Word counts: 2,982

Number of tables/figures: 6

Number of supplementary tables: 17

# Transdiagnostic Polygenic Risk Models for Psychopathology and Comorbidity: Cross-Ancestry Analysis in the *All of Us* Research Program

Phil H. Lee,[1,2,3*], Jae-Yoon Jung,[4] Brandon T. Sanzo,[1] Rui Duan,[4] Tian Ge,[1,2,3]
23andMe Research Team,[5] Irwin Waldman,[6] Jordan W. Smoller[1,2,3],
Ted Schwaba,[7] Elliot M. Tucker-Drob,[8] Andrew D. Grotzinger[9,10]

1. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Mass General Brigham, Boston, MA, USA
2. Department of Psychiatry, Mass General Brigham and Harvard Medical School, Boston, MA, USA
3. Stanly Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
4. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
5. 23andMe, Inc., San Francisco, CA, USA
6. Department of Psychology, Emory University, Atlanta, GA, USA
7. Department of Psychology, Michigan State University, MI, USA
8. Department of Psychology, University of Texas at Austin, TX, USA
9. Institute for Behavioral Genetics, University of Colorado at Boulder, CO, USA
10. Department of Psychology and Neuroscience, University of Colorado at Boulder, CO, USA

***Correspondence:**
Phil H. Lee, PhD
Center for Genomic Medicine
Simches Research Building
185 Cambridge St.
Boston, MA 02114
Phone: 617-643-9304
Fax: 617-643-3080
Email: PLEE0@mgh.harvard.edu

**Running Title:** Transdiagnostic Polygenic Risk Models for Psychopathology in *All of Us*

**Keywords:** transdiagnostic, comorbidity, psychiatric genomics, cross-ancestry, *All of Us* Research Program, psychopathology

## ABSTRACT

Psychiatric disorders exhibit substantial genetic overlap, raising questions about the utility of transdiagnostic genetic risk models. Using data from the *All of Us* Research Program (N=102,091), we evaluated common psychiatric genetic (CPG) factor-based polygenic risk scores (PRSs) compared to standard disorder-specific PRSs. The CPG PRS consistently outperformed disorder-specific scores in predicting individual disorder risk, explaining 1.07 to 24.6 times more phenotypic variance across 11 psychiatric conditions. Meanwhile, many disorder-specific PRSs retained independent but smaller contributions, highlighting the complementary nature of shared and disorder-specific genetic risk. While alternative multi-factor models improved model fit, the CPG PRS provided comparable or superior predictive performance across most disorders, including overall comorbidity burden. Cross-ancestry analyses however revealed notable limitations of European-centric GWAS datasets for other populations due to ancestral differences in genetic architecture. These findings underscore the potential value of transdiagnostic PRSs for psychiatric genetics while highlighting the need for more equitable genetic risk models.

## INTRODUCTION

The classification of psychiatric disorders remains one of medicine's most challenging puzzles.[1] While Emil Kraepelin's foundational work established categorical diagnostic systems,[2] clinical reality reveals overlapping symptoms and frequent comorbidities, complicating both diagnosis and treatment.[3] This raises a fundamental question: should we conceptualize psychiatric disorders as discrete entities, or as interrelated manifestations of broader dimensions of psychopathology? For example, the Hierarchical Taxonomy of Psychopathology (HiTOP)[4,5] and Research Domain Criteria (RDoC)[6,7] propose that psychiatric disorders exist along continuous dimensions rather than discrete categories, better capturing shared and comorbid symptomatology.

Recent advances in psychiatric genomics have provided new tools to examine this question through the lens of inherited risk. Large-scale genomic studies have revealed extensive genetic overlap among disorders, suggesting shared biological underpinnings.[8-11] Genomic Structural Equation Modeling (GenomicSEM)[12] enables quantification of this shared risk through transdiagnostic polygenic risk scores (PRSs), which capture genetic vulnerability shared across multiple disorders. However, the practical utility of transdiagnostic PRSs remains uncertain. Do they improve risk stratification compared to standard disorder-specific PRSs? Can they help explain psychiatric comorbidity? Does the intrinsic nature of transdiagnostic PRSs make them more generalizable across diverse populations, especially given the Eurocentric bias in psychiatric genetic research?

In this study, we address these questions by employing a common psychiatric genetic factor (CPG) approach to derive a transdiagnostic PRS designed to capture shared genetic risk across psychiatric disorders. The CPG model serves as a conceptual framework for comparing broad transdiagnostic genetic risk with disorder-specific PRSs. While our main analyses primarily utilize the CPG model, we also compare its performance with alternative transdiagnostic approaches of varying complexity[10,13] as a secondary analysis to provide a comparative assessment of transdiagnostic risk models.

To conduct these analyses, we leverage data from the *All of Us* Research Program,[14,15] one of the largest and most ancestrally diverse genomic initiatives to date. By integrating comprehensive health survey data on psychiatric diagnoses with participants' genetic data, *All of Us* provides an unprecedented opportunity to examine how well transdiagnostic PRSs explain

phenotypic variance in psychiatric burden and comorbidity across diagnostic categories and diverse ancestries. Specifically, we investigate the:

1. Comparative performance of CPG- versus disorder-specific PRS models across diagnostic categories.

2. Association of CPG PRSs with comorbidity burden, measured by the number of concurrent diagnoses.

3. Ancestry-related differences in predictive performance and implications for equitable risk stratification.

We hypothesize that CPG PRS will enhance risk stratification by capturing overall psychiatric burden more effectively than disorder-specific PRSs, both for individual disorders and for comorbidity burden. This improvement is expected due to the increased statistical power derived from aggregating genetic associations across multiple psychiatric GWASs, as well as from capturing shared genetic mechanisms contributing to multiple disorders.[16] Given known limitations of PRS generalizability derived from largely European-ancestry GWASs, we anticipate that ancestry-related attenuation will reduce the predictive accuracy in non-European populations.[17,18] However, the CPG PRS may demonstrate relatively better cross-ancestry performance compared to disorder-specific PRSs due to its broader representation of psychiatric vulnerability that is conserved across disorders and populations. By addressing these questions, we aim to inform ongoing efforts to integrate polygenic risk into precision psychiatry, clarify the potential utility of transdiagnostic PRSs for clinical risk stratification, and contribute to the development of more equitable approaches in psychiatric genomics.

# RESULTS

### Prevalence of Psychiatric Disorders and Comorbidity

Our analysis included 102,091 individuals from the *All of Us* Research Program who completed comprehensive health surveys for psychiatric disorders (**Fig 1**). Of these participants, 30,962 participants (25.96%) reported receiving clinical care for at least one of 13 psychiatric conditions (**Table 1**).

Mood and anxiety disorders were the most prevalent, with depression affecting 24.07% (95% CI, 23.80%- 24.34%) and anxiety reaction/panic disorders affecting 19.70% (95% CI, 19.44%-19.96%) of the cohort. Other conditions with notable prevalence rates included PTSD at 7.29% (95% CI, 7.11%-7.47%), ADHD at 5.20% (95% CI, 5.04%-5.36%), and bipolar disorder at 4.50% (95% CI, 4.35%-4.65%). Autism spectrum disorders exhibited the lowest prevalence at 0.38% (95% CI, 0.376%-0.383%).

Significant levels of comorbidity were evident across psychiatric disorders, as indicated by tetrachoric correlations ranging from modest ($r = 0.16$) to very strong ($r = 0.78$) (**Supplementary Table 1**). The strongest comorbid relationships were identified between anxiety reaction/panic disorder and depression ($r = 0.78$) and between drug use disorder and alcohol use disorder ($r = 0.72$). Depression also showed strong correlations with PTSD ($r = 0.67$), personality disorder ($r = 0.65$), and bipolar disorder ($r = 0.59$). Conversely, autism spectrum disorder demonstrated the weakest correlations overall, particularly with substance use disorders ($r < 0.2$).

Among participants with any psychiatric condition, 53% (16,539/30,962) reported multiple ($\geq 2$) disorders requiring current clinical care (**Supplementary Table 2**). These findings underscore the pervasive nature of comorbidity, reinforcing the need for integrative approaches in diagnosis, treatment planning, etiological models, and risk prediction.

**Associations of the CPG and Disorder-Specific PRSs with Psychiatric Disorders**

Of the 13 psychiatric conditions surveyed in the *All of Us* Program (**Table 1**), GWAS data were closely matched for 11 disorders, allowing for direct comparisons of disorder-specific and CPG genetic risks. These included depression, ADHD, alcohol use disorder, anxiety reaction/panic disorder, autism spectrum disorder, bipolar disorder, drug use disorder, eating disorder, PTSD, schizophrenia, and social phobia. For these 11 disorders, we assessed associations with three types of PRSs in European ancestry participants ($n$ = 78,937): (1) the CPG PRS, representing shared genetic risk across 15 psychiatric disorders calculated using the multivariate GWAS summary statistics from GenomicSEM[12]; (2) standard single-disorder PRSs, calculated for each individual disorder; and (3) disorder-specific PRSs, which isolate genetic effects unique to each disorder by removing shared variance captured by the CPG factor using residualization (see **Methods, Supplementary Tables 3-5**).

All standard, univariate disorder PRSs were significantly associated with their corresponding psychiatric disorders (all FDR $p < 0.001$, **Fig 2A, Supplementary Table 6**). The CPG PRS also showed significant positive associations with all 11 psychiatric disorders (all FDR $p < 0.001$). Moreover, the CPG PRS demonstrated significantly larger effect sizes compared to both standard single-disorder PRSs and disorder-specific PRSs: the median odds ratios (OR) for CPG PRSs was 1.58, compared to 1.42 for standard disorder PRSs (two-sided Wilcoxon rank-sum test $p=1.14\times10^{-2}$) and 1.17 for disorder-specific PRSs (two-sided Wilcoxon rank-sum test $p=1.42\times10^{-5}$).

