Journal of Affective Disorders 317 (2022) 287–297

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

General Anxiety Disorder-7 Questionnaire as a marker of low socioeconomic status and inequity

Julio C. Nunes [a,b,1,*], Megan K. Carroll [c,1], Kenneth W. Mahaffey [a], Robert M. Califf [c], P. Murali Doraiswamy [d,e], Sarah Short [c], Svati H. Shah [f,g], Susan Swope [a], Donna Williams [a], Adrian F. Hernandez [g,h], David S. Hong [i]

[a] Stanford Center for Clinical Research, Stanford University School of Medicine, Stanford, CA, USA
[b] Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA
[c] Verily Life Sciences (Alphabet), San Francisco, CA, USA
[d] Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA
[e] Duke Institute for Brain Sciences, Duke University School of Medicine, Durham, NC, USA
[f] Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC, USA
[g] Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, NC, USA
[h] Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, USA
[i] Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA

ARTICLE INFO

Keywords:
Generalized Anxiety Disorder-7
Measures of health and disease
Effective clinical intervention
Social determinants of health

ABSTRACT

Background: The General Anxiety Disorder-7 (GAD-7) questionnaire is a standard tool used for screening and follow-up of patients with Generalized Anxiety Disorder (GAD). Although it is generally accepted that anxiety correlates with clinical and psychosocial stressors, precise quantitative data is limited on the relations among GAD-7, traditional biomarkers, and other measures of health. Further research is needed about how GAD-7 relates to race, ethnicity, and socioeconomic status (SES) as an assembly. We determined how multiple demographic and socioeconomic data correlate with the participants' GAD-7 results when compared with laboratory, physical function, clinical, and other biological markers.

Methods: The Project Baseline Health Study (BHS) is a prospective cohort of adults representing several populations in the USA. We analyzed a deeply phenotyped group of 2502 participants from that study. Measures of interest included: clinical markers or history of medical diagnoses; physical function markers including gait, grip strength, balance time, daily steps, and echocardiographic parameters; psychometric measurements; activities of daily living; socioeconomic characteristics; and laboratory results.

Results: Higher GAD-7 scores were associated with female sex, younger age, and Hispanic ethnicity. Measures of low SES were also associated with higher scores, including unemployment, income $\leq$\$25,000, and $\leq$12 years of education. After adjustment for 158 demographic, clinical, laboratory, and symptom characteristics, unemployment and overall higher SES risk scores were highly correlated with anxiety scores. Protective factors included Black race and older age.

Limitations: Correlations identified in this cross-sectional study cannot be used to infer causal relationships; further, we were not able to account for possible use of anxiety treatments by study participants.

Conclusions: These findings highlight the importance of understanding anxiety as a biopsychosocial entity. Clinicians and provider organizations need to consider both the physical manifestations of the disorder and their patients' social determinants of health when considering treatment pathways and designing interventions.

* Corresponding author at: Stanford Center for Clinical Research, Stanford University, 3180 Porter Dr, Palo Alto, CA 94304, USA.
E-mail addresses: jcnunes@stanford.edu, julio.nunes@yale.edu (J.C. Nunes).

[1] Nunes JC and Carroll MK equally contributed to this manuscript and should be considered first authors, if possible.

https://doi.org/10.1016/j.jad.2022.08.085
Received 8 December 2021; Received in revised form 14 August 2022; Accepted 22 August 2022
Available online 27 August 2022

1. Introduction

Anxiety disorders are among the most common psychiatric diagnoses, with a lifetime prevalence of 32 % in the United States (Kessler et al., 2005). Worldwide, the World Health Organization estimates that about 300 million people are affected (Kessler et al., 2007). In the United States, anxiety disorders may contribute $110 million to $600 million per million inhabitants in indirect costs to the health care system (Konnopka et al., 2009). They also carry a large individual cost-burden due to their high prevalence, and may lead to school absences, work underperformance, socialization impairment, and overall disability (Swinson et al., 2006). Anxiety is closely related to multiple chronic disorders, such as cardiovascular disease, diabetes mellitus type 2, hypertension, arthritis, and chronic obstructive pulmonary disease (Sareen et al., 2005; Smyth, 2009; Olver and Hopwood, 2013).

Generalized Anxiety Disorder (GAD) is a chronic recurring subtype of anxiety, defined by persistent, excessive, difficult-to-control, and intrusive thoughts that may manifest comorbidly with somatic symptoms, sleep disturbances, and mental exhaustion. It frequently co-occurs with other psychiatric conditions, with at least 60 % of the patients reporting having another disorder, mainly major depression, substance use disorder, attention deficit-hyperactivity disorder, and other types of anxiety (Kessler et al., 2005).

Spitzer et al. (2006) created the first standardized and widely available measurement tool for symptoms of GAD, the General Anxiety Disorder-7 (GAD7). Consisting of 7 questions scored from 0 to 3, this tool is easily usable by clinicians of all specialties and levels. It has since been validated in primary care, general population, outpatient, and inpatient psychiatric settings (Ruiz et al., 2011; Löwe et al., 2008; Johnson et al., 2019). The tool's 7 components assess for 1) feeling nervous or on edge, 2) capacity to control worries and thoughts, 3) difficulty relaxing, 4) excessive worry about multiple themes, 5) restlessness, 6) irritability, and 7) constant worry or fear that unpleasant things will soon happen.

Several studies have examined possible correlations between higher degrees of anxiety and changes in laboratory measures, particularly biomarkers and peptides, related to serotonergic (e.g. 5-HT reuptake binding density, 5-HT plasma concentration), noradrenergic (e.g., platelet alpha-2 adrenergic peripheral binding density), and GABAergic (e.g., lymphocytes peripheral BDZ binding sites) systems (Iny et al., 1994; Schneider et al., 1987; Cameron et al., 1990; Hernández et al., 2002; Weizman et al., 1987; Sevy et al., 1989). Oxidative and immunological systems are also of interest, and previous studies have shown increased levels of inflammatory markers and reactive oxygen radicals in patients experiencing intense degrees of anxiety (Costello et al., 2019; Maes et al., 2018; Fındıklı et al., 2018; Emhan et al., 2015; Ercan et al., 2017). Vismara et al. (2020) recently reviewed and summarized the relationship between anxiety disorders and these biomarkers. However, despite these significant advances in understanding the physiology of anxiety, more research is needed about how it correlates with wide-ranging population health markers, demographic characteristics, and social determinants of health.

