Prevalence, Comorbidity, and Sociodemographic Correlates of Psychiatric Disorders in the All Of Us Biobank

Peter B. Barr, Ph.D.^{1,2}, Tim B. Bigdeli, Ph.D.^{1,2}, Jacquelyn M. Meyers, Ph.D.^{1,2}

¹ Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY

² VA New York Harbor Healthcare System, Brooklyn, NY

Corresponding author:

Peter B. Barr

Department of Psychiatry and Behavioral Sciences

SUNY Downstate Health Sciences University

450 Clarkson Ave, MSC 1203, Brooklyn, NY 11203

e-mail: peter.barr@downstate.edu.

Running title: Psychiatric Disorders in the All Of Us

Word count (abstract/text body): 279/3,099

Figures and Tables: 3 Tables, 2 Figures

ABSTRACT

Importance: All of Us is a landmark initiative for population-scale research into the etiology of psychiatric disorders and disparities across various sociodemographic categories.

Objective: To estimate the prevalence, comorbidity, and demographic covariates of psychiatric and substance use disorders in the All of Us biobank.

Design, Setting, and Participants: We estimated prevalence, overlap, and demographic correlates for psychiatric disorders derived from electronic health records in the All of Us biobank (release 5; N = 331,380)

Exposures: Social and demographic covariates.

Main Outcome and Measures: Psychiatric disorders derived from ICD10CM codes and grouped into phecodes across six broad domains: mood disorders, anxiety disorders, substance use disorders, stress-related disorders, schizophrenia, and personality disorders.

Results: The prevalence of various disorders ranges from approximately 15% to less than 1%, with mood and anxiety disorders being the most common, followed by substance use disorders, stress-related disorders, schizophrenia, and personality disorders. There is substantial overlap among disorders, with a large portion of those with a disorder (~57%) having two or more registered diagnoses and tetrachoric correlations ranging from 0.43 – 0.74. The prevalence of disorders across demographic categories demonstrates that non-Hispanic whites, those of low socioeconomic status, women and those assigned female at birth, and sexual minorities are at greatest risk for most disorders.

Conclusions and Relevance: Although the rates of disorders in All of Us are lower than rates for disorders in the general population, there is considerable variation, comorbidity, and differences across social groups. Large-scale resources like All of Us will prove to be invaluable for understanding the causes and consequences of psychiatric conditions. As we move towards an era of precision medicine, we must work to ensure it is delivered in an equitable manner.

INTRODUCTION

Psychiatric and substance use disorders remain an ever-present challenge to public health, incurring significant cost to society, affected individuals, and their families. Mental illness is estimated to contribute ~13% to the global burden of disease¹ and the human cost can be staggering: with over 49,000 Americans dead as the result of an opioid overdose and over 47,000 deaths attributable to suicide in 2019, alone². Psychiatric disorders incur significant financial cost, with estimates in the United States ranging from ~\$320 billion annually for serious mental illnesses³ to ~\$78 billion annually for opioid use⁴. Importantly, these disorders rarely manifest in isolation, showing strong patterns of comorbidity and shared risk factors⁵⁻⁷.

Historically, estimating the prevalence of psychiatric disorders has relied on nationally representative or population-based surveys often conducted by trained lay interviewers⁸⁻¹⁰, or through clinical ascertainment¹¹. More recently, large-scale biobanks and healthcare systems are becoming increasingly used for research purposes. To date, resources such as the UK Biobank¹², the Million Veteran Program¹³, Biobank Japan¹⁴, and FinnGen, among others, have proved to be an invaluable resource in the understanding of the causes of various psychiatric and related disorders. While valuable, these resources may not reflect the unique context within the US. Additionally, much of the research conducted in this area has focused primarily on individuals of European ancestries^{15,16}. This is despite the previously observed differences in rates and patterns of psychiatric disorders by race-ethnicity, as well as known risk factors that are especially pertinent in the US given its sordid history with racism and inequality. In order to improve diversity in health-focused epidemiological research, further interrogate existing health disparities, and deliver on the promise of equitable precision medicine, the National Institutes of Health launched the All of Us program, a historic effort to collect and study data from at least one million people living in the United States¹⁷. Beginning in 2018, participants were able to complete surveys, provide genotypic data, and link their electronic health records to help build a comprehensive database of their health-related data. The goal of the All of Us program is to use

this data to understand how biology, lifestyle and social determinants come together to affect health, and ultimately to treat and prevent illness.

The current analysis focuses on characterizing the prevalence, comorbidity, and sociodemographic disparities for psychiatric and substance use disorders in the All of Us database. We compare these estimates to those from nationally representative and population-based samples to help better understand the etiology of psychiatric disorders in All of Us, and the participants that make up this new sample.