For six disorders—ADHD, autism spectrum disorder, alcohol use disorder, depression, drug use disorder, and PTSD— both shared genetic risk (captured by the CPG PRS) and disorder-specific genetic risk made significant independent contributions. For instance, the CPG PRS for ADHD had an OR of 1.58 (95% CI: 1.52–1.64), while the disorder-specific ADHD PRS independently retained an OR of 1.32 (95% CI: 1.28–1.37). Similarly, for autism spectrum disorder, the CPG PRS showed an OR of 1.52 (95% CI: 1.34–1.72), and the disorder-specific ASD PRS exhibited an independent OR of 1.29 (95% CI: 1.14–1.47). These results suggest the complementary contributions of shared and disorder-specific genetic risk underlying psychiatric outcomes.

When evaluating phenotypic variance ($R^2$) explained by these genetic scores, the CPG PRS consistently outperformed disorder-specific PRSs across all psychiatric conditions (two-sided

Wilcoxon rank-sum test $p$=2.45x10$^{-3}$). **Fig 2B** demonstrates that incorporating both the CPG and disorder-specific PRSs provides additional predictive power beyond standard single-disorder PRSs (**Supplementary Table 7**). This improvement was particularly pronounced for eating disorder (increasing variance explained from 0.05% to 1.23%, a 24.6-fold increase), social phobia (from 0.71% to 6.07%), drug use disorder (from 3.64% to 9.18%), and bipolar disorder (from 1.10% to 4.47%), where standard single-disorder PRSs explained limited phenotypic variance, highlighting the utility of the CPG PRS for improving risk prediction.

To ensure robustness, sensitivity analyses were conducted by varying diagnostic criteria and control group definitions. Both lifetime and current diagnoses based on medication produced consistent findings (**Supplementary Tables 8,9**). In contrast, analyses of lifetime diagnoses without current clinical care showed weaker genetic associations compared to those under current care, which could reflect either attenuation due to recall bias or differences in the persistence of psychiatric conditions over time (**Supplementary Table 10**). Additionally, analyses utilizing control groups with no corresponding disorder with current clinical care yielded consistently significant findings, albeit with modestly decreased PRS effects than our primary findings (one-sided pairwise t-test $p$ = 6.72x10$^{-8}$; **Supplementary Tables 11**).

**Associations of the CPG and Disorder-Specific PRSs with Psychiatric Comorbidity**

We next examined the associations between PRSs and the burden of psychiatric comorbidity, measured by the number of concurrent psychiatric disorders under clinical care (**Fig 3, Supplementary Table 12**). The CPG PRS demonstrated a strong association with comorbidity burden (standardized beta coefficient β = 0.26, SE = 0.006, $R^2$ = 3.41%, FDR p < 0.05). Among standard single-disorder PRSs, all except the obsessive-compulsive disorder PRS, were significantly associated with psychiatric comorbidity. The depression PRS exhibited the strongest effect (β = 0.263, SE = 0.006, $R^2$ = 3.53%), followed by ADHD (β = 0.202, SE = 0.006, $R^2$ = 2.1%) and PTSD (β = 0.146, SE = 0.006, $R^2$ = 1.09%).

When the CPG component was removed, most disorder-specific effects were substantially attenuated. Three disorders however retained significant positive disorder-specific effects, with independent contributions to variance explained ($R^2$): depression (β = 0.11, SE = 0.006, $R^2$ = 0.63%), ADHD (β = 0.063, SE = 0.006, $R^2$ =0.21%), and PTSD (β = 0.039, SE = 0.006, $R^2$ =0.08%). These significant disorder-specific contributions likely reflect both the high prevalence of these conditions in the cohort (**Table 1**) and the large, well-powered GWAS datasets used to derive their PRSs (**Supplementary Table 3).** Taken together with the CPG PRS, these PRSs explained a total of 4.38% of the variance in psychiatric comorbidity (**Supplementary Table 12**).

**Cross-Ancestry Evaluation of the CPG and Disorder-Specific PRSs**

We examined the performance of the CPG and disorder-specific PRSs in predicting individual psychiatric conditions and comorbidity burden across AFR- and AMR-like participants, comparing these results to the EUR cohort presented earlier (**Fig 4, Supplementary Tables 13-14**).

For individual disorder prediction (**Fig 4A, Supplementary Table 13**), PRS effects were weakest in the AFR group, with few significant associations for the CPG PRS or disorder-specific PRSs, and the ORs showing minimal correlations with those observed in EUR ($r = 0.10$, $p = 6.13 \times 10^{-1}$). In contrast, PRS performance was more consistent in AMR, with highly correlated ORs for both CPG and disorder-specific PRSs ($r = 0.85$, $p = 2.57 \times 10^{-9}$).

When examining comorbidity burden (**Fig 4B, Supplementary Table 14**), the CPG PRS showed significant associations in all three ancestry groups, including AFR, and effect sizes (β coefficients) were significantly correlated between AFR and EUR ($r = 0.69$, $p = 2.0 \times 10^{-5}$). Correlations were stronger for AMR, with β coefficients closely aligned with EUR ($r = 0.84$, $p = 2.4 \times 10^{-9}$). Across all ancestry groups, the CPG PRS explained more variance in psychiatric risk (mixed effects $\beta=0.28$, SE=0.09, p-value=$3.78 \times 10^{-3}$) and comorbidity burden (mixed effects $\beta=1.48$, SE=0.30, p-value=$4.28 \times 10^{-6}$) than disorder-specific PRSs, indicating improved cross-ancestry transferability.

**Comparison with Alternative Transdiagnostic Models**

We examined whether more sophisticated transdiagnostic PRS models offer predictive advantages beyond the parsimonious CPG approach. We compared the CPG model with correlated four- and five-factor models building on the work by Grotzinger et al.[10,13] The four-factor model categorized 11 psychiatric disorders into compulsive/obsessive (F1), thought (F2), neurodevelopmental (F3), and internalizing (F4) disorders[10], while we further extended this framework to include a substance disorder (F5) dimension across the same 15 disorders[19-42] used to define the CPG (see **Methods**, detailed modeling results in **Supplementary Table 15**).

Our comparative analysis revealed that despite the superior model fit of the correlated multi-factor models in GenomicSEM (CFI=0.96 and 0.97 for four- and five-factor models respectively, compared to CFI=0.77 for CPG), the simpler CPG model demonstrated comparable or superior predictive performance across disorders (**Fig 5**, **Supplementary Tables 16-17**). The neurodevelopmental disorder factor (F3) showed the strongest performance for ADHD among the factor-specific PRSs, though still not surpassing CPG. Similarly, the substance use disorder factor (F5) showed comparable performance to CPG for alcohol use disorder, supporting the specificity of these latent factors. Notably, the internalizing factor (F4) demonstrated performance comparable to CPG across multiple disorders, exhibiting its broad transdiagnostic effects.

## DISCUSSION

In this study, we leveraged data from the *All of Us* Research Program,[14,15] a large and demographically diverse US cohort, to investigate the genetic risk underlying psychiatric disorders using a transdiagnostic approach. By combining comprehensive health survey data with robust genomic sample sizes, we demonstrate the utility of a CPG PRS for predicting both individual diagnoses and psychiatric comorbidity. To our knowledge, this is the first study to comprehensively evaluate how disorder-specific and transdiagnostic genetic risk contribute to psychiatric risk prediction within this national cohort.

Our findings underscore the potential utility of transdiagnostic risk models in psychiatric genomics and their implications for precision psychiatry.[43-45] The CPG PRS, reflecting shared genetic risk across 15 psychiatric disorders, outperformed standard individual disorder-based PRSs in predicting nearly all conditions. In particular, the CPG PRS explained significantly more variance beyond disorder-specific PRSs, including autism spectrum disorders (variance explained increased from 1.37% to 3.02%), bipolar disorder (1.10% to 4.47%), drug use disorder (3.64% to 8.18%), and social phobia (0.71% to 6.07%). These results highlight how integrating cross-disorder genetic information can substantially improve predictive accuracy beyond traditional, single-disorder models.

Although the CPG PRS accounted for a substantial proportion of genetic risk, disorder-specific PRSs retained independent contributions, albeit more modest in comparison. For ADHD (disorder-specific $R^2$ = 1.68%), autism spectrum disorder (disorder-specific $R^2$ = 0.82%), alcohol use disorder (disorder-specific $R^2$ = 1.98%, the strongest disorder-specific effect), and depression (disorder-specific $R^2$ = 0.95%), disorder-specific PRSs contributed significantly even after adjusting for the CPG component. This suggests that individual disorders harbor unique genetic mechanisms not fully captured by a purely transdiagnostic model. As GWAS sample sizes continue to grow, the precision of disorder-specific PRS estimates will likely improve, offering a more fine-grained understanding of genetic architecture underlying psychiatric risk.[46]

Another key observation was the robust association between the CPG PRS and psychiatric comorbidity. Individuals with higher CPG scores had significantly increased odds of carrying multiple psychiatric diagnoses, reinforcing the genetic basis of the high comorbidity observed in clinical settings.[47-49] This finding suggests that transdiagnostic genetic risk models could help identify individuals at heightened risk for multiple disorders, supporting their potential value for

early intervention and screening. Additionally, for disorders with limited GWAS discovery samples—such as eating disorders—the CPG PRS provides a practical advantage by leveraging statistical power from better-characterized conditions. However, it's important to acknowledge that the current effect sizes, while statistically significant, remain modest and would require substantial improvement before implementation in routine clinical practice. At present, these findings represent an important proof-of-concept rather than immediately actionable clinical tools.

