

## Original Article

**Cite this article:** Coombes BJ *et al* (2023). The genetic contribution to the comorbidity of depression and anxiety: a multi-site electronic health records study of almost 178 000 people. *Psychological Medicine* 1–7. <https://doi.org/10.1017/S0033291723000983>

Received: 6 May 2022  
Revised: 22 March 2023  
Accepted: 27 March 2023


### Keywords:

Anxiety; depression; electronic health records; polygenic risk

### Corresponding author:

Brandon J Coombes,  
E-mail: [coombes.brandon@mayo.edu](mailto:coombes.brandon@mayo.edu)

# The genetic contribution to the comorbidity of depression and anxiety: a multi-site electronic health records study of almost 178 000 people

Brandon J Coombes<sup>1</sup> , Isotta Landi<sup>2,3,4,5</sup>, Karmel W Choi<sup>6,7,8</sup>, Kritika Singh<sup>9,10</sup>, Brian Fennessy<sup>3</sup>, Greg D Jenkins<sup>1</sup>, Anthony Batzler<sup>1</sup>, Richard Pendegraft<sup>1</sup>, Nicolas A Nunez<sup>11</sup>, Y Nina Gao<sup>12</sup>, Euijung Ryu<sup>1</sup>, Priya Wickramaratne<sup>12</sup>, Myrna M Weissman<sup>12</sup>, Regeneron Genetics Center, Jyotishman Pathak<sup>13,14</sup>, J John Mann<sup>12</sup>, Jordan W Smoller<sup>6,7,15</sup>, Lea K Davis<sup>9,10</sup>, Mark Olfson<sup>12</sup>, Alexander W Charney<sup>2,3,4</sup> and Joanna M Biernacka<sup>1,11</sup>

<sup>1</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA; <sup>2</sup>Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA; <sup>3</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; <sup>4</sup>Mount Sinai Clinical Intelligence Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA; <sup>5</sup>Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA; <sup>6</sup>Department of Psychiatry, Center for Precision Psychiatry, Massachusetts General Hospital, Boston, MA, USA; <sup>7</sup>Psychiatric & Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; <sup>8</sup>Department of Psychiatry, Harvard Medical School, Boston, MA, USA; <sup>9</sup>Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; <sup>10</sup>Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; <sup>11</sup>Department of Psychiatry & Psychology, Mayo Clinic, Rochester, MN, USA; <sup>12</sup>Department of Psychiatry, Columbia University and New York State Psychiatric Institute, New York, USA; <sup>13</sup>Department of Population Health Sciences, Weill Cornell Medicine, New York, New York, USA; <sup>14</sup>Clinical and Translational Science Center, Weill Cornell Medicine, New York, New York, USA and <sup>15</sup>Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA

## Abstract

**Background.** Depression and anxiety are common and highly comorbid, and their comorbidity is associated with poorer outcomes posing clinical and public health concerns. We evaluated the polygenic contribution to comorbid depression and anxiety, and to each in isolation. **Methods.** Diagnostic codes were extracted from electronic health records for four biobanks [ $N = 177\,865$  including 138 632 European (77.9%), 25 612 African (14.4%), and 13 621 Hispanic (7.7%) ancestry participants]. The outcome was a four-level variable representing the depression/anxiety diagnosis group: neither, depression-only, anxiety-only, and comorbid. Multinomial regression was used to test for association of depression and anxiety polygenic risk scores (PRSs) with the outcome while adjusting for principal components of ancestry. **Results.** In total, 132 960 patients had neither diagnosis (74.8%), 16 092 depression-only (9.0%), 13 098 anxiety-only (7.4%), and 16 584 comorbid (9.3%). In the European meta-analysis across biobanks, both PRSs were higher in each diagnosis group compared to controls. Notably, depression-PRS (OR 1.20 per s.d. increase in PRS; 95% CI 1.18–1.23) and anxiety-PRS (OR 1.07; 95% CI 1.05–1.09) had the largest effect when the comorbid group was compared with controls. Furthermore, the depression-PRS was significantly higher in the comorbid group than the depression-only group (OR 1.09; 95% CI 1.06–1.12) and the anxiety-only group (OR 1.15; 95% CI 1.11–1.19) and was significantly higher in the depression-only group than the anxiety-only group (OR 1.06; 95% CI 1.02–1.09), showing a genetic risk gradient across the conditions and the comorbidity. **Conclusions.** This study suggests that depression and anxiety have partially independent genetic liabilities and the genetic vulnerabilities to depression and anxiety make distinct contributions to comorbid depression and anxiety.

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

## Introduction

Depression and anxiety disorders are among the most common psychiatric disorders with a lifetime prevalence of 17% and 30%, respectively (Global Burden of Disease Study 2013 Collaborators, 2015). Symptoms of depression and anxiety often co-occur with nearly half of adults with anxiety also reporting depressive symptoms (Kessler *et al.*, 2008). Comorbid depression and anxiety represents a significant clinical and public health concern and is associated with greater severity and earlier age of onset, more severe depression, increased suicidal ideation, poorer antidepressant response, psychosocial impairment, worse course of illness,

and increased comorbidity with substance use disorders (Joffe, Bagby, & Levitt, 1993; Nam, Kim, & Roh, 2016; Pollack, 2005; Zhou *et al.*, 2017). Proposed risk factors associated with the comorbidity of depression and anxiety include female sex, younger age, higher educational level, and childhood trauma (de Graaf, Bijl, Smit, Vollebergh, & Spijker, 2002) as well as genetics (Wray *et al.*, 2018).

