There is also a genetic overlap with other major psychiatric disorders, including SCZ (15). For example, two papers have

Stratifying depression by schizophrenia risk

shown genetic correlation between MDD and SCZ using both genome-wide association analysis identifying shared risk loci (15), or an approach estimating genetic correlation from sampled SNPs (16). These findings have been interpreted as showing that there are genetic variants that raise liability to both MDD and SCZ. Recently, however, a second possibility has been proposed – that the genetic correlation between SCZ and MDD is caused not by pleiotropy, but by the misclassification of SCZ cases as MDD. This possibility has recently been tested using a novel method aimed at using genotype data to distinguish pleiotropy from heterogeneity, where early evidence suggested that some cases of SCZ may have been misclassified as MDD, though this did not reach significance in this proof of technique study (17). If confirmed, this would imply that some cases of MDD would be a hidden ‘forme fruste’ of SCZ, or that some individuals with MDD may be on a trajectory towards SCZ, but have yet to manifest the clinical phenotype.

Effects of polygenic risk for SCZ have also been demonstrated on a number of related behavioural and cognitive phenotypes. It has been consistently demonstrated for example that polygenic risk for schizophrenia is associated with decreased cognitive ability in healthy control samples (18-21). However, since most studies have examined these relationships in unaffected controls, it is unclear whether the associations between SCZ risk and behaviour/cognitive ability are ameliorated or magnified by the presence of depression. Clarifying this relationship is important because, if MDD consists of a subtype which is closely related to schizophrenia, then that subtype would be expected to show associations with behaviour and cognitive ability that resembled those of SCZ itself.

Stratifying depression by schizophrenia risk

In the current paper we sought to stratify depression using a continuous scale of SCZ polygenic risk score, hypothesising that case-control differences would depend on genetic risk for schizophrenia, in other words, that there would be a significant interaction between SCZ-PGRS and MDD status. We made two predictions, (1) based on the premise that deficits in cognition are greater in SCZ than in MDD, we first predicted that the high-SCZ-PGRS MDD case-control differences for cognitive variables would be greater compared to those with lower SCZ-PGRS scores. (2) Since neuroticism and psychological distress differ to a greater extent in MDD than SCZ (22) we also predicted that SCZ-PGRS and MDD status interactions would be found for these variables, but that they would be the reverse direction to cognition (attenuated in those with high SCZ-PGRS). We sought to test these predictions in a population based cohort study, the Generation Scotland cohort: The Scottish Family Health Study sample (GH:SFHS) (23, 24). Replication of any significant findings was sought in UK Biobank.

Stratifying depression by schizophrenia risk

Materials and Methods

Main Study Participants

Generation Scotland: the Scottish Family Health Study (GS:SFHS)

GS:SFHS is a family- and population-based cohort study. Individuals were recruited at random through general medical practices across Scotland. Initial data collection for GS:SFHS took place between February 2006 and March 2011. The complete study protocol has been described in detail elsewhere (23-25). Details of the mental health and related assessments are summarised below. Ethical approval was provided by NHS Tayside Research Ethics Committee (reference 05/S1401/89). Written consent for the use of data was obtained from all participants.

The full cohort consists of 23,690 individuals who were over 18 years at the time of recruitment and 21,516 of these attended the research clinic. The present study includes 19,351 individuals on whom genome-wide genotype data were available (demographic details presented in Table 1). Pedigree information was available for all participants, and this has subsequently been validated against estimates of relatedness estimated using genome-wide single nucleotide polymorphism (SNP) data on 19,995 individuals.

Assessment of MDD in GS:SFHS

MDD was diagnosed using the structured clinical interview for the Diagnostic and Statistical Manual of Mental Disorders (SCID) (26). A brief questionnaire was administered to participants to screen for MDD. Participants were asked “Have you ever seen anybody for emotional or psychiatric problems?” and “Was there ever a time when you, or someone else, thought you should see

Stratifying depression by schizophrenia risk

someone because of the way you were feeling or acting?” If they answered yes to either of these questions (21.7% screened positive), they were asked to complete the SCID (26). If they answered no to both of these questions, they were assigned control status. Those who completed the SCID but did not meet the criteria for MDD or other major psychiatric disorder were also defined as controls. Individuals who declined to complete the screening questionnaire or the SCID had their MDD status set to missing. Individuals with a diagnosis of bipolar disorder were removed for the purposes of the current investigation.

Cognitive assessment in GS:SFHS

Cognitive abilities were assessed using four tests. Immediate and delayed scores from the recall section of one story of the Wechsler Logical Memory III ^{UK} test were summed to provide a measure of verbal declarative memory (27). The Wechsler Digit Symbol Coding test was used to measure processing speed (27). Verbal ability was assessed using the Mill Hill Vocabulary Scale, junior and senior synonyms (28). Executive function was measured using the letter-based phonemic verbal fluency test (letters C, F and L, for one minute each) (29). All test scores were scaled to a mean of zero and standard deviation of one. General cognitive ability (g) scores were obtained by conducting principal component analysis of the tests, and saving the scores on the first un-rotated principal component, on which all tests had substantial loadings (30)

Measures of neuroticism and psychological distress GS:SFHS

Stratifying depression by schizophrenia risk

Measures of psychological distress were assessed using the 28-item General Health Questionnaire (GHQ-28) (31), which consists of 4 subscales designed to assess: (A) somatic symptoms, (B) anxiety and insomnia, (C) social dysfunction and (D) 'severe depression'. A single measure of global affective symptoms was derived from item responses and the standard Likert scoring totals were used in the current study. The Eysenck Personality Questionnaire-Revised Short Form provided a measure of neuroticism (32). For numbers of individuals completing each measure see Table 2.

