

1 **Phenotype Risk Scores: moving beyond ‘cases’ and ‘controls’ to classify psychiatric disease**  
2 **in hospital-based biobanks.**

3 Dannielle S. Lebovitch<sup>1-4</sup>, Jessica S. Johnson<sup>1-4</sup>, Hillary R. Dueñas<sup>1</sup>, Laura M. Huckins<sup>1-6\*</sup>

4

5 **Affiliations:**

- 6 1. Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount  
7 Sinai, New York, NY 10029, USA
- 8 2. Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY  
9 10029, USA
- 10 3. Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New  
11 York, NY 10029, USA
- 12 4. Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount  
13 Sinai, New York, NY 10029, USA
- 14 5. Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount  
15 Sinai, New York, NY 10029, USA
- 16 6. Mental Illness Research, Education and Clinical Centers, James J. Peters Department of  
17 Veterans Affairs Medical Center, Bronx, NY 10468, USA

18

19 **\*Corresponding author. Correspondence should be addressed to**

20 [laura.huckins@mssm.edu](mailto:laura.huckins@mssm.edu); 212-659-1613; Icahn School of Medicine at Mount Sinai, 1

21 **Gustave Levy Place, NYC, NY, 10029.**

22 **Abstract**

23 Current phenotype classifiers for large biobanks with coupled electronic health records EHR and  
24 multi-omic data rely on ICD-10 codes for definition. However, ICD-10 codes are primarily  
25 designed for billing purposes, and may be insufficient for research. Nuanced phenotypes  
26 composed of a patients' experience in the EHR will allow us to create precision psychiatry to  
27 predict disease risk, severity, and trajectories in EHR and clinical populations. Here, we create a  
28 phenotype risk score (PheRS) for major depressive disorder (MDD) using 2,086 cases and  
29 31,000 individuals from Mount Sinai's biobank BioMe™. Rather than classifying individuals as  
30 'cases' and 'controls', PheRS provide a whole-phenome estimate of each individual's likelihood  
31 of having a given complex trait. These quantitative scores substantially increase power in EHR  
32 analyses and may identify individuals with likely 'missing' diagnoses (for example, those with  
33 large numbers of comorbid diagnoses and risk factors, but who lack explicit MDD diagnoses).

34 Our approach applied ten-fold cross validation and elastic net regression to select comorbid ICD-  
35 10 codes for inclusion in our PheRS. We identified 158 ICD-10 codes significantly associated  
36 with Moderate MDD (F33.1). Phenotype Risk Score were significantly higher among individuals  
37 with ICD-10 MDD diagnoses compared to the rest of the population (Kolgorov-Smirnov  $p < 2.2e-$   
38  $16$ ), and were significantly correlated with MDD polygenic risk scores ( $R^2 > 0.182$ ). Accurate  
39 classifiers are imperative for identification of genetic associations with psychiatric disease;  
40 therefore, moving forward research should focus on algorithms that can better encompass a  
41 patient's phenome.

42

43

## 44 **Introduction**

45 Substantial advances have been made in understanding the genetic architecture of psychiatric  
46 disease through the collection of very large case-control samples. Maximizing homogeneity  
47 within these samples (for example, applying stringent inclusion/exclusion criteria to define cases  
48 and controls), and minimizing variability in analytical design and diagnostic definitions, has led  
49 to the discovery of hundreds of genomic loci significantly associated with psychiatric disease.  
50 For example, the Psychiatric Genomics Consortium (PGC) has collated > 200,000 samples,  
51 across 11 psychiatric disorders, to identify 500 risk loci associated with these disorders<sup>1</sup>. The  
52 addition of biobank data to these analyses has substantially increased sample size, and identified  
53 yet more disease-associated variants; however, these findings have yet to yield meaningful  
54 insights into patient outcomes. To date, we lack the ability to stratify patients in clinically useful  
55 ways (for example, according to severity or chronicity), or to pinpoint appropriate treatments, or  
56 expected treatment outcomes, for each patient.

57  
58 Arguably, as a field we must now enter the ‘translational’ stage of our research. Rather than  
59 exclusively focusing on assembling new, large cohorts, we must now consider the presentation  
60 of psychiatric disease within our clinics; and, by extension, within our medical records.  
61 Understanding the role of genetic associations within the clinical population may better elucidate  
62 the genetic architecture and biological mechanisms of disease; more importantly, identifying  
63 variants that are associated with disease outcomes or trajectories within our clinics may lead to  
64 more readily interpretable or actionable findings.

65

66 Extending our genetic analyses to incorporate EHR and biobank-based data will require careful  
67 consideration of phenotype definition. The binary concepts of ‘cases’ and ‘controls’ may become  
68 inappropriate when considering lifelong, common, episodic psychiatric traits, or traits that are  
69 characterized by a constellation of symptoms and behaviours. By contrast, hospital-based  
70 biobanks offer the opportunity to study the whole phenome over the lifespan of a patient<sup>2-5</sup>.  
71 Rather than pursuing single phenotype associations assessed at single time points, we can include  
72 trajectories of disease, comorbidities, and behaviours.

73  
74 Here, we describe the creation of a ‘phenotype risk score’ for psychiatric traits, adapting the  
75 approach outlined by Bastarache et al<sup>6</sup>. Construction and use of phenotype risk scores (PheRS)  
76 may increase power in EHR studies, and provide whole-phenome perspectives of disease.  
77 Initially, this approach allowed symptom-based scoring of children with syndromic and  
78 mendelian traits in order to identify overlap and subtypes among disorders<sup>6</sup>. We extend this  
79 approach to psychiatric traits, and demonstrate application to major depressive disorder (MDD),  
80 a serious psychiatric disorder presenting with heterogeneous constellations of symptoms and  
81 comorbidities<sup>7</sup>. In particular, MDD is commonly comorbid with secondary psychiatric disorders  
82 (eg, substance use disorder<sup>8</sup>, generalized anxiety disorder<sup>9</sup>), as well as physical symptoms and  
83 traits (for example, cardiovascular disease<sup>10</sup>). In order to create our MDD-PheRS, we identify  
84 and quantify the comorbid traits and disorders commonly presenting with MDD; we term this the  
85 ‘comorbid phenome’, or *co-phenome*. MDD-PheRS scores are then calculated for each  
86 individual in our EHR as a weighted sum of co-phenome codes. Individuals with high scores  
87 have many of the traits and disorders that have been found to be commonly comorbid with

88 MDD. Individuals with low scores display none, or almost none, of the traits and symptoms  
89 commonly comorbid with MDD.

90

91 We propose that our phenotype risk score may be particularly useful in instances where patients  
92 have restricted access to healthcare, or face financial or social barriers in seeking out or  
93 accessing mental health treatment<sup>11</sup>. Whereas assessing these individuals based purely on ICD  
94 diagnoses may misclassify them as ‘controls’; a whole-phenome perspective may instead reveal  
95 that these patients exhibit very high levels of comorbid traits, or even meet some diagnostic  
96 criteria of the trait in question. Genetically, and biologically, we might therefore expect these  
97 individuals to fall closer to a ‘case’ definition; wrongly assigning them as ‘controls’ would  
98 significantly reduce power in our studies<sup>12</sup>.

99

100 Whereas GWAS have historically focused on European populations, to the detriment of cross-  
101 ancestry insights and translatability<sup>13-16</sup>, hospital-based biobanks provide a more representative  
102 overview of the local population. EHR-based analyses therefore provide an opportunity to  
103 advance insights into disease aetiology across populations. In particular, our Mount Sinai  
104 BioMe<sup>TM</sup> biobank participants<sup>16-18</sup> are deliberately sampled to include equal proportions of  
105 patients with African American, European American, and Hispanic American ancestries. This  
106 provides us with a unique opportunity to ensure that our algorithms are valid, and useful, for our  
107 entire patient population, without discrimination by ancestry. Although we do not expect to find  
108 substantial biological differences in disease aetiology between groups, we do expect differences  
109 in experience, environmental exposures, and clinician diagnostic bias to affect diagnoses. Indeed,

110 EHR- and epidemiological analyses identify multiple racial disparities in diagnosis, presentation,  
111 and treatment<sup>19-22</sup> of MDD<sup>19-29</sup>.

