

# **Genetic comorbidity between major depression and cardio-metabolic disease, stratified by age at onset of major depression.**

Hagenaars SP<sup>1,2</sup>, Coleman JRI<sup>1,2</sup>, Choi S<sup>1</sup>, Gaspar H<sup>1,2</sup>, Adams MJ<sup>3</sup>, Howard D<sup>3</sup>, Hodgson K<sup>1,2</sup>, Traylor M<sup>4</sup>, Air TM<sup>5</sup>, Andlauer TFM<sup>6,7</sup>, Arolt V<sup>8</sup>, Baune BT<sup>8,9,10</sup>, Binder EB<sup>11,12</sup>, Blackwood DHR<sup>29</sup>, Boomsma DI<sup>13</sup>, Campbell A<sup>29</sup>, Cearns M<sup>9</sup>, Czamara D<sup>14</sup>, Dannlowski U<sup>8</sup>, Domschke K<sup>15,16</sup>, de Geus EJC<sup>13,17</sup>, Hamilton SP<sup>18</sup>, Hayward C<sup>29,30</sup>, Hickie I<sup>19</sup>, Hottenga JJ<sup>13</sup>, Ising M<sup>14</sup>, Jones I<sup>20</sup>, Jones LA<sup>21</sup>, Kutalik Z<sup>22,23</sup>, Lucae S<sup>14</sup>, Martin NG<sup>24</sup>, Milaneschi Y<sup>25</sup>, Mueller-Myhsok B<sup>6,26,27</sup>, Owen MJ<sup>20</sup>, Padmanabhan S<sup>28,29</sup>, Penninx BWJH<sup>25</sup>, Pistis G<sup>31</sup>, Porteous DJ<sup>29,41,42</sup>, Preisig M<sup>31</sup>, Ripke S<sup>32,33,34</sup>, Shyn SI<sup>35</sup>, Sullivan PF<sup>36,37,38</sup>, Whitfield J<sup>24</sup>, Wray NR<sup>39,40</sup>, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, MEGASTROKE consortium, McIntosh AM<sup>3,41,29</sup>, Deary IJ<sup>41,43</sup>, Breen G<sup>1,2</sup>, Lewis CM<sup>1,2</sup>

Corresponding author:

Saskia P Hagenaars, [saskia.hagenaars@kcl.ac.uk](mailto:saskia.hagenaars@kcl.ac.uk); Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK

Short title: Genetic overlap between MDD and cardio-metabolic traits

1 Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

2 NIHR Biomedical Research Centre, South London and Maudsley NHS Trust, London, UK

3 Division of Psychiatry, University of Edinburgh, Edinburgh, UK

4 Stroke Research Group, Division of Clinical Neurosciences, University of Cambridge, UK

5 Discipline of Psychiatry, University of Adelaide, Adelaide, SA, AU

6 Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, DE

- 7 Department of Neurology, Klinikum rechts der Isar, Technical University of Munich, Munich, DE
- 8 Department of Psychiatry, University of Münster, Münster, Nordrhein-Westfalen, DE
- 9 Department of Psychiatry, Melbourne Medical School, University of Melbourne, Melbourne, AU
- 10 Florey Institute for Neuroscience and Mental Health, University of Melbourne, Melbourne, AU
- 11 Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, DE
- 12 Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, US
- 13 Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL
- 14 Max Planck Institute of Psychiatry, Munich, DE
- 15 Department of Psychiatry and Psychotherapy, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, DE
- 16 Center for NeuroModulation, Faculty of Medicine, University of Freiburg, Freiburg, DE
- 17 Amsterdam Public Health Institute, Vrije Universiteit Medical Center, Amsterdam, NL
- 18 Psychiatry, Kaiser Permanente Northern California, San Francisco, CA, US
- 19 Brain and Mind Centre, University of Sydney, Sydney, NSW, AU
- 20 MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, GB
- 21 Department of Psychological Medicine, University of Worcester, Worcester, GB
- 22 Institute of Social and Preventive Medicine (IUMSP), University Hospital of Lausanne, Lausanne, VD, CH
- 23 Swiss Institute of Bioinformatics, Lausanne, VD, CH
- 24 Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, AU
- 25 Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, NL
- 26 Munich Cluster for Systems Neurology (SyNergy), Munich, DE
- 27 University of Liverpool, Liverpool, GB
- 28 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
- 29 Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
- 30 MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
- 31 Department of Psychiatry, University Hospital of Lausanne, Prilly, Vaud, CH
- 32 Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, US
- 33 Department of Psychiatry and Psychotherapy, Universitätsmedizin Berlin Campus Charité Mitte, Berlin, DE
- 34 Medical and Population Genetics, Broad Institute, Cambridge, MA, US
- 35 Behavioral Health Services, Kaiser Permanente Washington, Seattle, WA, US
- 36 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE
- 37 Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, US
- 38 Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, US
- 39 Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU
- 40 Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU

41 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh,  
Edinburgh, GB

42 Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of  
Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

43 Department of Psychology, University of Edinburgh, Edinburgh, UK

Corresponding author: Saskia P Hagenaars, [saskia.hagenaars@kcl.ac.uk](mailto:saskia.hagenaars@kcl.ac.uk); Social, Genetic and  
Developmental Psychiatry Centre; Institute of Psychiatry, Psychology & Neuroscience;  
King's College London; London SE5 8AF; UK.

## **Abstract**

## **Introduction**

It's imperative to understand the specific and shared aetiologies of major depression and cardio-metabolic disease, as both traits are frequently comorbid and each represents a major burden to society. This study examined whether there is a genetic association between major depression and cardio-metabolic traits and if this association is stratified by age at onset for major depression.

## **Methods**

Polygenic risk scores analysis and linkage disequilibrium score regression was performed to examine whether differences in shared genetic aetiology exist between depression case control status (N cases = 40,940, N controls = 67,532), earlier (N = 15,844), and later onset depression (N = 15,800) with body mass index, coronary artery disease, stroke, and type 2 diabetes in eleven data sets from the Psychiatric Genomics Consortium, Generation Scotland, and UK Biobank.