Genotyping and Polygenic Profiling in GS:SFHS

Blood samples were collected using standard operating procedures and stored at the Wellcome Trust Clinical Research Facility Genetics Core, Edinburgh (www.wtcrf.ed.ac.uk) where they were genotyped using the IlluminaHumanOmniExpressExome-8v1.0 BeadChip and Infinium chemistry (33). The genotypes were then processed using the IlluminaGenomeStudio Analysis software v2011.1. The details of blood collection and DNA extraction are provided elsewhere (34).

Polygenic risk scores were calculated using the method described previously (35) using summary data from the most recent GWAS studies from the Psychiatric Genomics Consortium (PGC, 2014). Briefly, SNPs were excluded if they had a minor allele frequency (MAF) < 1%, deviated significantly from Hardy-Weinberg equilibrium ($p < 1 \times 10^{-6}$) in the total sample of founder individuals, or had a call rate < 99%. Clump based linkage disequilibrium pruning (r^2 0.2, 300kb window) was performed to create a SNP-set in linkage equilibrium. Prior to creating polygenic risk scores all strand-

Stratifying depression by schizophrenia risk

ambiguous SNPs were removed from the GS:SFHS dataset and the polygenic risk score for each individual was calculated using the cumulative sum of the number of each SNP allele multiplied by the log of the odds ratio for SCZ across their whole genome. Five SNP set scores were generated using p-value threshold cut-offs of 1, 0.5, 0.1, 0.05 and 0.01. Our primary analyses concerned those SNPs from the PGC data that met a significance level of $p=5 \times 10^{-8}$ consistent with previous studies (35). Findings for other thresholds are contained within Supplementary Material.

Replication sample: UK Biobank

This study includes replication data from the UK Biobank Study (<http://www.ukbiobank.ac.uk>). UK Biobank received ethical approval from the Research Ethics Committee (REC, reference: 11/NW/0382). UK Biobank is a health research resource that aims to improve the prevention, diagnosis, and treatment of a range of illnesses. Through UK Biobank, 502 655 community-dwelling participants were recruited between 2006 and 2010 in the United Kingdom (36). They underwent cognitive, neuroticism and physical assessments, provided blood, urine, and saliva samples for future analysis, gave detailed information about their backgrounds and lifestyles, and agreed to have their health followed longitudinally. For the present study, genome-wide genotyping data were available on 33,525 individuals. This dataset comprised individuals who survived the quality control process, were independent of individuals within GS:SFHS, who were unrelated, and who did not meet criteria for another major psychiatric disorder. Individuals were

Stratifying depression by schizophrenia risk

assigned to a narrow definition of MDD including cases of moderate or recurrent MDD (see Supplementary Material for further information).

Cognitive assessment in UK Biobank

Three cognitive tests were used in the present study to generate measures of general intelligence 'g'. These tests, which cover the cognitive domains of Reaction Time, Memory and Verbal-numerical Reasoning were collected at the baseline assessment. In addition, data from a second assessment which provided measures of processing speed (similar to the Digit Symbol Coding test, the Symbol Digit Substitution test) were also assessed (further details in Supplementary Material). All test scores were scaled to a mean of zero and standard deviation of one. General cognitive ability (g) was estimated using the principal component analysis approach described above (30).

Measures of neuroticism and psychological distress in UK Biobank

In UK Biobank, participants completed the Neuroticism scale of the Eysenck Personality Questionnaire-Revised Short Form (EPQ-R Short Form) (37). As a measure of psychological distress UK Biobank participants undertook the Patient Health Questionnaire (PHQ) which is a multipurpose instrument for screening, diagnosing, monitoring and measuring the severity of depression. As in GS:SFHS scores were scaled to a mean of zero and standard deviation of one.

Stratifying depression by schizophrenia risk

Genotyping and Polygenic Profiling in UK Biobank

Details of the array design, genotyping, quality control and imputation have been described previously (38). Quality control included removal of participants based on missingness, relatedness, gender mismatch, and non-British ancestry. Polygenic profiles were created for SCZ in all genotyped participants using PRSice (39). PRSice calculates the sum of alleles associated with the phenotype of interest across many genetic loci, weighted by their effect sizes estimated from a GWAS of that phenotype in an independent sample. Prior to creating the scores, SNPs with a minor allele frequency < 1% were removed and clumping was used to obtain SNPs in linkage equilibrium with an $r^2 < 0.25$ within a 200bp window. Multiple scores were then created for each phenotype containing SNPs selected according to the significance of their association with the phenotype. The GWAS summary data was used to create polygenic profiles for SCZ in the UK Biobank participants, at thresholds of $p < 0.01$, $p < 0.05$, $p < 0.1$, $p < 0.5$ and all SNPs. As in the main GS:SFHS analysis, current results focus on $P < 0.5$.

Statistical Analysis

All analyses were conducted in ASReml-R (www.vsni.co.uk/software/asrem/, version 3.0). Associations were examined between PGRSs for SCZ and cognition (cognitive factor 'g', and contributing tests), psychological distress (GHQ total), and neuroticism (from the Eysenck personality questionnaire) using mixed linear model association analysis in GS:SFHS. Age, age², sex, four multidimensional scaling ancestry components and polygenic profile scores were entered as fixed effects. Since GS:SFHS is a family-based study,

Stratifying depression by schizophrenia risk

to control for relatedness family structure was fitted as a random effect by creating an inverse relationship matrix using pedigree kinship information. SCZ-PGRS and MDD status were entered as main effects in the model and PGRS * MDD status effects were examined by including the interaction term. Where significant interactions were found, further tests were conducted in controls and MDD groups separately. Wald's conditional F-test was used to calculate the significance of fixed effects. P-values presented are raw p-values uncorrected for multiple testing, results were considered significant if they were reported in both cohorts ($p < 0.05$). Reported beta values are standardised values. The proportion of phenotypic variance explained by polygenic risk score was calculated by multiplying the profile score by its corresponding regression coefficient and estimating its variance. This value was then divided by the variance of the observed phenotype to yield a coefficient of determination between 0 and 1 and converted into a percentage (40).