METHODS

The All of Us Biobank

The All of Us Research Program is a prospective cohort study aiming to recruit at least one million individuals in the United States, with the overall goal of providing a unique resource to study the effects of lifestyle, environment and genomics on health and health outcomes. Participant recruitment is predominantly done through participating health care provider organizations and in partnership with Federally Qualified Health Centers, with an emphasis on recruiting persons affiliated with those centers. Interested potential participants can also enroll in the program as direct volunteers, visiting community-based enrollment sites. Initial enrollment, informed consent (including consent to share EHRs), and baseline health surveys are done digitally through the All of Us program website (<https://joinallofus.org>). Once this step is completed, the participant is invited to undergo a basic physical exam and biospecimen collection at the affiliated healthcare site. Participant follow-up is done in two ways, passively via linkage with EHR and actively by periodic follow-up surveys. For this study, we included data from participants enrolled in the study between May 6, 2018 and April 1st, 2021 (All of Us release 5, N = 331,380). This work was performed on data collected by the previously described All of Us Research Program using the All of Us Researcher Workbench, a cloud-based platform

where approved researchers can access and analyze All of Us data. All analyses were in concordance with the ethical guidelines outlined in the All of Us Code of Conduct.

Psychiatric Diagnoses

We derived diagnoses based on phecodes from registered ICD10CM codes in the All of Us system. In the case where participants had multiple documentations of the same code, we used the most recent instance. First, we selected all relevant ICD codes for the more prevalent psychiatric disorders, including mood, anxiety, substance use (excluding nicotine), stress-related, psychotic, personality, eating, and other disorders. We then clustered these ICD codes using phecodes¹⁸ from the *PheWAS* package in R¹⁹. Phecodes provide a convenient way to cluster codes in EHR based data²⁰. A full list of the EHR descriptions, ICD10CM codes, and corresponding phecodes is provided in eTables 1 and 2 in the supplement.

Sociodemographic Risk Factors

All of Us provides extensive measures of social and demographic characteristics. Our analyses included covariates for age, sex assigned at birth (male, female, and no answer), gender identity (man, woman, or no answer), sexual orientation (straight/heterosexual vs. all others), race-ethnicity (non-Hispanic white, Black/African-American, Hispanic or Latino/a/x, Asian, other race-ethnicity, and multiracial), educational attainment (less than HS, HS or equivalent, some college, and college degree or more), household income (ranges of categories from under \$25K to \$100K or more annually), access to health insurance, and country of origin (US vs foreign born) Given previous research on the widespread mental health disparities across these social factors^{8,21}, it is important to assess whether there are similar patterns in the All of Us data to ensure that it accurately reflects the broader US population.

Analytic Strategy

Our analytic strategy was threefold. First, after cleaning and collating the EHR data, we estimated the baseline lifetime prevalence of each disorder, comparing this prevalence to estimate of population lifetime prevalence (where available) from other published reports of nationally representative or similar data. Second, we estimated the conditional odds ratios (OR) using logistic regression models with each of the social and demographic covariates included. To correct for the number of independent models, we used a Bonferroni correction ($p < .05/6 = .0083$) Finally, we calculated the cooccurrence of each of the disorders and estimated the tetrachoric correlations to characterize overlap in disorders and patterns of comorbidities.

RESULTS

Sample Characteristics and Prevalence of Psychiatric Disorders

Table 1 presents the sample demographic characteristics. All of Us included questions for sex assigned at birth (38.1% Male, 60.7% Female, 1.3% Neither/skipped) and gender identity (37.8% Man, 60.2% Woman, 2.1% Neither/skipped). The vast majority of respondents identified as heterosexual (~89%). The sample is highly educated, with approximately 40% of the sample reporting a college degree or higher. However, when we compare this distribution to household income, we see that there is considerably more variation, with ~20% reporting an annual household income below \$25K and ~20% reporting an annual household of \$100K or more (though ~20% skipped this question). One of the goals of All of Us was to recruit a more racially and ethnically diverse sample, and this is reflected in the sample compositions (52% non-Hispanic White, 21% Black/African American, 18% Hispanic or Latino/a/x, 3% Asian, and 3% who identify as some other race-ethnicity or multiracial). Most participants were born in the United States (~84%).

Table 2 presents lifetime prevalence for psychiatric disorders. Within the All of Us EHR data, mood disorders are the most common (14.31%), with major depressive disorder being the most common form of mood disorder (12.61%), followed by bipolar disorder (3.21%), and dysthymic disorder (0.89%). Followed closely in prevalence to mood disorders, are those with any anxiety disorder (13.87%). Within anxiety disorders, anxiety disorder (typically unspecified) is the most prevalent (12.86%), followed by generalized anxiety disorder (3.87%), social anxiety disorder/agoraphobia (1.63%), and other phobias (0.38%). In terms of substance use disorder (SUD), 5.91% have a documented SUD, with 2.96% having a documented alcohol use disorder (AUD) and 4.35% having a documented SUD related to some other form of substance (see eTable 3 for breakdown by substance). For stress related disorders, 2.6% had a documented adjustment disorder and 2.32% received a post-traumatic stress disorder (PTSD) diagnosis, and a total of 4.54% have any documented stress-related disorder. Roughly 2% of All of Us participants have a documented sleep disorder. Attention-deficit hyperactivity disorder (ADHD) and schizophrenia had an estimated prevalence of ~1%, Personality disorders, eating disorders, and obsessive-compulsive disorder each had a prevalence below 1%. In almost every case, the prevalence of these disorders is lower than the prevalence estimated using nationally representative samples^{6,8,21–27}.