## **Results**

All cardio-metabolic polygenic risk scores were associated with depression status. Significant genetic correlations were found between depression and body mass index, coronary artery disease, and type 2 diabetes. Higher polygenic risk for body mass index, coronary artery disease and type 2 diabetes was associated with both early and later onset depression, while higher polygenic risk for stroke was associated with later onset depression only. Significant genetic correlations were found between body mass index and later onset depression, and between coronary artery disease and both early and late onset depression.

## **Conclusions**

The phenotypic associations between major depression and cardio-metabolic traits may partly reflect their overlapping genetic aetiology irrespective of the age depression first presents.

## Introduction

Major depressive disorder (MDD) and cardio-metabolic traits are both major causes of morbidity and mortality in high-income countries. Epidemiological studies have shown a well-established association between them(1): MDD increases the risk of cardio-metabolic disease onset and mortality, but cardio-metabolic disease itself can also increase risk of developing MDD. Specifically, a meta-analysis of 124,509 individuals across 21 studies showed that depression is associated with an 80% increased risk for developing coronary artery disease(2). Mezuk et al(3) showed that MDD predicted a 60% increased risk of type 2 diabetes, while type 2 diabetes predicted a 15% increased risk in MDD. MDD is associated with an increased risk of developing stroke (HR 1.45)(4), but meta-analyses have also shown that ~30% of stroke survivors suffered from MDD(5, 6). Risk factors for cardio-metabolic disease, such as obesity, are also linked to depression. Milaneschi et al(7) showed a bidirectional association between depression and obesity, where obesity increases risk for depression and depression increases risk for subsequent obesity. A more detailed review of the comorbidity between depression and cardio-metabolic traits is given in Penninx, Milaneschi (8).

Multiple mechanisms have been proposed to explain the association between MDD and cardio-metabolic diseases, for example biological dysregulation, or an unhealthy lifestyle(1). However, it remains unclear to what extent these mechanisms influence the association, as most studies examining the mechanisms were based on epidemiological observational study designs. The association between MDD and cardio-metabolic disease could also be due to shared genetic factors. Twin studies have shown that genetic factors contribute ~40% of the variation in liability to both MDD(9) and coronary artery disease(10, 11), 72% to type 2

diabetes(12), and 40 - 70% to BMI(13). No twin heritability estimates are available for stroke, as such studies have very limited numbers of twins. The most recent published GWAS for major depression from the Psychiatric Genomics Consortium, including 135,458 cases and 344,901 controls, identified 44 loci that were significantly associated with major depression, highlighting the highly polygenic nature of major depression(14). The heritability for major depression based on single nucleotide polymorphisms ( $h^2_{\text{SNP}}$ ) was estimated to be 8.7% on the liability scale, based on a lifetime risk of 0.15. Similarly, recent efforts to identify common variants associated with cardio-metabolic disease have shown these traits to be highly polygenic(15-18).

Several studies have shown genetic overlap between MDD and cardio-metabolic traits, in particular with coronary artery disease(14). Findings for other cardio-metabolic traits have been inconsistent. Previous studies have identified genetic overlap between MDD and BMI using polygenic risk scores(19), but not based on genetic correlations(20). However, more recent GWAS studies now have the power to detect a genetic correlation between MDD and BMI(14). Studies using twin data also showed that the phenotypic association between MDD and type 2 diabetes was partly due to genetic effects(21). This finding has been replicated using a polygenic risk score approach(22), but not using genetic correlations(22, 23). The Brainstorm Consortium did not find a significant genetic correlation between MDD and stroke(24), while Wassentheil-Smoller et al(25) showed that higher polygenic risk for MDD was associated with increased risk for stroke, in particular small vessel disease.

The inconsistency in results is likely due to differences in methodological approaches or in summary statistics, but could also be due to the heterogeneity of MDD. MDD onset can occur at any stage of life, but the factors associated with MDD are often age specific or age

restricted(26). Increased genetic risk for major depression is associated with earlier AAO compared to later AAO(14). Earlier AAO MDD has a higher heritability and is associated with increased risk for MDD in relatives. On the other hand, vascular disease and its risk factors are linked to a later age at onset (AAO) for MDD(27-29). A large study of Swedish twins showed that a later AAO for MDD in one twin was associated with a higher risk for vascular disease in the other twin(30). To date, no molecular genetic studies have examined the association between late onset MDD and cardio-metabolic traits and in the current study we have more power to replicate and further investigate this association leveraging both summary statistics and common genetic variant information.

The main aim of the present study is to examine the genetic association between MDD and cardio-metabolic traits using polygenic risk scores and genetic correlations. Secondly, we will examine the association stratified by AAO for MDD to test whether a higher genetic predisposition for cardio-metabolic traits is associated with a later AAO for MDD.

## Methods and Materials

### Samples

This study was performed using data from the Psychiatric Genomics Consortium (PGC) MDD working group (PGC-MDD), Generation Scotland: The Scottish Family Health Study (GS:SFHS), and UK Biobank.

#### *PGC*

Full details of the studies that form the PGC-MDD have previously been published(14). In summary, a subset of eleven studies from the full PGC-MDD analysis were included in the current study, based on the availability of AAO for MDD. All cases were required to have a lifetime diagnosis of MDD based on international consensus criteria (DSM-IV, ICD-9, or ICD-10)(31-33). This was ascertained using structured diagnostic instruments from direct interview by trained interviewers or clinician administered checklists. In most studies (10/11), controls were randomly selected from the general population and were screened for absence of lifetime MDD. This led to a total of 9,518 cases and 11,557 controls with genotype data and AAO information.

#### *GS:SFHS*

GS:SFHS is a family-based study consisting of 23,690 participants recruited from the population via general medical practices across Scotland. Sample characteristics and recruitment protocols have been described elsewhere(34, 35). In summary, MDD diagnosis was based on the Structured Clinical Interview for DSM-IV disorders (SCID)(36). Participants who answered positively to two mental health screening questions were invited to complete the full SCID to ascertain MDD diagnosis. Cases were further refined through NHS linkage. Controls were defined as participants who answered negatively to the two

screening questions or participants who did complete the SCID but did not meet criteria for MDD. This resulted in 1947 cases and 4858 controls, based on unrelated individuals.