Significant association and interaction effects were also explored within the independent sample UK Biobank where equivalent measures were available. In UK Biobank polygenic risk profile scores were examined for their association with observed phenotypes and for group interaction effects in ASReml-R using the same methods as above (details in supplementary material), but without the inclusion of a genetic relationship matrix due to the large data set and unrelated nature of the filtered UKB study population used in the current investigation.

Stratifying depression by schizophrenia risk

Results

Demographic, cognitive, trait-related features of MDD and PGRS scores in GS:SFHS.

Demographic details are presented in Table 1 for GS:SFHS. Thresholds for case-control group significance were defined as $p < 2 \times 10^{-5}$. Groups defined according to depression status were significantly different in terms of sex, psychological distress (GHQ) and for neuroticism (Table 1). The depression group scored higher on the latter two measures. Differences between the groups did not reach significance for the individual cognitive tests. MDD cases and controls also differed significantly on the SCZ PGRS score, with MDD individuals scoring higher.

Demographic details of individuals from UK Biobank included in the current study are presented in Supplementary Table 1. There were significant differences between the groups in terms of age and gender ($p < 2 \times 10^{-5}$). The groups also differed according to psychological distress and neuroticism, with the MDD cases scoring higher than controls. Differences between the groups did not reach significance for the individual cognitive tests and MDD cases and controls differed significantly on the SCZ PGRS score, with MDD individuals scoring higher.

Interaction between MDD status and polygenic risk for schizophrenia, in GS:SFHS and UK Biobank.

Results for tests of interaction are presented in Table 2 and Figures 1a,b. Significant interaction effects seen across both cohorts were PGRS SCZ * group interactions for measures of psychological distress ($\beta = -0.0402$,

Stratifying depression by schizophrenia risk

$p=1.44 \times 10^{-2}$, $\beta=-0.0903$, $p=8.03 \times 10^{-4}$, GS:SFHS and UK Biobank) and for neuroticism ($\beta=-0.0483$, $p=1.96 \times 10^{-3}$, $\beta=-0.0602$, $p=2.28 \times 10^{-2}$, GS:SFHS and UK Biobank). A significant interaction was also reported for measures of processing speed in GS:SFHS ($\beta=-0.0313$, $p=4.31 \times 10^{-2}$), but this was not replicated in the UK Biobank sample.

In terms of the direction of associations within diagnostic groups, in both GS:SFHS and UK Biobank samples there was a significant positive association in the control group for measures of psychological distress ($\beta=0.0400$, $p=1.46 \times 10^{-8}$, $\beta=0.0827$, $p=3.25 \times 10^{-14}$, in GS:SFHS and UK Biobank) and for neuroticism ($\beta=0.0369$, $p=9.15 \times 10^{-6}$, $\beta=0.0539$, $p=4.79 \times 10^{-19}$, GS:SFHS and UK Biobank) (see Table 3). There were no significant associations within the MDD groups for either measure, in either cohort (see Table 3, Figures 1 a,b). In terms of psychological distress the proportion of variance explained was 0.15% and 0.22% for control individuals, for GS:SFHS and UK Biobank respectively, and 0.02%, 0.01% for MDD cases. For measures of neuroticism values were 0.17% and 0.30% in controls and <0.01% in MDD cases (see Tables 3 and 4 and Figure 2). For the measure of processing speed in GS:SFHS there was a significant negative relationship in both controls and cases, but the MDD cases demonstrated greater effect size and greater variance explained (1.09%) versus that seen in control individuals (0.33%) (see Table 3). For both distress and neuroticism the SCZ-PGRS * MDD status interactions remained significant after modelling whether individuals were experiencing a current depressive episode at the time of assessment according to the SCID, and when modelling whether individuals

Stratifying depression by schizophrenia risk

were taking antidepressant medication at the time of assessment according to self-report, full details in supplementary material.

Stratifying depression by schizophrenia risk

Discussion

Consistent with our predictions we report larger MDD case-control differences in cognition in the context of higher SCZ-PGRS scores in GS:SFHS, although this was not replicated in UKB. We also found MDD case-control differences in neuroticism and distress that were attenuated by higher SCZ-PGRS scores. Lastly, we demonstrate that the greatest variance explained for neuroticism and distress is higher in controls rather than in MDD cases, supporting the suggestion that the MDD group contains a more heterogeneous group comprising differing aetiological subtypes.

PGRS SCZ associations and interactions with cognition

Previous studies have consistently reported a negative relationship between cognitive function and SCZ-PGRS, indicating shared genetic risk between SCZ and deficits in cognition (18-20, 41). With the exception of one previous study where genetic loading for SCZ (based on proximal family history) was reported to have a substantial negative impact on neurocognition in mood disorder patients (42), we are not aware of any other studies examining this relationship in the context of MDD. In the current study, data from the four individual subtests that contributed to the general factor indicated a significant interaction between SCZ-PGRS and MDD case control status for the digit symbol coding test. This test is a measure of processing speed, which is known to be significantly impaired in SCZ patients, even in relation to other neuropsychological measures (43, 44). Within GS:SFHS, significant negative associations between SCZ-PGRS and processing speed were reported in both controls and MDD cases. In other words, individuals in GS:SFHS with

Stratifying depression by schizophrenia risk

MDD demonstrated greater deficits in tests of processing speed in those who had higher genetic loading for SCZ versus those at low risk. However, the lack of formal replication means these results should be viewed with caution.