Comparisons between the CPG model and more complex multi-factorial models underscore the distinction between explanatory and predictive frameworks.[50] Although correlated-factor models provide a more detailed representation of genetic architecture[8-11] with improved model fit in GenomicSEM,[12] the simpler CPG PRS performed comparably—or better—in predicting both individual disorders and broad psychiatric comorbidity. While multi-factorial models help elucidate nuanced inter-disorder relationships and we anticipate them to offer superior predictive performance with increasing GWAS sample sizes, simpler transdiagnostic PRSs provide a pragmatic approach to risk prediction with current data.[51,52]

Two primary mechanisms may explain the strong predictive performance of the CPG PRS. First, it likely captures biological risk factors that cut across broad diagnostic categories,[53] potentially reflecting fundamental neurobiological processes underlying psychiatric disorders. Supporting this, we observed significant genetic correlations across all psychiatric disorders and latent dimensions in factor models, part of which the CPG risk model may capture more efficiently. Interestingly, the internalizing factor PRS (reflecting shared risk for mood and anxiety disorders) performed similarly to the CPG PRS, suggesting that this dimension may be particularly salient for broad psychiatric vulnerability.[54] Second, by aggregating GWAS data across multiple disorders, the CPG PRS benefits from enhanced statistical power, enabling more precise detection of cross-disorder genetic effects than would be possible with single-disorder PRSs. This aggregation combines numerous small effects—each relatively noisy on its own—into a more reliable measure of genetic liability, resulting in a more robust transdiagnostic predictor that captures pleiotropic effects often obscured in disorder-specific approaches.

Our findings also illuminate the cross-ancestry performance of transdiagnostic PRSs. Although the CPG PRS retained relative predictive utility across populations, its effect sizes were notably diminished outside of EUR cohorts. For individual disorders, the correlation between AFR and EUR effect sizes was weak ($r = 0.10$), yet moderate for comorbidity ($r = 0.69$). By contrast, the

AMR group showed stronger alignment with EUR (r = 0.85 for individual disorders and r = 0.84 for comorbidity). These disparities likely reflect ancestry-specific linkage disequilibrium (LD) patterns,[55] as well as distinct environmental or clinical factors,[56] underscoring the critical need to expand diverse discovery datasets to improve the transferability of genetic risk prediction.[17,57]

Several limitations warrant consideration. First, the *All of Us* cohort is US-based, which may limit the global generalizability of our findings. Second, grouping individuals into broad ancestry groups (AFR, AMR, and EUR) may oversimplify the nuanced and continuous nature of human genetic variation. Third, reliance on self-reported data may introduce misclassification and recall bias,[58] though our prevalence and comorbidity rates aligned with prior EHR-based reports,[59] and various sensitivity analyses confirmed the robustness of our findings. Fourth, while the CPG PRS showed comparable or superior performance compared to multi-factor models, this may partly reflect greater statistical power from aggregating 15 GWAS datasets rather than more accurately capturing underlying genetic architecture. Finally, the predictive performance of current PRSs remains inadequate for clinical use,[60] especially in non-European populations. Continued improvements in GWAS coverage and PRS methodologies are essential for reliable clinical application.

In conclusion, our study highlights the hierarchical genetic architecture of psychiatric disorders, where a transdiagnostic PRS captures a significant portion of shared liability across conditions while disorder-specific PRSs added additional, though smaller, risk information. Cross-ancestry analyses further emphasize the urgent need for more diverse genomic datasets to improve predictive accuracy across populations.[61] Together, these findings advocate for an integrated strategy[62,63]—incorporating both transdiagnostic and disorder-specific genetic risk—to enhance risk assessment and advance precision psychiatry.

# ONLINE METHODS

### *All of Us* Research Program

The *All of Us* Research Program[14] is a nationwide, longitudinal study aimed at advancing precision medicine by recruiting a diverse cohort of participants across the U.S. The study collects extensive health-related data, including genetic information, electronic health records, lifestyle factors, medical history surveys, and environmental exposures. Our analysis utilizes the curated data v7 (2022Q4R9), which includes enrollment participant data up to July 1, 2022. The present study was performed between November 2023 and February 2025 using the *All of Us* Researcher Workbench, a cloud-based platform where approved researchers can analyze *All of Us* data.

### Outcome Measures

Psychiatric disorder diagnoses were obtained from the *All of Us Personal and Family Health History* survey (https://www.researchallofus.org/data-tools/survey-explorer/). Participants self-reported whether they had been diagnosed with each of the following mental health and substance use conditions: alcohol use disorder, anxiety reaction/panic disorder, attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder, bipolar disorder, depression, drug use disorder, eating disorder, personality disorder, post-traumatic stress disorder (PTSD), schizophrenia, social phobia, and other mental health conditions. To improve case validity, participants were classified as cases if they reported both a diagnosis and current clinical care (affirmative response to "*still seeing a doctor or health care provider?*"). Controls were those with no psychiatric disorder diagnoses.

We also computed an overall measure of concurrent psychiatric comorbidity as the total number of the aforementioned psychiatric disorders for each participant. These cumulative counts ranged from 0 (no psychiatric disorder) to 13 (participants with all surveyed conditions).

### Common Genetic Factor of Psychopathology (CPG)

To model shared genetic variance across psychiatric disorders, we assembled GWAS datasets for 15 psychiatric disorders: ADHD, alcohol use disorder, anorexia nervosa, anxiety disorder, autism spectrum disorder, bipolar disorder, major depression, PTSD, cannabis use disorder, cocaine addiction, nicotine dependence, obsessive compulsive disorder, opioid use disorder, Tourette's syndrome, schizophrenia (**Supplementary Table 3**). These datasets, derived from European-ancestry populations due to limited representation of other ancestries, included

conditions directly assessed in the *All of Us* survey data (e.g., ADHD, alcohol use disorder, schizophrenia, bipolar disorder) or closely aligned with survey-derived measures (e.g., substance use disorders (e.g., cannabis, cocaine, nicotine, and opioid) for drug use disorder). The inclusion of additional GWAS datasets was intended to enhance the robustness and generalizability of the common psychiatric genetic factor.

We fitted a common factor model in GenomicSEM[12] to estimate shared genetic variance across 15 psychiatric disorders. We selected this parsimonious approach for its conceptual clarity and to facilitate direct comparisons between the transdiagnostic effects and disorder-specific effects. Further details, including disorder loadings onto the common factor, are provided in **Supplementary Table 4**.

**Polygenic Risk Score Predictors**

Genotyping, imputation, and quality control (QC) procedures for the *All of Us* data have been detailed elsewhere.[15] PRS predictors were separately calculated for participants of African/African American (AFR), admixed Hispanic/Latino American (AMR), and European descent (EUR) using PRS-CS[64] and PLINK,[65] leveraging European LD reference panels from the 1000 Genomes Project to match discovery GWASs. Using the genome-wide genetic data estimated from the common factor model, we calculated $PRS_{CPG}$ as follows:

$$PRS_{CPG} = \sum_{j=1}^{M} \beta_j G_{ij}$$

where *i* indexes the individual, *j* indexes each variant, *M* is the total number of genetic variants included in the score, the weight ($\beta_j$ ) is the posterior effect obtained from PRS-CS[64] for variant *j*, $G_{ij}$ is the number of effect alleles, 0,1, and 2. To remove potential population substructure, we regressed PRSs on the top ten genetic PCs as below:

$$PRS_{CPG} = \beta_0 + \sum_{k=1}^{K} \gamma_k PC_{ik} + \epsilon$$

where $\beta_0$ is the intercept, $PC_{ik}$ are the *K* ancestry-specific genetic PCs, $\gamma_k$ are the regression coefficients for the k-th PC, and $\epsilon$ represents the residuals. After fitting the regression, we used the residual as the $PRS_{CPG}$.

Similarly, we calculated two types of PRSs for each of the 15 psychiatric disorders. The standard disorder PRS for a specific disorder (e.g., $PRS_{SCZ}$) was calculated in the same way as described previously using GWAS summary statistics specific to the corresponding disorder.

Additionally, we calculated disorder-specific PRS (e.g., $PRS_{SCZ\text{-}CPG}$) that indexes disorder-specific genetic risk independent of the shared genetic variance captured by the common factor. We used a linear regression approach where the standard disorder PRS was modeled as a function of $PRS_{CPG}$ and ancestry-specific genetic principal components (PCs). The residuals from this regression represent the disorder-specific PRS:

$$\mathrm{PRS}_{\mathrm{SCZ}} = \beta_0 + \beta_1 \cdot \mathrm{PRS}_{\mathrm{CPG}} + \sum_{k=1}^{K} \gamma_k \cdot \mathrm{PC}_k + \epsilon$$

where $\beta_0$ is the intercept, $\beta_1$ is the regression coefficient for $PRS_{CPG}$, $PC_k$ are *K* ancestry-specific genetic PCs, $\gamma_k$ are the regression coefficients for the k-th PC, and $\epsilon$ represents the residuals.

To account for ancestry-specific genetic architecture, derivation of genetic principal components (PCs) and PRS calculation were performed separately within each ancestry group.

**Association Analysis of PRS and Outcome Measures**

For binary outcome measures (individual disorder diagnosis), we examined the associations between each PRS predictor and outcome measure using logistic regression via the *glm* function in the *lme4* R package (version 4.4.2). For comorbidity variables, which represent the count of concurrent psychiatric diagnoses, we used negative binomial regression. This approach extends Poisson regression by allowing for overdispersion, a feature noticed in our count data, and was tested using the *glm.nb* function from the MASS R package. All regression models included the following covariates: age, self-reported sex, genotyping batch, and the top ten genetic PCs to adjust for within-ancestry population structure. Continous covariates were standardized.