The Baseline Health Study (BHS) (Arges et al., 2020) is an ongoing prospective cohort study that aims to determine overall health biomarkers, biopsychosocial status, and demographics of a group of participants whose sex, race, and ethnicities are representative of the broader US adult population. In this study patients are deeply phenotyped, meaning that various demographic, laboratory, clinical, physical function, psychometric, and imaging findings are obtained annually during longitudinal follow-up visits. The BHS represents a unique opportunity to assess multiple types of anxiety correlates.

In this study, we focus on how multiple demographic and socioeconomic data correlate with GAD-7 results when compared with laboratory, physical function, clinical, and other biological markers. Our goal is to understand what role socioeconomic status (SES) and multiple other determinants of health may play in the biopsychosocial approach of anxiety disorders. Although some of the individual correlations of such socioeconomic factors have previously been reported, we add to the current knowledge by demonstrating the full range of these correlations in a single population, due to the comprehensive nature of data collection in the BHS.

2. Methods

2.1. The baseline health study

The Baseline Health Study is an ongoing multicenter longitudinal observational study. Multimodal data, including different demographic, clinical, imaging, and laboratory measurements, were collected for 2502 adult participants. The BHS design has been previously described (Arges et al., 2020), including its data collection plan, inclusion and exclusion criteria, institutional review board regulation, and other main components.

Study participants were recruited and enrolled via an online registration platform. The deeply phenotyped cohort used for this analysis was selected through an algorithm designed to produce a cohort that corresponds proportionately to U.S. adult population demographic data. Persons with various health statuses were included in the study, ranging from healthy controls to patients living with advanced disease. Due to the BHS methodology, the sampling was designed to over-represent persons with heart disease or cancer.

For this study, we performed a cross-sectional analysis that encompasses information of the baseline visit, when GAD-7 and the majority of the clinical data were collected. The GAD-7 was completed either at the in-person visit or digitally after the visit.

The BHS is funded by Verily Life Sciences (San Francisco, CA) and conducted by a collaborative group including Stanford University (Stanford, CA), Duke University (Durham, NC), and the California Health and Longevity Institute (Westlake Village, CA). The enrolling sites locations are Palo Alto, CA; Durham, NC; Kannapolis, NC; and Los Angeles, CA.

2.2. Statistical analysis

Demographic and clinical characteristics were described across the five GAD-7 severity categories (0, 1–4, 5–9, 10–14, 15+), separately for women and men, in the population of participants who completed the GAD-7 survey at baseline. To test associations, the Cochrane-Armitage trend test was used for binary variables, while the Spearman rank correlation test was used for continuous variables. Given the exploratory nature of the study, we did not adjust for multiple tests. Data were considered non-missing if they were collected within 200 days of a participants' baseline in-person study visit.

2.3. SES risk score

Socioeconomic status variables were combined into a single "SES risk score" variable that was used in subsequent LASSO regressions and on Fig. 2. The variable was a sum score of 5 individual items coded as 1/0: 1) reported high school education or less, 2) reported income <$25,000, 3) marital status reported as unmarried, 4) employment status reported as not working, and 5) insurance status reported as uninsured. A higher score was considered as having "higher risk" SES. We opted to have a composite score comprised of the different socioeconomic score domains to illustrate the importance of this assembly of socioeconomic factors on anxiety symptoms.

2.4. Imputation of missing data

All categorical predictors were converted into indicator variables via dummy coding. Multiple imputation by chained equations (MICE) was used to address missing data, as LASSO regression techniques require complete data on the input dataset. MICE assumes that all data are

missing at random (MAR), meaning that the probability of a data point being missing depends only on observed data. Further details on the methodology behind the imputation process have been previously described (Chen and Wang, 2013). This analysis employed predictive mean matching to calculate the missing values of both continuous and categorical target variables. Predictive mean matching ensures that missing data are assigned a value that has already been observed in the data (Azur et al., 2011; Chen and Wang, 2013). The imputation process was repeated a total of five times, with five iterations for each imputation.

2.5. LASSO regression

LASSO procedures addressing multiply imputed data by using group methods (MI-LASSO) were conducted to select models of baseline characteristics that may be predictive of GAD-7 score (logarithm of GAD-7 + 1). All variables were standardized to allow for meaningful comparison of coefficients. In the MI-LASSO method, an L_1-norm (the sum of the absolute values of the regression coefficients) penalization is placed on the regression model; this penalizes the inclusion of additional variables and forces the coefficients of variables that do not contribute to the model to zero. In the presence of multiply imputed data, LASSO can produce different results for each imputed dataset; MI-LASSO fits models on all datasets jointly while considering each set of estimated coefficients ($\widehat{\beta}_{1,j}$, …, $\widehat{\beta}_{D,j}$, where D is the number of imputed datasets) as a group, and applying the group LASSO penalty to the model. Shrinkage of the coefficients is controlled by the tuning parameter (λ), which is optimized by selecting the model that minimizes the Bayesian Information Criterion (BIC) (Chen and Wang, 2013).

Analyses were performed using R v3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). Data were imputed using the 'MICE' package v3.13.0 (van Buuren and Groothuis-Oudshoorn, 2011). MI-LASSO regressions were run using the MI.LASSO function developed by Chen and Wang (2013). Figures were created using the 'ggplot2' package v3.3.0 (Wickham and Sievert, 2016).

3. Results

A total of 2453 participants completed baseline GAD-7 assessments, of whom 1343 (55 %) were female and 1110 (45 %) were male. Associations between GAD-7 scores and demographic data are shown in Table 1. Younger age, Hispanic ethnicity, and identifying as other race (defined as one of: Native Hawaiian or Pacific Islander, American Indian or Alaska Native, or Other race) were associated with higher GAD-7 scores for both men and women. The study's population racial-ethnic distribution represents that of the US census, with the exception of Hispanic persons, who are underrepresented (11.3 % of the study population compared with 18.4 % according to 2019 estimates from the US Census).

Table 2 describes associations between markers of socioeconomic status and GAD-7 scores. Less education, lower income, unmarried status, not working for at least 1 year, and uninsured status were all associated with higher anxiety scores in both men and women. In general, higher GAD-7 scores were associated with lower SES across multiple types of measurement.