PGRS SCZ associations and interactions with clinical and trait features

Neuroticism is considered to be a stable heritable trait (45) characterized by high tension, irritability, dissatisfaction, shyness, low mood and reduced self-confidence (32). Psychological distress is a more generalised measure of psychiatric well-being or psychological health than a specific psychopathological categorisation. These two traits have been reported to share strong genetic links (7), and there has been considerable support for overlapping genetic risk factors affecting neuroticism and MDD (46). Direct links between genetic risk for schizophrenia and these measures however are less commonly reported than associations with cognitive deficits, as described above.

These measures demonstrated significant interaction effects between SCZ-PGRS and MDD case-control status, seen in both GS:SFHS and UK Biobank cohorts. In addition, the variance explained for these measures was greater in control individuals (~0.20%) than in MDD cases (~0.01%). We suggest that this may be attributable to the aetiologically diverse subgroups within MDD cases. The origins of these interactions were an attenuation of case-control differences on a background of higher genetic risk of SCZ, seen in both cohorts. Neuroticism and psychological distress may therefore be less closely related to the aetiology of depression in the context of high genetic risk for schizophrenia (see Fig 1). It is, however, also possible that the absence of

Stratifying depression by schizophrenia risk

a similar gradient of effect in MDD cases could be due to the substantial effects of depression on distress and neuroticism, leading to a ceiling effect whereby the additional effect of SCZ-PGRS on these phenotypes is undetectable. However, we consider that this is unlikely since the variance of these measures are *not* in fact smaller in the MDD groups in comparison to the control group (which would be the case if there were ceiling effects, see Table 1), and so truncation of these scores at the extreme end seems unlikely. Together, these findings suggest that depression on a background of high genetic risk for schizophrenia may represent a somewhat causally distinct form of the condition.

Although this is a large and arguably well powered study with independent replication, there are important limitations that should be considered. Of note is that the data used to derive polygenic scores do not presently fully account for all heritability attributable to common variation, nor do these measures reflect the contribution of rare variation and copy-number variants. In addition, the degree of variance explained by the SCZ-PGRS score for these phenotypes is relatively modest (0.5-1%). It should be considered however that they are consistent with other previously reported figures of 0-2% for similar phenotypes and in similarly unselected population-based cohorts, and that the higher figures reported in the literature (7-18%) generally relate to the phenotype of SCZ case-control status (see (20)).

Another limitation relates to consistency of testing between these large datasets. Particularly relevant are the differences in administration of the tests of processing speed. Within GS:SFHS, a pen and paper version of the Digit Symbol Coding task from the Wechsler Adult Intelligence Scale III was used,

Stratifying depression by schizophrenia risk

where participants had a two minute time limit to complete the task. In UK Biobank the format of the task was a computerised version where the participant had to click on the number which matched the symbol shown and the participant had one minute to complete the test. Importantly, there were also fewer individuals included in the DSC assessments in UK Biobank (n=7799 controls and n=1767 MDD individuals) than for other cognitive measures, and in relation to GS:SFHS (n=16,566 controls and n=2566 MDD individuals) since this test was conducted at a follow-up assessment rather than at baseline.

In summary, these findings are consistent with a model in which genetic risk for schizophrenia predicts depressive traits in the general population, but that neuroticism and psychological distress may be less closely related to the aetiology of depression on a background of high genetic risk for schizophrenia. This may represent a somewhat causally distinct form of MDD more closely related to SCZ. The study of the genetic basis of variation in such measures is likely to further the understanding of mechanisms by which SCZ genes affect neural function in the context of health and depressive illness.

Stratifying depression by schizophrenia risk

Acknowledgements

This study is supported by a Wellcome Trust Strategic Award “Stratifying Resilience and Depression Longitudinally” (STRADL) (Reference 104036/Z/14/Z). Generation Scotland (GS:SFHS) received core support from the Chief Scientist Office (CSO) of the Scottish Government Health Directorates [grant number CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the UK’s Medical Research Council and the Wellcome Trust. Ethics approval for the study was given by the NHS Tayside committee on research ethics (reference 05/S1401/89). We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. This research has also been conducted using the UK Biobank Resource.

HCW is supported by a JMAS SIM fellowship from the Royal College of Physicians of Edinburgh and by an ESAT College Fellowship from the University of Edinburgh. Part of the work was undertaken in The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE), part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1); funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. Age UK (The Disconnected Mind project) also provided support for the work undertaken at CCACE.

Stratifying depression by schizophrenia risk

Conflicts of Interest

AMM has previously received grant support from Pfizer, Lilly and Janssen. These studies are not connected to the current investigation. SML has previously received research grant support from Pfizer, Roche, Abbvie and Sunovion, as well as personal fees from Roche, Sunovion and Janssen, outside the present work. Remaining authors report no conflicts of interest.

References

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003 Jun 18; **289**(23): 3095-3105.
2. Kennedy SH. Core symptoms of major depressive disorder: relevance to diagnosis and treatment. *Dialogues Clin Neurosci* 2008; **10**(3): 271-277.
3. Austin MP, Mitchell P, Goodwin GM. Cognitive deficits in depression: possible implications for functional neuropathology. *Br J Psychiatry* 2001 Mar; **178**: 200-206.
4. Navrady LB, Ritchie SJ, Chan SWY, Kerr DM, Adams MJ, Hawkins E *et al.* Intelligence and neuroticism in relation to depression and psychological distress: evidence of interaction using data from Generation Scotland: Scottish Family Health Study and UK Biobank. *bioRxiv doi: 101101/039545* 2016.
5. Glahn DC, Curran JE, Winkler AM, Carless MA, Kent JW, Jr., Charlesworth JC *et al.* High dimensional endophenotype ranking in the search for major depression risk genes. *Biol Psychiatry* 2012 Jan 1; **71**(1): 6-14.