The etiology of comorbid depression and anxiety disorders remains poorly understood. Epidemiological data suggest that depression and anxiety commonly co-occur because they share underlying liabilities or transdiagnostic factors (Blanco *et al.*, 2013; Kotov *et al.*, 2017; Krueger, 1999). Latent variable techniques applied to the correlational structure of psychiatric comorbidities have consistently identified a broad internalizing factor that includes depression and anxiety disorders (Blanco *et al.*, 2013; Kotov *et al.*, 2017; Krueger, 1999). Clinically, depression and anxiety also have significant overlap with anxiety symptoms being incorporated into some diagnostic criteria for depression (American Psychiatric Association, 2022). Further, dysregulation of the hypothalamic–pituitary–adrenal axis and the role of common neurotransmitters have been implicated in both (Ressler & Nemeroff, 2000; Zorn *et al.*, 2017). These observations, together with shared clinical responses to similar pharmacological (Nutt, 2004) and psychosocial interventions (Craske, 2012) as well as shared genetic risk factors across internalizing disorders, support the common liability hypothesis of depression and anxiety disorders (Blanco, Wall, Feng, & Olfson, 2021).

An alternative possibility is that mood and anxiety have partially or even largely overlapping, but distinct, risk factors and clinical characteristics. Using Research Domain Criteria (RDoC), for example, depression is described as a disorder of impaired reward response, learning, and valuation (Dillon *et al.*, 2014; Greenebaum & Nierenberg, 2020). In contrast, anxiety is described as a dysfunction of threat detection (Dillon *et al.*, 2014; Greenebaum & Nierenberg, 2020). A recent twin study sought to explain both the comorbidity and distinctive nature of the disorders and found that both depression and anxiety have a positive genetic correlation with behavioral inhibition (response to negative stimuli), but only depression has a negative genetic correlation with behavioral activation (response to positive stimuli) (Takahashi, Yamagata, Ritchie, Barker, & Ando, 2021). Risk factors for each disorder also appear to be at least partly distinct such as neuroticism being more strongly linked to depression, arguing against the idea of a fully shared underlying neurobiological liability (Moscati, Flint, & Kendler, 2016).

Both depression and anxiety are moderately heritable disorders, with heritability estimates of 30–40% (Sullivan, Neale, & Kendler, 2000) and 20–60% (Polderman *et al.*, 2015), respectively. Genome-wide association studies (GWAS) of depression and anxiety have largely been performed separately and without explicitly accounting for the comorbidity (Giannakopoulou, Lin, Meng, & Su, 2021; Howard *et al.*, 2019; Levey *et al.*, 2020, 2021; Purves *et al.*, 2020). Given the high rate of comorbidity between depression and anxiety, these GWAS likely capture shared genetic liability for the two disorders, but may also capture distinct genetic risk factors of each disorder. Prior studies show a robust genetic correlation between depression and anxiety with estimates ranging from 80% to 95% (Mei *et al.*, 2022). However, none of these studies specifically investigated the differences between the two disorders. A recent study showed substantial genetic overlap between the two but also trait-specific influences (Thorp *et al.*, 2021).

However, in the same study, polygenic risk scores (PRSs) did not demonstrate trait-specificity to depression and anxiety.

Electronic Health Record (EHR)-linked biobanks provide access to data from large collections of patients presenting for clinical care and have become a valuable resource for genetic studies. While EHR-derived data may not capture patients' diagnoses as well as direct diagnostic interviews, EHR-based studies enable investigation of real-world diagnoses, including comorbidity between diagnoses, such as anxiety and depression. Biobanks also often include larger and more diverse cohorts than typical disease-specific case–control research samples. In this multi-ancestry study, we leveraged four EHR-linked biobanks from the USA to assess the contribution of genetic risk to the comorbidity of depression and anxiety. We first assessed whether PRSs for depression and anxiety, derived using results from large GWAS of patients of European ancestries primarily assessed in clinical research studies, were predictive of the respective disorders defined using diagnosis codes from EHR. Next, we evaluated how the genetic risk factors for depression and anxiety jointly contribute to the comorbidity of depression and anxiety.

## Methods

### *Hospital-based biobanks*

Data for this study were obtained from four different health care system biobanks: Mayo Clinic Biobank (MCB) linked to the Mayo Clinic hospital system (Bielinski *et al.*, 2011); BioMe linked to the Mount Sinai Health System in New York City, New York (Belbin *et al.*, 2017); and two biobanks from the PsycheMERGE (electronic MEdical Record and GENomics) Network (Zheutlin *et al.*, 2019): BioVU linked to Vanderbilt University Medical Center (VUMC) in Nashville, Tennessee (Roden *et al.*, 2008); and the Mass General Brigham (MGB) Biobank linked to the MGB hospital system in Boston, Massachusetts (Karlson, Boutin, Hoffnagle, & Allen, 2016). Patients enrolled in all four biobanks gave informed consent for use of their EHR data linked to their genetic data. Each site obtained institutional review board approval for the EHR-biobank research. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

The EHR includes information on patients such as demographics, medications/prescriptions, laboratory values, billing codes from the International Classification of Diseases, 9th and 10th editions (ICD-9 and ICD-10-Clinical Modifications) (WHO & World Health Organization, 1993), and Current Procedural Terminology (CPT) codes. For the current study, structured EHR data for participants were extracted at the four sites during 2021 and included all patient data before that date (details are provided in online Supplementary Methods).