Five datasets were imputed from the observed data using 5 cycles per imputation; the 5 datasets were simultaneously entered in the MI-LASSO model. Results from the full MI-LASSO regression models, which included 158 total predictors and were stratified by sex, are presented in Fig. 1. Predictors classically described as somatic symptoms of anxiety per DSM-V criteria (e.g., headaches) were not included in the final model to allow for the observation of factors (e.g., symptoms of other medical conditions; sociodemographic data) not yet established as components of the syndrome of anxiety. A full list of variables entered into the MI-LASSO regression is available in Appendix A. In both men

Table 1
Summary of demographic characteristics across GAD-7 severity scores at baseline, stratified by sex. Data are presented as N (%) or median (IQR). *P*-values for trend were calculated with the use of Spearman Correlation or Cochrane-Armitage tests, where appropriate.

Demographics	Overall	Female GAD-7						Male GAD-7					
	N = 2453	0 (*N* = 404)	1–4 (*N* = 531)	5–9 (*N* = 257)	10–14 (*N* = 103)	15+ (*N* = 48)	*p*-Value	0 (*N* = 418)	1–4 (*N* = 479)	5–9 (*N* = 137)	10–14 (*N* = 54)	15+ (*N* = 22)	*p*-Value
Age, median (IQR)	50.0 (35.3–64.1)	57.2 (43.3–68.7)	49.7 (35.3–62.4)	47.5 (34.3–58.2)	44.1 (32.7–56.0)	42.8 (30.0–54.2)	<0.0001	59.0 (43.0–70.4)	46.5 (31.8–62.9)	41.9 (33.6–55.6)	31.8 (26.1–50.7)	42.1 (29.5–53.2)	<0.0001
Race/ethnicity													
White, n (%)	1561 (63.6)	266 (65.8)	347 (65.3)	159 (61.9)	60 (58.3)	31 (64.6)	0.1842	272 (65.1)	300 (62.6)	90 (65.7)	28 (51.9)	8 (36.4)	0.0204
Black, n (%)	390 (15.9)	81 (20.0)	77 (14.5)	41 (16.0)	18 (17.5)	6 (12.5)	0.1680	62 (14.8)	68 (14.2)	18 (13.1)	10 (18.5)	9 (40.9)	0.0846
Asian, n (%)	254 (10.4)	29 (7.2)	49 (9.2)	25 (9.7)	6 (5.8)	3 (6.2)	0.9947	56 (13.4)	68 (14.2)	12 (8.8)	5 (9.3)	1 (4.5)	0.0943
Native Hawaiian or Pacific Islander, N (%)	27 (1.1)	3 (0.7)	7 (1.3)	5 (1.9)	0 (0)	0 (0)	0.6227	4 (1.0)	6 (1.3)	2 (1.5)	0 (0)	0 (0)	0.6063
American Indian or Alaska Native, n (%)	28 (1.1)	2 (0.5)	7 (1.3)	5 (1.9)	4 (3.9)	1 (2.1)	0.0142	4 (1.0)	3 (0.6)	1 (0.7)	1 (1.9)	0 (0)	0.6238
Other race, n (%)	193 (7.9)	23 (5.7)	44 (8.3)	22 (8.6)	15 (14.6)	7 (14.6)	0.0017	20 (4.8)	34 (7.1)	14 (10.2)	10 (18.5)	4 (18.2)	<0.0001
Hispanic ethnicity, n (%)	279 (11.3)	38 (9.4)	58 (10.9)	36 (14.0)	23 (22.3)	12 (25.0)	<0.0001	33 (7.9)	50 (10.4)	17 (12.4)	8 (14.8)	4 (18.2)	0.0146
Site													
West Lake, n (%)	468 (19.1)	45 (11.1)	97 (18.3)	58 (22.6)	29 (28.2)	15 (31.2)	<0.0001	60 (14.4)	114 (23.8)	30 (21.9)	18 (33.3)	2 (9.1)	0.0069
Durham, n (%)	479 (19.5)	101 (25.0)	112 (21.1)	50 (19.5)	16 (15.5)	7 (14.6)	0.0087	86 (20.6)	66 (13.8)	19 (13.9)	14 (25.9)	8 (36.4)	0.7588
Kannapolis, n (%)	511 (20.8)	99 (24.5)	111 (20.9)	45 (17.5)	22 (21.4)	16 (33.3)	0.6918	96 (23.0)	78 (16.3)	30 (21.9)	7 (13.0)	7 (31.8)	0.4171
Palo Alto, n (%)	995 (40.6)	159 (39.4)	211 (39.7)	104 (40.5)	36 (35.0)	10 (20.8)	0.0912	176 (42.1)	221 (46.1)	58 (42.3)	15 (27.8)	5 (22.7)	0.0758

GAD-7, Generalized Anxiety Disorder Questionnaire-7.

Table 2
Summary of socioeconomic status across GAD-7 severity scores at baseline, stratified by sex. Data are presented as N (%) or median (IQR). P-values for trend were calculated with the use of Spearman Correlation or Cochrane-Armitage tests, where appropriate.