Association of Social and Demographic Covariates with Psychiatric Disorders

Table 3 includes the estimates for the associations between demographic covariates and six categories of psychiatric disorders: any mood disorder (MOOD), any anxiety disorders (ANX), any substance use disorder (SUD), any stress-related disorders (STRESS), schizophrenia (SCZ), and any personality disorder (PERS).

Similar patterns emerge across sex and gender, where those assigned male at birth and those who identify as men have lower odds of mood ($OR_{SEX} = 0.85$; $OR_{GENDER} = 0.84$) and anxiety ($OR_{SEX} = 0.88$; $OR_{GENDER} = 0.75$) disorders, but increased risk for SUDs ($OR_{SEX} = 1.55$;

$OR_{\text{GENDER}} = 1.42$) and schizophrenia ($OR_{\text{GENDER}} = 1.87$). Those who did not identify with hegemonic categories for sex assigned at birth or gender, and those who did not identify as straight/heterosexual were at increased risk across SUDs ($OR_{\text{SEX}} = 1.46$; $OR_{\text{GENDER}} = 1.19$; $OR_{\text{SEX. ORIEN.}} = 1.20$), stress-related disorders ($OR_{\text{GENDER}} = 1.25$; $OR_{\text{SEX. ORIEN.}} = 1.30$), schizophrenia ($OR_{\text{GENDER}} = 1.67$; $OR_{\text{SEX. ORIEN.}} = 1.52$), and personality disorders ($OR_{\text{GENDER}} = 1.61$; $OR_{\text{SEX. ORIEN.}} = 1.87$).

Compared to those aged 18 to 29, older individuals are at increased risk for almost all of the disorders included in this analysis, with odds ratios ranging from 1.10 to 2.27. The exception is for those aged 65 and older being at reduced risk for personality disorders ($OR = 0.39$). Compared to Non-Hispanic White participants, individuals of all other racial ethnic categories are at reduced risk for mood, anxiety, substance use, stress-related, and personality disorders, to varying degree. However, for schizophrenia, Black/African American ($OR = 1.19$) participants, those of other racial-ethnic groups ($OR = 1.63$), and those who identify as multiracial were all at increased risk ($OR = 1.69$). Those who were born outside the US were at reduced risk for every type of disorder ($ORs = 0.34 - 0.68$).

Socioeconomic variation provides a more complicated picture. Those with a college degree or higher have lower risk of all six categories of disorder relative to those with less than a high school education ($ORs = 0.35 - 0.71$), but those with some college or the equivalent of a high school diploma only have reduced risk of mood disorders, SUDs, and schizophrenia, and the effect sizes tended to be smaller ($ORs = 0.58 - 0.93$). For income, however, individuals in *any* income category have a relatively lower risk of psychiatric disorder across any of the six categories ($ORs = 0.05 - 0.70$). Those reporting no health insurance had much lower odds of having any disorder ($ORs = 0.33 - 0.60$), which likely reflect their inability to access care. Those who are never married, cohabiting, separated, divorced, or widowed had much higher risk of every class of disorder ($ORs = 1.08 - 2.21$).

Figure 1 presents the prevalence by each of the demographic categories included in the analyses, above (see supplemental S4 for exact values). The differences across groups align almost entirely with the results from the conditional models, with a few notable exceptions. The raw prevalence for any SUD is ~8% in among those who are Black/African American, compared to ~ 5% in those who are non-Hispanic White, the opposite pattern seen in the conditional analyses. The reversed direction of association for Black/African Americans, from being at increased risk in bivariate associations to being at reduced risk in the conditional models, is accounted for by the inclusion of educational attainment, household income, and health insurance, suggesting that much of the risk is related to socioeconomic inequalities.

Patterns of Comorbidity in the All of Us Data

Figure 2 presents the tetrachoric correlations and patterns of comorbidity for each of the six classes of disorders included in the current analysis. In keeping with the All of Us data use requirements, we filtered all cells with fewer than 20 respondents. In terms of correlations, there is significant correlations across all disorders ranging from $r = 0.43$ to $r = 0.75$. In terms of specific patterns of overlap, 57% of the approximately 70,000 diagnoses involved some overlap and 30,798 (~43%) occurred in isolation. Of those with two registered diagnoses ($n = 22,797$), the vast majority had some configuration of mood and anxiety disorder. Those with three disorders ($n = 10,968$) mostly had a mood disorder, an anxiety disorder, and either an SUD or stress-related disorder. The last pattern of comorbidity that occurs relatively frequently was among those with a registered diagnosis of a mood, anxiety, substance use, and stress-related disorder. Overall, these patterns of comorbidity demonstrate the fact that psychiatric disorders more frequently occur in conjunction than on their own.

DISCUSSION

The goal of the All of Us program is to create a resource for health research that can benefit *all* of the population in the United States, with a particular emphasis in improving health equity and representation in medical research. In the current analysis, we provided an overview of the currently available data related to psychiatric and substance use disorders, how these disorders vary across social and demographic factors, and how these disorders cluster together.