Stratifying depression by schizophrenia risk

6. Kendler KS, Gatz M, Gardner CO, Pedersen NL. Personality and major depression: a Swedish longitudinal, population-based twin study. *Arch Gen Psychiatry* 2006 Oct; **63**(10): 1113-1120.
7. Luciano M, Huffman JE, Arias-Vasquez A, Vinkhuyzen AA, Middeldorp CM, Giegling I *et al.* Genome-wide association uncovers shared genetic effects among personality traits and mood states. *Am J Med Genet B Neuropsychiatr Genet* 2013 Sep; **159B**(6): 684-695.
8. Middeldorp CM, de Moor MH, McGrath LM, Gordon SD, Blackwood DH, Costa PT *et al.* The genetic association between personality and major depression or bipolar disorder. A polygenic score analysis using genome-wide association data. *Transl Psychiatry* 2011; **1**: e50.
9. Barch DM, Sheffield JM. Cognitive impairments in psychotic disorders: common mechanisms and measurement. *World Psychiatry* 2014 Oct; **13**(3): 224-232.
10. Bagby RM, Bindseil KD, Schuller DR, Rector NA, Young LT, Cooke RG *et al.* Relationship between the five-factor model of personality and unipolar, bipolar and schizophrenic patients. *Psychiatry Res* 1997 May 5; **70**(2): 83-94.
11. Van Os J, Jones PB. Neuroticism as a risk factor for schizophrenia. *Psychol Med* 2001; **31**: 1129-1134.

Stratifying depression by schizophrenia risk

12. McGue M, Christensen K. Genetic and environmental contributions to depression symptomatology: evidence from Danish twins 75 years of age and older. *J Abnorm Psychol* 1997 Aug; **106**(3): 439-448.
13. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000 Oct; **157**(10): 1552-1562.
14. Lubke GH, Hottenga JJ, Walters R, Laurin C, de Geus EJ, Willemsen G *et al.* Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biol Psychiatry* 2012 Oct 15; **72**(8): 707-709.
15. Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, Nurnberger JI *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013 Apr 20; **381**(9875): 1371-1379.
16. Hong Lee S, Ripke S, Consortium CDGotPG. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**: 984-994.
17. Han B, Pouget JG, Slowikowski K, Stahl E, Lee HL, Diogo D *et al.* Using genotype data to distinguish pleiotropy from heterogeneity:

Stratifying depression by schizophrenia risk

deciphering coheritability in autoimmune and neuropsychiatric diseases. *bioRxiv* 2016.

18. Lencz T, Knowles E, Davies G, Guha S, Liewald DC, Starr JM *et al.* Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consortium (COGENT). *Mol Psychiatry* 2014 Feb; **19**(2): 168-174.
19. McIntosh AM, Gow A, Luciano M, Davies G, Liewald DC, Harris SE *et al.* Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. *Biol Psychiatry* 2013 May 15; **73**(10): 938-943.
20. Liebers DT, Pirooznia M, Seiffudin F, Musliner KL, Zandi PP, Goes FS. Polygenic Risk of Schizophrenia and Cognition in a Population-Based Survey of Older Adults. *Schizophr Bull* 2016 Feb 12.
21. Jones HJ, Stergiakouli E, Tansey KE, Hubbard L, Heron J, Cannon M *et al.* Phenotypic Manifestation of Genetic Risk for Schizophrenia During Adolescence in the General Population. *JAMA Psychiatry* 2016 Mar 1; **73**(3): 221-228.
22. Fanous AH, Kendler KS. The genetic relationship of personality to major depression and schizophrenia. *Neurotox Res* 2004; **6**(1): 43-50.

Stratifying depression by schizophrenia risk

23. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM *et al.* Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* 2012 Jul 10.
24. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ *et al.* Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet* 2006; **7**: 74.
25. Kerr SM, Campbell A, Murphy L, Hayward C, Jackson C, Wain LV *et al.* Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. *BMC Med Genet* 2013; **14**(1): 38.
26. First MB, Spitzer RL, Miriam G, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P)*. Biometrics Research, New York State Psychiatric Institute.: New York, 2002.
27. Weschler D. *WAIS-III Administration and Scoring Manual*. Psychological Corporation: London, UK, 1998.
28. Raven J. *The Mill Hill vocabulary scale*. H.K.Lewis: London, 1965.

Stratifying depression by schizophrenia risk

29. Lezak MD. *Neuropsychological Assessment*. Oxford University Press: New York, 1995.
30. Johnson W, Bouchard TJJ, Krueger RF, McGue M, Gottesman II. Just one g: consistent results from three test batteries. *Intelligence* 2004; **32**: 95-107.
31. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. *Psychol Med* 1979 Feb; **9**(1): 139-145.
32. Eysenck HJ, Eysenck SBG. *Manual for the Eysenck personality inventory*. University of London Press: London, 1964.
33. Gunderson KL. Whole-genome genotyping on bead arrays. *Methods Mol Biol* 2009; **529**: 197-213.
34. Clarke TK, Hall LS, Fernandez-Pujals AM, MacIntyre DJ, Thomson P, Hayward C *et al*. Major depressive disorder and current psychological distress moderate the effect of polygenic risk for obesity on body mass index. *Transl Psychiatry* 2015; **5**: e592.
35. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF *et al*. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009 Aug 6; **460**(7256): 748-752.