### *UK Biobank*

UK Biobank is a large resource for identifying determinants of diseases in middle aged and older healthy individuals ([www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk))(37). A total 502,655 community-dwelling participants aged between 37 and 73 years were recruited between 2006 and 2010 in the United Kingdom, and underwent extensive testing including mental health assessments. MDD status in UK Biobank was derived from the online mental health questionnaire as previously described(38, 39). Briefly, MDD cases were defined as individuals meeting lifetime criteria for MDD based on questions from the Composite International Diagnostic Interview. Individuals reporting previous self-reported diagnosis of schizophrenia (or other psychosis) or bipolar disorder were excluded as MDD cases. Controls were defined as individuals who did not have any self-reported diagnosis of mental illness, did not take any anti-depressant medications, had not previously been hospitalised with a mood disorder, and did not meet previously defined criteria for a mood disorder(40). For the current study, this led to 29,475 cases and 51,243 controls.

### **Age at onset (AAO)**

AAO was defined as follows, based on previous work by Power et al(26) developed to account for the substantial by-study heterogeneity in the measure. Heterogeneity in AAO within the PGC MDD cohorts has been extensively investigated. Responses depend on the specific setting in which AAO is asked, and may reflect age at first symptoms, first visit to general practitioner or first diagnosis(26). Using cut offs for AAO (e.g. onset under 30) does not capture the variance in this measure, and we therefore followed Power et al(26) to use the

within study distribution to define early and later onset depression. This approach assumes that all cases were recruited from the same age at onset distribution with differences due to study-specific parameters. Cases reporting AAO older than the recorded age at interview were excluded from each study. Within each study, cases were ordered by AAO and divided into equal octiles (O1 – O8). The first three octiles (O1 – O3) were combined into the early AAO group, the last three octiles (O6 – O8) into the late AAO group. Splitting into equal octiles can result in individuals with the same AAO being arbitrarily placed in different octiles. To address this, cases in O4 with the same AAO as the maximum AAO in O3 were assigned to the early AAO group. Similarly, cases in O5 with the same AAO as the minimum AAO in O6 were assigned to the late AAO group. This led to a total of 15,844 MDD cases with early AAO and 15,800 MDD cases with late AAO, and 67,532 controls.

### **Genotyping and quality control**

Genotyping procedures have been described in the original analysis for each study(14, 41, 42). All analysis were based on individuals from European ancestry only.

PGC cohorts were genotyped following their local protocols. Individual genotype data for all PGC cohorts was processed using the PGC “Ricopili” pipeline for standardised quality control, imputation and analysis (<https://github.com/Nealelab/ricopili>)(43). Further details on the default parameters used for the current study can be found in the Supplementary material.

For GS:SFHS, genotyping and imputation procedures have previously been published by Nagy et al(41), a brief summary can be found in the Supplementary Material. Monogenic variants and variants with low imputation quality (INFO < 0.4) were removed.

Two highly-overlapping custom genotyping arrays (UK BiLEVE Axiom array and UK Biobank Axiom array, ~800,000 markers) were used to genotype all UK Biobank participants (N = 487,410)(42). Detailed QC procedures are described in the Supplementary Material.

## Statistical analysis

### *Polygenic risk scores*

Six polygenic risk scores (PRS) were calculated based on GWAS summary statistics for body mass index (BMI), coronary artery disease (CAD)(16), stroke and two stroke subtypes(15), and type 2 diabetes (T2D)(18) based on external GWAS p-value < 0.5. Table 1 provides further detail on the GWAS summary statistics.

PRS were calculated in all genotyped participants in each study using PRSice v2 (<https://github.com/choishingwan/PRSice>)(45). Prior to creating the scores, clumping was used to obtain SNPs in linkage disequilibrium with an  $r^2 < 0.25$  within a 250kb window. Due to the large number of overlapping samples between published BMI GWAS(13) and the PGC studies, a new GWAS was performed in the UK Biobank. The analysis used BGENIE v1.2 (42) on 291,684 individuals in UK Biobank that did not complete the Mental Health Questionnaire and were therefore independent of prediction samples, as MDD case control status was defined from the Mental Health Questionnaire. The Manhattan and QQ plot of the BMI GWAS can be found in the Supplementary Figure 2.

Logistic regression was used to test the associations between the five PRS and four different MDD case-control sets (all MDD cases vs. control subjects, early AAO cases vs. control subjects, late AAO cases vs. control subjects, and late AAO cases vs. early AAO cases). These analyses test our two hypotheses; comparing controls and MDD cases tests the

association of genetic risk for cardiometabolic traits with MDD, and whether this association differs by AAO. In contrast, comparing late and early AAO cases tests the association of genetic risk for cardiometabolic traits with AAO itself. Analyses were performed separately for each study, adjusting for relevant covariates in each study (PGC & GS:SFHS: 5 genetic principal components for population stratification; UK Biobank: 6 genetic principal components for population stratification, assessment centre, and genotyping batch). All PRS were standardised with each study across samples. Meta-analyses of the results across studies for each PRS – MDD combination were then conducted to synthesize the findings for maximum statistical power and to check for heterogeneity. Fixed-effects models were used in which the standardized regression coefficients were weighted by the inverse of their squared standard error. We tested for the presence of between-study heterogeneity using Cochran's Q. We corrected for multiple testing across all 24 meta-analysis models (4 phenotypes x 6 PRS) using the Benjamini Hochberg false discovery rate method(46), using a critical p-value of 0.0026, which means that all p-values equal to or below the critical p-value are considered significant.

#### *GWAS meta-analysis and genetic correlations*

Genome wide association analyses for three AAO-stratified MDD case-control subsets were performed within each study, with adjustments for population stratification. Study-specific covariates—for example, site or familial relationships—were also fitted as required (see Supplementary Material). The GWAS for GS:SFHS included all individuals (compared to unrelated individuals in the PRS analysis) to maximise power. Quality control of the study-level summary statistics was performed using the EasyQC software(47), which implemented the exclusion of SNPs with imputation quality <0.8 and minor allele count <25. P-value based meta-analyses, with genomic control, were then performed for each of the three

outcome measures, using the METAL package(48). SNPs with a combined sample size of less than 1000 participants were excluded.