In terms of raw prevalence, the disorders captured in the All of Us EHR occur at lower frequency than estimates from nationally representative samples. There are two possible explanations. First, is that survey-based instruments overestimate the prevalence of disorder. However, the more likely explanation is that there is ascertainment bias in All of Us and among individuals seeking treatment for psychiatric disorders more generally. Other biobanks have known issues with ‘healthy participant’ bias²⁸. While biobanks provide a useful resource for many medical conditions, household survey approaches may be better able to access hard-to-reach populations, like those with severe mental illnesses. As expected from prior research^{7,8,29}, we demonstrated widespread comorbidity between disorders. Overall, these disorders were strongly correlated with one another, though the patterns ranged from very common (mood + anxiety disorders) to less common (4 + disorders).

The pattern of associations across social and demographic factors is similar to those from other randomly drawn samples. Differences in risk across gender identity and sex assigned at birth are generally similar and in line with prior research, whereby women and those assigned female at birth are at increased risk for mood and anxiety disorders, but at lower risk of SUDs²¹. Importantly, All of Us includes a sizeable number of individuals whose identity does not fall into dominant categories in terms of sex assigned at birth, gender identity, and sexual orientation. Persons in these categories were at increased risk virtually each of the classes of psychiatric disorders, similar to prior research^{30,31}. The increase prevalence of disorders among

these persons is likely the consequence of increased exposure to discrimination and other stressors associated with holding a marginalized identity, similar to those in other marginalized social positions^{32,33}. Prior research has demonstrated that states with more protective policies towards LGBTQ+ persons saw reduced associations between sexual minority status and psychiatric disorder³¹. Further work into policies that can reduce the disparities within these groups is necessary.

In addition to diversity across sex, gender, and sexuality, All of Us also contains substantial diversity across race and ethnicity. Compared to non-Hispanic White participants, those who identify as Black/African American, Hispanic or Latino/a/x, Asian, multiracial, or other racial ethnic category were at *reduced* risk of most disorders, a pattern observed across most psychiatric disorders^{34,35}. Given racial health disparities in other medical conditions attributed to racism, racial discrimination, and racial inequality more broadly^{36,37}, the opposite direction of association for mental health and psychiatric disorders has proven difficult to explain. It is possible that the items used to assess disorders do not transport well across cultural or racial-ethnic contexts³⁸, and we need disorders/items that accurately capture different racialized experiences³⁹. Other possibilities include increased unhealthy behaviors used as a coping mechanism that consequently worsen physical health disparities but mitigate against mental health disparities⁴⁰, or the increased consumption of pharmaceuticals by non-Hispanic Whites that include mental health related side effects⁴¹, among others. The one exception to the above pattern was in regard to schizophrenia. Like previous findings, the increased prevalence of schizophrenia among those who identify as Black/African American is also a well-documented pattern in social epidemiology³⁴. This disparity has been attributed to social antecedents (e.g., racism, lack of access to care)⁴² as well as racial bias in the diagnostic category for schizophrenia itself⁴³. Finally, those who are born outside of the US have consistently lower rates of each of the disorders included. These decreased rates could reflect a healthy immigrant effect⁴⁴ or other protective cultural factors among immigrants.

Socioeconomic disparities in all of the disorders are apparent in both education and income. Compared to those with a college degree, all of the other levels of educational attainment are at increased risk of each class of disorder. The gradient in psychiatric disorders is most apparent in household income, with those in the >\$25K annually category being at consistently highest risk. These consistent patterns could be due to the increased stress and reduced resources experienced by those at the lower end of the socioeconomic continuum^{45–47}, or the reduced upward social mobility of those experience psychiatric disorders (e.g., social drift)⁴⁸, though it is likely a mixture of both.

Our analysis is not without limitations. First, we grouped many different ICD10 codes and diagnoses into larger categories. We did this for simplicity, as the number of available diagnoses in the EHR is tremendous. Future iterations can take a more granular approach for certain diagnostic subcategories. Second, we focused on lifetime diagnoses of disorders. Next steps should also focus on the occurrence of more recent onset disorders and better tease apart time ordering in regard to psychiatric disorders and other health conditions. Future work can also leverage the extensive EHR data to model the course of treatment for specific disorders, and how this may vary across social and demographic contexts.

The All of Us biobank presents a vital resource in better characterizing the etiology of psychiatric conditions and advancing health equity. In the current analysis we have demonstrated the rich data available on psychiatric disorders. Future research will be able to harness the vast genetic, lifestyle, and other social data to better understand the causes and consequences of psychiatric conditions. As we move towards an era of precision medicine, biobanks such as All of Us will be an invaluable resource.

ACKNOWLEDGMENTS

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. In addition, the All of Us Research Program would not be possible without the partnership of its participants.