112

113

114

115

116 **Methods**

117 **Participants**

118 Participants were included from Mount Sinai’s BioMe™ biobank. BioMe™ participants agree to  
119 the sharing of clinical electronic health record (EHR) and genotype data. The Charles Bronfman  
120 Institute for Personalized Medicine at Mount Sinai Medical Center (MSMC) BioMe™  
121 BioBank<sup>18</sup> is an EMR-linked biorepository drawing from more than 70,000 inpatients and  
122 800,000 outpatients annually. The MSMC serves the diverse local communities of upper  
123 Manhattan, including Central Harlem (86% African American), East Harlem (88%  
124 Hispanic/Latino), and the Upper East Side (88% European American) with broad health  
125 disparities across these groups. BioMe™ enrolled over 50,000 participants between September  
126 2007 and January 2020, with proportionally 21% African American, 34% Hispanic/Latino, and  
127 31% European American individuals. The BioMe™ population reflects community-level disease  
128 burden and health disparities with broad public health impact. Biobank operations are fully  
129 integrated in clinical care processes, with direct recruitment from clinical sites, waiting areas,  
130 and phlebotomy stations by dedicated Biobank recruiters, independent of clinical care providers,  
131 prior to or following a clinician standard of care visit. Recruitment currently occurs at over 30  
132 clinical care sites.

133

134 Electronic health records for individuals in BioMe™ include International Classification of  
135 disease (ICD) 9 and 10 codes, with an average of 31.8 ICD-10 diagnoses per individual, an  
136 average EHR-length of 2,942 days, and an average of 81 independent visits<sup>30</sup>. There are 2,086  
137 individuals with an ICD-10 diagnosis of F33 MDD (**Table 1**), defined as “Major Depressive

138 Disorder, recurrent”, of which the largest single subgroup is “MDD, recurrent, moderate”  
139 (ICD10 code F33.1, 1, N=156; **Table 1**).

140

#### 141 **Identifying the MDD co-Phenome**

142 The MDD co-phenome represents the typical diagnostic history of an individual with MDD in  
143 our healthcare system. In order to maximize power, we focused here on the most common MDD  
144 ICD-10 code (F33.1, “Moderate Depression”). To calculate the MDD co-phenome, we first  
145 removed rare ICD-10 codes (codes occurring in <5 individuals; 8,565 out of 15,495 codes  
146 removed). Next, we identified all codes significantly comorbid with MDD F33.1 using Fisher's  
147 exact test.

148

#### 149 **Creating a phenotype risk score (PheRS)**

150 ICD-10 codes employ a hierarchical structure to describe processes, states, traits and disorders.  
151 Given the high degree of similarity between codes, the longitudinal nature of electronic health  
152 records, and the difficult, iterative process of arriving at psychiatric diagnoses, many ICD codes  
153 may be co-linear or redundant. Failure to correct for this overlap may result in significant skew  
154 in our phenotype risk score estimates.

155

156 In order to adjust for this potential redundancies and co-linearity between codes, we adopted a  
157 penalized regression approach (elastic net<sup>31-33</sup>). Elastic Net Regression (ENR) is a combination  
158 of two regularization regression methods, least absolute shrinkage and selection operator  
159 (LASSO) and Ridge Regression. LASSO has a shrinkage function to reduce collinearity and  
160 select parameters, while Ridge Regression maintains the equation from becoming underfitted.

161 We employed a ten-fold cross-validation approach to avoid overfitting, and used the “cvglmnet”  
 162 function included in the glmnet 2.0-16 package in R version 3.5.3<sup>31</sup>.

163 We calculate our MDD-phenotype risk score (MDD-PheRS) as a weighted sum of MDD-co-  
 164 phenome ICD codes using two methods:

- 165 1. PheRS<sub>m1</sub>: Weights are calculated as the log-inverse prevalence of each associated ICD-  
 166 10 diagnosis code with MDD in BioMe™ (**Box 1**), analogous to Bastarache et al<sup>6</sup>.
- 167 2. PheRS<sub>m2</sub>: Weights are defined as effect sizes (betas) obtained from the ten-fold cross  
 168 validation elastic net model (**Box 1**).

169

### 170 **Box 1: PheRS Calculation**

	$PheRS_{m1}$	$PheRS_{m2}$
Base Formula	$PheRS = \sum_{i=1}^{N_p} w_p x_{i,p}$ $x_{i,p} = \begin{cases} 1 & \text{if individual } i \text{ has phenotype } p \\ 0 & \text{otherwise} \end{cases}$ $N_p = \text{number of comorbid ICD codes}$	
Weights Calculation:	$w_p = \log \frac{N}{n_p}$ $N = \text{population size}$ $n_p = \text{number of patients with ICD code}$	$w_p = \beta_{elastic\ net}$

171

172 We compared PheRS distributions between (i) individuals with- and without- MDD F33.1 codes  
 173 and (ii) individuals with- and without- **any** MDD diagnostic codes using kolgorov-smirnov tests  
 174 for significance.

175

176

## 177 **Comparison to MDD Polygenic Risk Scores**

178 We calculated MDD-polygenic risk scores (MDD-PRS) for all individuals in our biobank, using  
179 the most recent PGC summary statistics<sup>34</sup> (142,000+ individuals; **45,396 cases and 97,250**  
180 **controls**). PGC cases were identified using self-report, clinician interview and/or medical record  
181 diagnoses. We used PRSice-2<sup>35,36</sup> to calculate MDD-PRS for all BioMe<sup>TM</sup> participants at ten  
182 standard p-value thresholds:  $1 \times 10^{-4}$ ,  $1 \times 10^{-3}$ , 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1.

183

184 We partitioned individuals into MDD-PRS deciles, and tested for case enrichment in the top  
185 decile compared to (i) the bottom decile and (ii) all other deciles. In particular, we tested for  
186 enrichment in the top decile of (1) MDD cases/control status as defined by ICD code F33.1; and  
187 fold change of (2) PheRSm1 and (3) PheRSm2.

188

189 PRS analyses are particularly vulnerable to ancestry differences between ‘base’ and ‘target’  
190 GWAS<sup>14,37-40</sup>; i.e., the ancestry of individuals from which the PRS is derived must closely match  
191 that to which the PRS is applied. To date, no large-scale MDD-GWAS have been performed in  
192 individuals of African-American or Hispanic descent. Therefore, we restricted this section of our  
193 analysis to individuals of European descent.

194

## 195 **Cross-ancestry specificity of phenotype risk scores**

196 Our MDD co-phenome assesses the whole phenome of each individual with MDD in our  
197 healthcare system. These may reflect biological similarities or comorbidities between disorders,  
198 pleiotropy of genetic risk factors, diagnostic bias or mis-diagnosis, or simply patterns of  
199 healthcare usage that reflect, but are not directly related to, an underlying health condition<sup>11</sup>.

200 Modeling the whole EHR without accounting for potential differences between ancestries risks  
201 masking crucial differences that may affect the score, and ultimately may produce a less useful  
202 algorithm for understanding disease risk and clinical outcomes.

203  
204 In short, we expect that a person's race likely significantly shapes their access to and experience  
205 of healthcare systems<sup>11</sup>. Therefore, we repeated our co-phenome analysis, partitioning  
206 individuals by self-defined race/ethnicity. We included racial and ethnic groups with >1,000  
207 individuals, and >100 MDD-F33.1 diagnosis counts (**S. Table 1**; categories with sufficient data  
208 for analysis are shown in **bold**).

209  
210 For each group with sufficient individuals and cases, we repeated our co-phenome analysis using  
211 the same elastic net regression approach. In instances where co-phenomes differed significantly,  
212 we repeated construction of our MDD-PheRS. We compared distributions of MDD-PheRS using  
213 overall- and group-specific- co-phenome input and assessed significance using Kolgorov-  
214 smirnov tests.