Genetic correlations between the cardio-metabolic traits and the four MDD summary statistics were calculated using Linkage Disequilibrium score regression (LDSC) (49) using the default HapMap LD reference. To maximise power (and given genetic correlation analyses in LDSC are robust to sample overlap), the largest available summary statistics were used. These included the previously described summary statistics for coronary artery disease(16), stroke(15), and type 2 diabetes(18), as well as summary statistics for major depression from the PGC-MDD and 23andMe GWAS, including UK Biobank(14), and the most recent BMI GWAS, including both UK Biobank and the GIANT consortium(17). All results were corrected for multiple testing using the Benjamini Hochberg false discovery rate method (critical P-value = 0.0064). Differences between genetic correlations were assessed using block-jackknife (Supplementary Material)(50-52).

## Results

### *Summary of AAO*

The total sample consisted of 40,940 cases, including 15,844 cases classified as early AAO, 15,800 cases as late AAO, 6296 general cases not included in AAO-specific analysis, and 67,532 controls. The mean AAO (SD) across samples was 32.36 (15.22) years. The mean AAO per study ranged from 21.22 years to 37.23 years, with the German PGC-MDD samples, GS:SFHS and UK Biobank having a higher mean AAO. Figure 1 shows the distribution of AAO and the stratification in the early and late AAO groups in each of the studies. Supplementary Figure 1 shows the cumulative distribution of AAO in each study.

### *Polygenic risk analysis of cardio-metabolic disease*

The meta-analyses across studies showed that PRS for BMI, coronary artery disease, all stroke, ischaemic stroke, small vessel disease, and type 2 diabetes were significantly associated with MDD case control status (Figure 2, Supplementary Table 2). Study specific results can be found in Supplementary Table 3.

When stratifying by AAO and comparing with controls, higher PRS for BMI was more strongly associated with late AAO (OR = 1.07, 95% CI = 1.06 – 1.09,  $P = 4.10 \times 10^{-15}$ ) than early AAO (OR = 1.05, 95% CI = 1.03 – 1.07,  $P = 2.60 \times 10^{-7}$ ), however the difference between them was not significant (Figure 2). No significant difference was found between BMI PRS and late vs early AAO within MDD cases (OR = 1.02, 95% CI = 1.00 – 1.05,  $P = 0.04$ ). Higher PRS for CAD was associated with both late (OR = 1.05, 95% CI = 1.03 – 1.07,  $P = 8.66 \times 10^{-7}$ ) and early AAO (OR = 1.05, 95% CI = 1.03 – 1.07,  $P = 1.50 \times 10^{-7}$ ) compared to controls, while no significant difference was found between late and early AAO in MDD cases only (OR = 1.00, 95% CI = 0.97 – 1.02,  $P = 0.776$ ). Higher PRS for all stroke was only

associated with late AAO compared to controls (OR = 1.03, 95% CI = 1.01 – 1.05, P = 0.0006). Higher PRS for type 2 diabetes was associated with both late (OR = 1.03, 95% CI = 1.01 – 1.05, P = 0.001) and early (OR = 1.03, 95% CI = 1.01 – 1.05, P = 0.0004) AAO compared to controls, the association between late and early AAO in MDD cases only was non-significant (OR = 1.00, 95% CI = 0.97 – 1.02, P = 0.82).

### *Genetic correlations with cardio-metabolic traits*

Genetic correlations were calculated between MDD and cardio-metabolic traits based on the largest available GWAS (Figure 3, Supplementary Table 4). Significant genetic correlations were identified between MDD and BMI ( $r_g = 0.13$ , SE = 0.02,  $p = 3.43 \times 10^{-11}$ ), coronary artery disease ( $r_g = 0.12$ , SE = 0.03,  $p = 8.47 \times 10^{-6}$ ), and type 2 diabetes ( $r_g = 0.11$ , SE = 0.03,  $p = 0.0001$ ). The genetic correlation between ischaemic stroke and MDD was 0.10 (SE = 0.04,  $p = 0.026$ ), however this correlation did not survive the correction for multiple testing ( $p \leq 0.0064$ ). No significant genetic correlations were identified for all stroke and small vessel disease.

In order to calculate genetic correlations between the AAO stratified MDD traits and cardio-metabolic traits, we ran GWAS on the AAO stratified traits. Supplementary Figure 3-5 shows the Manhattan and QQ plots for these traits. Supplementary Table 4 provides further detail on the heritability estimates of the stratified MDD traits. One genome-wide significant variant was identified in the GWAS for early AAO vs controls (rs2789313 on chromosome 10, Z-score = 5.56,  $P = 2.74 \times 10^{-8}$ ). This variant is located in the *MALRDI* gene, which is involved in hepatic bile acid metabolism and lipid homeostasis(53). This gene has not previously been associated with depression. Supplementary Figure 3 shows a locus zoom plot of the region on

chromosome 10 including rs2789313. Supplementary Tables 5 – 7 provide summary statistics on all suggestive ( $P < 1 \times 10^{-5}$ ) variants for the stratified GWAS.

The results for the stratified AAO traits showed a significant genetic correlation between BMI and late AAO vs controls ( $r_g = 0.19$ ,  $SE = 0.04$ ,  $p = 2.42 \times 10^{-5}$ ), no significant genetic correlations were identified for either early AAO vs controls ( $r_g = 0.05$ ,  $SE = 0.03$ ,  $p = 0.059$ ) or late vs early AAO ( $r_g = 0.11$ ,  $SE = 0.05$ ,  $p = 0.037$ ). A stronger genetic correlation was found between late AAO and CAD ( $r_g = 0.19$ ,  $SE = 0.06$ ,  $p = 0.002$ ) compared to early AAO and CAD ( $r_g = 0.11$ ,  $SE = 0.04$ ,  $p = 0.006$ ), however the difference between the two genetic correlations was non-significant (block jackknife test for comparison of  $r_g$   $p = 0.227$ ). No significant genetic correlations were found between any of the AAO stratified MDD outcomes and all stroke, ischaemic stroke, small vessel disease, or type 2 diabetes (Figure 3, Supplementary Table 4).