### *Identifying cases with depression and anxiety*

Depression and anxiety were defined using an initial list of ICD9/10-CM codes mapped to phcodes [i.e. higher order group of diagnoses, as described and validated by the Phcode map 1.2b1 (Denny *et al.*, 2010)] available from <https://phewascatalog.org/phcodes>. (He *et al.*, 2019; Wei *et al.*, 2017) These definitions were then modified through expert curation from the authors. Depression was defined as having at least two depression-related

ICD9/10 codes, using an initial list of ICD9/10-CM codes mapped to the phecode for depression (296.2 and 296.22) (He et al., 2019; Wei et al., 2017), with the addition of dysthymic disorder [ICD9:300.4; ICD10:F34.1], depressive type psychosis [ICD9:298.0], and atypical depressive disorder [ICD9:296.82]. Anxiety was defined as having at least two anxiety-related ICD9/10 codes, using an initial list of ICD9/10-CM codes mapped to phecodes for anxiety disorders, generalized anxiety disorders, social phobias and panic disorders, and phobias (300.1, 300.11, 300.12, 300.13, respectively) (He et al., 2019; Wei et al., 2017) with the addition of separation anxiety disorder [ICD9: 309.21; ICD10: F93.0] and removing hysteria (ICD9: 300.1) and overanxious disorder (ICD9: 313). The complete list of ICD9/10-CM codes is presented in online Supplementary Tables S1 and S2. Patients with only one diagnostic code were excluded from all analyses and controls were patients who had no documented ICD codes for depression and anxiety disorders. No exclusions were made for diagnoses of other psychiatric disorders.

### Genetic data

DNA samples from blood obtained from study participants were genotyped using genome-wide arrays, and the genetic data were imputed and processed for quality control at each site as described in the eMethods section. Briefly, single nucleotide polymorphisms were excluded using filters for call rate, minor allele frequency (<1%), and Hardy–Weinberg equilibrium. Individuals were excluded for missingness rate, sex errors, heterozygosity, and relatedness. The study included three ancestries: European, African, and Hispanic. MCB and MGB included only patients from European ancestries in their analyses due to sample size limitations of other ancestries. Genotype imputation was performed after the initial quality control and converted to best-guess genotypes for all markers with high-quality imputation (dosage- $R^2 > 0.8$ ). All subsequent analyses were adjusted for principal components (PCs) within each ancestral group to reduce confounding by population substructure.

### Polygenic risk scores

To estimate the genetic risk for depression and anxiety, we calculated PRSs using the summary statistics from the Psychiatric Genomic Consortium (PGC) GWAS from major depressive disorder (MDD) (Howard et al., 2019) and anxiety (ANX) (Purves et al., 2020) working groups, respectively. Both GWASs were performed in European ancestry samples. The PRS calculations were performed separately for each ancestral group at each site. The PRSs were calculated using LDpred2 (Privé, Arbel, & Vilhjálmsson, 2020) for MCB, PRSice2 (Choi & O'Reilly, 2019) using a PRS-PCA approach (Coomes, Ploner, Bergen, & Biernacka, 2020) for BioMe and MGB, and PRS-CS (Ge, Chen, Ni, Feng, & Smoller, 2019) for BioVU. The PRS-PCA approach was implemented in BioMe and MGB to avoid the bias of searching for the best  $p$  value threshold with PRSice2.

### Statistical analysis

All analyses were stratified by site and ancestral group. Logistic regression was first used to assess whether each PRS (depression or anxiety) was associated with the matching EHR-defined diagnosis while adjusting for PCs. Next, we used multinomial

regression to assess how well the combination of the MDD-PRS and ANX-PRS jointly predict the comorbidity of depression and anxiety by including either one PRS at a time or both MDD-PRS and ANX-PRS in the same model. The categorical outcome used in the multinomial model classified patients into one of four different groups: neither depression nor anxiety (controls), depression-only, anxiety-only, and both. Post-hoc pairwise comparisons were used to evaluate differences between groups. Effect sizes for PRS analyses were represented by odds ratios and percent of variation explained on the liability scale (assuming a 20% prevalence of both depression and anxiety in the general population). Meta-analyses within ancestry groups were performed using a fixed-effects meta-analysis with the meta R package.

## Results

Figure 1 presents the frequencies of depression and anxiety diagnoses defined from the EHR at each site (total  $N = 177\,865$ ). Around 75% of each biobank sample had no ICD codes for depression or anxiety with the exception of MGB which had a higher prevalence of anxiety-related codes (25%) as seen in previous analyses of the data (Lee et al., 2022) and thus a lower proportion of controls. When comparing diagnostic rates across self-reported race and ethnicity in BioMe, African American and Hispanic participants had higher rates of depression codes (16% and 21%, respectively) than self-identified White participants (12%). The full demographic characteristics and EHR summaries by site are shown in online Supplementary Table S3.