To evaluate the proportion of variance explained by each PRS, we calculated *Nagelkerke's pseudo* $R^2$ for binary outcomes and adjusted $R^2$ for continuous outcomes. As observed $R^2$ were not directly comparable between disorders with different sample prevalence, we calculated liability-based $R^2$ as follows[66]:

$$R^2_{\text{liability}} = R^2_{\text{obs}} \times \frac{K(1-K)}{p(1-p)} \times \frac{\phi\left(\Phi^{-1}(K)\right)^2}{\left[K(1-K)\right]^2},$$

where $K$ is the population prevalence, $p$ is the sample prevalence, $\Phi^{-1}(K)$ is the liability threshold for prevalence $K$ under a standard normal distribution, and $\phi$ is its probability density function.

The statistical significance of variance uniquely attributable to PRS predictors was assessed using likelihood ratio tests (LRTs). LRTs compared full models (including the PRS) to null models (excluding the PRS), providing a robust measure of the incremental predictive power of PRS predictors. All p-values were corrected for multiple comparisons using the false discovery rate (FDR) to control for type I error across the large number of tests conducted.

Pairwise comparisons of odds ratios (ORs) between the three groups (CPG vs standard disorder vs. disorder-specific PRSs) were conducted using Wilcoxon rank-sum tests. Statistical significance was set at $p < 0.05$ after multiple testing correction. Group medians were calculated to quantify the magnitude of differences between pairs.

Given that the GWAS datasets used for PRS derivation were exclusively based on European-ancestry individuals, the primary analyses were conducted on participants of European descent to minimize potential confounding due to ancestry-specific differences in genetic architecture. To examine the cross-ancestry generalizability of PRS predictors, separate analyses were conducted on AFR and AMR participants and these results are presented in a dedicated section.

**Sensitivity Analyses**

To ensure robustness, we conducted sensitivity analyses addressing diagnostic criteria, control group definitions, and comorbidity measures. Associations were tested using both lifetime and current diagnoses to evaluate the impact of temporal scope on PRS-outcome relationships. Control group heterogeneity was assessed by comparing results using individuals without current diagnoses versus those without any lifetime diagnoses. For comorbidity outcomes, winsorization was applied to cap extreme values at the 95th percentile, mitigating the influence of outliers and ensuring consistency in findings across transformations.

**Correlated Factor Models**

To assess the predictive performance of the CPG risk model in comparison to more complex latent factor models, we developed a five correlated factor model using the same 15 psychiatric disorders (**Supplementary Table 3**) analyzed in our CPG approach using GenomicSEM.[12] First, we estimated genetic correlations between all disorder pairs using LD Score Regression (LDSC).[67] We then performed exploratory factor analysis (EFA) on the genetic correlation matrix to determine the optimal number of factors, evaluating models with 2-6 factors. Model selection was guided by comparative fit indices (CFI) and standardized root mean square residual (SRMR). The five-factor solution demonstrated superior fit compared to models with fewer factors (p-value of $\chi^2 = 4.49 \times 10^{-95}$; CFI = 0.97, SRMR = 0.78) while remaining parsimonious and interpretable. The resulting five correlated factors were characterized as: compulsive/obsessive disorders (F1_5), schizophrenia/bipolar disorders (F2_5), neurodevelopmental disorders (F2_5), internalizing disorders (F4_5), and substance use disorders (F5_5). The factor loadings and inter-factor correlations are fully detailed in **Supplementary Table 15**. Additionally, we compared our model with the previously published four-factor model by Grotzinger et al. (2022),[10] which identified factors representing compulsive/obsessive disorders (F1_4), thought disorders (F2_4), neurodevelopmental disorders (F3_4), and internalizing disorders (F4_4). We refer to original publication[10] for further details of the four-factor model.

For both the five-factor and four-factor models, we generated factor-specific PRSs using the GenomicSEM.[12] We conducted multivariate genome-wide association analyses for each factor while accounting for the correlations among factors. This approach allowed us to identify genetic variants specifically associated with each latent dimension while controlling for shared genetic influences across dimensions. For each factor-specific GWAS summary statistics, we employed PRS-CS[64] and PLINK[65] to calculate PRSs using the same setting as the CPG model.

**Ethics Review**

The *All of Us* Research Program conducted data collection under centralized Institutional Review Board (IRB) approval, with informed consent obtained from participants. This study adhered to the ethical guidelines outlined in the *All of Us* Code of Conduct. The Massachusetts General Hospital IRB determined that this study was exempt from human subjects research as it involves a secondary analysis of de-identified data from the *All of Us* Research Program.

## Tables

**Table 1. Prevalence of 13 Psychiatric Conditions Assessed in the *All of Us* Personal and Family Health Survey (N=102,091).** Participants were classified as cases if they reported both a diagnosis and current clinical care for a given condition. Sample prevalence represents the proportion of participants meeting these criteria within the study sample. Population prevalence (12-month) estimates are provided for reference from external epidemiological studies. "-" indicates that a population prevalence estimate was not available.

| Surveyed Conditions | Case Number | Sample Prevalence (%) | Population Prevalence (12 month) (%) |
|---|---|---|---|
| Depression | 22,546 | 24.07 | 8.3 |
| Anxiety Reaction/Panic Disorder | 17,451 | 19.7 | 2.7 |
| Post-Traumatic Stress Disorder | 5,590 | 7.29 | 3.6 |
| Attention-Deficit Hyperactivity Disorder | 3,899 | 5.2 | 4.4 |
| Bipolar Disorder | 3,349 | 4.5 | 2.8 |
| Social Phobia | 1,350 | 1.86 | 7.1 |
| Alcohol Use Disorder | 1,197 | 1.66 | 10.2 |
| Eating Disorder | 1,152 | 1.59 | 1.2 |
| Other Mental Health or Substance Use Condition | 998 | 1.38 | - |
| Personality Disorder | 868 | 1.21 | 1.4 |
| Drug Use Disorder | 809 | 1.12 | 3.9 |
| Schizophrenia | 449 | 0.63 | 0.64 |
| Autism Spectrum Disorder | 273 | 0.38 | 2.21 |

## FIGURE and FIGURE CAPTIONS

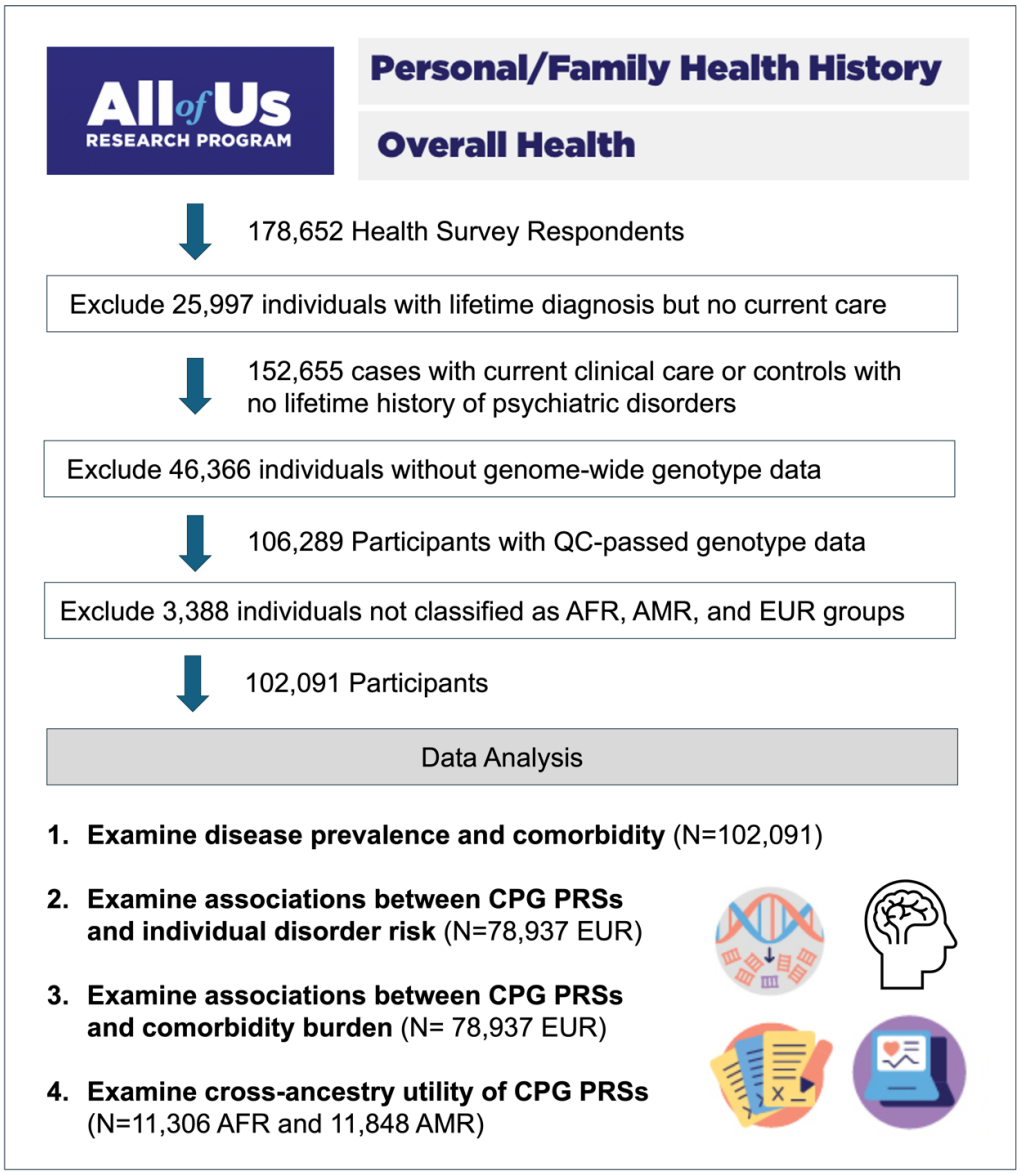
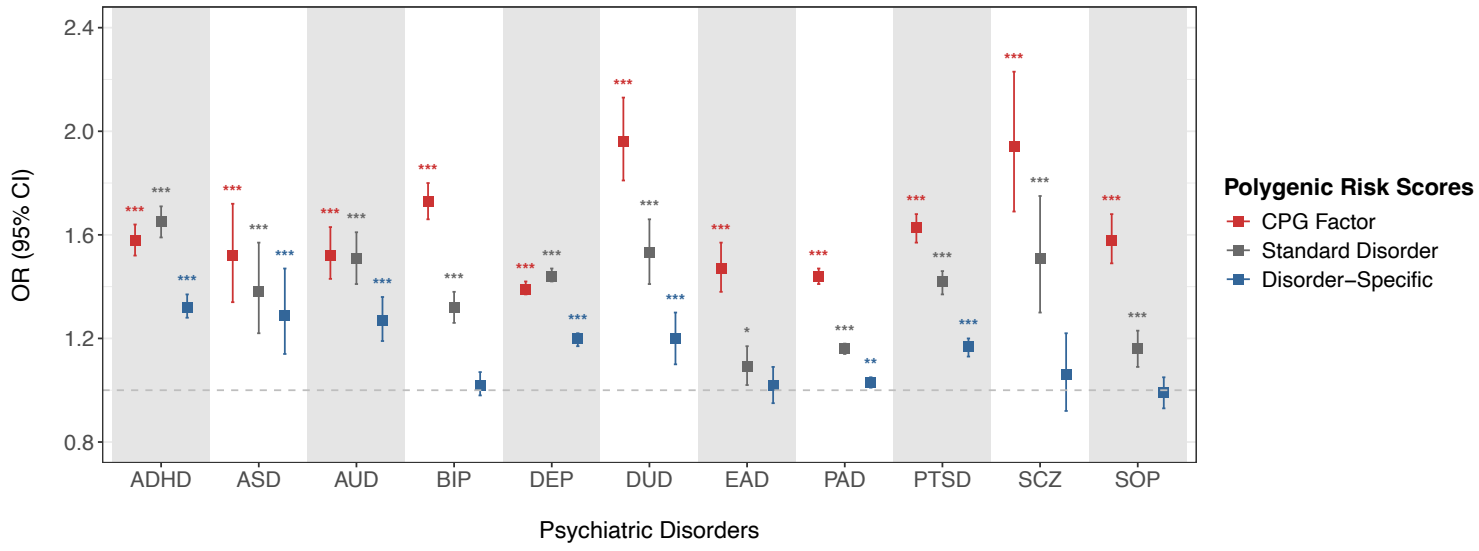