REFERENCES

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. *The Lancet Psychiatry*. 2016;3(2):171-178. doi:10.1016/S2215-0366(15)00505-2
2. Center for Disease Control and Prevention, Centers for Disease Control and Prevention (US). *Mortality Data—National Vital Statistics System*. Vol 2019.; 2019. <https://www.cdc.gov/nchs/nvss/deaths.htm>
3. Insel TR. Assessing the economic costs of serious mental illness. *Am J Psychiatry*. 2008;165(6):663-665. doi:10.1176/appi.ajp.2008.08030366
4. Florence CS, Zhou C, Luo F, Xu L. The Economic Burden of Prescription Opioid Overdose, Abuse, and Dependence in the United States, 2013. *Med Care*. 2016;54(10):901-906. doi:10.1097/MLR.0000000000000625
5. Kessler R, Chiu W. Prevalence, Severity, and Comorbidity of Twelve-month DSM-IV

- Disorders in the National Comorbidity Survey Replication (NCS- R). *Arch Gen* 2005;62(6):617-627. doi:10.1001/archpsyc.62.6.617.Prevalence
6. Grant BF, Goldstein RB, Saha TD, et al. Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry*. 2015;72(8):757-766. doi:10.1001/jamapsychiatry.2015.0584
 7. Grant BF, Saha TD, Ruan WJ, et al. Epidemiology of DSM-5 Drug Use Disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. *JAMA Psychiatry*. 2016;73(1):39. doi:10.1001/jamapsychiatry.2015.2132
 8. Hasin DS, Grant BF. The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) Waves 1 and 2: review and summary of findings. *Soc Psychiatry Psychiatr Epidemiol*. 2015;50(11):1609-1640. doi:10.1007/s00127-015-1088-0
 9. Kessler RC, Merikangas KR. The National Comorbidity Survey Replication (NCS-R): Background and aims. *Int J Methods Psychiatr Res*. 2004;13(2):60-68. doi:10.1002/mpr.166
 10. Demyttenaere K, Bruffaerts R, Posada-Villa J, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *J Am Med Assoc*. 2004;291(21):2581-2590. doi:10.1001/jama.291.21.2581
 11. Warden D, Rush AJ, Trivedi MH, Fava M, Wisniewski SR. The STAR*D project results: A comprehensive review of findings. *Curr Psychiatry Rep*. 2007;9(6):449-459. doi:10.1007/s11920-007-0061-3
 12. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
 13. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214-223. doi:10.1016/j.jclinepi.2015.09.016

14. Nagai A, Hirata M, Kamatani Y, et al. Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol.* 2017;27(3):S2-S8. doi:10.1016/j.je.2016.12.005
15. Mills MC, Rahal C. A scientometric review of genome-wide association studies. *Commun Biol.* 2019;2(1):9. doi:10.1038/s42003-018-0261-x
16. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51(4):584-591. doi:10.1038/s41588-019-0379-x
17. The All of Us Research Program Investigators. The “All of Us” Research Program. *N Engl J Med.* 2019;381(7):668-676. doi:10.1056/nejmsr1809937
18. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol.* 2013;31(12):1102-1110. doi:10.1038/nbt.2749
19. Carroll RJ, Bastarache L, Denny JC. R PheWAS: Data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics.* 2014;30(16):2375-2376. doi:10.1093/bioinformatics/btu197
20. Wei W-Q, Bastarache LA, Carroll RJ, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS One.* 2017;12(7):e0175508.
21. Kessler RC, Berglund P, Demler O, et al. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry.* 2005;62(6):593. doi:10.1001/archpsyc.62.6.593
22. Galmiche M, Déchelotte P, Lambert G, Tavalacci MP. Prevalence of eating disorders over the 2000-2018 period: A systematic literature review. *Am J Clin Nutr.* 2019;109(5):1402-1413. doi:10.1093/ajcn/nqy342
23. Trull TJ, Jahng S, Tomko RL, Wood PK, Sher KJ. Revised NESARC personality disorder diagnoses: Gender, prevalence, and comorbidity with substance dependence disorders. *J*

Pers Disord. 2010;24(4):412-426. doi:10.1521/pedi.2010.24.4.412

24. Ohayon MM. Prevalence of DSM-IV diagnostic criteria of insomnia: Distinguishing insomnia related to mental disorders from sleep disorders. *J Psychiatr Res.* 1997;31(3):333-346. doi:10.1016/S0022-3956(97)00002-2
25. Perälä J, Suvisaari J, Saarni SI, et al. Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Arch Gen Psychiatry.* 2007;64(1):19-28. doi:10.1001/archpsyc.64.1.19
26. Goldstein RB, Smith SM, Chou SP, et al. The epidemiology of DSM-5 posttraumatic stress disorder in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. *Soc Psychiatry Psychiatr Epidemiol.* 2016;51(8):1137-1148. doi:10.1007/s00127-016-1208-5
27. Grant BF, Saha TD, Ruan WJ, et al. Epidemiology of DSM-5 Drug Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions—III. *JAMA Psychiatry.* 2016;73(1):39-47. doi:10.1001/jamapsychiatry.2015.2132
28. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol.* 2017;186(9):1026-1034. doi:10.1093/aje/kwx246
29. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, Severity, and Comorbidity of Twelve-month DSM-IV Disorders in the National Comorbidity Survey Replication (NCS-R). *Arch Gen Psychiatry.* 2005;62(6):617-627. doi:10.1001/archpsyc.62.6.617
30. Boyd CJ, Veliz PT, McCabe SE. Severity of DSM-5 cannabis use disorders in a nationally representative sample of sexual minorities. *Subst Abus.* 2020;41(2):191-195. doi:10.1080/08897077.2019.1621242
31. Hatzenbuehler ML, Keyes KM, Hasin DS. State-level policies and psychiatric morbidity in lesbian, gay, and bisexual populations. *Am J Public Health.* 2009;99(12):2275-2281. doi:10.2105/AJPH.2008.153510