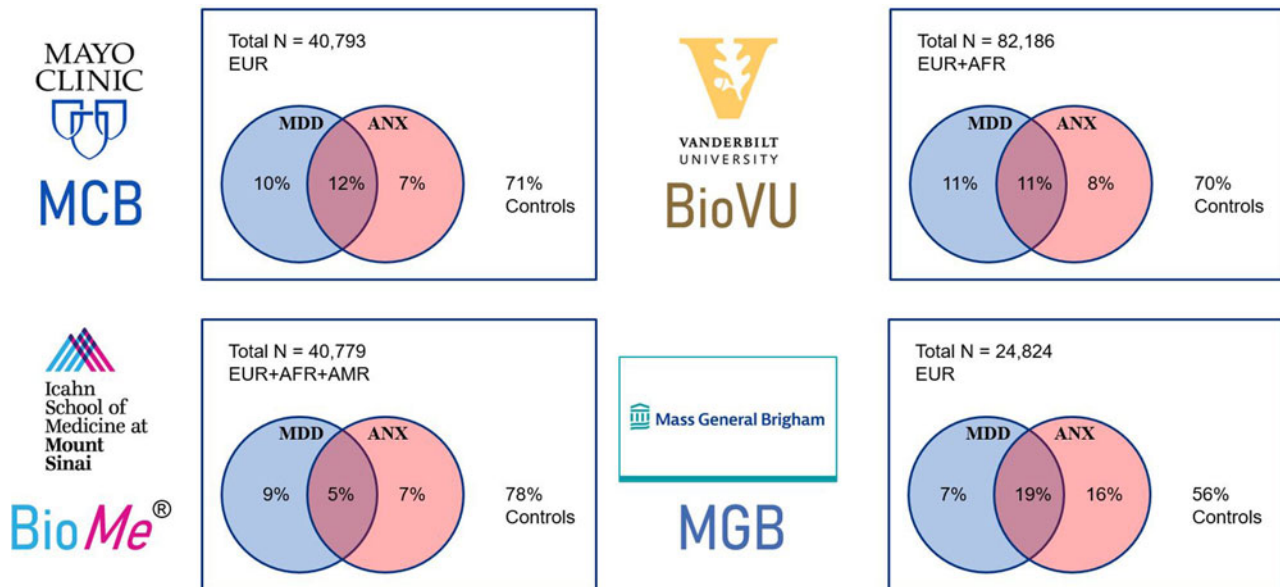
### PRS prediction of depression and anxiety separately

We first tested whether the MDD-PRS and ANX-PRS, trained on samples from European ancestries that are largely clinically ascertained, were associated with depression and anxiety diagnoses in the EHR, respectively. Figure 2 shows the PRS predictive performance for each site and ancestry group as measured by the percent variation explained on the liability scale. Among those of European ancestry, the MDD-PRS explained a significant proportion of variation in the liability of depression (0.6–1.8% across sites). While the MDD-PRS was also associated with depression in the African American and Hispanic participants from BioVU and BioMe, the predictive performance was markedly attenuated, explaining only 0.10–0.40% of variation in the liability of depression.

In comparison to the performance of the MDD-PRS, the ANX-PRS explained less variation in the liability of anxiety in European cohorts (0–0.9% across sites) and was not predictive of anxiety in BioMe ( $p = 0.24$ ). Similar to MDD-PRS, the performance of the ANX-PRS was reduced in the non-European cohorts and was only significant in the BioVU African American sample ( $R^2 = 0.19\%$ ;  $p = 0.014$ ). The full results are shown in online Supplementary Table S4.

### Joint PRS prediction of depression and anxiety comorbidity

Next, we tested whether both PRSs were associated with the comorbidity of depression and anxiety using a multinomial outcome specifying whether a participant had neither diagnosis ( $N = 132\,960$ ), depression-only ( $N = 16\,092$ ), anxiety-only ( $N = 13\,098$ ), or comorbid ( $N = 16\,584$ ). In the combined

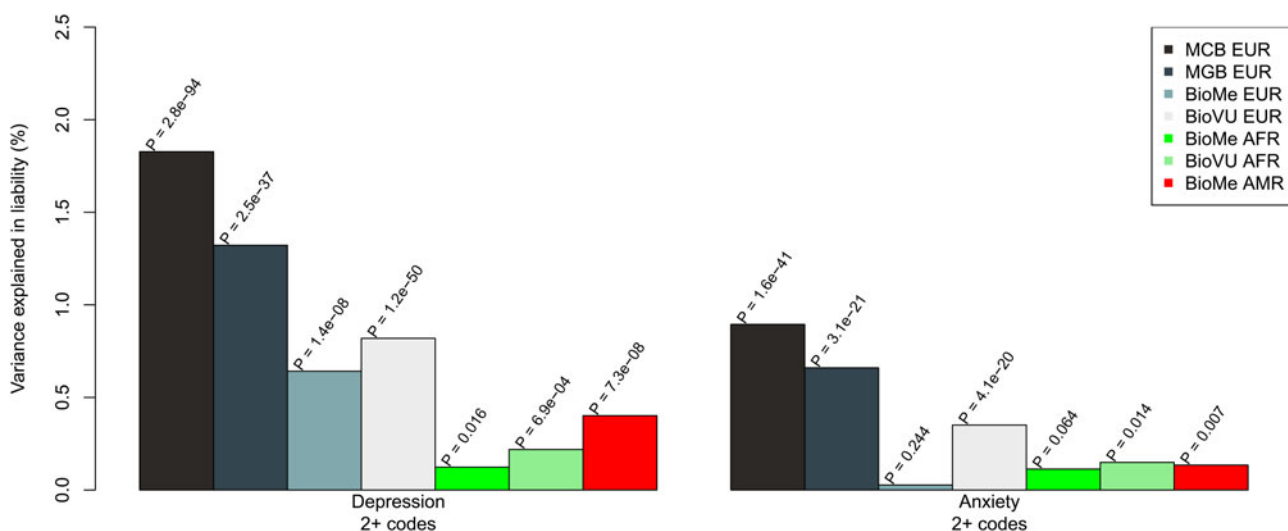


**Figure 1.** Distribution of depression (MDD) and anxiety (ANX) diagnoses at each site specified by at least two diagnosis codes from the EHR. Each site's sample size and ancestries (EUR, European; AFR, African/African American; AMR, Hispanic) are included in the top left of each site plot.