Socioeconomic status	Overall	Female GAD-7						Male GAD-7					
	N = 2453	0 (N = 404)	1–4 (N = 531)	5–9 (N = 257)	10–14 (N = 103)	15+ (N = 48)	*p*-Value	0 (N = 418)	1–4 (N = 479)	5–9 (N = 137)	10–14 (N = 54)	15+ (N = 22)	p-Value
Education													
High school or less, n (%)	176 (7.2)	28 (6.9)	32 (6.0)	17 (6.6)	15 (14.6)	9 (18.8)	0.0024	25 (6.0)	20 (4.2)	15 (10.9)	11 (20.4)	4 (18.2)	<0.0001
Some college, n (%)	455 (18.5)	70 (17.3)	103 (19.4)	71 (27.6)	25 (24.3)	11 (22.9)	0.0076	65 (15.6)	70 (14.6)	24 (17.5)	10 (18.5)	6 (27.3)	0.2192
College, n (%)	634 (25.8)	120 (29.7)	178 (33.5)	68 (26.5)	24 (23.3)	11 (22.9)	0.0721	95 (22.7)	105 (21.9)	24 (17.5)	8 (14.8)	1 (4.5)	0.0192
Graduate degree or higher, n (%)	670 (27.3)	123 (30.4)	140 (26.4)	58 (22.6)	13 (12.6)	4 (8.3)	<0.0001	144 (34.4)	144 (30.1)	37 (27.0)	6 (11.1)	1 (4.5)	<0.0001
Income													
<$25,000, n (%)	193 (7.9)	24 (5.9)	40 (7.5)	34 (13.2)	20 (19.4)	9 (18.8)	<0.0001	20 (4.8)	24 (5.0)	6 (4.4)	10 (18.5)	6 (27.3)	<0.0001
$25,000–50,000, n (%)	256 (10.4)	43 (10.6)	52 (9.8)	39 (15.2)	15 (14.6)	9 (18.8)	0.0186	34 (8.1)	44 (9.2)	14 (10.2)	6 (11.1)	0 (0.0)	0.7462
$50,000–100,000, n (%)	493 (20.1)	86 (21.3)	142 (26.7)	53 (20.6)	26 (25.2)	6 (12.5)	0.5255	70 (16.7)	77 (16.1)	24 (17.5)	7 (13.0)	2 (9.1)	0.4488
$100,000–150,000, n (%)	308 (12.6)	57 (14.1)	61 (11.5)	28 (10.9)	2 (1.9)	2 (4.2)	0.0006	64 (15.3)	71 (14.8)	19 (13.9)	2 (3.7)	2 (9.1)	0.0678
$150,000–200,000, n (%)	197 (8.0)	35 (8.7)	50 (9.4)	16 (6.2)	5 (4.9)	2 (4.2)	0.0643	43 (10.3)	30 (6.3)	12 (8.8)	3 (5.6)	1 (4.5)	0.1106
>$200,000, n (%)	358 (14.6)	64 (15.8)	77 (14.5)	27 (10.5)	8 (7.8)	3 (6.2)	0.0028	79 (18.9)	75 (15.7)	20 (14.6)	5 (9.3)	0 (0.0)	0.0077
Marital status													
Married, n (%)	1054 (43.0)	179 (44.3)	242 (45.6)	97 (37.7)	31 (30.1)	11 (22.9)	0.0002	232 (55.5)	200 (41.8)	44 (32.1)	13 (24.1)	5 (22.7)	<0.0001
Divorced, n (%)	180 (7.3)	42 (10.4)	49 (9.2)	30 (11.7)	12 (11.7)	4 (8.3)	0.7771	20 (4.8)	15 (3.1)	4 (2.9)	2 (3.7)	2 (9.1)	0.7563
Formerly in long term relationship, n (%)	97 (4.0)	12 (3.0)	22 (4.1)	11 (4.3)	3 (2.9)	0 (0.0)	0.9021	16 (3.8)	24 (5.0)	5 (3.6)	3 (5.6)	1 (4.5)	0.6686
Living together, n (%)	205 (8.4)	26 (6.4)	46 (8.7)	31 (12.1)	9 (8.7)	8 (16.7)	0.0084	23 (5.5)	34 (7.1)	19 (13.9)	8 (14.8)	1 (4.5)	0.0049
Never in long term relationship, n (%)	255 (10.4)	38 (9.4)	60 (11.3)	22 (8.6)	13 (12.6)	9 (18.8)	0.1884	26 (6.2)	54 (11.3)	21 (15.3)	9 (16.7)	3 (13.6)	0.0004
Separated, n (%)	48 (2.0)	9 (2.2)	6 (1.1)	14 (5.4)	5 (4.9)	1 (2.1)	0.0378	6 (1.4)	4 (0.8)	3 (2.2)	0 (0.0)	0 (0.0)	0.8138
Widowed, n (%)	75 (3.1)	29 (7.2)	23 (4.3)	5 (1.9)	3 (2.9)	1 (2.1)	0.0025	3 (0.7)	9 (1.9)	2 (1.5)	0 (0.0)	0 (0.0)	0.4172
Employment status													
Employed for wages, n (%)	1147 (46.8)	189 (46.8)	273 (51.4)	121 (47.1)	40 (38.8)	20 (41.7)	0.1892	175 (41.9)	228 (47.6)	71 (51.8)	23 (42.6)	7 (31.8)	0.4526
Homemaker, n (%)	67 (2.7)	13 (3.2)	25 (4.7)	18 (7.0)	5 (4.9)	4 (8.3)	0.0372	0 (0.0)	0 (0.0)	1 (0.7)	1 (1.9)	0 (0.0)	0.0128
Unable to work, n (%)	65 (2.6)	1 (0.2)	11 (2.1)	11 (4.3)	13 (12.6)	4 (8.3)	<0.0001	5 (1.2)	8 (1.7)	5 (3.6)	4 (7.4)	3 (13.6)	<0.0001
Not working ≥1 year, n (%)	49 (2.0)	5 (1.2)	11 (2.1)	8 (3.1)	4 (3.9)	2 (4.2)	0.0281	4 (1.0)	7 (1.5)	6 (4.4)	1 (1.9)	1 (4.5)	0.0279
Retired, n (%)	457 (18.6)	106 (26.2)	104 (19.6)	25 (9.7)	7 (6.8)	5 (10.4)	<0.0001	121 (28.9)	72 (15.0)	12 (8.8)	5 (9.3)	0 (0.0)	<0.0001
Student, n (%)	56 (2.3)	9 (2.2)	11 (2.1)	11 (4.3)	2 (1.9)	1 (2.1)	0.5223	5 (1.2)	13 (2.7)	2 (1.5)	2 (3.7)	0 (0.0)	0.2892
Not working <1 year, n (%)	58 (2.4)	11 (2.7)	9 (1.7)	7 (2.7)	5 (4.9)	1 (2.1)	0.4999	9 (2.2)	11 (2.3)	4 (2.9)	1 (1.9)	0 (0.0)	0.8428
Self-employed, n (%)	254 (10.4)	34 (8.4)	41 (7.7)	35 (13.6)	14 (13.6)	2 (4.2)	0.1543	46 (11.0)	61 (12.7)	14 (10.2)	4 (7.4)	3 (13.6)	0.7879
Health insurance status													
Insured, n (%)	1809 (73.7)	321 (79.5)	436 (82.1)	192 (74.7)	67 (65.0)	32 (66.7)	0.0003	315 (75.4)	313 (65.3)	93 (67.9)	33 (61.1)	7 (31.8)	<0.0001
Uninsured, n (%)	110 (4.5)	17 (4.2)	15 (2.8)	20 (7.8)	7 (6.8)	3 (6.2)	0.0391	11 (2.6)	24 (5.0)	7 (5.1)	2 (3.7)	4 (18.2)	0.0083
Smoking status													
Current smoker, n (%)	331 (13.5)	34 (8.4)	56 (10.5)	46 (17.9)	25 (24.3)	14 (29.2)	<0.0001	43 (10.3)	62 (12.9)	23 (16.8)	17 (31.5)	11 (50.0)	<0.0001
Former smoker, n (%)	539 (22.0)	73 (18.1)	119 (22.4)	55 (21.4)	21 (20.4)	8 (16.7)	0.6857	109 (26.1)	105 (21.9)	31 (22.6)	11 (20.4)	7 (31.8)	0.4639
Nonsmoker, n (%)	1583 (64.5)	297 (73.5)	356 (67.0)	156 (60.7)	57 (55.3)	26 (54.2)	<0.0001	266 (63.6)	312 (65.1)	83 (60.6)	26 (48.1)	4 (18.2)	0.0005

GAD-7, Generalized Anxiety Disorder Questionnaire-7.

