SUPPLEMENTARY INFORMATION

Comorbidity Profiles of Posttraumatic Stress Disorder Across the Medical Phenome

Hicks *et al.*

Detailed sample information for each Biobank

VUMC's BioVu. Diagnostic codes from the International Classification of Disease versions 9 or 10 (ICD-9/10), were obtained from VUMC's EHR database of >3.2 million individuals. These data are linked to VUMC's biobank, BioVU, with DNA collected from >250,000 individuals. The VUMC Institutional Review Board (IRB) approved this study (protocol #190418), and all participants provided written informed consent. At the time of analysis, genotype data were available for 73,488 unrelated individuals of European and African ancestry. Details on the quality control procedures are described elsewhere(1).

MSSM's Bio*Me*. ICD-9/10 codes from MSSM's EHR database were obtained for 31,704 individuals enrolled in the Bio*Me* biobank. Bio*Me* enrollment is unrestricted (non-selective in terms of gender, race, ethnicity, age, medical condition, or disease status) and all participants provide genetic data derived from blood samples at the time of enrollment. BioMe was approved by the MSSM IRB (#07-0529) and all BioMe participants provided written informed consent. Patient information was de-identified in accordance with HIPAA's "safe harbor" method prior to access. Individuals were genotyped on the Illumina Global Screening Array; quality control (QC), imputation of the genotype data, and genetically inferred ancestry assignment for Bio*Me* is described elsewhere(2,3).

MGB Biobank. ICD-9/10 codes were obtained for 29,885 individuals participating in the MGB (formerly Partners Healthcare) Biobank(4). Patients included met a minimum data floor criteria of at least 3 visits after 2005, each more than 30 days apart, with at least one clinical note of any kind. DNA was isolated from blood and standard quality control procedures, including ancestry assignment, were performed on the genomic data. The MGB IRB (protocol #2018P002642) approved this study, and all participants provided written informed consent.

## Adjusting for healthcare utilization rates in PheWAS model

Our main analyses do not adjust for healthcare utilization rates, which likely inflate our type I error. In sensitivity analyses, we found correction for record density as a crude proxy of healthcare utilization decreases the number of phenotypic associations with PTSD from 1137 to 618 and 794 to 325 significant phecodes for PTSD$_{Broad}$ and PTSD$_{Narrow}$, respectively. (Summary statistics for healthcare utilization-adjusted models: Supplementary Table 5). However, we highlight that this adjustment may introduce risk of collider bias, distorting associations between PTSD and comorbidities, and therefore are not included in our main analyses.

Supplemental References

1. Dennis JK, Sealock JM, Straub P, Lee YH, Hucks D, Actkins K, *et al.* (2021): Clinical laboratory test-wide association scan of polygenic scores identifies biomarkers of complex disease. *Genome Med* 13: 6.

2. Zheutlin AB, Dennis J, Karlsson Linnér R, Moscati A, Restrepo N, Straub P, *et al.* (2019): Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. *Am J Psychiatry* 176: 846–855.

3. Johnson JS, Cote AC, Dobbyn A, Sloofman LG, Xu J, Cotter L, *et al.* (2023): Mapping anorexia nervosa genes to clinical phenotypes. *Psychol Med* 53: 2619–2633.