215  
216 Next, we constructed three additional PheRS, specific to each large ancestry group, and  
217 compared predictive accuracy within and across ancestries, and to an approach grouping across  
218 all ancestries (ie, our original PheRS). We first derived PheRS effect sizes using a ten-fold cross-  
219 validation approach in each ancestry group separately (the "base" population). Next, we applied  
220 these PheRS, and our overall PheRS, to each of the three ancestry groups in turn (the "target"  
221 populations), and tested F33.1 prediction accuracy across all base-target combinations.

## 222 **Results**

### 223 **Characterizing the MDD co-phenome**

224 We define the ‘co-phenome’ as the set of diagnoses and traits that are comorbid with depression  
225 within our healthcare system. First, we identified all depression-related codes in our biobank  
226 (**Table 1**). To maximize power, we selected the most frequent code (F33.1, Moderate  
227 Depression) to represent ‘depression’ within our EHR. We identified 158 ICD-10 codes that  
228 were significantly comorbid with F33.1-MDD (**Table 2**), including many well-studied MDD risk  
229 factors (e.g., childhood sexual abuse<sup>41</sup>, homelessness<sup>42</sup>), comorbidities (e.g., BMI<sup>43</sup>, respiratory  
230 distress<sup>44</sup>) and clinical features (e.g., suicidal ideation<sup>45</sup>, opioid abuse<sup>46</sup>), as well as some less  
231 expected associations (e.g., guillain-barre syndrome, rheumatic heart failure).

232

### 233 **Creation of an MDD-Phenotype Risk Score**

234 Phenotype risk scores (PheRS) quantify the amount of MDD co-phenome codes present in each  
235 individuals’ diagnostic history. Higher PheRS scores imply higher whole-phenome similarity to  
236 MDD cases. In order to account for the significant diagnostic and biological overlap between  
237 ICD codes, the hierarchical, clustered nature of ICD-10 codes, the imprecision inherent in ICD-  
238 10 code usage, the longitudinal nature of EHR, and the difficult, iterative process of arriving at  
239 psychiatric diagnoses, we applied a penalized regression framework (elastic net regression) to  
240 automatically select codes for inclusion in our PheRS (**Figure 1**) using a ten-fold cross-  
241 validation framework.

242

243 Our model includes 158 ICD codes (**Table 3**), and significantly predicts F33.1 MDD in a ten-  
244 fold cross-validation framework ( $R^2 = 0.367$ ). The largest contributing effect size of all 158

245 codes identified is respiratory failure (J96.91, OR= 7.41), followed by personal history of  
246 physical or sexual abuse in childhood (Z62.810, OR= 6.03). Of the 158 codes identified to be  
247 predictive of F33.1 MDD, 30 are related to psychiatric disorders and traits; of these, the largest  
248 effect size is from PTSD (ICD-10 code F3.10, OR= 4.40). PheRS scores for individuals with (i)  
249 F33.1-MDD and (ii) broadly defined MDD (F32, F33) are significantly higher than those for  
250 patients without MDD diagnoses (kolgorov-smirnov test,  $p < 2.2 \times 10^{-16}$ ; **Figure 2**).

251

### 252 **MDD-PheRS correlate with polygenic risk scores**

253 In order to test whether individuals with higher MDD phenotype risk scores also have higher  
254 genetic risk for MDD, we calculated MDD-PRS<sup>34</sup> for all 10,792 European BioMe<sup>TM</sup>  
255 individuals<sup>18</sup> using PRSice-2<sup>35,36</sup>. Individuals in the top decile of MDD-PRS had significantly  
256 higher MDD-PheRS scores compared to individuals in the bottom decile (fold change=3.57, 2.46  
257 for PheRSm1, PheRSm2, respectively; **Figure 3**). This enrichment persists even when removing  
258 individuals with any MDD diagnostic codes (ie, F32, F33) from our analysis (fold change=3.58,  
259 3.21, respectively; **Figure 3**). Next, we compared the amount of phenotypic variance explained  
260 by MDD-PRS for both our phenotype risk scores, and MDD F33.1 case-control status. We find  
261 substantially more variance explained by MDD-PRS for both phenotype risk scores ( $R^2=0.217$ ,  
262  $0.182$ , respectively) than for F33.1 case-control status ( $R^2=0.051$ ).

263

### 264 **PheRS vary across ancestries**

265 Although we do not expect substantial biological aetiological differences between racial/ethnic  
266 groups, it is likely that race and ethnicity plays a significant role in shaping our experiences of  
267 the healthcare system<sup>11</sup>. As such, we hypothesized that co-phenome profiles may vary when

268 partitioning patients according to race. In **Figure 4i-iii**, we considered three potential scenarios.  
269 First- if bias plays only a small role in diagnoses, i.e, if the majority of diagnoses are assigned  
270 accurately, we expect highly overlapping comorbidity profiles and risk-factors for MDD between  
271 patient groups (**Figure 4i**). However, if bias affects diagnostic accuracy in one group compared  
272 to the others, we might expect less overlap in co-phenomes between groups (**Figure 4ii**);  
273 alternatively, bias may lead to differentially missing data between groups (for example, bias may  
274 lead to fewer lab tests or follow-up appointments; **Figure 4iii**).

275  
276 To test these hypotheses, we partitioned patients into three groups, according to self-reported  
277 race/ethnicity: European (N=10,792), African-American (N=8,046), Hispanic-American  
278 (N=11,047). We identified significantly different co-phenome profiles specifically among  
279 Hispanic-American patients within our system (**S. Table 2**). In particular, Hispanic-American  
280 individuals with MDD had very large numbers of co-morbid ICD codes, many of which (608)  
281 were only identified among this population. In addition, 495 codes were significant among  
282 African-American and Hispanic-American individuals with MDD, but were not co-morbid with  
283 MDD among European-Americans. By contrast, almost all comorbid codes among European-  
284 Americans were shared with both other groups (**Figure 4iv**).

285  
286 Next, we constructed three additional PheRS, specific to each large ancestry group, and  
287 compared predictive accuracy of PheRS within and across ancestries (ie, our original PheRS),  
288 resulting in nine PheRS scores. For each score, the ‘base’ dataset describes the ancestry group  
289 used to derive the score, and the ‘target’ describes the dataset to which the score is applied; e.g.,  
290 PheRS<sub>AA-HA</sub> denotes a score developed in African-American patients and applied in Hispanic-

291 American patients. We found significant correlation in PheRS scores across all base-target  
292 combinations (spearman rho > 0.58), with highest correlations for AA/HA base/target  
293 combinations (spearman rho > 0.79). We did not observe any effect of base population on  
294 predictive accuracy when applied to the full cohort ( $R^2$  range 0.196-0.211; **S. Table 3**); however,  
295 we do observe differences when applying across specific groups (eg: base-EA target-AA).  
296 As expected, we observe the highest predictive accuracy when base and target populations are  
297 matched (**S. Table 3**); followed by AA-HA base/target combinations. Scores derived in solely  
298 EA individuals performed noticeably worse in AA/HA populations, and vice-versa.

299

300

301

## 302 **Discussion**

303           Large-scale, collaborative approaches to amass large numbers of cases and controls have  
304 yielded great insights into complex trait architecture. In particular, concerted efforts across the  
305 psychiatric genetics field has resulted in >500 genome-wide significant associated loci across 15  
306 psychiatric disorders and traits. However, such case/control definitions match neither the lived  
307 experiences of patients, nor of clinicians. In particular, these study designs cannot account for  
308 comorbidities or whole phenome perspectives, nor account for changing comorbidity patterns  
309 and diagnostic journeys throughout the lifespan.