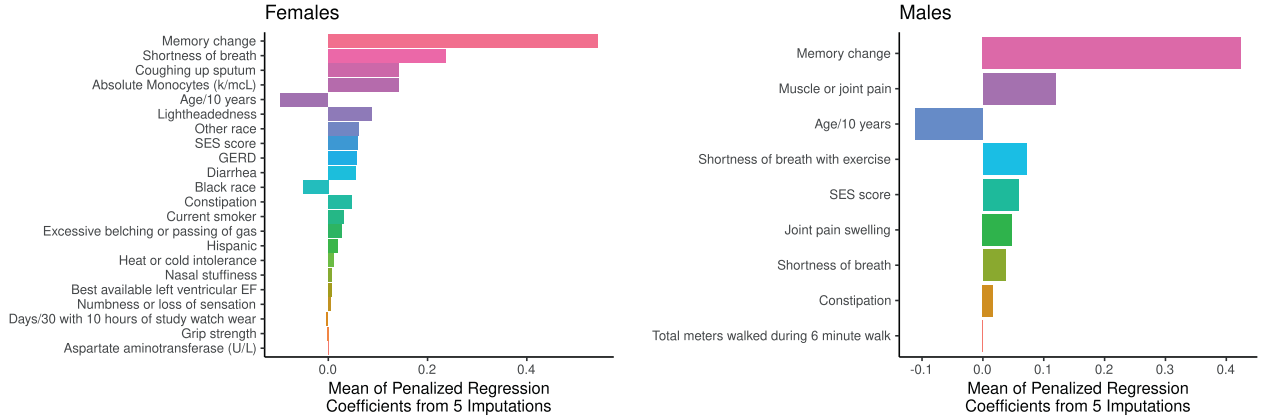

Fig. 1. Factors associated with GAD-7 score after MI-LASSO regression. The mean of the MI-LASSO regression coefficients from 5 imputed datasets that predict GAD-7 severity are presented for both male and female study participants. Coefficients are sorted in descending order of magnitude.

Fig. 2. Associations between socioeconomic status (SES) and GAD-7 score, stratified by race or ethnicity. Each panel describes the interaction of GAD-7 severity and SES risk score within one of 43 racial grouping or one ethnicity grouping. Groups were not mutually exclusive, as those who reported Hispanic or Latino ethnicity could have also reported White, Black, or Asian race. Those who reported "Other" races were omitted from the figure. Column percentages were calculated as the counts within an individual cell divided by the total count of participants given that cells' GAD-7 severity score. Individual cells were shaded based on the column percentage; darker cells indicate a greater column percentage. Fisher's Exact Tests for count data with simulated p-value were performed within each race/ethnicity category; significant overall associations between SES risk score and GAD-7 severity for all groupings (White: $p < 0.001$; Black or African American: $p = 0.012$; Asian: $p = 0.010$; Hispanic or Latino ethnicity: $p = 0.005$).

and women, memory change was the strongest positive predictor of GAD-7 severity. In women, the coefficient was more than twice that of the next strongest positive predictor (shortness of breath); in men, the coefficient was >3 times that of the next strongest positive predictor (muscle or joint pain). For women, we also observed positive associations between GAD-7 severity and other physical health-related symptoms (e.g., coughing up sputum, lightheadedness, diarrhea, and constipation), absolute monocyte levels, reporting Other race, and SES score. There were negative relationships between GAD-7 severity and both age and Black race. In men, there were positive associations between GAD-7 and other physical function measures including shortness of breath with exercise, joint pain or swelling, shortness of breath, and constipation. Similar to findings observed in women, GAD-7 severity decreased as age increased in men.

Fig. 2 presents a heat map figure of associations between SES risk scores and GAD-7 severity, stratified by race. Overall associations between SES risk score and GAD-7 severity were significant for all races. Black participants tended to have higher concentrations in the high-SES risk, high-GAD-7 severity corner of the table; as SES risk score increased, the likelihood of having a higher GAD-7 severity score also increased. Black participants were also most likely to have an SES score of at least 3. White participants had higher concentrations in the low-SES risk, low GAD-7 corner of the table; although there was a similar trend toward higher GAD-7 scores as SES risk score increased, white participants had lower SES risk scores overall. Asian participants were more likely to have low SES risk, and few participants ($n = 13$) had a GAD-7 severity of at least 10. Hispanic or Latino participants were more likely to have lower SES risk scores, however those with higher SES risk scores were more likely to have higher GAD-7 severity scores.

4. Discussion

Our study enrolled a large, deeply phenotyped cohort from the Baseline Health Study which allowed us to assess a broad array of anxiety correlates, spanning physiological, medical, psychological, and social factors. Our findings confirm and highlight previous literature regarding GAD-7 and different measurements of SES. Here, we demonstrate the importance of understanding GAD-7 scores within a psychosocial contextual, as it raises important concerns not only about psychological symptoms, but also about how anxiety scores are associated with broader social determinants of health. Although some of the individual correlations are expected, their overall assembly underscores the importance of the biopsychosocial model when managing mental health disorders.

Previous studies have demonstrated how low income, unemployment, fewer years of education, lack of insurance, and smoking correlate with higher levels of anxiety. Hinz et al. (2017) validated the psychometric properties of the GAD-7 in Germany and found that those social determinants of health were correlated with higher anxiety scores. Finegan et al. (2019) investigated the influence of socioeconomic deprivation and neighborhood violence on anxiety and found similar results but also observed that low neighborhood average income and high crime rates were likewise correlated with higher GAD-7 scores. Chen et al. (2019) reported similar findings in an analysis of 13,775 medical records.