**Figure 1. Data preparation and analysis workflow for the investigations of the common psychiatric genetic factor (CPG), psychiatric disorder risk, and comorbidity burden.** The analysis cohort included a total of 102,091 participants, representing AFR (African/African American), AMR (Hispanic/Latin American), and EUR (European) ancestry groups. Given that the GWAS datasets used to generate polygenic risk scores (PRSs) were derived from EUR populations, our primary analysis focused on EUR participants (n=78,937), with separate cross-ancestry analyses conducted for AFR (n=11,306) and AMR (n=11,848) ancestry groups.

**A.**


OR (95% CI)
2.4
2.0
1.6
1.2
0.8
ADHD
ASD
AUD
BIP
DEP
DUD
EAD
PAD
PTSD
SCZ
SOP
Psychiatric Disorders
Polygenic Risk Scores
CPG Factor
Standard Disorder
Disorder–Specific

**B.**

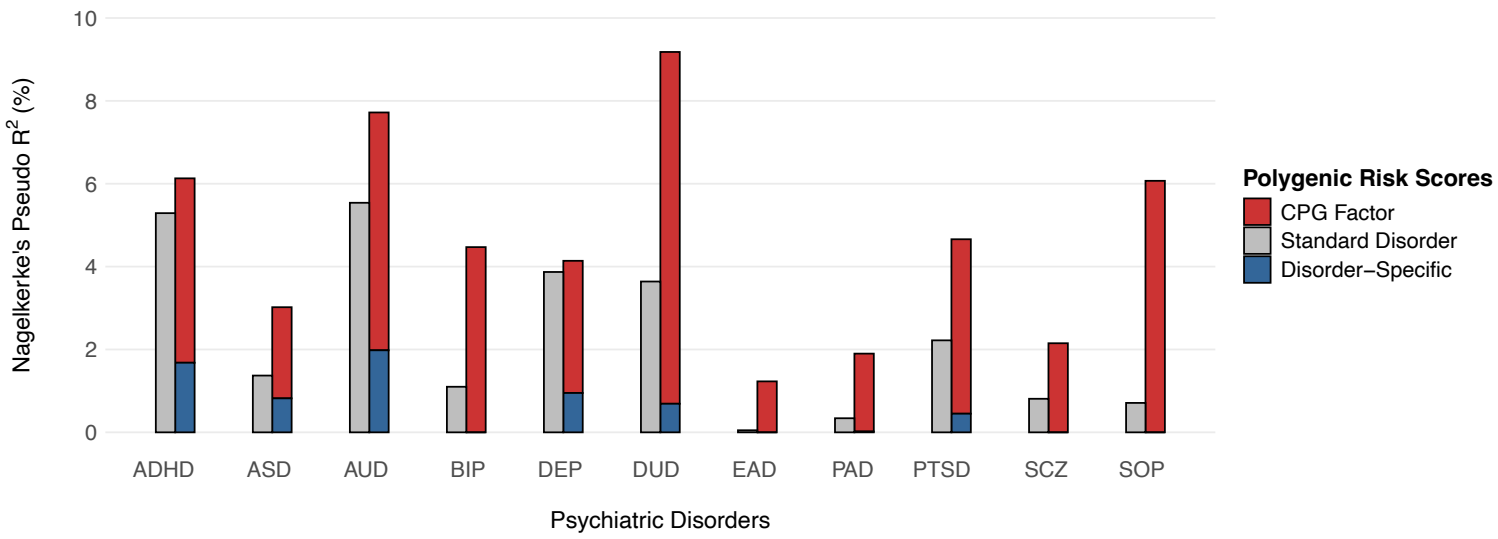

Nagelkerke's Pseudo $R^2$ (%)
10
8
6
4
2
0
ADHD
ASD
AUD
BIP
DEP
DUD
EAD
PAD
PTSD
SCZ
SOP
Psychiatric Disorders
Polygenic Risk Scores
CPG Factor
Standard Disorder
Disorder–Specific

**Figure 2. Association findings of the common psychiatric genetic factor (CPG) and disorder-targeted polygenic risk scores (PRSs) across 11 psychiatric disorders in the European ancestry population**. A. Odds ratios (OR) plot. The forest plot displays OR with 95% confidence intervals (CI) for associations between psychiatric disorders and three types of polygenic risk scores (PRS): standard disorder PRS (gray), disorder-specific PRS (dark blue), and CPG PRS (red). Significance levels are indicated by asterisks: * FDR (false discovery rate) $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. B. Phenotypic variance explained by PRSs. The y-axis shows liability-based Nagelkerke's pseudo $R^2$ values (%) for three PRS categories: standard disorder PRS (gray), disorder-specific PRS (dark blue), and CPG PRS (red). For each disorder, individual PRS effects are shown in the left bar, while the stacked bars on the right display the additive effects of the shared (CPG) and disorder-specific PRS components. Disorders are arranged by alphabetical orders. ADHD: attention-deficit/hyperactivity disorder; ASD: autism spectrum disorder; AUD: alcohol use disorder; BPD: bipolar disorder; EAD: eating disorder; PAD: anxiety reaction/ panic disorder; PTSD: post-traumatic stress disorder; SCZ: schizophrenia; SOP: social phobia.