310           Electronic Health Records (EHR) provide an attractive opportunity to design studies that  
311 more closely match the experiences of our patients. In this study, we demonstrate application of  
312 phenotype risk scores to quantify EHR-MDD comorbidities. We suggest that these PheRS may  
313 be applied for two key goals. First, to increase statistical power in EHR analyses by assigning a  
314 quantitative ‘MDD-PheRS’ value to all individuals, rather than a binary case/control designation.  
315 Second, our analysis may allow us to identify individuals with ‘missing’ diagnoses; for example,  
316 those who exhibit a large number of comorbid diagnoses or risk factors, but who lack explicit  
317 MDD diagnoses within their records. Such missing diagnoses may stem from a lack of ability to  
318 afford treatment; an unwillingness to seek treatment due to stigma or shame discussing mental  
319 health<sup>47</sup>; diagnostic oversight or bias<sup>48</sup>; or difficulty in reaching an accurate diagnosis due to the  
320 lack of reliable tests or biomarkers for psychiatric disease<sup>49</sup>. Here, we suggest that such  
321 individuals may be identified using PheRS; for example we identified people with high PheRS  
322 and high MDD PRS but never received a formal diagnoses.

323           As we might expect for a complex, episodic trait, we find a wide range of associated  
324 traits and diagnoses contributing to our score. These associations can be grouped into a few

325 broad categories. First, known biological risk factors for MDD, for example; Type II Diabetes  
326 (E11.65)<sup>50</sup>. Second, we identify a number of environmental factors (e.g., homelessness<sup>51</sup>), which  
327 may precipitate or occur as a consequence of MDD. Third, we identified a number of psychiatric  
328 diagnoses and traits (e.g. Alcohol Dependence; F10.21). These may represent truly distinct  
329 diagnoses (for example, patients may truly have distinct PTSD and MDD diagnoses); or, they  
330 may provide evidence of complex diagnostic journeys for patients within our study. For  
331 example, early diagnoses of MDD may later be amended to diagnoses of bipolar disorder as  
332 more diagnostic information becomes available to clinicians. More complex evaluations with  
333 larger psychiatric EHR will be required to disentangle these patterns.

334 Finally, we identified a set of unexpected associations (moderate persistent asthma,  
335 J45.40; Gastro-esophageal reflux disease, K21.9). These may be representative of many factors.  
336 It is possible that some unknown or understudied shared biological aetiology exists between  
337 these traits; (for example, between lower leg pain (M79.604) and diabetes or homelessness, both  
338 of which are comorbid with MDD in our sample). It is also possible that some unknown  
339 ‘healthcare pleiotropy’ exists between traits; that is, traits that commonly co-present within the  
340 EHR for factors unrelated to biology- for example, sets of tests commonly performed jointly at  
341 annual check-ups.

342 Our analysis identifies a substantial number of individuals with high MDD PheRS and  
343 PRS values, but with no MDD diagnoses. Similarly, we identified some individuals with very  
344 low PheRS and PRS scores, but with multiple ICD-MDD diagnoses. Traditional EHR  
345 approaches, which typically require 1-2 diagnostic codes in order to form a diagnosis, would  
346 misclassify these individuals. For example, among individuals falling in the top two PheRS and  
347 PRS deciles, we identified 383 individuals with no F33.1 MDD diagnoses.

348           In order to ascertain whether our PheRS captures biologically relevant information about  
349 MDD, we tested whether PheRS scores within our biobank also correlated with MDD-PRS  
350 derived from the latest PGC-MDD GWAS. We found that our PheRS scores correlate  
351 significantly with MDD-PRS (Figure. 5), and explained more variance in the MDD-PheRS than  
352 in the directly observed MDD diagnosis. Notably, excluding individuals with MDD from our  
353 MDD-PRS analysis does not remove the association between PRS and PheRS, implying that this  
354 effect is not driven by individuals with MDD and high PheRS. Therefore, we hypothesize that  
355 either (i) our PheRS captures undiagnosed MDD cases, driving the PRS-PheRS associations; or  
356 (ii) that MDD-PRS is associated with the constellation of symptoms and traits associated with  
357 MDD (i.e., the MDD co-phenome), even in the absence of MDD.

358           Finally, we tested cross-ancestry applicability of our PheRS across three key  
359 racial/ethnic groups in our biobank. We urge caution in interpreting these results. Since race is a  
360 social rather than biological concept, we do not expect that true biological differences exist  
361 between self-defined racial groups; nor do we expect the genetic aetiology of MDD to differ  
362 according to race. However, we do expect that access to, and experience of, healthcare will differ  
363 according to race. In large part, these differences will stem from structural inequality, bias and  
364 racism experienced by our patients- in differing access to expert clinicians and insurance; in  
365 biased diagnostic practices and therapeutic approaches; and in differing environmental exposures  
366 and stressors. While some of these factors may be quantified within our medical records (for  
367 example, the number of visits, referrals to specialists or life-threatening traumas experienced),  
368 others will be implicit and unquantifiable (for example, biased or stereotyped diagnostic  
369 guidelines, or long-term stress from racist microaggressions). Sample sizes are insufficient here  
370 to probe effects of intersectionality (eg, compound bias experienced by Black women), although

371 we expect these compound factors, as well as additional unknown biases inherent across  
372 healthcare systems, to further bias comorbidity profiles and diagnostic accuracy within EHR.

373 Indeed, we found significantly different comorbidity profiles between all three  
374 racial/ethnic groups included in our analysis; in particular, we found far more ICD codes co-  
375 morbid with MDD among Hispanic-Americans overall, and a large number of codes comorbid  
376 only among Hispanic-Americans. By contrast, almost no codes were comorbid uniquely among  
377 European-Americans: rather, all codes significantly co-morbid among European-Americans were  
378 also significantly comorbid among other groups. These differences imply that MDD diagnoses  
379 are applied differently across racial and ethnic groups. It is possible that these differences stem  
380 from power imbalances- that is, the increased number of MDD cases among Hispanic-Americans  
381 enables us to identify significantly more associations; however, this would not account for the  
382 large number of codes shared only between African-American and Hispanic-American  
383 individuals. It is also possible that the diagnosis in question (F33.1) is given more readily- that is,  
384 at some lower clinical threshold- among HA patients than in other groups. This broader, more  
385 general use of the diagnosis would also account for the larger and more diffuse set of  
386 comorbidities, and is in line with our previous research regarding diagnostic bias in psychiatry<sup>11</sup>.

387 We tested which approach is most appropriate using a cross-ancestry base/target design.  
388 We derived PheRS using (i) all samples; (ii) each of the three racial/ethnic groups only, and  
389 tested predictive accuracy ( $R^2$ ) in the full cohort, and in each ancestry separately. We observe  
390 highest accuracy when base and target populations are matched (as would be expected), and only  
391 slight reductions in accuracy with HA/AA base/target combinations. Importantly, these patterns  
392 do not match expectations based on sample size. By default, larger discovery (base) populations  
393 should result in more accurate phenotype risk scores. However, AA- and HA-base PheRS

394 perform similarly, despite significant differences in sample size; and both outperform EA-base  
395 PheRS when applied to the full cohort. Our results indicate that it is not appropriate to assume  
396 that a European-default derived algorithm would be sufficient or appropriate.

397         However, two scenarios may justify construction of a PheRS using all samples jointly,  
398 regardless of race/ethnicity. First, it is possible that the power gained from increasing sample size  
399 (ie, by jointly analyzing all samples) outweighs the loss of specificity. However, it will be vital  
400 to ensure that power is gained evenly across all individuals and groups- ie, that predictive  
401 accuracy is not just increased significantly among a subset of participants. Alternatively, it is  
402 possible that diagnostic accuracy is higher among some racial/ethnic groups than others- that is,  
403 the MDD diagnosis underlying our model is more reliably applied in one group. In this case,  
404 PheRS derived from all samples, or from this single group, may perform best. More research into  
405 racial and ethnic biases in psychiatric diagnoses is warranted to probe these questions further. In  
406 particular, we caution that MDD diagnoses may be substantially less biased across ancestries  
407 than other psychiatric diagnoses. For example, we have previously shown that psychotic  
408 diagnosis and treatment are significantly different between racial and ethnic groups in our  
409 sample; similarly, previous studies have shown that diagnostic patterns differ greatly for eating  
410 disorders, autism spectrum disorder, and ADHD, among others. For these traits, race and  
411 ethnicity may be even more relevant in creating a useful phenotype risk score.