Utilizing a regularized regression approach in a comprehensively phenotyped dataset, our findings substantiate the role of socioeconomic variables as an integral construct in anxiety scores. Indeed, we found that “Socioeconomic Risk” was among the 7 most important anxiety correlates for women and among the 5 most important correlates for men in the LASSO regressions, highlighting the importance of the assembly of these non-physiological factors on anxiety. Indeed, social determinants of health have been found to have equal or greater impact on mortality relative to physiological and clinical indicators (Adler and Stead, 2015; Pantell et al., 2013), making it unsurprising that we identified similar associations in anxiety scores.

With this framework in mind, clinicians and policy makers must consider socioeconomic issues, access to care, and broader social context when designing interventions for addressing anxiety in clinical settings. Different treatment modalities may be more effective if employed using holistic approaches that incorporate more than just biochemical principles or assessment of cognitive function. Furthermore, SES may limit the types of pharmacological interventions and therapy that patients may receive, thus, clinicians and health systems should be focused on innovative methods to ensure people with anxiety have equal access to the standard of care treatments (Giebel et al., 2020; Evans-Lacko et al., 2017).

Findings also indicated potential relationships between anxiety scores and self-reported race/ethnicity status. Identifying as a person of Hispanic/Latinx ethnicity was correlated with higher GAD-7 scores, a result consistent with previously published findings. Factors that may contribute to this correlation include higher rates of unemployment, job insecurity, lack of access to health care, distance from family, use of English as a second language, and systemic racism and prejudice (Alvarez et al., 2018; Georgiades et al., 2018). Additionally, concerns about migration and visa status could lead certain ethnic groups to not enroll, which may be reflected in the underrepresentation of Latinx populations in our sample. We did not ask participants about their visa status, which is a limitation of this study. This aspect is important, as insecurity around residency status has been found to contribute to higher degrees of psychosocial distress among Hispanic populations (Alif et al., 2020; Ross et al., 2019).

After adjustment for multiple factors, Black race was correlated with lower overall GAD-7 scores, a finding consistent with previous studies suggesting that Black individuals may have higher levels of resiliency when facing adversity (Assari, 2016). However, when focusing on the relationships between race, SES risk score, and GAD-7 score (Fig. 2), Black participants with high SES risk mostly have higher scores, suggesting that poor SES may be a significant associated factor in the development of anxiety in this group. This contrasts with White participants, who had higher levels of anxiety distributed across all levels of SES. Nevertheless, this unexpected finding should be considered preliminary and warrants further investigation.

We also observed higher levels of anxiety among young unmarried women. These findings were previously reported in the literature and most recently updated during the COVID-19 pandemic. In studies of temporal trends and demographic disparities of mental health disorders, sex, age, and marital status were predictors of higher deterioration of mental health status, particularly when stricter social isolation measures were in place (Santabárbara et al., 2021; Kantor and Kantor, 2020).

The BHS is longitudinally assessing a large cohort of patients for markers of social, biological, clinical, and behavioral function, and once it is completed, will provide an opportunity to better understand these relationships across time. Additionally, by the end of the study, race- and ethnicity-specific analyses may allow a more nuanced understanding of whether social determinants of health interact with anxiety in specific racial and ethnic populations. The representation of each demographic group is a priority for the BHS study and its final population pool aims to reflect every race/ethnicity cohort per the US census. The BHS is also an opportunity to understand the validity of behavioral data collection in the context of digital health.

4.1. Limitations

Several limitations to this study should be noted. Cross-sectional studies cannot determine causality. Therefore, this study can demonstrate anxiety correlations but does not attempt to explain basic underlying causal effects. In addition, we lack details about anxiety treatments, which could be an unaccounted modifying factor. In future longitudinal follow-up studies, we plan to thoroughly consider behavioral treatments and medications and their associations.

4.2. Conclusion

In conclusion, higher GAD-7 scores were correlated with multiple measures across the spectrum of biological, clinical, behavioral, and social measures. The relationship with SES is particularly striking, highlighting the importance of considering social determinants of health when designing interventions for anxiety. Focusing solely on the biological treatment of anxiety or the personal cognitive issues may have a limited effectiveness when compared with holistic approaches to care. Clinicians and policy makers are likely to be more effective if they are aware of social and contextual circumstances unique to the patient combined with social factors affecting groups. Additionally, when systems are set in place aiming to improve SES of individuals and populations, anxiety should be considered as a factor that may limit the capacity of these individuals to participate and respond to those interventions.

CRediT authorship contribution statement

Julio C. Nunes – Drafted and wrote the paper, performed the analysis.

Megan K. Carroll – Contributed data or analysis tools, performed the analysis, wrote the paper.

Kenneth W. Mahaffey – Conceived and designed the analysis, reviewed the paper.

Robert M. Califf – Conceived and designed the analysis, reviewed the paper.

P. Murali Doraiswamy – Conceived and designed the analysis, reviewed the paper.

Sarah Short – Contributed data or analysis tools, performed the analysis, reviewed the paper.

Svati H. Shah – Contributed data or analysis tools, reviewed the paper.

Susan Swope – Collected the data, reviewed the paper.

Donna Williams – Collected the data, reviewed the paper.

Adrian F. Hernandez – Conceived and designed the analysis, reviewed the paper.

David S. Hong – Conceived and designed the analysis, wrote and reviewed the paper, coordinated writing and analysis efforts.

Disclosures

Dr. Califf is an employee of Verily Life Sciences and Google Health; he also serves as an advisory board/science advisory board member for Human Health and Potential (Singapore), Basking Biosciences, Launch and Scale Speedometer, and Medicxi Ventures. He is also a board member for AMYRIAD Pharma, Centessa Pharmaceuticals, Clinetic, One Fifteen, Portola, and Cytokinetics. Ms. Carroll and Ms. Short are employees of Verily Life Sciences. Dr. Hernandez notes external relationships with Amgen, AstraZeneca, Bayer, Biofourmis Singapore, Boehringer Ingelheim, Boston Scientific, Cytokinetics, Eidos Therapeutics, Eli Lilly, Intercept, Luitpold Pharmaceuticals, Merck, MyoKardia, Novartis, Pfizer, Relypsa, and Verily. Dr. Doraiswamy notes external relationships with Advera Health, Apollo Hospitals, Danone, Evidation Health, Live Love Laugh, Lumos Labs, MarvelBiome, Neuroglee, Transposon Therapeutics, Turtle Shell Technologies, UMethod, VitaKey, and Vivly. Dr. Shah reports external relationships with American Heart Association, AstraZeneca, Baylis Medical, Biosense Webster, Cardivia Medical, Lilly, McGraw-Hill Publishing, the National Institutes of Health, NewPace, and Project Baseline, LLC. Dr. Mahaffey's financial disclosures can be viewed at http://med.stanford.edu/profiles/kenneth-mahaffey.