32. Pearlin LI, Lieberman MA, Menaghan EG, et al. The stress process. *J Health Soc Behav.* 1981;22(4):337-356. doi:10.2307/2136676
33. Lennon MC. The Structural Contexts of Stress. *J Health Soc Behav.* 1989;30(3):261. doi:10.2307/2136958
34. Williams DR, Earl TR. Commentary: Race and mental health - More questions than answers. *Int J Epidemiol.* 2007;36(4):758-760. doi:10.1093/ije/dym114
35. Breslau J, Aguilar-Gaxiola S, Kendler KS, Su M, Williams D, Kessler RC. Specifying race-ethnic differences in risk for psychiatric disorder in a USA national sample. *Psychol Med.* 2006;36(1):57-68. doi:10.1017/S0033291705006161
36. Williams DR, Sternthal M. Understanding racial-ethnic disparities in health: sociological contributions. *J Heal Soc Behav.* 2010;51 Suppl(1 suppl):S15-27. doi:10.1177/0022146510383838
37. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. *Public Heal Rep.* 2001;116(5):404-416. <http://www.ncbi.nlm.nih.gov/pubmed/12042604>
38. Kim G, DeCoster J, Huang CH, Chiriboga DA. Race/Ethnicity and the Factor Structure of the Center for Epidemiologic Studies Depression Scale: A Meta-Analysis. *Cult Divers Ethn Minor Psychol.* 2011;17(4):381-396. doi:10.1037/A0025434
39. Brown TN. Critical Race Theory Speaks to the Sociology of Mental Health: Mental Health Problems Produced by Racial Stratification. *J Health Soc Behav.* 2003;44(3):292-301.
40. Jackson JS, Knight KEM, Rafferty JA. Race and Unhealthy Behaviors: Chronic Stress, the HPA Axis, and Physical and Mental Health Disparities Over the Life Course. *Am J Public Health.* 2010;100(5):933-939. doi:10.2105/AJPH.2008.143446)
41. Schnittker J, Do D. Pharmaceutical Side Effects and Mental Health Paradoxes among Racial-Ethnic Minorities. *J Health Soc Behav.* 2020;61(1):4-23. doi:10.1177/0022146519899115

42. Schwartz RC, Blankenship DM. Racial disparities in psychotic disorder diagnosis: A review of empirical literature. *World J psychiatry*. 2014;4(4):133-140. doi:10.5498/wjp.v4.i4.133
43. Metzl JM. *The Protest Psychosis: How Schizophrenia Became a Black Disease*. Beacon Press; 2010.
44. Salas-Wright CP, Vaughn MG, Goings TC, Miller DP, Schwartz SJ. Immigrants and mental disorders in the united states: New evidence on the healthy migrant hypothesis. *Psychiatry Res*. 2018;267:438-445. doi:10.1016/j.psychres.2018.06.039
45. Link BG, Phelan JC. Social Conditions as a Fundamental Cause of Disease. *J Health Soc Behav*. 1995;35:80-94.
46. Lloyd DA, Turner RJ. Cumulative lifetime adversities and alcohol dependence in adolescence and young adulthood. *Drug Alcohol Depend*. 2008;93(3):217-226. doi:10.1016/j.drugalcdep.2007.09.012
47. Turner RJ, Avison WR. Status variations in stress exposure: implications for the interpretation of research on race, socioeconomic status, and gender. *J Heal Soc Behav*. 2003;44(4):488-505. <http://www.ncbi.nlm.nih.gov/pubmed/15038145>
48. Breslau J, Lane M, Sampson N, Kessler RC. Mental disorders and subsequent educational attainment in a US national sample. *J Psychiatr Res*. 2008;42(9):708-716. doi:10.1016/j.jpsychires.2008.01.016

Table 1: All of Us Sample Characteristics (N = 329,038)

		N	%
Sex assigned at birth	Female	199,658	60.68%
	Male	125,184	38.05%
	Not male, not female, no matching concept, prefer not to answer, or skipped	4,196	1.28%
Gender	Woman	197,961	60.16%
	Man	124,202	37.75%
	Not man only, not woman only, no matching concept, prefer not to answer, or skipped	6,875	2.09%
Sexual orientation	Straight/Heterosexual	290,176	88.19%
	All others	38,862	11.81%
Age	18-29	48,438	14.72%
	30-44	74,353	22.60%
	45-64	128,463	39.04%
	65+	77,784	23.64%
Race-ethnicity	Non-Hispanic White	172,753	52.50%
	Black/African American	69,084	21.00%
	Hispanic or Latino/a/x	59,348	18.04%
	Asian	10,747	3.27%
	Other race-ethnicity	5,697	1.73%
	Multiracial	5,282	1.61%
	Missing	6,127	1.86%
Education	Less than high school	32,178	9.78%
	High school or equivalent	65,214	19.82%
	Some college	83,518	25.38%
	College grad/advanced degree	140,542	42.71%
	Missing	7,586	2.31%
Household Income	Less than 25k	90,190	27.41%
	25k to 50k	49,190	14.95%
	50k to 75k	33,696	10.24%
	75k 100k	25,867	7.86%
	more than 100k	66,389	20.18%
	Missing	63,706	19.36%
Health insurance	Yes	296,627	90.15%
	No	23,235	7.06%
	Missing	9,176	2.79%
Marital status	Married	137,597	41.82%
	Living With Partner	21,906	6.66%
	Divorced	46,496	14.13%
	Never Married	85,479	25.98%
	Separated	11,655	3.54%
	Widowed	16,779	5.10%
	Missing	9,126	2.77%
Birthplace	USA	275,464	83.72%
	Foreign born	50,237	15.27%
	Missing	3,337	1.01%