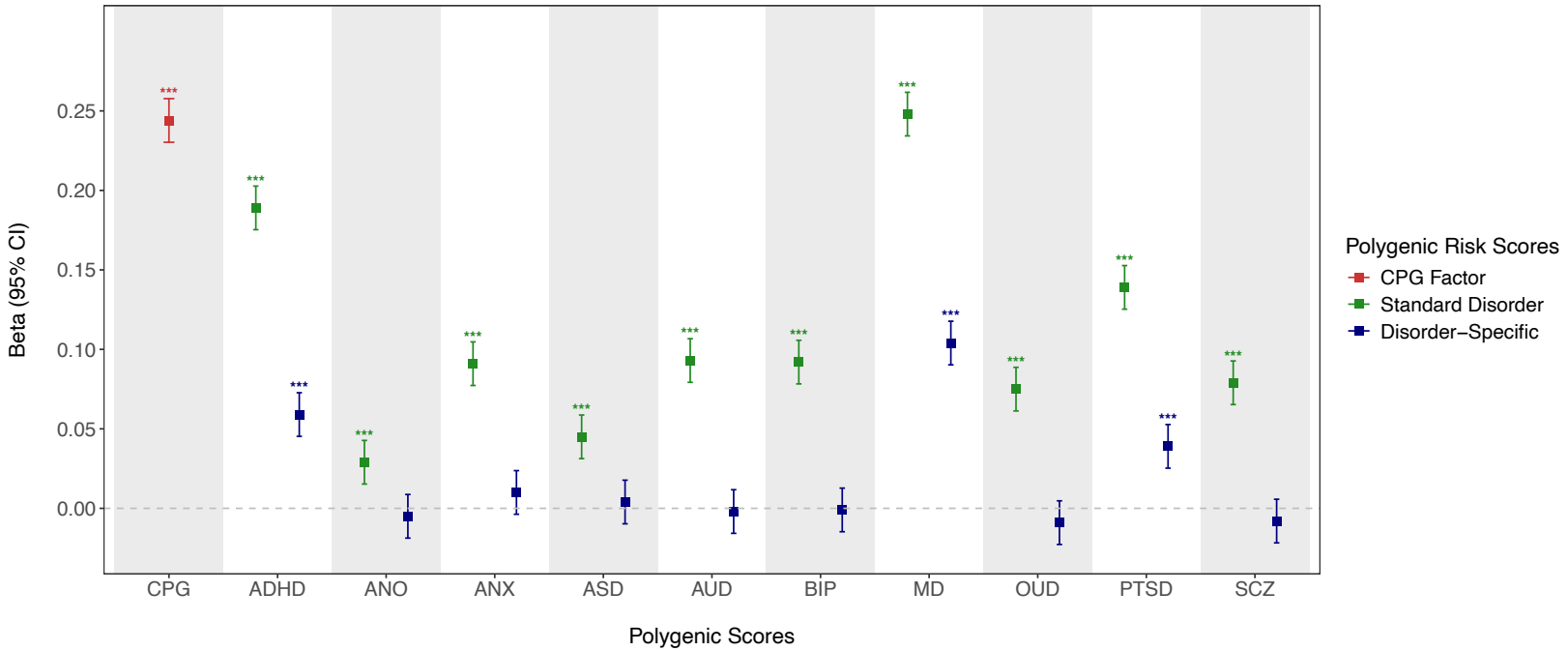



**Figure 3. Differential associations of the common psychiatric genetic factor (CPG) and disorder-specific polygenic risk scores (PRSs) with psychiatric comorbidity in the European ancestry population.** The plot displays beta coefficients and with 95% confidence intervals (CI) for associations between PRSs and psychiatric comorbidity. Three types of PRS are compared: standard disorder PRS (gray), disorder-specific PRS with CPG component removed (navy), and CPG PRS (red). The x-axis shows different psychiatric disorder PRSs: MD (major depression), ADHD (attention-deficit/hyperactivity disorder), PTSD (post-traumatic stress disorder), AUD (alcohol use disorder), BIP (bipolar disorder), ANX (anxiety), SCZ (schizophrenia), OUD (opioid use disorder), CUD (cannabis use disorder), NID (nicotine dependence), ASD (autism spectrum disorder), CAD (cocaine addiction), TS (Tourette syndrome), ANO (anorexia nervosa), and OCD (obsessive-compulsive disorder). Significance levels are indicated by asterisks: * FDR (false discovery rate) $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

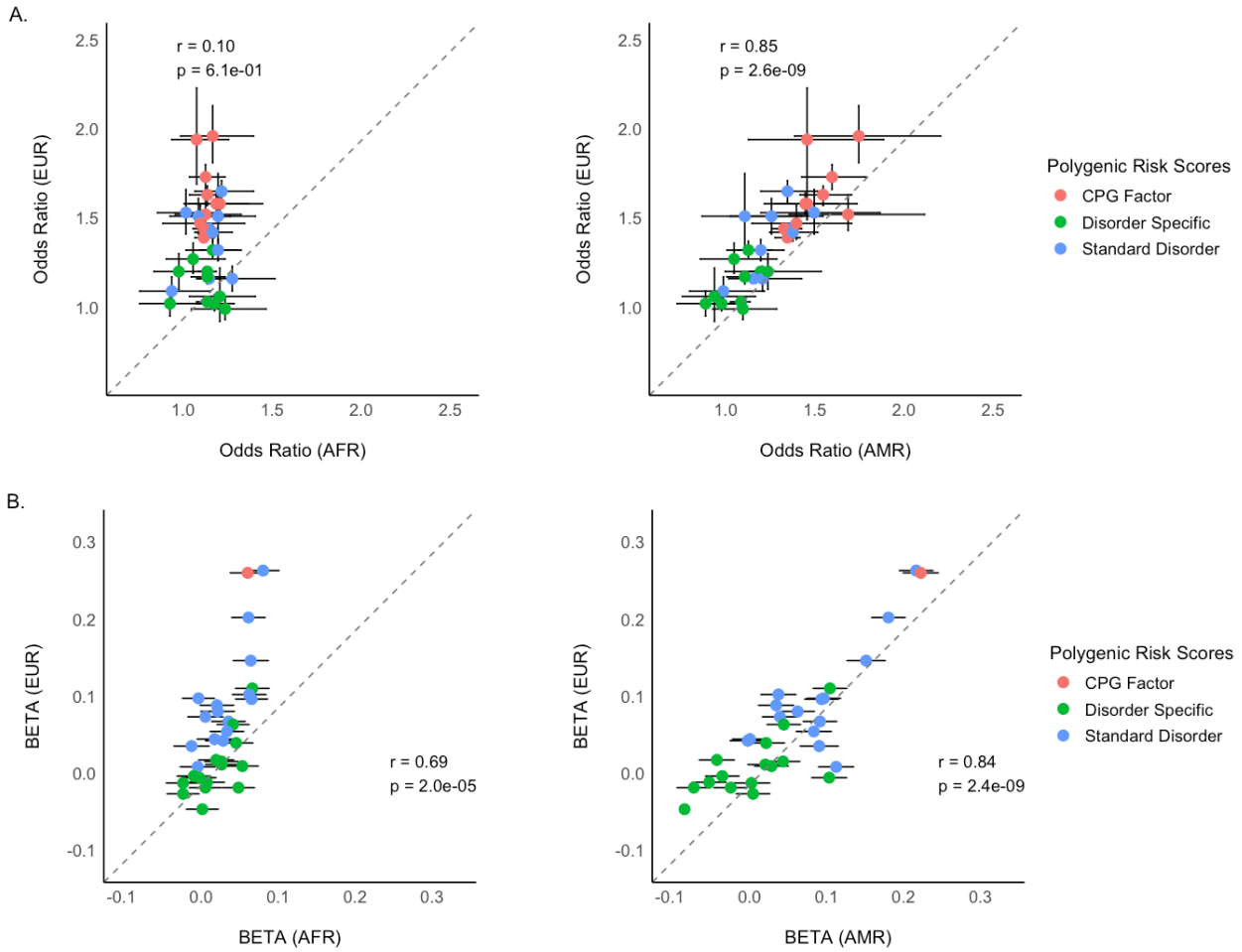



**Figure 4. Cross-ancestry applicability of European-derived polygenic risk scores (PRSs) in African/African American (AFR) and Hispanic/Latin American (AMR) ancestry groups.** **A.** Odds ratio (OR) plot showing the associations between PRSs (CPG factor, disorder-specific, and standard disorder PRSs) and 11 psychiatric disorders surveyed in the *All of Us* data. ORs for each ancestry group are compared against those derived from the European (EUR) cohort. The dotted line represents the unity line, with correlations (r) and significance (p-values) shown for AFR and AMR relative to EUR. **B.** Regression beta plot assessing the associations between the CPG and 15 psychiatric disorder PRSs and psychiatric comorbidity burden, measured as the number of concurrent psychiatric conditions. Effect sizes (β coefficients) for AFR and AMR groups are compared with those from the EUR cohort. The CPG PRS consistently explains the largest phenotypic variance in comorbidity burden across ancestry groups, with significant cross-ancestry correlations observed for both AMR and AFR populations.