## Discussion

This study explored whether the association between MDD and cardio-metabolic traits is partly due to genetic factors. Using data from publicly available GWAS and from the PGC, UK Biobank, and Generation Scotland, we showed significant genetic overlap between MDD and BMI, coronary artery disease, stroke, and type 2 diabetes from polygenic risk scores and genetic correlations.

Coronary artery disease and type 2 diabetes showed similar patterns of association. Both were associated with MDD in general and with late and early AAO, however the association with late vs early AAO was null. This indicates that both traits are associated with MDD irrespective of the AAO, and not associated with AAO itself. The current study is based on an extension of the strategies applied by Power et al. (26), who showed significant associations between polygenic risk for coronary artery disease and MDD, as well as with early and late onset MDD compared to controls. The results in the current study are based on a larger sample and more recent effect estimates for coronary artery disease, and have therefore greater power to detect an effect.

This is the first study to show genetic overlap between MDD and type 2 diabetes using both polygenic risk scores and genetic correlations. Twin studies have previously shown genetic overlap between MDD and type 2 diabetes, but evidence from genetic studies has been weaker. The current study, however, does show evidence for a shared genetic aetiology between MDD and type 2 diabetes based on both polygenic risk scores and genetic correlations (21-23).

The polygenic risk score analysis showed that individuals with a stronger genetic liability for stroke (and its subtypes) were more likely to suffer from MDD. Clinical and neuroimaging studies have shown that stroke could predispose the development of MDD by disrupting frontostriatal circuits in the brain involved in mood regulation(54-56), specifically in older individuals with MDD or those with a later AAO for MDD. We did not find a significant association between genetic risk for stroke and AAO. This could be due to a lack of power, as the stroke outcomes had low heritability estimates (observed  $h^2 \sim 1\%$ , Supplementary Table 4B) and the sample size of the stratified MDD outcomes is smaller than the overall MDD case-control comparison.

When stratifying by AAO, higher polygenic risk for BMI was more strongly associated with late AAO than early AAO compared to controls, while a significant genetic correlation was only found between late AAO vs control and BMI. A causal association using Mendelian randomization has previously been identified between BMI and depression, showing a 1.12 fold increase in depression for each SD increase in BMI(14). Similarly, Vogelzangs et al(57) have shown that over a five year period both overall and abdominal obesity were associated with an increased risk for MDD onset in men. Although the differences observed in this study are not significant following multiple testing correction, in the context of previous findings, they tentatively suggest that the aetiological processes underlying later onset MDD are linked to vascular pathology and its risk factors such as obesity.

LD score regression produced fewer significant associations than the polygenic risk score analysis. Genetic correlations using LD score regression are based on summary statistics only and could therefore be less powerful than a polygenic risk score approach based on raw genotype data. However, the direction of effect was the same across both analysis approaches

and significant genetic correlations were corroborated by significant polygenic risk score associations.

The current study has a number of limitations. The measures for AAO for MDD rely on self-report and are assessed differently across cohorts. We addressed this by stratifying AAO into octiles relative to the mean in each study, therefore assuming that each study recruited AAO from the same distribution with differences between studies due to differences in ascertainment measure. It should be noted that the mean AAO for late onset cases was 47.85 years, which is below what is generally considered to be late onset or geriatric depression (onset > 60 years). More pronounced overlap might exist between late onset MDD and cardio-metabolic traits when focussing on cases with a later AAO, however this study only had a small number of these included. However, this study does show that the effect of genetic risk for somatic conditions on depression is not limited to late onset or geriatric depression. By standardising the PRS in all models we have possibly introduced bias into our results. When the sample prevalence of a trait is not equal to the population prevalence, the mean of the PRS will not represent the mean of the PRS in the population, and thus introducing bias which could lead to inflated effect estimates.

Future studies could focus on the effects of medication or disease status for the cardio-metabolic traits, as the current study was unable to adjust for this due to lack of information on these variables. Pathway analysis might provide further insight into the biology underlying the association between cardio-metabolic traits and MDD. In order to further dissect the comorbidity between MDD and cardio-metabolic disease, it is imperative to have cleaner phenotypes, in particular with regards to AAO. Biobanks with electronic health records might

aid in providing improved phenotyping, but this still does not provide the same information as one gets from clinical research studies.

In summary, this study showed genetic overlap between MDD and cardio-metabolic traits based on polygenic risk scores and genetic correlations. These associations were largely irrespective of AAO for MDD, in particular for coronary artery disease and type 2 diabetes. The association with BMI showed some evidence for a stronger link with later onset MDD, however this finding needs to be replicated. Cardio-metabolic traits and its risk factors are to open to preventative strategies, therefore further understanding of the shared genetic aetiology between cardio-metabolic traits and MDD could play an important role in the prevention or management of MDD.

## Acknowledgements

SPH is funded by the Medical Research Council (MR/S0151132). CML is funding by the Medical Research Council (N015746/1). This study presents independent research supported by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, NIHR, Department of Health or King's College London. We thank participants and scientists involved in making the UK Biobank resource available (<http://www.ukbiobank.ac.uk/>). UK Biobank data used in this study were obtained under approved application 16577. We are deeply indebted to the investigators who comprise the PGC, and to the hundreds of thousands of subjects who have shared their life experiences with PGC investigators. The PGC has received major funding from the US National Institute of Mental Health and the US National Institute of Drug Abuse (U01 MH109528 and U01 MH1095320). Acknowledgements and funding sources for each of the primary studies in the PGC are contained in supplementary materials. Statistical analyses were carried out on the NL Genetic Cluster Computer (<http://www.geneticcluster.org/>) hosted by SURFsara, and the King's Health Partners High Performance Compute Cluster funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980). We would like to thank the research participants and employees of 23andMe for making this work possible. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [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 Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award “Stratifying Resilience and Depression Longitudinally” (STRADL) Reference 104036/Z/14/Z). Ethics

approval for the Generation Scotland was given by the NHS Tayside committee on research ethics (reference 15/ES/0040), and all participants provided written informed consent for the use of their data. The MEGASTROKE project received funding from sources specified at <http://www.megastroke.org/acknowledgments.html>