Table 2: Prevalence of Psychiatric Diagnoses in the All of US EHR (N = 329,038)

Phenotype	Cases	Prev.	95% CI		Pop. Prev.	
Major depressive disorder	41,493	12.61%	12.29%	12.93%	16.6%	²¹
Bipolar	10,577	3.21%	2.88%	3.55%	3.9%	
Dysthymic disorder	2,924	0.89%	0.55%	1.23%	2.5%	²¹
Any mood disorder	47,100	14.31%	14.00%	14.63%	20.8%	²¹
Anxiety disorder (unspecified)	42,309	12.86%	12.54%	13.18%	-	
Generalized anxiety disorder	12,736	3.87%	3.54%	4.21%	5.7%	²¹
Social anxiety disorder/Agoraphobia	5,365	1.63%	1.29%	1.97%	12.1/1.4%	²¹
Phobia	1,236	0.38%	0.03%	0.72%	12.5%	²¹
Any anxiety disorder	45,649	13.87%	13.56%	14.19%	28.8%	²¹
Alcohol use disorder	9,725	2.96%	2.62%	3.29%	29.1%	⁶
Substance use disorder	14,301	4.35%	4.01%	4.68%	9.9%	²⁷
Any substance use disorder	19,440	5.91%	5.58%	6.24%	32.3%	⁸
Adjustment disorder	8,565	2.60%	2.27%	2.94%	-	
Post-traumatic stress disorder	7,643	2.32%	1.99%	2.66%	6.1%	²⁶
Any stress-related disorder	14,929	4.54%	4.20%	4.87%	-	
Schizophrenia	3,562	1.08%	0.74%	1.42%	0.9%	²⁵
Sleep disorder	6,591	2.00%	1.66%	2.34%	5.6%	²⁴
Attention-deficit/hyperactivity disorder	3,588	1.09%	0.75%	1.43%	8.1%	²¹
Antisocial personality/Borderline personality disorder	1,396	0.42%	0.08%	0.77%	3.8/2.7%	²³
Shizotypal personality disorder	83	0.03%	-0.32%	0.37%	0.6%	²³
Any personality disorder	2,327	0.71%	0.37%	1.05%	0.3 - 2.7%	²³
Eating Disorder	1,014	0.31%	-0.03%	0.65%	8.4/2.2%	²²
Obsessive compulsive disorder	942	0.29%	-0.05%	0.63%	1.9%	²³