4. Boutin NT, Schecter SB, Perez EF, Tchamitchian NS, Cerretani XR, Gainer VS, *et al.* (2022): The Evolution of a Large Biobank at Mass General Brigham. *J Pers Med* 12: 1323.
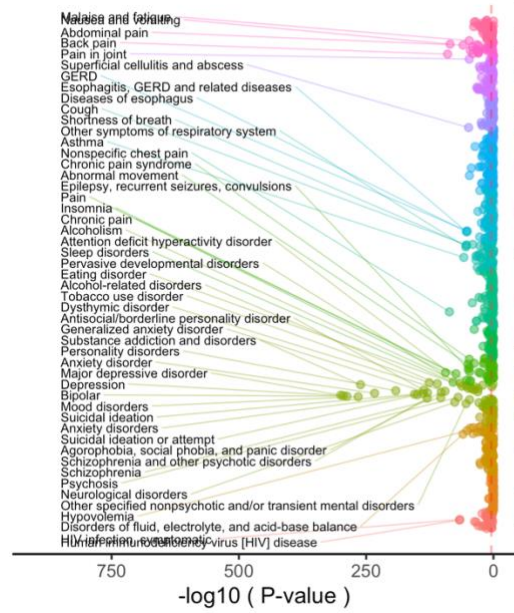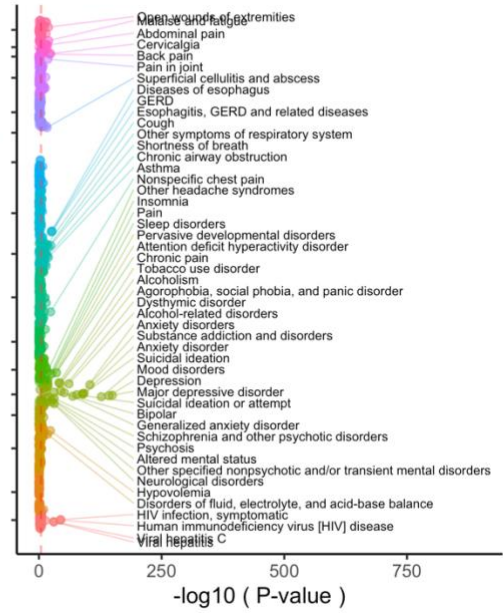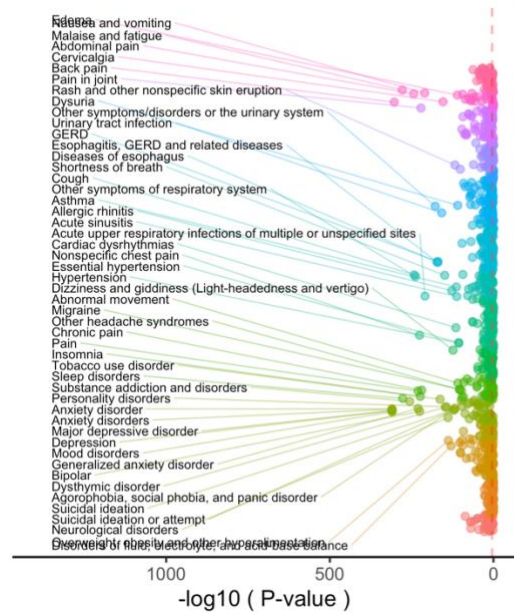
Supplemental Figure 1. Phenome-wide associations with narrowly-defined PTSD in females (A) and males (B).
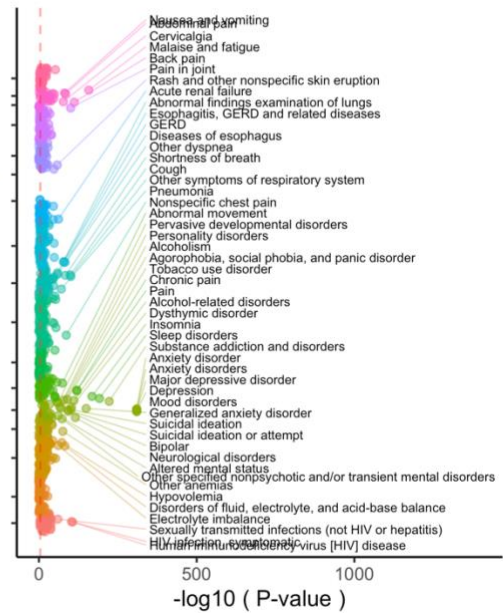
Supplemental Figure 2. Phenome-wide associations with broadly-defined PTSD in females (A) and males (B).