## References

1. Penninx BWJH (2017): Depression and cardiovascular disease: Epidemiological evidence on their linking mechanisms. *Neuroscience & Biobehavioral Reviews*. 74:277-286.
2. Nicholson A, Kuper H, Hemingway H (2006): Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. *European Heart Journal*. 27:2763-2774.
3. Mezuk B, Eaton WW, Albrecht S, Golden SH (2008): Depression and Type 2 Diabetes Over the Lifespan. *A meta-analysis*. 31:2383-2390.
4. Pan A, Sun Q, Okereke OI, Rexrode KM, Hu FB (2011): Depression and risk of stroke morbidity and mortality: A meta-analysis and systematic review. *JAMA*. 306:1241-1249.
5. Hackett ML, Pickles K (2014): Part I: Frequency of Depression after Stroke: An Updated Systematic Review and Meta-Analysis of Observational Studies. *International Journal of Stroke*. 9:1017-1025.
6. Ayerbe L, Ayis S, Wolfe CDA, Rudd AG (2013): Natural history, predictors and outcomes of depression after stroke: systematic review and meta-analysis. *British Journal of Psychiatry*. 202:14-21.
7. Milaneschi Y, Simmons WK, van Rossum EFC, Penninx BWJH (2018): Depression and obesity: evidence of shared biological mechanisms. *Molecular Psychiatry*.
8. Penninx BW, Milaneschi Y, Lamers F, Vogelzangs N (2013): Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. *BMC Medicine*. 11:129.
9. Sullivan PF, Neale MC, Kendler KS (2000): Genetic epidemiology of major depression: review and meta-analysis. *The American journal of psychiatry*. 157:1552-1562.
10. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. (2011): Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics*. 43:333.
11. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. (2015): Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*. 47:702.
12. Willemsen G, Ward KJ, Bell CG, Christensen K, Bowden J, Dalgård C, et al. (2015): The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. *Twin Research and Human Genetics*. 18:762-771.
13. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. (2015): Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 518:197.
14. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. (2018): Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*. 50:668-681.
15. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. (2018): Multi-ancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 50:524-537.
16. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. (2015): A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 47:1121-1130.
17. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. (2018): Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *bioRxiv*.

18. Scott RA, Scott LJ, Mägi R, Marullo L, Gaulton KJ, Kaakinen M, et al. (2017): An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes*. 66:2888-2902.
19. Milaneschi Y, Lamers F, Peyrot WJ, et al. (2017): Genetic association of major depression with atypical features and obesity-related immunometabolic dysregulations. *JAMA Psychiatry*. 74:1214-1225.
20. Wong BC-F, Chau CK-L, Ao F-K, Mo C-H, Wong S-Y, Wong Y-H, et al. (2018): Differential associations of depression-related phenotypes with cardiometabolic risks: Polygenic analyses and exploring shared genetic variants and pathways. *Depression and Anxiety*. 0.
21. Kan C, Pedersen NL, Christensen K, Bornstein SR, Licinio J, MacCabe JH, et al. (2016): Genetic overlap between type 2 diabetes and depression in Swedish and Danish twin registries. *Molecular Psychiatry*. 21:903.
22. Wong BC-F, Chau CK-L, Ao F-K, Mo C-H, Wong S-Y, Wong Y-H, et al. (2017): Differential associations of various depression-related phenotypes with cardiometabolic risks: Identification of shared genetic factors and implications for drug repositioning. *bioRxiv*.
23. Clarke T-K, Obsteter J, Hall LS, Hayward C, Thomson PA, Smith BH, et al. (2016): Investigating shared aetiology between type 2 diabetes and major depressive disorder in a population based cohort. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 174:227-234.
24. Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, Duncan L, et al. (2018): Analysis of shared heritability in common disorders of the brain. *Science*. 360.
25. Wassertheil-Smoller S, Qi Q, Dave T, Mitchell BD, Jackson RD, Liu S, et al. (2018): Polygenic Risk for Depression Increases Risk of Ischemic Stroke: From the Stroke Genetics Network Study. *Stroke*. 49:543-548.
26. Power RA, Tansey KE, Buttenschøn HN, Cohen-Woods S, Bigdeli T, Hall LS, et al. (2017): Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. *Biological Psychiatry*. 81:325-335.
27. Alexopoulos GS, Meyers BS, Young RC, Campbell S, Silbersweig D, Charlson M (1997): 'Vascular depression' hypothesis. *Archives of general psychiatry*. 54:915-922.
28. Taylor WD, Aizenstein HJ, Alexopoulos GS (2013): The vascular depression hypothesis: mechanisms linking vascular disease with depression. *Molecular Psychiatry*. 18:963.
29. Naismith SL, Norrie LM, Mowszowski L, Hickie IB (2012): The neurobiology of depression in later-life: Clinical, neuropsychological, neuroimaging and pathophysiological features. *Progress in Neurobiology*. 98:99-143.
30. Kendler KS, Fiske A, Gardner CO, Gatz M (2009): Delineation of Two Genetic Pathways to Major Depression. *Biological Psychiatry*. 65:808-811.
31. World Health Organization (1978): The International Classification of Diseases: Ninth Revision (ICD-9). *Annals of Internal Medicine*. 88:424-426.
32. World Health Organization (1992): *International Classification of Diseases*. Geneva: World Health Organization.
33. American Psychiatric Association (1994): *Diagnostic and Statistical Manual of Mental Disorders*. Washington DC: American Psychiatric Association.
34. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. (2013): Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *International Journal of Epidemiology*. 42:689-700.

35. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. (2006): Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Medical Genetics*. 7:74.
36. First MB, Spitzer RL, Gibbon M, Williams JB (1997): *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)*. Washington, DC, USA: American Psychiatric Publishing, Inc.
37. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. (2015): 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.
38. Coleman JRI, Purves KL, Davis KAS, Rayner C, Choi SW, Hübel C, et al. (2018): Genome-wide gene-environment analyses of depression and reported lifetime traumatic experiences in UK Biobank. *bioRxiv*.
39. Davis KAS, Coleman JRI, Adams M, Allen N, Breen G, Cullen B, et al. (2018): Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *BJPsych Open*. 4:83-90.
40. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. (2013): Prevalence and Characteristics of Probable Major Depression and Bipolar Disorder within UK Biobank: Cross-Sectional Study of 172,751 Participants. *PLOS ONE*. 8:e75362.
41. Nagy R, Boutin TS, Marten J, Huffman JE, Kerr SM, Campbell A, et al. (2017): Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Medicine*. 9:23.
42. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. (2018): The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 562:203-209.
43. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, et al. (2019): RICOPILI: Rapid Imputation for Consortias PipeliNe. *bioRxiv*.587196.
44. The UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, et al. (2015): The UK10K project identifies rare variants in health and disease. *Nature*. 526:82.
45. Euesden J, Lewis CM, O'Reilly PF (2015): PRSice: Polygenic Risk Score software. *Bioinformatics*. 31:1466-1468.
46. Benjamini Y, Hochberg Y (1995): Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 57:289-300.
47. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, et al. (2014): Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols*. 9:1192.
48. Willer CJ, Li Y, Abecasis GR (2010): METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 26:2190-2191.
49. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*. 47:291.
50. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. (2015): An atlas of genetic correlations across human diseases and traits. *Nature Genetics*. 47:1236.
51. Tukey J (1958): Bias and confidence in not quite large samples. *Ann Math Statist*. 29:614.
52. Quenouille MH (1956): Notes on Bias in Estimation. *Biometrika*. 43:353-360.
53. Vergnes L, Lee Jessica M, Chin Robert G, Auwerx J, Reue K (2013): Diet1 Functions in the FGF15/19 Enterohepatic Signaling Axis to Modulate Bile Acid and Lipid Levels. *Cell Metabolism*. 17:916-928.
54. Alexopoulos GS (2017): Depression and Cerebrovascular Disease: What is to be Done? *The American Journal of Geriatric Psychiatry*. 25:129-130.

55. van Sloten TT, Sigurdsson S, van Buchem MA, Phillips CL, Jonsson PV, Ding J, et al. (2015): Cerebral Small Vessel Disease and Association With Higher Incidence of Depressive Symptoms in a General Elderly Population: The AGES-Reykjavik Study. *American Journal of Psychiatry*. 172:570-578.
56. Herrmann LL, Le Masurier M, Ebmeier KP (2008): White matter hyperintensities in late life depression: a systematic review. *Journal of Neurology, Neurosurgery & Psychiatry*. 79:619-624.
57. Vogelzangs N, Kritchevsky SB, Beekman ATF, Brenes GA, Newman AB, Satterfield S, et al. (2010): Obesity and Onset of Significant Depressive symptoms: Results from a community-based cohort of older men and women. *The Journal of Clinical Psychiatry*. 71:391-399.

**Table 1.** Information about GWAS summary statistics, based on European individuals only, to create polygenic risk scores in PGC-MDD, GS:SFHS, and UK Biobank.

Trait	Origin	N cases/N controls	SNPs with p-value < $5*10^{-8}$
Body mass index	UK Biobank	291,683 individuals	27,164
Coronary artery disease	CARDIoGRAM(16)	60,801 cases – 123,504 controls	2,213
Stroke: all	MEGASTROKE(15)	40,585 cases – 406,111 controls	371
Stroke: ischaemic	MEGASTROKE(15)	34,217 cases – 406,111 controls	307
Stroke: small vessel disease	MEGASTROKE(15)	5,386 cases – 406,111 controls	0
Type 2 diabetes	DIAGRAM(18)	26,676 cases – 132,532 controls	1,788

**Figure 1.** Distribution of Age at onset for MDD per sample. Scatterplot represents the stratification of the early and late AAO groups within each sample.

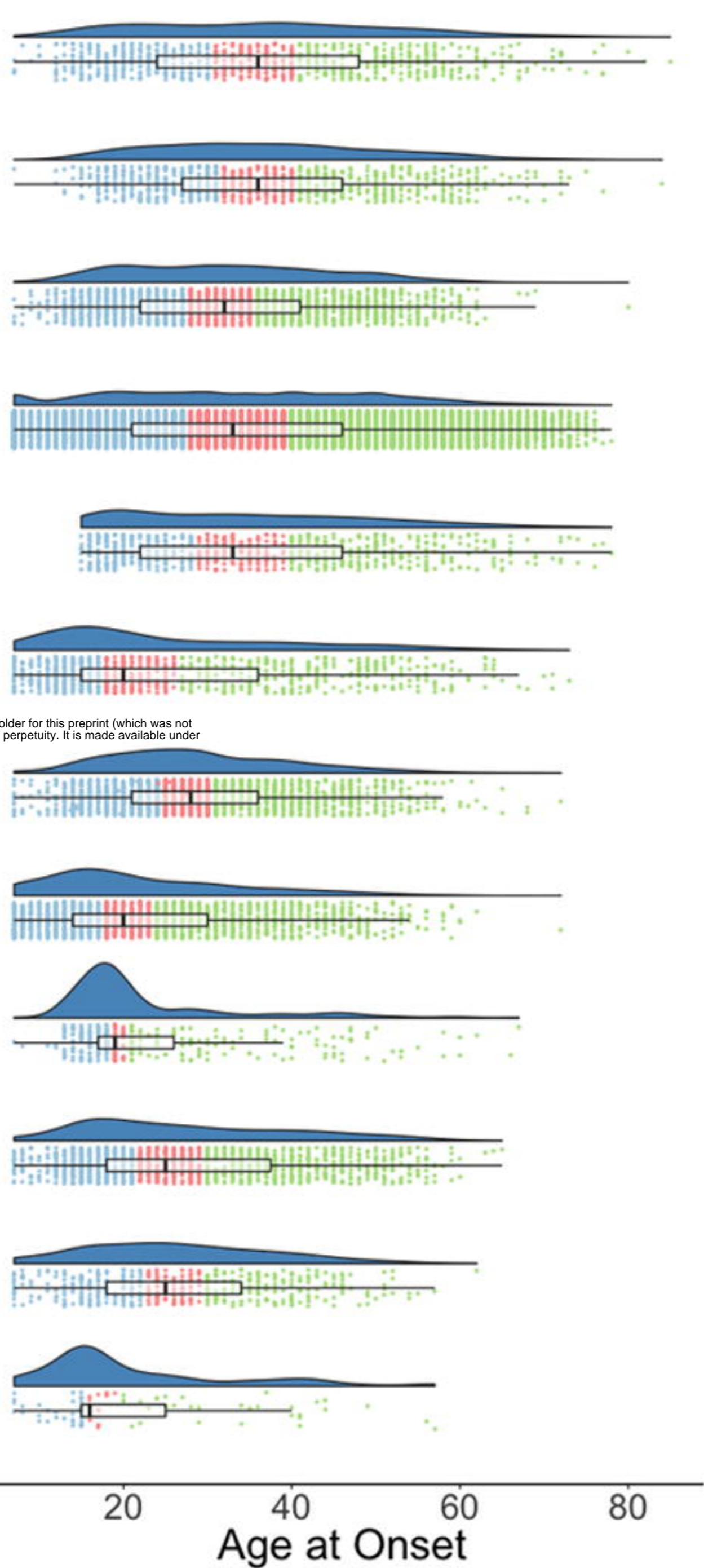
**Figure 2.** Meta-analysis of associations between cardio-metabolic PRS and stratified MDD outcomes. OR, Odds Ratio; 95% CI, 95% confidence interval; PRS, polygenic risk score.