Prev. = prevalence; Pop. Prev. = population prevalence

Table 3: Adjusted Estimates of Risk for Psychiatric Disorders

	MOOD		ANX				SUD		STRESS		SCZ		PERS	
	<u>OR</u>	<u>SE</u>		<u>OR</u>	<u>SE</u>		<u>OR</u>	<u>SE</u>		<u>OR</u>	<u>SE</u>		<u>OR</u>	<u>SE</u>
Female (ref)	-	-		-	-		-	-		-	-		-	-
Male	0.85	0.06	*	0.88	0.06		1.55	0.08	*	0.89	0.09		1.02	0.16
Neither male nor female selected	0.96	0.05		0.96	0.05		1.46	0.07	*	0.98	0.08		1.10	0.15
Woman (ref)	-	-		-	-		-	-		-	-		-	-
Man	0.84	0.06	*	0.76	0.06	*	1.42	0.08	*	0.97	0.09		1.87	0.16
Neither man nor women selected	0.99	0.04		0.94	0.04		1.19	0.06	*	1.25	0.06	*	1.67	0.13
Straight/Heterosexual	-	-		-	-		-	-		-	-		-	-
Other than straight/heterosexual	1.28	0.02	*	1.19	0.02	*	1.20	0.02	*	1.30	0.02	*	1.52	0.04
														*
Age 18-29 (ref)	-	-		-	-		-	-		-	-		-	-
Age 30-44	1.60	0.02	*	1.49	0.02	*	1.98	0.03	*	1.55	0.03	*	2.27	0.06
Age 45-64	1.88	0.02	*	1.47	0.02	*	2.00	0.03	*	1.61	0.03	*	2.16	0.06
Age 65 and older	1.48	0.02	*	1.10	0.02	*	0.97	0.03		1.16	0.04	*	0.90	0.10
														*
Non-Hispanic White (ref)	-	-		-	-		-	-		-	-		-	-
Black/African American	0.53	0.02	*	0.45	0.02	*	0.62	0.02	*	0.70	0.02	*	1.19	0.05
Hispanic/Latino/a/x	0.77	0.02	*	0.75	0.02	*	0.86	0.03	*	0.80	0.03	*	1.12	0.06
Asian	0.47	0.05	*	0.48	0.04	*	0.42	0.10	*	0.60	0.07	*	0.69	0.22
Other race-ethnicity	0.92	0.04		0.91	0.04		0.97	0.06		1.16	0.06		1.63	0.12
Multiracial	0.91	0.04		0.87	0.04	*	0.96	0.06		1.02	0.06		1.69	0.12
No race-ethnicity listed	0.87	0.04	*	0.84	0.04	*	0.98	0.05		0.92	0.06		1.62	0.10
														*
US born (ref)	-	-		-	-		-	-		-	-		-	-
Foreign born	0.68	0.02	*	0.65	0.02	*	0.34	0.04	*	0.67	0.03	*	0.44	0.08
No birthplace reported	0.94	0.05		0.95	0.05		1.18	0.06	*	0.97	0.08		0.99	0.13
Less than HS (ref)	-	-		-	-		-	-		-	-		-	-
HS diploma or equivalent	0.92	0.02	*	0.98	0.02		0.81	0.02	*	1.00	0.03		0.78	0.05
Some college	0.93	0.02	*	0.98	0.02		0.64	0.02	*	1.00	0.03		0.58	0.05
College degree or higher	0.65	0.02	*	0.71	0.02	*	0.28	0.03	*	0.75	0.03	*	0.35	0.07
No education reported	0.82	0.04	*	0.84	0.04	*	0.95	0.04		0.84	0.06	*	0.86	0.09
Income less than 25K (ref)	-	-		-	-		-	-		-	-		-	-
Income 25-50K	0.70	0.02	*	0.76	0.02	*	0.50	0.03	*	0.71	0.03	*	0.32	0.07
Income 50-75K	0.56	0.02	*	0.64	0.02	*	0.34	0.04	*	0.60	0.03	*	0.16	0.13
Income 75-100K	0.47	0.02	*	0.57	0.02	*	0.27	0.05	*	0.49	0.04	*	0.10	0.20
Income 100K or more	0.35	0.02	*	0.43	0.02	*	0.19	0.04	*	0.38	0.04	*	0.05	0.19
No income reported	0.72	0.01	*	0.77	0.02	*	0.74	0.02	*	0.74	0.02	*	0.74	0.04
														*
Have health insurance (ref)	-	-		-	-		-	-		-	-		-	-
No health insurance	0.42	0.03	*	0.42	0.03	*	0.47	0.03	*	0.50	0.04	*	0.33	0.08
No health insurance info reported	0.67	0.03	*	0.68	0.04	*	0.78	0.04	*	0.68	0.06	*	0.65	0.09
														*
Married (ref)	-	-		-	-		-	-		-	-		-	-

Cohabiting	1.14	0.02	*	1.08	0.02	*	1.90	0.03	*	1.12	0.04	*	1.59	0.09	*	1.19	0.10	*
Divorced	1.40	0.02	*	1.25	0.02	*	2.02	0.03	*	1.44	0.03	*	2.18	0.07	*	1.96	0.07	*
Never married	1.24	0.02	*	1.09	0.02	*	1.91	0.03	*	1.17	0.03	*	2.42	0.06	*	1.57	0.07	*
Separated	1.41	0.03	*	1.20	0.03	*	2.21	0.04	*	1.40	0.04	*	2.25	0.09	*	1.82	0.10	*
Widowed	1.30	0.02	*	1.17	0.02	*	1.76	0.04	*	1.46	0.04	*	1.89	0.10	*	1.27	0.12	*
No marital info reported	1.22	0.03	*	1.12	0.03	*	1.81	0.04	*	1.29	0.05	*	2.35	0.09	*	1.86	0.12	*

* $p < .05/6 = .0083$

OR = odds ratio; SE = standard error; MOOD = any mood disorder; ANX = any anxiety disorder; SUD = any substance use disorder; STRESS = any stress-related disorder;

SCZ = schizophrenia; PERS = any personality disorder

All estimates conditional on all other covariates included in the model

FIGURE CAPTIONS

Figure 1: Prevalence of Psychiatric Disorders across Social and Demographic Covariates

Unconditional prevalence of each category of psychiatric disorder across social and demographic covariates. Error bars represent 95% confidence intervals. MOOD = any mood disorder; ANX = any anxiety disorder; SUD = any substance use disorder; STRESS = any stress-related disorder; SCZ = schizophrenia; PERS = any personality disorder

Figure 2: Correlations and Patterns of Comorbidity across Six Classes of Disorders

Tetrachoric correlations (left panel) and overlap between classes of disorders (right panel). All correlations significant at $p < .001$ threshold. Count for each combination of disorder presented at the top of the bar chart in right panel. Bottom right panel denotes the combination of disorders for counts in the above plot. MOOD = any mood disorder; ANX = any anxiety disorder; SUD = any substance use disorder; STRESS = any stress-related disorder; SCZ = schizophrenia; PERS = any personality disorder.