412         In summary, shallow phenotypes composed of a single data point from EHR do not  
413 accurately capture a patient's experience with disease. Algorithms comprising multiple data  
414 points, such as PheRS, allow us to consider heterogeneous patient experience, comorbidity, bias,  
415 and ancestry, as well as a patient's whole phenome, in order to arrive at accurate and robust  
416 phenotypes. With this tool, we begin to use our currently available biobanks to look for

417 translational opportunities to better predict trajectories of disease, comorbidities, and behaviours

418 to find multi-omic associations to understand disease etiology.

419

420

421 **Acknowledgements**

422 This work was supported by a Faculty Scholar Award from the Seaver Foundation: "Analytical  
423 Genomics of Vulnerable Populations".

424 This work was supported in part through the resources and staff expertise provided by the  
425 Charles Bronfman Institute for Personalized Medicine and The BioMe™ Biobank Program at the  
426 Icahn School of Medicine at Mount Sinai.

427 We thank Dr. Eli Stahl, and Dr. Alex Charney for their thoughtful comments during lab meetings  
428 and poster sessions. We thank Dr. Inga Peter for her helpful comments and revisions to this  
429 manuscript, and for her assistance in examining Ms. Lebovitch's MPH thesis, which formed a  
430 preliminary version of this analysis.

431 We also acknowledge the supportive atmosphere of the Pamela Sklar Division of Psychiatric  
432 Genetics; our work is stronger for the collegiate and collaborative atmosphere of our division. As  
433 always, we are grateful for the advice, guidance and mentorship of our founder and namesake,  
434 Dr. Pamela Sklar.

435

436

437

438

439

440



442 **References**

- 443
- 444 1 Sullivan PF, Agrawal A, Bulik CM, *et al.* Psychiatric Genomics: An Update and an Agenda.  
445 *The American journal of psychiatry* 2018; **175**: 15–27.
- 446 2 Ritchie MD, Denny JC, Crawford DC, *et al.* Robust replication of genotype-phenotype  
447 associations across multiple diseases in an electronic medical record. *American journal of*  
448 *human genetics* 2010; **86**: 560–72.
- 449 3 Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance  
450 Precision Medicine. *Annual review of genomics and human genetics* 2016; **17**: 353–373.
- 451 4 Denny JC, Ritchie MD, Basford MA, *et al.* PheWAS: demonstrating the feasibility of a  
452 phenome-wide scan to discover gene-disease associations. *Bioinformatics (Oxford, England)*  
453 2010; **26**: 1205–10.
- 454 5 Pendergrass SA, Brown-Gentry K, Dudek SM, *et al.* The use of phenome-wide association  
455 studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy  
456 discovery. *Genetic epidemiology* 2011; **35**: 410–22.
- 457 6 Bastarache L, Hughey JJ, Hebring S, *et al.* Phenotype risk scores identify patients with  
458 unrecognized Mendelian disease patterns. *Science (New York, NY)* 2018; **359**: 1233–1239.
- 459 7 Fernandez-Pujals AM, Adams MJ, Thomson P, *et al.* Epidemiology and Heritability of Major  
460 Depressive Disorder, Stratified by Age of Onset, Sex, and Illness Course in Generation  
461 Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* 2015; **10**: e0142197.

- 462 8 Grant BF, Stinson FS, Dawson DA, *et al.* Prevalence and co-occurrence of substance use  
463 disorders and independent mood and anxiety disorders: results from the National  
464 Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry* 2004; **61**:  
465 807–16.
- 466 9 Ruscio AM, Hallion LS, Lim CCW, *et al.* Cross-sectional Comparison of the Epidemiology of  
467 *DSM-5* Generalized Anxiety Disorder Across the Globe. *JAMA Psychiatry* 2017; **74**: 465.
- 468 10 Penninx BWJH. Depression and cardiovascular disease: Epidemiological evidence on their  
469 linking mechanisms. *Neuroscience & Biobehavioral Reviews* 2017; **74**: 277–86.
- 470 11 Dueñas HR, Seah C, Huckins LM. Implicit Bias of Encoded Variables: Frameworks for  
471 addressing structured bias in EHR Data. *Human Molecular Genetics* 2020; **In Press**.
- 472 12 Visscher PM, Hemani G, Vinkhuyzen AAE, *et al.* Statistical power to detect genetic  
473 (co)variance of complex traits using SNP data in unrelated samples. *PLoS genetics* 2014; **10**:  
474 e1004269.
- 475 13 Popejoy A, Fullerton S. Genomics is failing on diversity. *Nature* 2016; **538**: 161–164.
- 476 14 Peterson RE, Kuchenbaecker K, Walters RK, *et al.* Genome-wide Association Studies in  
477 Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell*  
478 2019; **179**: 589–603.
- 479 15 Bustamante CD, De La Vega FM, Burchard EG. Genomics for the world. *Nature* 2011; **475**:  
480 163–5.

- 481 16 Wojcik GL, Graff M, Nishimura KK, *et al.* Genetic analyses of diverse populations improves  
482 discovery for complex traits. *Nature* 2019; **570**: 514–518.
- 483 17 Belbin GM, Odgis J, Sorokin EP, *et al.* Genetic identification of a common collagen disease  
484 in Puerto Ricans via identity-by-descent mapping in a health system. *eLife* 2017; **6**: e25060.
- 485 18 BioMe BioBank Program | Icahn School of Medicine. Icahn School of Medicine at Mount  
486 Sinai. <https://icahn.mssm.edu/research/ipm/programs/biome-biobank> (accessed July 2, 2020).
- 487 19 Lin H-C, Erickson SR, Balkrishnan R. Physician prescribing patterns of innovative  
488 antidepressants in the United States: the case of MDD patients 1993-2007. *Int J Psychiatry*  
489 *Med* 2011; **42**: 353–68.
- 490 20 Yucel A, Essien EJ, Sanyal S, *et al.* Racial/ethnic differences in the treatment of adolescent  
491 major depressive disorders (MDD) across healthcare providers participating in the medicaid  
492 program. *J Affect Disord* 2018; **235**: 155–61.
- 493 21 Wu C-H, Erickson SR, Piette JD, Balkrishnan R. Mental health resource utilization and  
494 health care costs associated with race and comorbid anxiety among Medicaid enrollees with  
495 major depressive disorder. *J Natl Med Assoc* 2012; **104**: 78–88.
- 496 22 Wu C-H, Erickson SR, Piette JD, Balkrishnan R. The association of race, comorbid anxiety,  
497 and antidepressant adherence among Medicaid enrollees with major depressive disorder. *Res*  
498 *Social Adm Pharm* 2012; **8**: 193–205.
- 499 23 López E, Steiner AJ, Manier K, *et al.* Quality of life and functioning of Hispanic patients  
500 with Major Depressive Disorder before and after treatment. *J Affect Disord* 2018; **225**: 117–22.