Funding

The Baseline Health Study and this analysis were funded by Verily Life Sciences, San Francisco, CA. This analysis was funded by Verily Life Sciences (San Francisco, CA)

in partnership with Stanford University and Duke University.

Role of the funding source

The employees from funding source contributed to the data analyses, interpretation, editing of the manuscript and are coauthors. The final decision to submit the manuscript was made by the academic authors.

Data availability statement

The didentified Project Baseline Health Study (PBHS) data corresponding to this study are available upon request for the purpose of examining its reproducibility. Interested investigators should direct requests to jsaiz@verily.com. Requests are subject to approval by PBHS governance.

Conflict of interest

None.

Acknowledgments

Baseline Health Study Team: American Society of Clinical Oncology, Alexandria, VA, USA: Richard L. Schilsky. Duke University, School of Medicine, Durham, NC, USA: Jennifer Allen, MaryAnn Anderson, Kevin Anstrom, Lucus Araujo, Kristine Arges, Kaveh Ardalan, Bridget Baldwin, Suresh Balu, Mustafa R. Bashir, Manju Bhapkar, Robert Bigelow, Tanya Black, Rosalia Blanco, Gerald Bloomfield, Durga Borkar, Leah Bouk, Ebony Boulware, Nikki Brugnoni, Erin Campbell, Paul Campbell, Larry Carin, Tammy Jo Cassella, Tina Cates, Ranee Chatterjee Montgomery, Victoria Christian, John Choong, Michael Cohen-Wolkowiez, Elizabeth Cook, Scott Cousins, Ashley Crawford, Nisha Datta, Melissa Daubert, James Davis, Jillian Dirkes, Isabelle Doan, Marie Dockery, P. Murali Doraiswamy, Pamela S. Douglas, Shelly Duckworth, Ashley Dunham, Gary Dunn, Ryan Ebersohl, Julie Eckstrand, Vivienne Fang, April Flora, Emily Ford, Lucia Foster, Elizabeth Fraulo, John French, Geoffrey S. Ginsburg, Cindy Green, Latoya Greene, Jeffrey Guptill, Donna Hamel, Jennifer Hamill, Chris Harrington, Rob Harrison, Lauren Hedges, Brooke Heidenfelder, Adrian F. Hernandez, Cindy Heydary, Tim Hicks, Lina Hight, Deborah Hopkins, Erich S. Huang, Grace Huh, Jillian Hurst, Kelly Inman, Gemini Janas, Glenn Jaffee, Janace Johnson, Tiffanie Keaton, Michel Khouri, Daniel King, Jennifer Korzekwinski, Lynne H. Koweek, Anthony Kuo, Lydia Kwee, Dawn Landis, Rachele Lipsky, Desiree Lopez, Carolyn Lowry, Kelly Marcom, Keith Marsolo, Paige McAdams, Shannon McCall, Robert McGarrah, John McGugan, Dani Mee, Sabrena Mervin-Blake, Prithu Mettu, Mathias Meyer, Justin Meyers, Calire N. Miller, Rebecca Moen, Lawrence H. Muhlbaier, Michael Murphy, Ben Neely, L. Kristin Newby, Jayne Nicoldson, Hoang Nguyen, Maggie Nguyen, Lori O'Brien, Sumru Onal, Jeremey O'Quinn, David Page, Neha J. Pagidipati, Kishan Parikh, Sarah R. Palmer, Bray Patrick-Lake, Brenda Pattison, Michael Pencina, Eric D. Peterson, Jon Piccini, Terry Poole, Tom Povsic, Alicia Provencher, Dawn Rabineau, Annette Rich, Susan Rimmer, Fides Schwartz, Angela Serafin, Nishant Shah, Svati Shah, Kelly Shields, Steven Shipes, Peter Shrader, Jon Stiber, Lynn Sutton, Geeta Swamy, Betsy Thomas, Sandra Torres, Debara Tucci, Anthony Twisdale, Susan A. Whitney, Robin Williamson, Lauren Wilverding, Charlene A. Wong, Lisa Wruck. Ellen Young Gemini Group, USA: Jane Perlmutter. Health Collaboratory and Cancer 101, New York, NY, USA: Sarah Krug. Rare Dots, Inc., USA: S. Whitney Bowman-Zatzkin. Society of Participatory Medicine, USA: Sarah Krug. Stanford University, School of Medicine, Stanford, CA, USA: Themistocles Assimes, Vikram Bajaj, Maxwell Cheong, Millie Das, Manisha Desai, Alice C. Fan, Dominik Fleischmann, Sanjiv S. Gambhir, Garry Gold, Francois Haddad, David Hong, Curtis Langlotz, Yaping J. Liao, Rong Lu, Kenneth W. Mahaffey, David Maron, Rebecca

McCue, Rajan Munshi, Fatima Rodriguez, Sumana Shashidhar, George Sledge, Susie Spielman, Ryan Spitler, Sue Swope, Donna Williams, Julio C. Nunes. University of Florida, College of Medicine, Gainesville, FL, USA: Carl J Pepine. University of Missouri, Children's Mercy Hospital, Kansas City, MO, USA: John D Lantos. University of Texas, Dell Medical School, Austin, TX, USA: Michael Pignone. University of Washington, Department of Biostatistics, Seattle, WA, USA: Patrick Heagerty. Vanderbilt University, School of Medicine, Nashville, TN, USA: Laura Beskow, Gordon Bernard. Verily Inc., South San Francisco, CA, USA: Kelley Abad, Giulia Angi, Robert M. Califf, Lawrence Deang, Joy Huynh, Manway Liu, Cherry Mao, Michael Magdaleno, William J. Marks, Jr., Jessica Mega, David Miller, Nicole Ong, Darshita Patel, Vanessa Ridaura, Scarlet Shore, Sarah Short, Michelle Tran, Veronica Vu, Celeste Wong. Harvard University, School of Medicine, Boston, MA, USA: Robert C. Green. Google Inc., Mountain View, CA, USA: John Hernandez. California Health and Longevity Institute, Westlake Village, CA, USA: Jolene Benge, Gislia Negrete, Gelsey Sierra, Terry Schaack