Stratifying depression by schizophrenia risk

36. Allen NE, Sudlow C, Peakman T, Collins R. UK biobank data: come and get it. *Sci Transl Med* Feb 19; **6**(224): 224ed224.
37. Deary IJ, Bedford A. Some origins and evolution of the EPQ-R (short form) Neuroticism and Extraversion items. . *Personality and Individual Differences* 2011; **50**(8): 1213-1217.
38. Hagenaars SP, Harris SE, Davies G, Hill WD, Liewald DC, Ritchie SJ *et al*. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Mol Psychiatry* 2016 Jan 26.
39. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics* 2015 May 1; **31**(9): 1466-1468.
40. Nakagawa S, Schielzeth H. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 2013; **4**(2): 133-142.
41. Hatzimanolis A, Bhatnagar P, Moes A, Wang R, Roussos P, Bitsios P *et al*. Common genetic variation and schizophrenia polygenic risk influence neurocognitive performance in young adulthood. *Am J Med Genet B Neuropsychiatr Genet* 2015 Jul; **168**(5): 392-401.

Stratifying depression by schizophrenia risk

42. Hori H, Matsuo J, Teraishi T, Sasayama D, Kawamoto Y, Kinoshita Y *et al.* Schizotypy and genetic loading for schizophrenia impact upon neuropsychological status in bipolar II and unipolar major depressive disorders. *J Affect Disord* 2012 Dec 15; **142**(1-3): 225-232.
43. Dickinson D, Ramsey ME, Gold JM. Overlooking the obvious: a meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Arch Gen Psychiatry* 2007 May; **64**(5): 532-542.
44. Zaninotto L, Guglielmo R, Calati R, Ioime L, Camardese G, Janiri L *et al.* Cognitive markers of psychotic unipolar depression: a meta-analytic study. *J Affect Disord* 2015 Mar 15; **174**: 580-588.
45. Bouchard TJ, Jr. Genes, environment, and personality. *Science* 1994 Jun 17; **264**(5166): 1700-1701.
46. de Moor MH, van den Berg SM, Verweij KJ, Krueger RF, Luciano M, Arias Vasquez A *et al.* Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry* 2015 Jul 1; **72**(7): 642-650.

Stratifying depression by schizophrenia risk

Figure Legends

Figure 1a,b: Scatterplot of SCZ-PGRS and neuroticism according to MDD status in GS:SFHS, and UK Biobank respectively

Figure 2: Bar chart of percentage variance explained for distress and neuroticism in GS:SFHS and UK Biobank

Table 1 Demographics, clinical, behavioural and temperament measures for GS:SFHS individuals in the current study.

	Controls (n=16,764)*		MDD cases (n=2587)*		Significance T / Chi 2 (p value)
	Mean	St dev	Mean	St dev	
Demographics					
Mean age (yrs)*	47.61	(15.27)	46.36 ^a	(12.88)	15.73 (7.31x10 ⁻⁵)
Gender (M:F)*	7218:9546	-	746:1841 ^a	-	186.54 (2.20x10 ⁻¹⁶)
Cognitive					
Logical memory (n=16,764;n=2587)	30.70	(8.48)	31.16	(7.89)	9.16 (2.48x10 ⁻³)
Digit symbol (n=16,566;n=2566)	72.32	(17.32)	71.59	(16.21)	1.47 (2.25x10 ⁻¹)
Mill Hill vocabulary (n=16,456;n=2563)	30.11	(4.72)	31.21	(4.79 ³)	1.70 (1.92x10 ⁻¹)
Verbal fluency (n=16,553;n=2575)	39.62	(11.67)	40.40	(11.91)	13.53 (2.35x10 ⁻⁴)
Clinical and trait-related features					
Psycholog distress (GHQ)	14.93	(7.55)	22.70 ^a	(12.62)	1797.00 (<1.00x10 ⁻¹⁰⁰)
Neuroticism	3.46	(2.94)	6.46 ^a	(3.32)	2155 (<1.00x10 ⁻¹⁰⁰)
PGRS measure (measures scaled)					
SCZ	-0.02	(1.00)	0.10	(1.02)	24.00 (9.69x10 ⁻⁷)

* Sample size as indicated unless otherwise specified in descriptive variable column, numbers for controls and MDD respectively. Controlled for age, sex, C1-4 and relatedness.

Table 2 PGRSs SCZ* MDD status interactions in both samples

	GS:SFHS (Controls n=16,764, MDD n=2587)*			UK Biobank (Controls n=27,476, MDD n=6049)*		
	Statistics (Beta z ratio)	p value	F stat	Statistics (Beta z ratio)	p value	F stat
Cognitive						
Composite 'g' factor	-0.0220 -1.1744	(2.67x10 ⁻¹)	1.1240	Composite 'g' factor	-0.0434 -1.3537	(1.75x10 ⁻¹) 1.8330
- Logical memory (n=16,764;n=2587)	-0.0028 -0.1800	(8.65x10 ⁻¹)	0.0290	-Memory	-0.0290 -0.9265	(3.54x10 ⁻¹) 0.8583
- Digit symbol (n=16,566;n=2566)	-0.0313 -2.3226	(4.31x10⁻²)	4.0900	- Symb digit substitution (n=7799;n=1767)	0.0197 0.4514	(6.52x10 ⁻¹) 0.2037
- Mill Hill vocabulary (n=16,456;n=2563)	0.0009 0.0673	(9.49x10 ⁻¹)	0.0040	-Reaction Time (n=27,299;n=6008)	0.0119 0.4603	(6.45x10 ⁻¹) 0.2118
- Verbal fluency (n=16,553;n=2575)	-0.0010 -0.0703	(7.42x10 ⁻¹)	0.1080	-Verb-Num Reasoning (n=26,847;n=5929)	-0.0282 -1.0530	(2.92x10 ⁻¹) 1.1090
Clinical and trait-related features						
Psycholog distress (GHQ)	-0.0402 -2.5872	(1.44x10⁻²)	5.9890	Psycholog distress (PHQ) (n=26,581;n=5969)	-0.0903 -3.3517	(8.03x10⁻⁴) 11.2300
Neuroticism	-0.0483 -3.2209	(1.96x10⁻³)	9.5890	Neuroticism	-0.0602 -2.2770	(2.28x10⁻²) 5.1850

Significant interactions highlighted in bold.