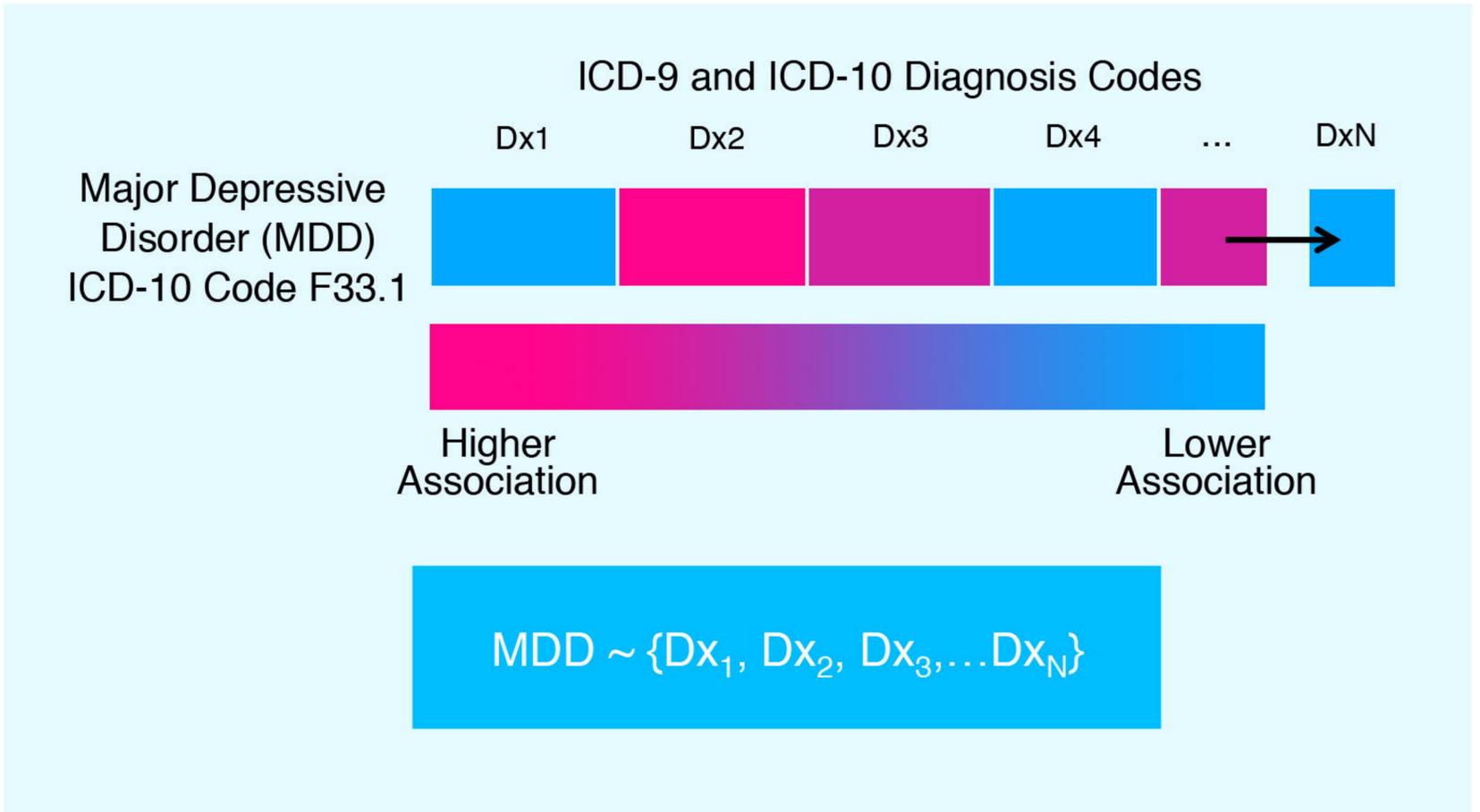
- 501 24 Bailey RK, Blackmon HL, Stevens FL. Major depressive disorder in the African American  
502 population: meeting the challenges of stigma, misdiagnosis, and treatment disparities. *J Natl*  
503 *Med Assoc* 2009; **101**: 1084–9.
- 504 25 Bailey RK, Patel M, Barker NC, Ali S, Jabeen S. Major depressive disorder in the African  
505 American population. *J Natl Med Assoc* 2011; **103**: 548–57.
- 506 26 Shao Z, Richie WD, Bailey RK. Racial and Ethnic Disparity in Major Depressive Disorder. *J*  
507 *Racial Ethn Health Disparities* 2016; **3**: 692–705.
- 508 27 Fyffe DC, Sirey JA, Heo M, Bruce ML. Late-life depression among black and white elderly  
509 homecare patients. *Am J Geriatr Psychiatry* 2004; **12**: 531–5.
- 510 28 Dunlop BW, Still S, LoParo D, *et al*. Somatic symptoms in treatment-naïve Hispanic and  
511 non-Hispanic patients with major depression. *Depress Anxiety* 2020; **37**: 156–65.
- 512 29 Assari S, Dejman M, Neighbors HW. Ethnic Differences in Separate and Additive Effects of  
513 Anxiety and Depression on Self-rated Mental Health Among Blacks. *J Racial Ethn Health*  
514 *Disparities* 2016; **3**: 423–30.
- 515 30 Zheutlin AB, Dennis J, Karlsson Linnér R, *et al*. Penetrance and Pleiotropy of Polygenic  
516 Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. *American*  
517 *Journal of Psychiatry* 2019; : [appi.ajp.2019.18091085](https://doi.org/10.1176/appi.ajp.2019.18091085).
- 518 31 Hastie HZ and T. elasticnet: Elastic-Net for Sparse Estimation and Sparse PCA. 2020  
519 <https://CRAN.R-project.org/package=elasticnet> (accessed July 2, 2020).

- 520 32 Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the*  
521 *Royal Statistical Society: Series B (Statistical Methodology)* 2005; **67**: 301–320.
- 522 33 Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models  
523 via Coordinate Descent. *Journal of Statistical Software; Vol 1, Issue 1 (2010)* 2010; published  
524 online Feb. <https://www.jstatsoft.org/v033/i01> <http://dx.doi.org/10.18637/jss.v033.i01>.
- 525 34 Wray NR, Ripke S, Mattheisen M, *et al.* Genome-wide association analyses identify 44 risk  
526 variants and refine the genetic architecture of major depression. *Nat Genet* 2018; **50**: 668–81.
- 527 35 Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics*  
528 2014; **31**: 1466–1468.
- 529 36 Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score  
530 analyses. *Nature Protocols* 2020; : 1–14.
- 531 37 Martin AR, Gignoux CR, Walters RK, *et al.* Human Demographic History Impacts Genetic  
532 Risk Prediction across Diverse Populations. *American journal of human genetics* 2017; **100**:  
533 635–649.
- 534 38 Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Current clinical use of  
535 polygenic scores will risk exacerbating health disparities. *bioRxiv* 2019; : 441261.
- 536 39 Bigdeli T, Ripke S, Peterson R, *et al.* Genetic effects influencing risk for major depressive  
537 disorder in China and Europe. *Translational psychiatry* 2017; **7**: e1074.
- 538 40 Duncan L, Shen H, Gelaye B, *et al.* Analysis of polygenic risk score usage and performance  
539 in diverse human populations. *Nature Communications* 2019; **10**: 3328.

- 540 41 Lindert J, von Ehrenstein OS, Grashow R, Gal G, Braehler E, Weisskopf MG. Sexual and  
541 physical abuse in childhood is associated with depression and anxiety over the life course:  
542 systematic review and meta-analysis. *Int J Public Health* 2014; **59**: 359–72.
- 543 42 Hodgson KJ, Shelton KH, van den Bree MBM, Los FJ. Psychopathology in Young People  
544 Experiencing Homelessness: A Systematic Review. *Am J Public Health* 2013; **103**: e24–37.
- 545 43 Opel N, Redlich R, Grotegerd D, *et al.* Obesity and major depression: Body-mass index  
546 (BMI) is associated with a severe course of disease and specific neurostructural alterations.  
547 *Psychoneuroendocrinology* 2015; **51**: 219–26.
- 548 44 Davydow DS, Desai SV, Needham DM, Bienvenu OJ. Psychiatric Morbidity in Survivors of  
549 the Acute Respiratory Distress Syndrome: A Systematic Review. *Psychosomatic Medicine*  
550 2008; **70**: 512–519.
- 551 45 Claassen CA, Trivedi MH, Rush AJ, *et al.* Clinical differences among depressed patients  
552 with and without a history of suicide attempts: findings from the STAR\*D trial. *Journal of*  
553 *Affective Disorders* 2007; **97**: 77–84.
- 554 46 Martins SS, Fenton MC, Keyes KM, Blanco C, Zhu H, Storr CL. Mood/Anxiety disorders  
555 and their association with non-medical prescription opioid use and prescription opioid use  
556 disorder: longitudinal evidence from the National Epidemiologic Study on Alcohol and Related  
557 Conditions. *Psychol Med* 2012; **42**: 1261–72.
- 558 47 Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB. Missing clinical and behavioral  
559 health data in a large electronic health record (EHR) system. *J Am Med Inform Assoc* 2016; **23**:  
560 1143–9.