meta-analysis across participants of European ancestry at all sites (total  $N = 138\,632$ ), both MDD-PRS and ANX-PRS were significantly associated with each case subgroup compared to controls after adjusting for one another (Table 1). Notably, both the MDD-PRS (OR 1.26 per s.d. increase in PRS; 95% CI 1.24–1.29;  $p = 1 \times 10^{-105}$ ) and the ANX-PRS (OR 1.09; 95% CI 1.07–1.11;  $p = 3 \times 10^{-15}$ ) had the largest effect size in the comorbid group when compared to controls after adjusting for the other PRS. Furthermore, the MDD-PRS was significantly higher in the comorbid group than the depression-only group (OR 1.09; 95% CI 1.06–1.12;  $p = 7 \times 10^{-8}$ ) and the anxiety-only group (OR 1.15; 95% CI 1.11–1.19;  $p < 1 \times 10^{-16}$ ) and was significantly higher in the depression-only group than the anxiety-only group (OR 1.06; 95% CI 1.02–1.09;  $p = 0.001$ ), showing a gradient of genetic

risk across the isolated conditions and the comorbidity (online Supplementary Table S6).

In the meta-analysis of African American participants from BioMe and BioVU ( $N = 25\,612$ ), after adjusting for the contribution of the ANX-PRS, the MDD-PRS was associated only with the comorbid ( $N = 1399$ ; OR 1.13; 95% CI 1.04–1.23;  $p = 0.005$ ) group when compared to controls ( $N = 20\,810$ ), and not with depression-only ( $N = 2339$ ; OR 1.06; 95% CI 0.99–1.13;  $p = 0.12$ ) and ANX-only ( $N = 1064$ ; OR 1.01; 95% CI 0.93–1.10;  $p = 0.81$ ). Among the Hispanic participants from BioMe ( $N = 9034$ ), the MDD-PRS was only predictive of the comorbid group ( $N = 1343$ ; OR 2.12; 95% CI 1.61–2.80;  $p = 1 \times 10^{-7}$ ) compared to controls ( $N = 10\,037$ ). After adjusting for the contribution of MDD-PRS, the ANX-PRS was not predictive of any case



**Figure 2.** PRS prediction of depression and anxiety separately. Site- and ancestry-specific association of MDD-PRS and ANX-PRS with MDD and ANX, respectively, defined by having at least two ICD codes from the EHR. Performance is measured by variance explained by the PRS on the liability scale (assuming 20% population prevalence for both disorders).  $p$  values for each association are listed above each bar.

**Table 1.** Joint PRS prediction of depression and anxiety comorbidity from the multinomial model

Site-ancestry ( <i>N</i> control)	Group	<i>N</i> case	MDD-PRS OR (95% CI)	MDD-PRS <i>p</i> value	ANX-PRS OR (95% CI)	ANX-PRS <i>p</i> value
Meta-EUR ( <i>N</i> = 102 113)	Depression-only	12 201	1.16 (1.14, 1.19)	$3 \times 10^{-39}$	1.04 (1.01, 1.06)	0.001
	Anxiety-only	10 299	1.10 (1.07, 1.13)	$1 \times 10^{-14}$	1.07 (1.04, 1.09)	$3 \times 10^{-7}$
	Comorbid	14 019	1.26 (1.24, 1.29)	$1 \times 10^{-105}$	1.09 (1.07, 1.11)	$3 \times 10^{-15}$
Meta-AFR ( <i>N</i> = 20 810)	Depression-only	2339	1.06 (0.99, 1.13)	0.117	1.00 (0.93, 1.07)	0.98
	Anxiety-only	1064	1.01 (0.93, 1.10)	0.807	1.02 (0.94, 1.11)	0.64
	Comorbid	1399	1.13 (1.04, 1.23)	0.005	1.05 (0.97, 1.14)	0.25
BioMe-AMR ( <i>N</i> = 10 037)	Depression-only	1552	1.34 (0.92, 1.95)	0.125	0.68 (0.49, 0.93)	0.015
	Anxiety-only	689	1.36 (1.05, 1.77)	0.02	0.80 (0.65, 1.00)	0.052
	Comorbid	1343	2.12 (1.61, 2.80)	$1 \times 10^{-7}$	0.87 (0.69, 1.10)	0.26

EUR, European ancestry; AFR, African ancestry; AMR, Hispanic ancestry; OR, odds ratio associated with 1 s.d. increase in the PRS; PRS, polygenic risk score; MDD, major depressive disorder; ANX, anxiety-related disorders.

subgroup compared to controls among African American and Hispanic participants. The full results by ancestry and by site are presented in online Supplementary Tables S5 and 6.

## Discussion

Prior GWASs of depression and anxiety disorders typically have not considered the common comorbidity between depression and anxiety. Therefore, little is known about how genetic risk impacts the comorbidity. To study this, we performed a multi-site and multi-ancestry PRS analysis across four different EHR-linked biobanks using genetic risk of depression and anxiety. The MDD-PRS predicted diagnosis of depression in the EHR in all ancestries and at all sites. However, ANX-PRS was less predictive of anxiety diagnoses, with the ANX-PRS having a smaller effect size than MDD-PRS and showing no significant associations with anxiety disorders among any ancestries in BioMe. Importantly though, in our model of the comorbidity in European ancestry samples, the MDD-PRS showed a gradient of risk with respect to diagnosis of depression and anxiety such that patients with the comorbidity had the highest MDD-PRS followed by those with depression-only, anxiety-only, and neither diagnosis. Although not statistically significant, anxiety genetic risk showed a similar trend with the comorbid group having the highest genetic risk for anxiety followed by those with anxiety-only, depression-only, and neither. The observed gradient of genetic risk across the diagnostic groups, with the MDD-PRS being significantly higher for patients with depression-only than those with anxiety-only after adjusting for ANX-PRS, suggests that the additive genetic factors may partially differ between the two disorders.