* Sample size as indicated unless otherwise specified in descriptive variable column, numbers for controls and MDD respectively. Controlled for age, sex, C1-4 and relatedness.

Table 3 Associations between PGRSs for SCZ with cognitive, clinical and trait-related features of MDD in cases/controls separately

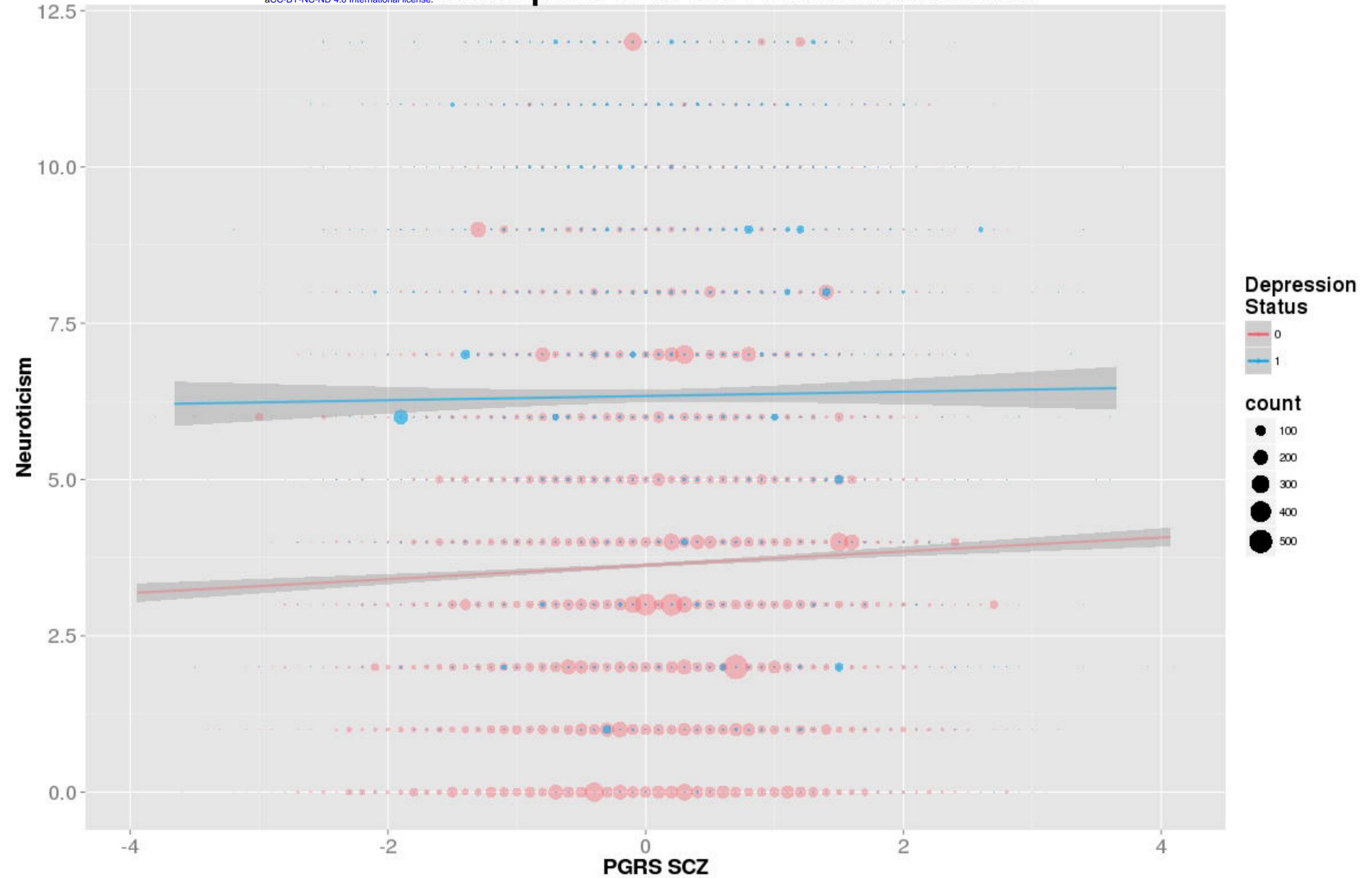
Generation Scotland: SFHS						
	Controls (n=16,764)			MDD cases (n=2587)		
	Beta; z ratio	p value	R ²	Beta; z ratio	p value	R ²
Cognitive						
Composite 'g' factor	-0.0773 -8.9629	(2.61x10 ⁻¹⁷)	0.3470	-0.0906 -4.1807	(6.50x10 ⁻⁵)	0.6671
- Logical memory	-0.0600 -7.3358	(2.33x10 ⁻¹³)	0.3119	-0.0764 -4.1814	(3.00x10 ⁻⁵)	0.8353
- Digit symbol	-0.0654 -9.2235	(3.27x10⁻²⁰)	0.3258	-0.0906 -5.3987	(7.35x10⁻⁸)	1.0900
- Mill Hill vocabulary	-0.0251 -3.2046	(1.36x10 ⁻³)	0.0379	-0.0131 -0.6736	(5.01x10 ⁻¹)	0.0173
- Verbal fluency	-0.0152 -1.8316	(6.70x10 ⁻²)	0.0190	-0.0064 -0.3132	(7.54x10 ⁻¹)	0.0110
Clinical and trait-related features						
Psycholog distress (GHQ)	0.0400 5.6700	(1.46x10⁻⁸)	0.1487	0.0111 0.3872	(6.99x10⁻¹)	0.0149
Neuroticism	0.0369 4.8657	(91.15x10⁻⁶)	0.1661	-0.0002 -0.0107	(9.91x10⁻¹)	0.0009
UK Biobank						
	Controls (n=27,476)			MDD cases (n=6049)		
	Beta; z ratio	p value	R ²	Beta; z ratio	p value	R ²
Cognitive						
Composite 'g' factor	-0.1203 -7.1856	(6.99x10 ⁻¹³)	0.3156	-0.1696 -5.8886	(4.12x10 ⁻⁹)	0.6162