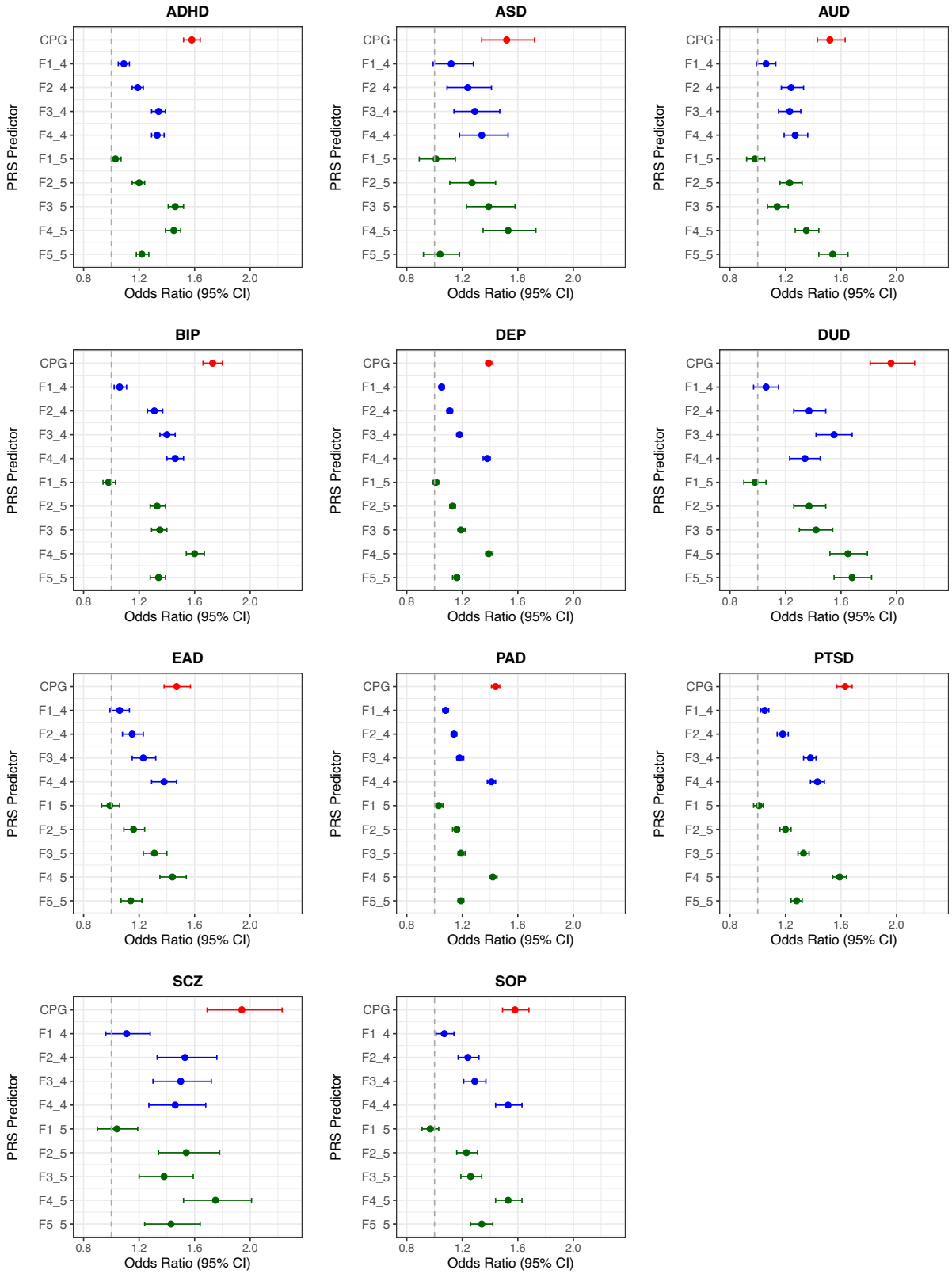

ADHD
ASD
AUD
BIP
DEP
DUD
EAD
PAD
PTSD
SCZ
SOP
PRS Predictor
CPG
F1_4
F2_4
F3_4
F4_4
F1_5
F2_5
F3_5
F4_5
F5_5
0.8
1.2
1.6
2.0
Odds Ratio (95% CI)

**Figure 5. Comparison of PRSs based on the common psychiatric genetic factor (CPG) and correlated multi-factor models across 11 psychiatric conditions.** The figure consists of 11 forest plots, each representing one of 11 psychiatric disorders: ADHD, ANX/PD (anxiety reaction/panic disorder), ASD (autism spectrum disorder), AUD (alcohol use disorder), BIP (bipolar disorder), DEP (depression), DUD (drug use disorder), ED (eating disorder), PTSD (post-traumatic stress disorder), SCZ (schizophrenia), and SOP (social phobia). Each plot displays odds ratios (OR) with 95% confidence intervals for different PRS predictors. The predictors include the Common Psychiatric Genomic PRS (CPG, shown in red), four correlated-factor model PRSs (F1_4 through F4_4, shown in blue), and five correlated-factor model PRSs (F1_5 through F5_5, shown in green). The four correlated factor model, introduced by Grotzinger et al. (2022)[10], represent: F1_4: compulsive/obsessive, F2_4: thought disorder, F3_4: neurodevelopmental disorder, and F4_4: internalizing disorder. The five correlated factor model, estimated in this study (Supplementary Table 15), represent: F1_5: compulsive/obsessive, F2_5: schizophrenia/bipolar disorder, F3_5: neurodevelopmental disorder, F4_5: internalizing disorder, and F5_5: substance use disorder. The x-axis represents OR ranging from 0.8 to 2.3, with a vertical line at 1.0 indicating no effect. Data points to the right of this line represent positive associations between genetic risk scores and disorder prevalence.

## ACKNOWLEDGEMENT

This work was supported by grants to PHL from the U.S. National Institute of Health (R01MH119243, R01GM148494) and by MassGeneral Brigham Department of Psychiatry. We gratefully acknowledge *All of Us* participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's *All of Us* Research Program for making available the participant data release (v7) examined in this study. The *All of Us* Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. This study used GWAS summary statistics obtained from the 23andMe, Inc. cohort. We would like to thank the research participants and employees of 23andMe, Inc. for making this work possible. GWAS summary statistics data used in this study were also obtained from the database of Genotypes and Phenotypes (dbGaP) under accession number phs001672.v12.p1. We thank the investigators who contributed the phenotype and genotype data to dbGaP. We thank Million Veteran Program (MVP) staff, researchers, and volunteers, who have contributed to MVP, and especially participants who previously served their country in the military and now generously agreed to enroll in the study. The MVP Program is detailed in the study by Gaziano, J.M. et al.[68] and is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration. The dbGaP dataset(s) used in this study were accessed with authorized approval from the National Institutes of Health (NIH). This study also used GWAS summary statistics obtained from the Psychiatric Genomics Consortium (PGC). We would like to thank the research participants and the investigators of these studies for making the data publicly available, which was essential to conducting this study possible.

## DECLARATION OF INTERESTS

All authors have no interests to declare.

## ROLE OF THE FUNDING SOURCE

The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. We have not received any financial gain from a pharmaceutical company or other agency to write this article. Authors were not precluded from accessing data in the study, and they accept responsibility to submit for publication.

## DATA AVAILABILITY

The *All of Us* Research Program data used in this study are available to authorized researchers through the *All of Us* Research website (https://allofus.nih.gov). For GWAS summary statistics from 23andMe, Inc. are available to the researchers through a separate permission process via https://research.23andme.com/dataset-access/. GWAS summary statistics from the MVP data are available on dbGaP (accession no. phs001672). For all other publicly available GWAS summary statistics, we provided the download links in the **Supplementary Table 3**.

## CODE AVAILABILITY

For all analyses, we used open-source software packages from R (v4.3) available in the *All of Us* Research Workbench (https://www.researchallofus.org/data-tools/workbench/), visualization using the R-studio desktop applications (https://posit.co/download/rstudio-desktop/), and publicly available methods GenomicSEM (https://github.com/GenomicSEM/GenomicSEM), PLINK v2.0 (https://www.cog-genomics.org/plink/2.0/), and PRS-CS (https://github.com/getian107/PRScs) as summarized in the Methods section.

## REFERENCES


1 Kendler, K. S. The nature of psychiatric disorders. *World Psychiatry* **15**, 5-12 (2016). https://doi.org:10.1002/wps.20292

2 Engstrom, E. J. & Weber, M. M. The Directions of psychiatric research by Emil Kraepelin. 1887. *Hist Psychiatry* **16**, 345-364 (2005). https://doi.org:10.1177/0957154X05056763

3 Adam, D. Mental health: On the spectrum. *Nature* **496**, 416-418 (2013). https://doi.org:10.1038/496416a

4 Cicero, D. C. *et al.* State of the Science: The Hierarchical Taxonomy of Psychopathology (HiTOP). *Behav Ther* **55**, 1114-1129 (2024). https://doi.org:10.1016/j.beth.2024.05.001

5 Kotov, R. *et al.* The Hierarchical Taxonomy of Psychopathology (HiTOP): A Quantitative Nosology Based on Consensus of Evidence. *Annu Rev Clin Psychol* **17**, 83-108 (2021). https://doi.org:10.1146/annurev-clinpsy-081219-093304

6 Insel, T. *et al.* Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* **167**, 748-751 (2010). https://doi.org:10.1176/appi.ajp.2010.09091379

7 Cuthbert, B. N. The role of RDoC in future classification of mental disorders . *Dialogues Clin Neurosci* **22**, 81-85 (2020). https://doi.org:10.31887/DCNS.2020.22.1/bcuthbert

8 Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* **45**, 984-994 (2013). https://doi.org:10.1038/ng.2711

9 Lee, P., Anttila V, Won H, Feng YA, Rosenthal J, Zhu Z, et al. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482 (2019).

10 Grotzinger, A., Mallard TT, Akingbuwa WA, Ip HF, Adams MJ, Lewise CM, McIntosh AM, Grove J, Dalsgaard S, Peter-Lesch K, Strom N, Meier SM, Mattheisen M, Borglum AD, Mors O, Breen G, Lee PH, Kendler KS, Smoller JS, Tucker-Drob EM, Nivard MG. Genetic Architecture of 11 Major Psychiatric Disorders at Biobehavioral, Functional, Genomic, and Molecular Genetic Levels of Analysis. *Nat Genet* **54**, 548-559 (2022).

11 Grotzinger, A. D. *et al.* The Landscape of Shared and Divergent Genetic Influences across 14 Psychiatric Disorders. *medRxiv*, 2025.2001.2014.25320574 (2025). https://doi.org:10.1101/2025.01.14.25320574

12 Grotzinger, A. *et al.* Genomic SEM provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav* **3**, 513-525 (2019).