We found that the comorbid group was at the highest level of genetic risk for depression and anxiety, which could indicate that genetic factors play a larger role in those with the comorbidity compared with only one diagnosed disorder. However, it is important to consider how our study design may also have impacted these results. In our study, depression and anxiety were defined with ICD codes that have only modest concordance with cases identified by structured or semi-structured clinical assessments used in research (Fiest et al., 2014). Thus, while our finding that the MDD + ANX group has the highest genetic risk of MDD could reflect true biological differences, it might

instead reflect the fact that the MDD + ANX comorbidity is more clinically conspicuous and therefore better captured in the EHR because it is associated with more severe depressive symptoms, higher levels of impairment, poorer clinical outcomes, and more suicidal ideation (Joffe et al., 1993; Nam et al., 2016; Pollack, 2005; Zhou et al., 2017). These patients may be more likely to be referred for specialty care resulting in a higher likelihood of a depression or anxiety diagnosis. If so, our PRS findings may still reflect biologically determined greater symptom severity in the MDD + ANX group. Consequently, repeating this study in a large sample with more rigorous diagnostic assessments is warranted.

A strength of this study is that a larger range of the phenotype can be easily and cost-effectively captured through the EHR. While non-EHR clinical studies may have more in-depth assessment of the phenotype of interest, they typically lack comprehensive assessment of important comorbidities and have smaller sample sizes. It is also important to acknowledge the limitations of EHR-based studies; most notably, reliance on ICD codes from the EHR can result in misclassification of cases and controls. A strategy to improve case/control ascertainment from the EHR is to use natural language processing (NLP) to incorporate clinical notes into the classification system (Ford, Carroll, Smith, Scott, & Cassell, 2016). Such an approach requires NLP algorithms that translate well across health systems, which is an active area of research. Heterogeneity of diagnosis prevalence between the different sites may have also contributed to differential power in the analyses. Given this, it is perhaps more striking that the PRSs demonstrated significant association despite heterogeneity of patient populations, diagnostic practices, and differences in EHR systems across sites, albeit with small amounts of variance explained for diagnosis of depression and anxiety. Notably, the MDD-PRS was less predictive of EHR diagnosis of MDD in the EHR (liability- $R^2$  ranging from 0.1% to 1.8%) than that seen in out-of-sample prediction in research cohorts (liability- $R^2$  ranging from 1.5% to 3.2%) (Howard et al., 2019) whereas the prediction performance of ANX-PRS for EHR diagnosis of anxiety was similar to that seen in the UK Biobank, ANGST, and iPSYCH (liability- $R^2$  ranging from 0.1% to 0.7%) (Purves et al., 2020).

Finally, this study of depression and anxiety was one of the first to include samples from ancestrally diverse groups. However, the sample sizes of African and Hispanic ancestry patients were much smaller than the European ancestry group,

resulting in lower power to detect associations with genetic risk. Furthermore, the constructed PRSs were based on cohorts with European ancestry, which are known to have poorer prediction in non-European ancestries (Martin et al., 2019). The African American and Hispanic patients also had higher rates of depression diagnoses which is contrary to what might be expected based on prior epidemiological (Breslau, Kendler, Su, Gaxiola-Aguilar, & Kessler, 2005) or clinical (Stockdale, Lagomasino, Siddique, McGuire, & Miranda, 2008) research. Moreover, diagnostic biases (Gara et al., 2012) and decreased healthcare access for people from marginalized communities (Bailey, Mokonogho, & Kumar, 2019) can contribute to the misclassification in these groups and thus further reduce power.

In our study, genetic risks of depression and anxiety contributed jointly to depression and anxiety, but each PRS showed stronger association with the corresponding diagnosis. Our results suggest that depression and anxiety may have partially independent genetic liabilities and the genetic vulnerabilities to depression and anxiety make distinct contributions to comorbid depression and anxiety. It is also important to note that the prior GWASs of MDD and anxiety from which the PRSs were computed did not exclude individuals with comorbid anxiety and depression and thus, it is expected that both the depression and anxiety PRSs also captured the shared liability limiting the unique contributions from these PRS to the comorbid condition. Future large separate GWASs of depression, anxiety, and their comorbidity are needed to fully explore the extent to which anxiety and depression have distinct liabilities. Nevertheless, even without these GWAS to create comorbid-specific PRS, we found that the MDD-PRS and ANX-PRS were independently associated with the comorbidity of depression and anxiety in the EHR, supporting the hypothesis that the correlated disorders represent partially distinct nosological entities.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291723000983>.

**Acknowledgements.** This study was supported by the National Institute of Mental Health (NIMH) grants: R01MH121924, R01MH121923, R01MH121922, R01MH121921, and R01MH118233. We acknowledge the Mayo Clinic Biobank (MCB) research team, as well as the patient-participants who consented to participate in this research program. We also acknowledge Regeneron Genetics Center for providing genetic data for MCB participants for the analysis. The Mayo Clinic Biobank and Project Generation were supported in part by Mayo Clinic's Center for Individualized Medicine. K. W. C. was supported in part by funding from the NIMH (K08MH127413) and a NARSAD Brain and Behavior Foundation Young Investigator Award. Y. N. G. was supported in part by a Moynihan Clinical Research Fellowship from the Leon Levy Foundation and Award Number R25MH086466 from the NIMH. L. K. D. was supported by R01MH118233 and R56MH120736. Vanderbilt EHR resources were supported by the National Center for Research Resources, Grant UL1 RR024975-01, and are now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Vanderbilt University Medical Center's BioVU is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH funded Shared Instrumentation Grant S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at <https://vict.vumc.org/biovu-funding/>. K. S. is funded by the American Heart Association Predoctoral Fellowship AHA827137.