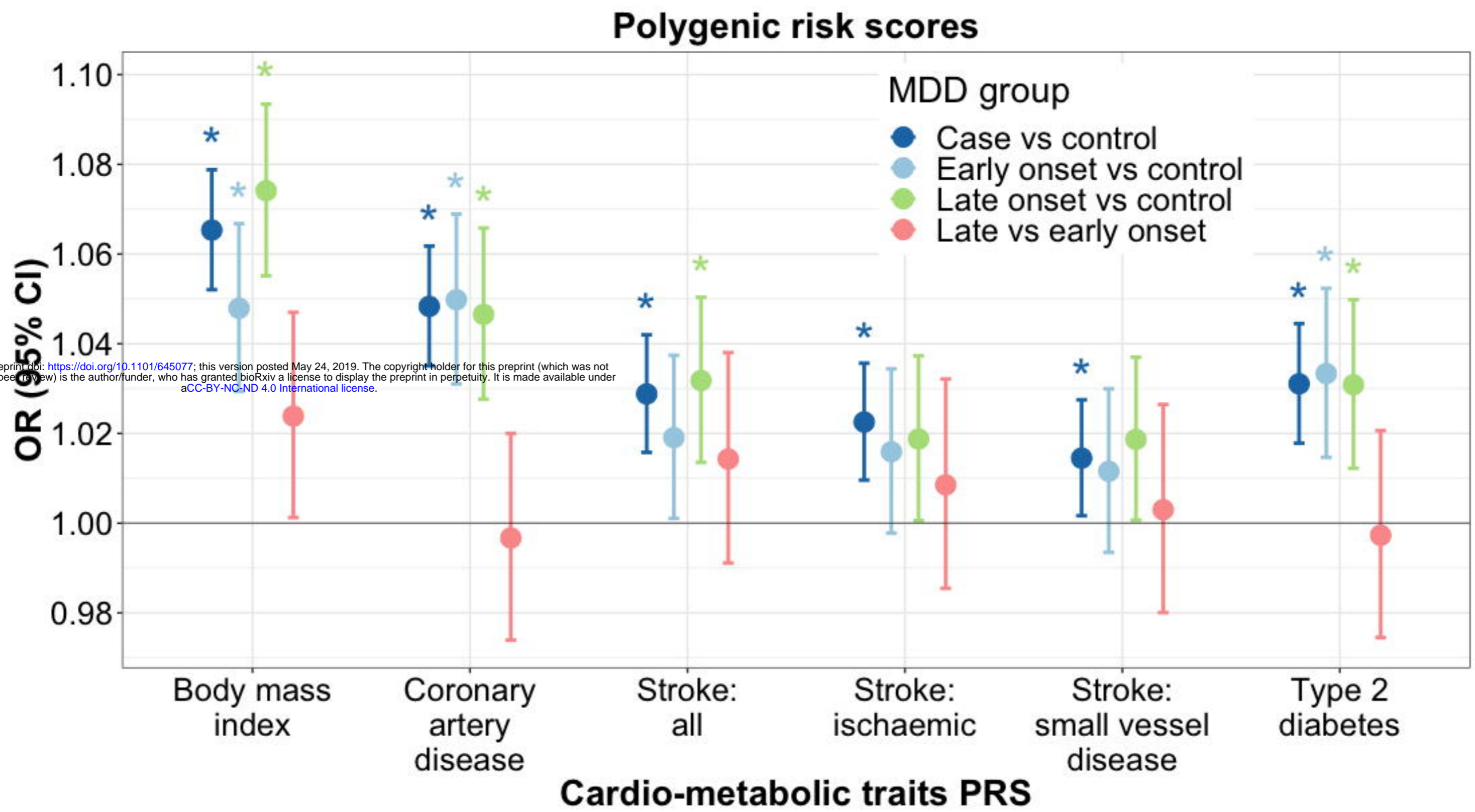
**Figure 3.** Genetic correlations (SE), derived using LD score regression, between cardio-metabolic traits and stratified MDD outcomes.



Studyname

MARS  
(N control = 888,  
N cases = 848)  
GSK  
(N control = 861,  
N cases = 880)  
GS:SFHS  
(N control = 4858,  
N cases = 2000)  
UK Biobank  
(N control = 51,243,  
N cases = 29,475)  
Muenster  
(N control = 834,  
N cases = 960)  
STAR\*D  
(N control = 934,  
N cases = 936)  
QIMR  
(N control = 1695,  
N cases = 1928)  
Radiant  
(N control = 2762,  
N cases = 2337)  
Edinburgh  
(N control = 258,  
N cases = 372)  
NESDA/NTR  
(N control = 1602,  
N cases = 1494)  
PsyCoLaus  
(N control = 1445,  
N cases = 507)  
CoFaMS  
(N control = 125,  
N cases = 120)





Genetic correlations