13 Grotzinger, A. D. *et al.* Transcriptome-Wide Structural Equation Modeling of 13 Major Psychiatric Disorders for Cross-Disorder Risk and Drug Repurposing. *JAMA Psychiatry* **80**, 811-821 (2023). https://doi.org:10.1001/jamapsychiatry.2023.1808

14 All of Us Research Program, I. *et al.* The "All of Us" Research Program. *N Engl J Med* **381**, 668-676 (2019). https://doi.org:10.1056/NEJMsr1809937

15 All of Us Research Program Genomics, I. Genomic data in the All of Us Research Program. *Nature* **627**, 340-346 (2024). https://doi.org:10.1038/s41586-023-06957-x

16 Allegrini, A. G. *et al.* The p factor: genetic analyses support a general dimension of psychopathology in childhood and adolescence. *J Child Psychol Psychiatry* **61**, 30-39 (2020). https://doi.org:10.1111/jcpp.13113

17 Martin, A., Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **51**, 584-591 (2019).

18 Prive, F. *et al.* Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am J Hum Genet* **109**, 373 (2022). https://doi.org:10.1016/j.ajhg.2022.01.007

19 Brasher, M. S. *et al.* Testing associations between human anxiety and genes previously implicated by mouse anxiety models. *Genes Brain Behav* **22**, e12851 (2023). https://doi.org:10.1111/gbb.12851

20 Burstein, D. *et al.* Genome-wide analysis of a model-derived binge eating disorder phenotype identifies risk loci and implicates iron metabolism. *Nat Genet* **55**, 1462-1470 (2023). https://doi.org:10.1038/s41588-023-01464-1

21 Cabana-Dominguez, J., Shivalikanjli, A., Fernandez-Castillo, N. & Cormand, B. Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions. *Prog Neuropsychopharmacol Biol Psychiatry* **94**, 109667 (2019). https://doi.org:10.1016/j.pnpbp.2019.109667

22 Demontis, D. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat Genet* **55**, 198-208 (2023). https://doi.org:10.1038/s41588-022-01285-8

23 Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75 (2019). https://doi.org:10.1038/s41588-018-0269-7

24 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444 (2019). https://doi.org:10.1038/s41588-019-0344-8

25 Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343-352 (2019). https://doi.org:10.1038/s41593-018-0326-7

26 International Obsessive Compulsive Disorder Foundation Genetics, C. & Studies, O. C. D. C. G. A. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry* **23**, 1181-1188 (2018). https://doi.org:10.1038/mp.2017.154

27 Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry* **7**, 1032-1045 (2020). https://doi.org:10.1016/S2215-0366(20)30339-4

28 Levey, D. F. *et al.* Reproducible Genetic Risk Loci for Anxiety: Results From approximately 200,000 Participants in the Million Veteran Program. *Am J Psychiatry* **177**, 223-232 (2020). https://doi.org:10.1176/appi.ajp.2019.19030256

29 Levey, D. F. *et al.* Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat Neurosci* **24**, 954-963 (2021). https://doi.org:10.1038/s41593-021-00860-2

30 Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet* **53**, 817-829 (2021). https://doi.org:10.1038/s41588-021-00857-4

31 Nievergelt, C. M. *et al.* International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat Commun* **10**, 4558 (2019). https://doi.org:10.1038/s41467-019-12576-w

32 Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry* **21**, 1391-1399 (2016). https://doi.org:10.1038/mp.2015.197

33 Polimanti, R. *et al.* Leveraging genome-wide data to investigate differences between opioid use vs. opioid dependence in 41,176 individuals from the Psychiatric Genomics Consortium. *Mol Psychiatry* **25**, 1673-1687 (2020). https://doi.org:10.1038/s41380-020-0677-9

34 Rovira, P. *et al.* Shared genetic background between children and adults with attention deficit/hyperactivity disorder. *Neuropsychopharmacology* **45**, 1617-1626 (2020). https://doi.org:10.1038/s41386-020-0664-5

35 Sanchez-Roige, S. *et al.* Genome-wide association study of problematic opioid prescription use in 132,113 23andMe research participants of European ancestry. *Mol Psychiatry* **26**, 6209-6217 (2021). https://doi.org:10.1038/s41380-021-01335-3

36 Sanchez-Roige, S. *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am J Psychiatry* **176**, 107-118 (2019). https://doi.org:10.1176/appi.ajp.2018.18040369

37 Stein, M. B. *et al.* Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the Million Veteran Program. *Nat Genet* **53**, 174-184 (2021). https://doi.org:10.1038/s41588-020-00767-x

38 Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508 (2022). https://doi.org:10.1038/s41586-022-04434-5

39 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348 (2019). https://doi.org:10.1038/s41588-019-0481-0

40 Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat Genet* **51**, 1207-1214 (2019). https://doi.org:10.1038/s41588-019-0439-2

41 Yu, D. *et al.* Interrogating the Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am J Psychiatry* **176**, 217-227 (2019). https://doi.org:10.1176/appi.ajp.2018.18070857

42 Zhou, H. *et al.* Association of OPRM1 Functional Coding Variant With Opioid Use Disorder: A Genome-Wide Association Study. *JAMA Psychiatry* **77**, 1072-1080 (2020). https://doi.org:10.1001/jamapsychiatry.2020.1206

43 Nierenberg, A. Precision Psychiatry and a Vision for the Future. *Psychiatric Annals* **51**, 206-206 (2021).

44 Coutts, F., Koutsouleris, N. & McGuire, P. Psychotic disorders as a framework for precision psychiatry. *Nat Rev Neurol* **19**, 221-234 (2023). https://doi.org:10.1038/s41582-023-00779-1

45 Salazar de Pablo, G. *et al.* Implementing Precision Psychiatry: A Systematic Review of Individualized Prediction Models for Clinical Practice. *Schizophr Bull* **47**, 284-297 (2021). https://doi.org:10.1093/schbul/sbaa120

46 Sullivan, P. F. & Geschwind, D. H. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162-183 (2019). https://doi.org:10.1016/j.cell.2019.01.015

47 Forbes, M., Tackett JL, Markon KE, Krueger RF. Beyond comorbidity: Toward a dimensional and hierarchal approach to understanding psychopathology across the lifespan. *Dev Psychopathol* **28**, 9711-9986 (2016).

48 Cerda, M., Sagdeo, A., Johnson, J. & Galea, S. Genetic and environmental influences on psychiatric comorbidity: a systematic review. *J Affect Disord* **126**, 14-38 (2010). https://doi.org:10.1016/j.jad.2009.11.006

49 Plana-Ripoll, O. *et al.* Exploring Comorbidity Within Mental Disorders Among a Danish National Population. *JAMA Psychiatry* **76**, 259-270 (2019). https://doi.org:10.1001/jamapsychiatry.2018.3658

50 Clapp Sullivan, M. L. *et al.* Beyond the factor indeterminacy problem using genome-wide association data. *Nat Hum Behav* **8**, 205-218 (2024). https://doi.org:10.1038/s41562-023-01789-1

51 Woodward, D. J. *et al.* Leveraging pleiotropy for the improved treatment of psychiatric disorders. *Mol Psychiatry* **30**, 705-721 (2025). https://doi.org:10.1038/s41380-024-02771-7

52 Plomin, R. & von Stumm, S. Polygenic scores: prediction versus explanation. *Mol Psychiatry* **27**, 49-52 (2022). https://doi.org:10.1038/s41380-021-01348-y

53 Lee, P. H., Feng, Y. A. & Smoller, J. W. Pleiotropy and Cross-Disorder Genetics Among Psychiatric Disorders. *Biol Psychiatry* **89**, 20-31 (2021). https://doi.org:10.1016/j.biopsych.2020.09.026

54 Konrad, A. C. *et al.* Risk factors for internalizing symptoms: The influence of empathy, theory of mind, and negative thinking processes. *Hum Brain Mapp* **45**, e26576 (2024). https://doi.org:10.1002/hbm.26576

55 Duncan, L., Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, Peterson R, Domingue B. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* **10**, 3328 (2019).

56 Shi, Y. *et al.* Racial Disparities in Diagnosis of Attention-Deficit/Hyperactivity Disorder in a US National Birth Cohort. *JAMA Netw Open* **4**, e210321 (2021). https://doi.org:10.1001/jamanetworkopen.2021.0321

57 Kachuri, L. *et al.* Principles and methods for transferring polygenic risk scores across global populations. *Nat Rev Genet* **25**, 8-25 (2024). https://doi.org:10.1038/s41576-023-00637-2

58 Cai, N. *et al.* Assessment and ascertainment in psychiatric molecular genetics: challenges and opportunities for cross-disorder research. *Mol Psychiatry* **30**, 1627-1638 (2025). https://doi.org:10.1038/s41380-024-02878-x

59 Barr, P. B., Bigdeli, T. B. & Meyers, J. L. Prevalence, Comorbidity, and Sociodemographic Correlates of Psychiatric Disorders Reported in the All of Us Research Program. *JAMA Psychiatry* **79**, 622-628 (2022). https://doi.org:10.1001/jamapsychiatry.2022.0685

60 Jayasinghe, D., Eshetie, S., Beckmann, K., Benyamin, B. & Lee, S. H. Advancements and limitations in polygenic risk score methods for genomic prediction: a scoping review. *Hum Genet* **143**, 1401-1431 (2024). https://doi.org:10.1007/s00439-024-02716-8

61 Li, B. S., Cai, T. & Duan, R. Targeting Underrepresented Populations in Precision Medicine: A Federated Transfer Learning Approach. *Ann Appl Stat* **17**, 2970-2992 (2023). https://doi.org:10.1214/23-AOAS1747

62 Gu, T., Han, Y. & Duan, R. Robust angle-based transfer learning in high dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology* (2024). https://doi.org:10.1093/jrsssb/qkae111

63 Li, R. *et al.* A One-Shot Lossless Algorithm for Cross-Cohort Learning in Mixed-Outcomes Analysis. *medRxiv* (2024). https://doi.org:10.1101/2024.01.09.24301073

64 Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* **10**, 1776 (2019). https://doi.org:10.1038/s41467-019-09718-5

65 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015). https://doi.org:10.1186/s13742-015-0047-8

66 Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* **36**, 214-224 (2012). https://doi.org:10.1002/gepi.21614

67 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295 (2015). https://doi.org:10.1038/ng.3211

68 Gaziano, J. M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-223 (2016). https://doi.org:10.1016/j.jclinepi.2015.09.016