**Competing interests.** None.

## References

- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association Publishing.
- Bailey, R. K., Mokonogho, J., & Kumar, A. (2019). Racial and ethnic differences in depression: Current perspectives. *Neuropsychiatric Disease and Treatment*, 15, 603–609.
- Belbin, G. M., Odgis, J., Sorokin, E. P., Yee, M.-C., Kohli, S., Glicksberg, B. S., ... Kenny, E. E. (2017). Genetic identification of a common collagen disease in Puerto Ricans via identity-by-descent mapping in a health system. *eLife*, 6. <https://doi.org/10.7554/eLife.25060>.
- Bielinski, S. J., Chai, H. S., Pathak, J., Talwalkar, J. A., Limburg, P. J., Gullerud, R. E., ... Chute, C. G. (2011). Mayo Genome Consortia: A genotype-phenotype resource for genome-wide association studies with an application to the analysis of circulating bilirubin levels. *Mayo Clinic Proceedings*. *Mayo Clinic*, 86(7), 606–614.
- Blanco, C., Krueger, R. F., Hasin, D. S., Liu, S.-M., Wang, S., Kerridge, B. T., ... Olfson, M. (2013). Mapping common psychiatric disorders. *JAMA Psychiatry* 70(2), 199. <https://doi.org/10.1001/jamapsychiatry.2013.281>
- Blanco, C., Wall, M. M., Feng, T., & Olfson, M. (2021). Evaluating the modified common liability hypothesis of psychiatric comorbidity. *Journal of Psychiatric Research*, 141, 9–15.
- Breslau, J., Kendler, K. S., Su, M., Gaxiola-Aguilar, S., & Kessler, R. C. (2005). Lifetime risk and persistence of psychiatric disorders across ethnic groups in the United States. *Psychological Medicine*, 35(3), 317–327.
- Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*, 8(7). <https://doi.org/10.1093/gigascience/giz082>
- Coombes, B. J., Ploner, A., Bergen, S. E., & Biernacka, J. M. (2020). A principal component approach to improve association testing with polygenic risk scores. *Genetic Epidemiology*, 44(7), 676–686.
- Craske, M. G. (2012). Transdiagnostic treatment for anxiety and depression. *Depression and Anxiety*, 29(9), 749–753.
- de Graaf, R., Bijl, R. V., Smit, F., Vollebergh, W. A. M., & Spijker, J. (2002). Risk factors for 12-month comorbidity of mood, anxiety, and substance use disorders: Findings from the Netherlands Mental Health Survey and Incidence Study. *The American Journal of Psychiatry*, 159(4), 620–629.
- Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., ... Crawford, D. C. (2010). PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics (Oxford, England)*, 26(9), 1205–1210.
- Dillon, D. G., Rosso, I. M., Pechtel, P., Killgore, W. D. S., Rauch, S. L., & Pizzagalli, D. A. (2014). Peril and pleasure: An RDoC-inspired examination of threat responses and reward processing in anxiety and depression. *Depression and Anxiety*, 31(3), 233–249.
- Fiest, K. M., Jette, N., Quan, H., St Germaine-Smith, C., Metcalfe, A., Patten, S. B., & Beck, C. A. (2014). Systematic review and assessment of validated case definitions for depression in administrative data. *BMC Psychiatry*, 14, 289.
- Ford, E., Carroll, J. A., Smith, H. E., Scott, D., & Cassell, J. A. (2016). Extracting information from the text of electronic medical records to improve case detection: A systematic review. *Journal of the American Medical Informatics Association*, 23(5), 1007–1015. <https://doi.org/10.1093/jamia/ocv180>
- Gara, M. A., Vega, W. A., Arndt, S., Escamilla, M., Fleck, D. E., Lawson, W. B., ... Strakowski, S. M. (2012). Influence of patient race and ethnicity on clinical assessment in patients with affective disorders. *Archives of General Psychiatry*, 69(6), 593–600.
- Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, 10(1), 1776.
- Giannakopoulou, O., Lin, K., Meng, X., Su, M.-H., Kuo, P.-H., Peterson, R.E., ... Major Depressive Disorder Working Group of the Psychiatric Genomics