- 561 48 Shrivastava A, Johnston M, Bureau Y. Stigma of Mental Illness-1: Clinical reflections. *Mens*  
562 *Sana Monogr* 2012; **10**: 70–84.
- 563 49 Abi-Dargham A, Horga G. The search for imaging biomarkers in psychiatric disorders.  
564 *Nature Medicine* 2016; **22**: 1248–55.
- 565 50 Kan C, Pedersen NL, Christensen K, *et al.* Genetic overlap between type 2 diabetes and  
566 depression in Swedish and Danish twin registries. *Molecular Psychiatry* 2016; **21**: 903–9.
- 567 51 Fazel S, Geddes JR, Kushel M. The health of homeless people in high-income countries:  
568 descriptive epidemiology, health consequences, and clinical and policy recommendations. *The*  
569 *Lancet* 2014; **384**: 1529–40.
- 570

## Phenome-wide associations in BioMe



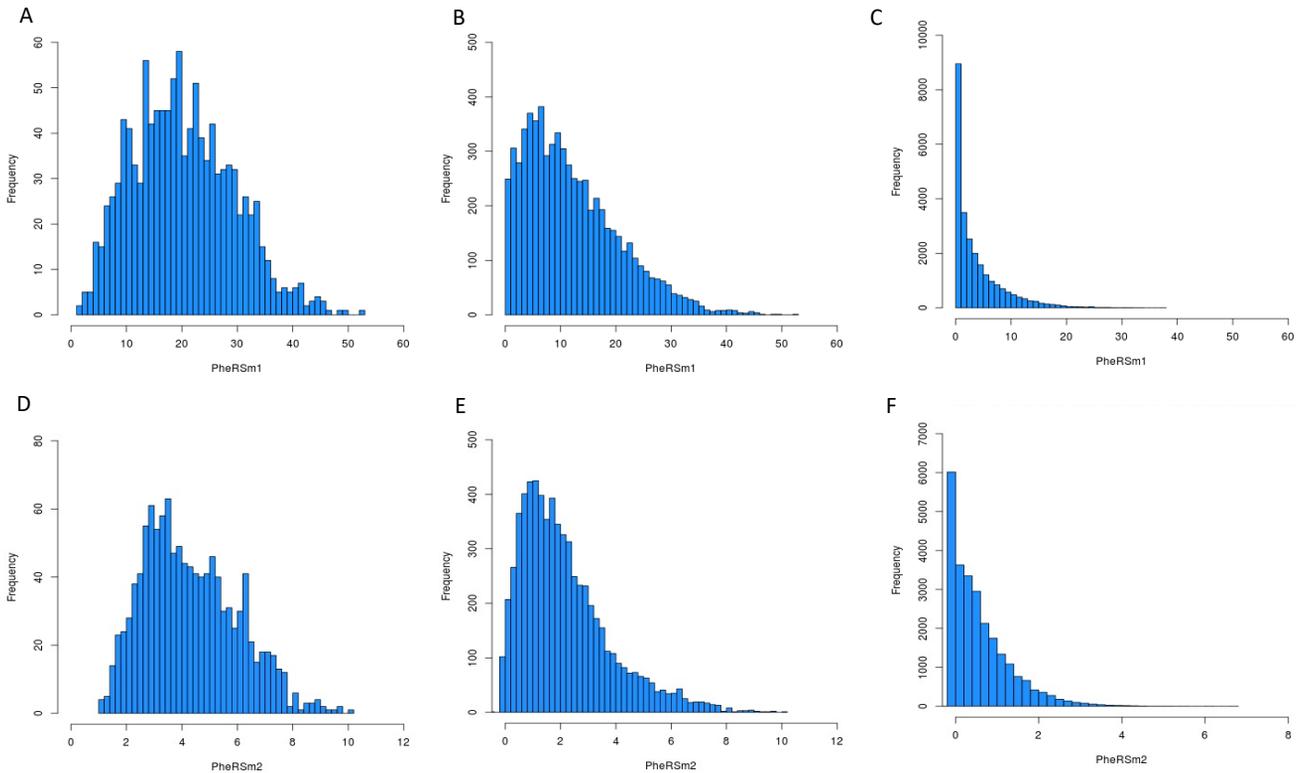
### Elastic Net Identifies Associations

$$\text{MDD} \sim \text{BMI}(w_p) + \text{Insomnia}(w_p) + \text{Chronic Pain}(w_p) + \dots$$

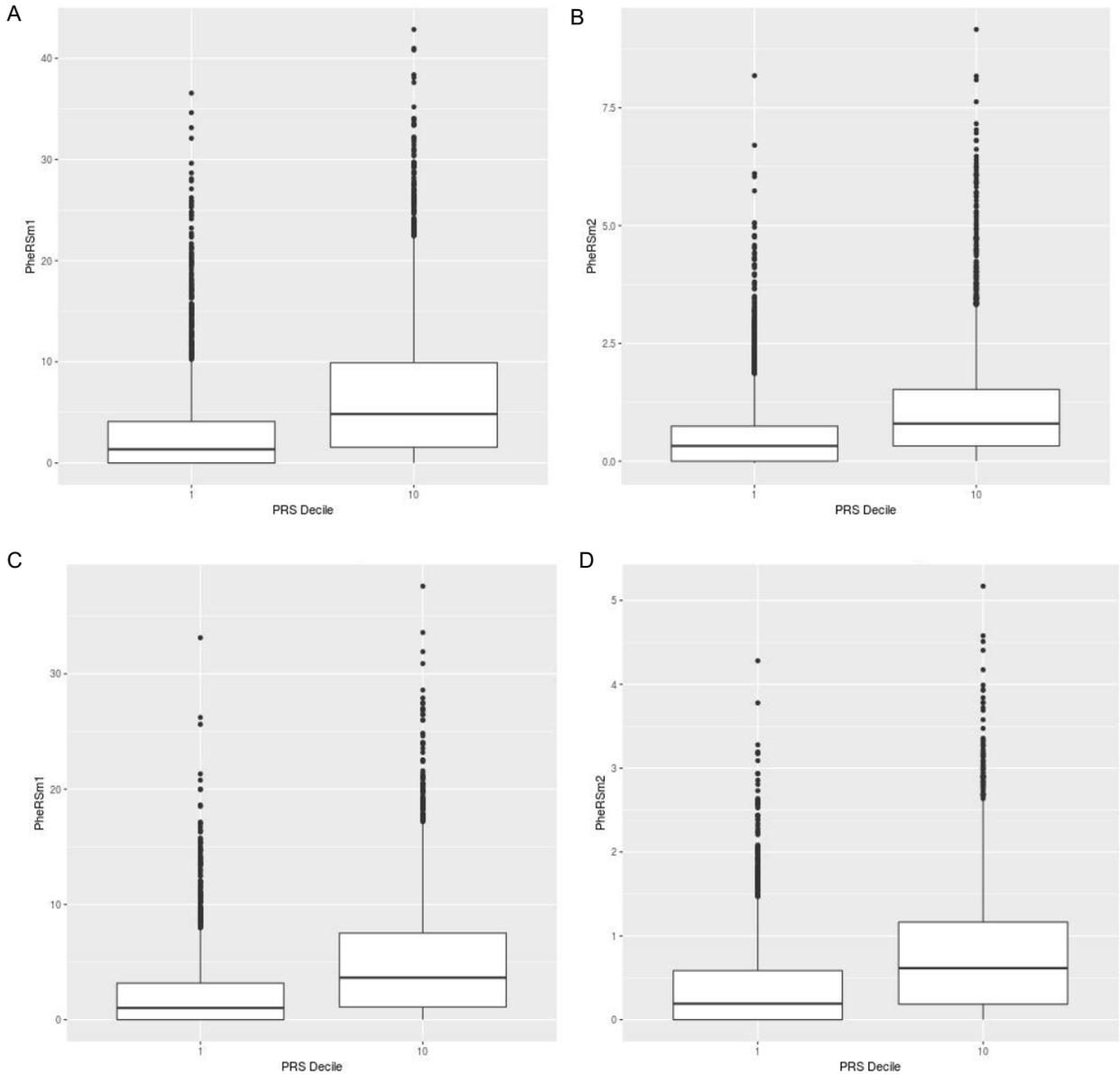


### Calculate PheRS in BioMe Population

	BMI	Insomnia	Chronic Pain		PheRS
BioMe Individuals	ID1	1	1	→	3
	ID2	0	0	→	1
	ID3	0	0	→	0
	ID4	1	0	→	2
	...	—	—	—	
	IDN	0	1	→	2



**Figure 2: MDD-Phenotype Risk Scores are significantly higher among individuals with histories of depression.** For both methods (PheRSm1- A-C; PheRSm2- D-F), we find that phenotype risk scores are highest among individuals with diagnoses of F33.1 (A, D); elevated among individuals with any depression diagnosis, excluding F33.1 (B, E), and lowest among individuals with no history of any depression diagnoses (C,F). All comparisons are significant; kolgorov-smirnov test  $p < 2.2 \times 10^{-16}$ .