Stratifying depression by schizophrenia risk

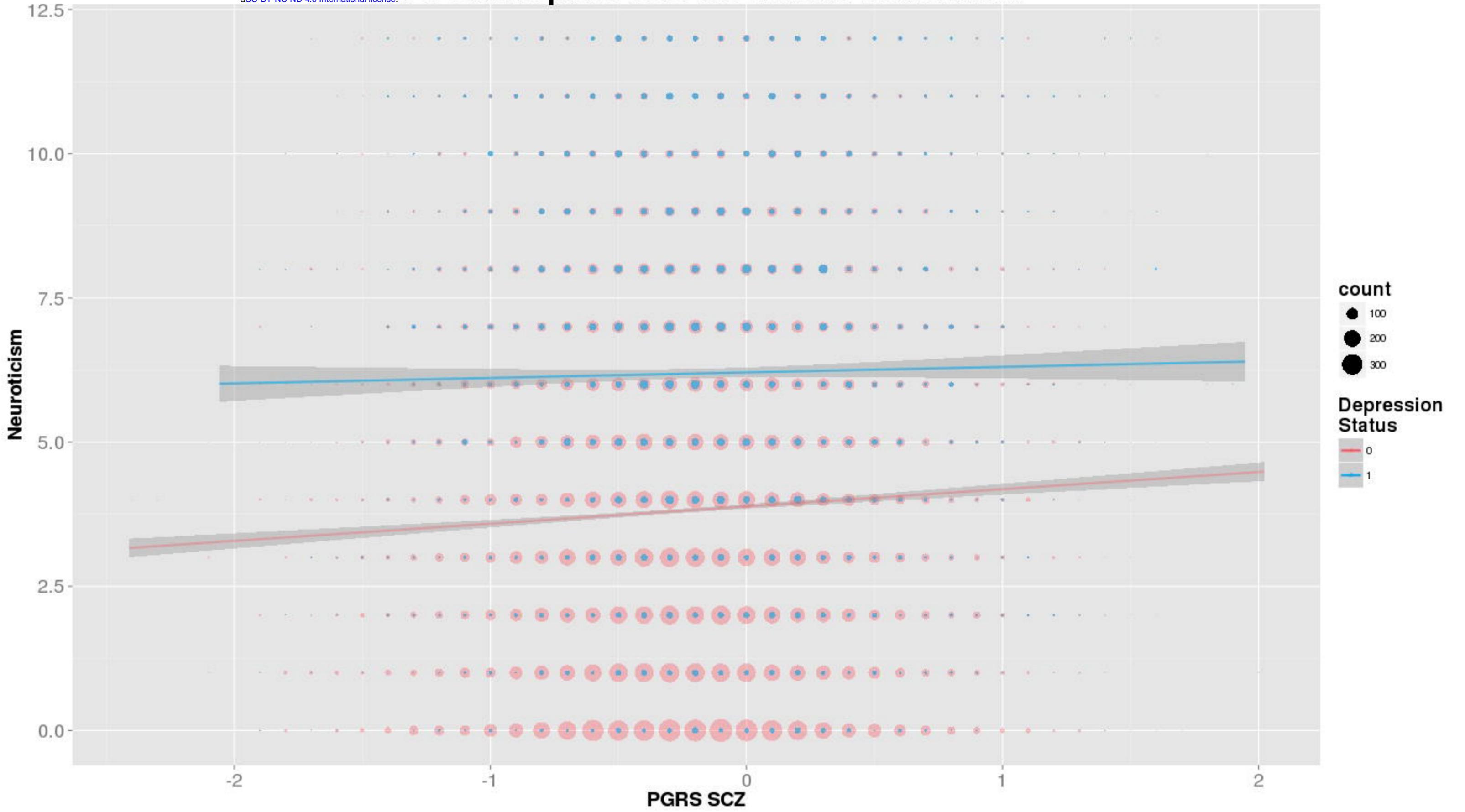
-Memory	0.0657 6.3921	(1.66x10 ⁻¹⁰)	0.1536	0.0490 2.2324	(2.56x10 ⁻²)	0.0863
- Symbol digit substitution	-0.1006 -5.1415	(2.79x10 ⁻⁷)	0.2816	-0.0785 -1.9623	(5.00x10 ⁻²)	0.1802
-Reaction Time	0.0390 3.4929	(4.79x10 ⁻⁴)	0.0422	0.0509 2.0881	(3.68x10 ⁻²)	0.0692
-Verbal-Numerical Reasoning	-0.1101 -9.5051	(2.16x10 ⁻²¹)	0.3467	-0.1390 -5.5560	(2.88x10 ⁻⁸)	0.5495
Clinical and trait-related features						
PHQ (health qu)	0.0827 7.5920	(3.25x10⁻¹⁴)	0.2218	-0.0259 -0.8072	(4.20x10⁻¹)	0.0111
Neuroticism	0.0539 8.9238	(4.79x10⁻¹⁹)	0.3026	0.0055 0.4103	(6.82x10⁻¹)	0.0029

R² = estimate of variance in trait explained by polygene score in %.

GSISFS: Scatterplot PGRS SCZ versus Neuroticism



UKB: Scatterplot PGRS SCZ versus Neuroticism



% Variance explained SCZ-PGRS for distress and neuroticism