- Consortium (2021). The genetic architecture of depression in individuals of East Asian ancestry: A genome-wide association study. *JAMA Psychiatry*, 78 (11), 1258–1269.
- Global Burden of Disease Study 2013 Collaborators (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, 386(9995), 743–800.
- Greenebaum, S. L. A., & Nierenberg, A. A. (2020). More on the NIMH RDoC initiative. *Bipolar Disorders*, 22(1), 11–12. <https://doi.org/10.1111/bdi.12874>
- He, M., Santiago Ortiz, A. J., Marshall, J., Mendelsohn, A. B., Curtis, J. R., Barr, C. E., ... Kim, S. C. (2019). PB140 ICD-9 to ICD-10 mapping for research in biogenetics and biosimilarity using administrative healthcare data. *Value in Health* 22, S55. <https://doi.org/10.1016/j.jval.2019.04.122>
- Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., ... McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*, 22(3), 343–352.
- Joffe, R. T., Bagby, R. M., & Levitt, A. (1993). Anxious and nonanxious depression. *The American Journal of Psychiatry*, 150(8), 1257–1258.
- Karlson, E., Boutin, N., Hoffnagle, A., & Allen, N. (2016). Building the partners HealthCare biobank at partners personalized medicine: Informed consent, return of research results, recruitment lessons and operational considerations. *Journal of Personalized Medicine*, 6(1), 2. <https://doi.org/10.3390/jpm6010002>
- Kessler, R. C., Gruber, M., Hettema, J. M., Hwang, I., Sampson, N., & Yonkers, K. A. (2008). Co-morbid major depression and generalized anxiety disorders in the National Comorbidity Survey follow-up. *Psychological Medicine*, 38(3), 365–374.
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., ... Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477.
- Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, 56(10), 921. <https://doi.org/10.1001/archpsyc.56.10.921>
- Lee, Y. H., Thaweethai, T., Sheu, Y.-H., Feng, Y.-C. A., Karlson, E. W., Ge, T., ... Smoller, J. W. (2022). Impact of selection bias on Polygenic Risk Score estimates in healthcare settings. <https://doi.org/10.1101/2022.07.20.22277710>.
- Levey, D. F., Gelernter, J., Polimanti, R., Zhou, H., Cheng, Z., Aslan, M., ... the Million Veteran Program. (2020). Reproducible genetic risk loci for anxiety: Results from ~200,000 participants in the Million Veteran Program. *American Journal of Psychiatry*, 177(3), 223–232. <https://doi.org/10.1176/appi.ajp.2019.19030256>
- Levey, D. F., Stein, M. B., Wendt, F. R., Pathak, G. A., Zhou, H., Aslan, M., ... Gelernter, J. (2021). Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nature Neuroscience*, 24(7), 954–963.
- Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, 51(4), 584–591.
- Mei, L., Gao, Y., Chen, M., Zhang, X., Yue, W., Zhang, D., & Yu, H. (2022). Overlapping common genetic architecture between major depressive disorders and anxiety and stress-related disorders. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 113, 110450.
- Moscati, A., Flint, J., & Kendler, K. S. (2016). Classification of anxiety disorders comorbid with major depression: Common or distinct influences on risk? *Depression and Anxiety*, 33(2), 120–127.
- Nam, Y.-Y., Kim, C.-H., & Roh, D. (2016). Comorbid panic disorder as an independent risk factor for suicide attempts in depressed outpatients. *Comprehensive Psychiatry*, 67, 13–18.
- Nutt, D. (2004). Anxiety and depression: Individual entities or two sides of the same coin? *International Journal of Psychiatry in Clinical Practice*, 8(Suppl 1), 19–24.
- Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, 47(7), 702–709.
- Pollack, M. H. (2005). Comorbid anxiety and depression. *The Journal of Clinical Psychiatry*, 66(Suppl 8), 22–29.
- Privé, F., Arbel, J., & Vilhjálmsson, B. J. (2020). LDpred2: Better, faster, stronger. *Bioinformatics*, 36(22–23), 5424–5431.
- Purves, K. L., Coleman, J. R. I., Meier, S. M., Rayner, C., Davis, K. A. S., Cheesman, R., ... Eley, T. C. (2020). A major role for common genetic variation in anxiety disorders. *Molecular Psychiatry*, 25(12), 3292–3303.
- Ressler, K. J., & Nemeroff, C. B. (2000). Role of serotonergic and noradrenergic systems in the pathophysiology of depression and anxiety disorders. *Depression and Anxiety*, 12(Suppl 1), 2–19.
- Roden, D. M., Pulley, J. M., Basford, M. A., Bernard, G. R., Clayton, E. W., Balsler, J. R., & Masys, D. R. (2008). Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clinical Pharmacology and Therapeutics*, 84(3), 362–369.
- Stockdale, S. E., Lagomasino, I. T., Siddique, J., McGuire, T., & Miranda, J. (2008). Racial and ethnic disparities in detection and treatment of depression and anxiety among psychiatric and primary health care visits, 1995–2005. *Medical Care*, 46(7), 668–677.
- Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry*, 157, (10), 1552–1562. <https://doi.org/10.1176/appi.ajp.157.10.1552>
- Takahashi, Y., Yamagata, S., Ritchie, S. J., Barker, E. D., & Ando, J. (2021). Etiological pathways of depressive and anxiety symptoms linked to personality traits: A genetically-informative longitudinal study. *Journal of Affective Disorders*, 291, 261–269.
- Thorpe, J. G., Campos, A. I., Grotzinger, A. D., Gerring, Z. F., An, J., Ong, J.-S., ... Derks, E. M. (2021). Symptom-level modelling unravels the shared genetic architecture of anxiety and depression. *Nature Human Behaviour*, 5(10), 1432–1442.
- Wei, W.-Q., Bastarache, L. A., Carroll, R. J., Marlo, J. E., Osterman, T. J., Gamazon, E. R., ... Denny, J. C. (2017). Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS ONE*, 12(7), e0175508.
- WHO, & World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. World Health Organization.
- Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., ... Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, 50(5), 668–681.
- Zheutlin, A. B., Dennis, J., Karlsson Linnér, R., Moscati, A., Restrepo, N., Straub, P., ... Smoller, J. W. (2019). Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 106,160 patients across four health care systems. *The American Journal of Psychiatry*, 176(10), 846–855.
- Zhou, Y., Cao, Z., Yang, M., Xi, X., Guo, Y., Fang, M., ... Du, Y. (2017). Comorbid generalized anxiety disorder and its association with quality of life in patients with major depressive disorder. *Scientific Reports*, 7(1). <https://doi.org/10.1038/srep40511>
- Zorn, J. V., Schür, R. R., Boks, M. P., Kahn, R. S., Joëls, M., & Vinkers, C. H. (2017). Cortisol stress reactivity across psychiatric disorders: A systematic review and meta-analysis. *Psychoneuroendocrinology*, 77, 25–36.