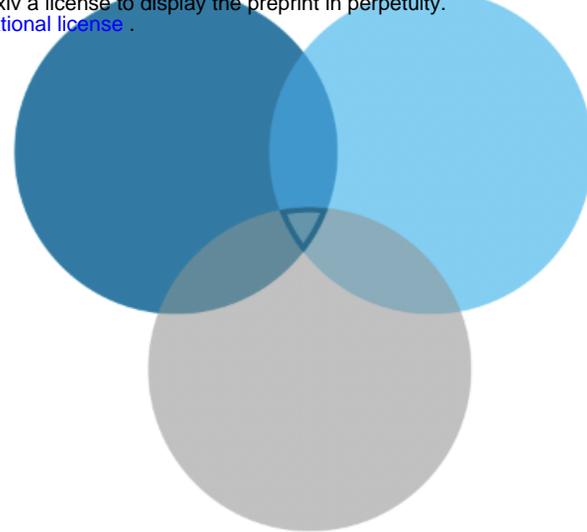
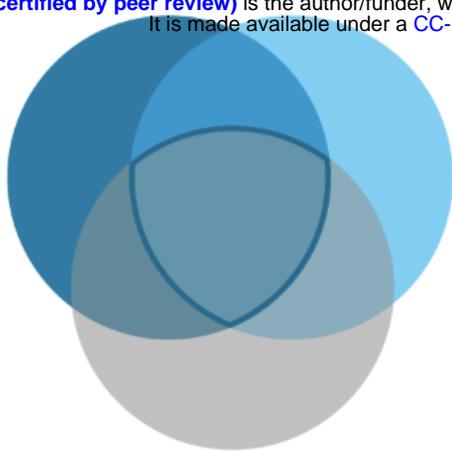


**Figure 3: MDD Polygenic Risk Scores predict higher MDD Phenotype Risk Scores.** Phenotype Risk Scores are significantly higher among individuals with the top decile of polygenic risk, compared to the bottom, for both methods (**A**- PheRSm1, fold change = 3.57; **B**-PheRSm2, fold change = 2.46). This effect persists even when removing individuals with F33.1 diagnoses (**C**- PheRSm1, fold change =3.58; **D**- PheRSm2, fold change=3.21).

i.

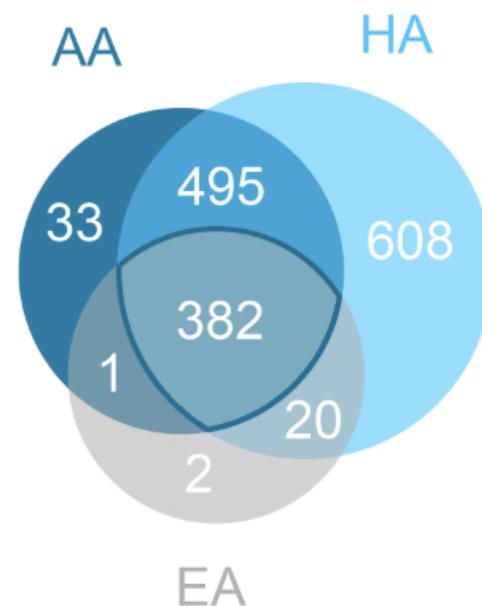
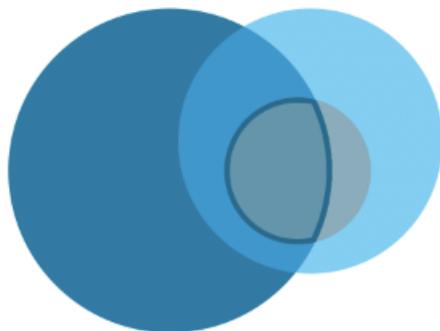
ii.

medRxiv preprint doi: <https://doi.org/10.1101/2021.01.25.21249615>; this version posted January 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).



iii.

iv.



**Figure 4: Hypothetical scenarios modelling the impact of diagnostic bias, compared to observed MDD-comorbidity profiles.**

- (i) Bias plays only a small role in diagnoses, and the majority of diagnoses are assigned accurately. We expect highly overlapping co-phenomes and risk-factors for MDD between patient groups.
- (ii) Diagnostic bias affects these groups unevenly; we expect less overlap in co-phenomes.
- (iii) Diagnostic bias results in under-diagnosis and missing data (ie, a smaller than expected co-phenome) in one group, and uneven overlap between groups.
- (iv) Observed overlap between MDD co-phenomes of African-American (AA), European-American (EA), Hispanic-American (HA) in our database.

Figure legends:

**Figure 1: Constructing Phenotype Risk Scores from EHR data for complex traits.**

Step 1: Identification of traits, disorders and phenotypes associated with MDD.

Step 2: Construction of a joint model predicting MDD from all comorbid traits, using 10-fold cross-validation and elastic net regression.

Step 3: Phenotype risk scores are calculated as weighted sums of traits and disorders present in each individual's phenome. Traits are weighted according to (i) prevalence (PheRSm1) or (ii) elastic net calculated effect size (PheRSm2).

**Figure 2: MDD-Phenotype Risk Scores are significantly higher among individuals with histories of depression.** For both methods (PheRSm1- A-C; PheRSm2- D-F), we find that phenotype risk scores are highest among individuals with diagnoses of F33.1 (A, D); elevated among individuals with any depression diagnosis, excluding F33.1 (B, E), and lowest among individuals with no history of any depression diagnoses (C,F). All comparisons are significant; kolgorov-smirnov test  $p < 2.2 \times 10^{-16}$ .

**Figure 3: MDD Polygenic Risk Scores predict higher MDD Phenotype Risk Scores.**

Phenotype Risk Scores are significantly higher among individuals with the top decile of polygenic risk, compared to the bottom, for both methods (PheRSm1-A; PheRSm2-B). This effect persists even when removing individuals with F33.1 diagnoses (PheRSm1-C. Fold change =3.58; PheRSm2-D. Fold change=3.21).

**Figure 4: Hypothetical scenarios modelling the impact of diagnostic bias, compared to observed MDD-comorbidity profiles.** (i) Bias plays only a small role in diagnoses, and the majority of diagnoses are assigned accurately. We expect highly overlapping co-phenomes and risk-factors for MDD between patient groups. (ii) Diagnostic bias affects these groups unevenly; we expect less overlap in co-phenomes. (iii) Diagnostic bias results in under-diagnosis and missing data (ie, a smaller than expected co-phenome) in one group, and uneven overlap between groups. (iv) Observed overlap between MDD co-phenomes of African-American (AA), European-American (EA), Hispanic-American (HA) in our database.

Table legends:

Table 1: Number of BioMe Participants with ICD10 MDD diagnostic codes

Table 2: ICD Codes significantly associated with F33.1

Table 3: ICD codes contributing to MDD-PheRS

Supplementary table legends:

S. Table 1: Number of F33.1 Diagnoses according to self-reported sex, and race/ethnicity

S. Table 2A: ICD codes significantly co-morbid with MDD among African American patients

S. Table 2B: ICD codes significantly co-morbid with MDD among European American patients

S. Table 2C: ICD codes significantly co-morbid with MDD among Hispanic American patients

Supplementary Table 3: Cross-Ancestry predictive accuracy