Appendix A. List of candidate covariates for multivariable LASSO regression

A.1. Demographic and baseline characteristics

Variable name	Definition	Description of derivation
Age_at_enrollmentd10	Age in years at baseline	(enrollment_date - DOB)/10
Female	Self-reported sex at birth	N/A
Race_asian	Self-reported race	N/A
Race_black		
Race_other		
Hispanic	Self-reported Hispanic ancestry (yes/no)	N/A
Current_smoker	Self-reported smoking status at baseline	N/A
Former_smoker		
Pack_years_smoked	Self-reported cigarettes smoked per day multiplied by years of regular smoking divided by 20	Cigarettes_per_day * years_regular_smoking / 20

A.2. On site assessments - vitals and physical health metrics

Variable name	Definition	Description of derivation
Systolic_blood_pressure	Average of 2 systolic blood pressure readings (mmHg) measured at baseline	(sbp1_mmhg + sbp2_mmhg) / 2
Diastolic_blood_pressure	Average of 2 diastolic blood pressure readings (mmHg) measured at baseline	(dbp1_mmhg + dbp2_mmhg) / 2
vs_pulse_bpm	Heart rate (bpm) measured at baseline	N/A
Body_mass_index	Body weight (kg) divided by the square of height (cm) at baseline	Weight_kg / (height_cm / 100)^2
vs_wc_cm	Waist circumference (cm) at baseline	N/A
vs_osat_pct	Oxygen saturation (%) at baseline	N/A
vs_rrate_bpm	Respiratory rate (bpm) at baseline	N/A
Ankle_brachial_index	Minimum of left and right ankle brachial index at baseline, using the maximum of left dorsalis pedis pressure and left posterior tibial pressure, maximum of right dorsalis pedis pressure and right posterior tibial pressure, and maximum of right and left brachial systolic pressure	Minimum between: Max (left_dp + left_pt) / max (right brachial_pressure + left_brachial_pressure) and max (right_dp + right_pt) / max (right brachial_pressure + left_brachial_pressure)
Ratio_of_forced_expiratory_volume	Ratio of forced expiratory volume (first third of forced breath) and forced vital capacity at baseline	FEV1/FVC
pperf_6mdis_m	Total meters walked during 6 min walk distance test at baseline	N/A
Ten_meter_walk_speed	Walking speed in meters per second using average of 3 fast walk trials during 10 m walk test at baseline	6/mean (fw1_sec + fw2_sec + fw3_sec)
Handgrip_dominant	Average of 3 hand grip trials of dominant hand in kilograms at baseline	Mean (right1_kg + right2_kg + right3_kg) or mean (left1_kg + left2_kg + left3_kg)
Single_legged_balance	Average of left and right leg trials in seconds during single-legged balance test at baseline	(slb_left_sec + slb_right_sec) / 2
Sitting_rising_score	Sum of sitting and rising scores during sitting-rising test at baseline	Sitting_score + rising_score
pperf_30sscr	Number of stands during 30 s chair stand test at baseline	N/A
Best_available_left_ventricular	Best available left ventricular EF from biplane, single and visual read data from resting echocardiogram at baseline	Biplane EF if available; if not then single plane EF; else visual EF
lvmassi	Left ventricular mass index from resting echocardiogram at baseline	N/A
ccs	Coronary calcium score from coronary calcium scan at baseline	N/A

A.3. On site assessments - medical conditions (60 most commonly reported), symptoms (50 most commonly reported), and allergies

Variable name	Definition	Description of derivation
Physical health-related medical conditions		
cc_oa	Self-reported osteoarthritis at baseline	N/A
cc_gerd	Self-reported gastroesophageal reflux disease at baseline	N/A
cc_htn	Self-reported hypertension at baseline	N/A
cc_asthma	Self-reported asthma at baseline	N/A

(*continued on next page*)

(*continued*)

Variable name	Definition	Description of derivation
cc_cataracts	Self-reported cataracts at baseline	N/A
cc_hypercholesterolemia	Self-reported hypercholesterolemia at baseline	N/A
cc_dm2	Self-reported type II diabetes at baseline	
cc_sleepapnea	Self-reported sleep apnea at baseline	N/A
cc_colonpolyps	Self-reported colon polyps at baseline	N/A
cc_pneumonia	Self-reported pneumonia	N/A
cc_hypothyroidism	Self-reported hypothyroidism at baseline	N/A
cc_hearingloss	Self-reported severe hearing loss at baseline	N/A
cc_kidney_or_bladder_stones	Self-reported kidney or bladder stones at baseline	N/A
cc_arrhythmia	Self-reported arrhythmia at baseline	N/A
cc_gallbladder	Self-reported gallbladder disease at baseline	N/A
cc_tinnitus	Self-reported tinnitus at baseline	N/A
cc_ibd	Self-reported irritable bowel disorder at baseline	N/A
cc_osteopenia	Self-reported osteopenia at baseline	N/A
cc_nonmelanoma	Self-reported non-melanoma skin cancer at baseline	N/A
cc_osteoporosis	Self-reported osteoporosis at baseline	N/A
cc_hemorrhoids	Self-reported hemorrhoids at baseline	N/A
cc_gout	Self-reported gout at baseline	N/A
cc_glaucoma	Self-reported glaucoma at baseline	N/A
cc_bph	Self-reported benign prostatic hyperplasia at baseline	N/A
cc_diverticulosis	Self-reported diverticulosis at baseline	N/A
cc_pud	Self-reported peptic ulcers at baseline	N/A
cc_melanoma	Self-reported melanoma skin cancer at baseline	N/A
cc_diverticulitis	Self-reported diverticulitis at baseline	N/A
cc_hxmi	Self-reported myocardial infarction at baseline	N/A
cc_breast_cancer	Self-reported breast cancer at baseline	N/A
cc_copd_emphysema	Self-reported COPD (with emphysema) at baseline	N/A
cc_psoriasis	Self-reported psoriasis at baseline	N/A
cc_cad	Self-reported coronary artery disease (including angina) at baseline	N/A
cc_fibromyalgia	Self-reported fibromyalgia at baseline	N/A
cc_ra	Self-reported rheumatoid arthritis at baseline	N/A
cc_pe_or_dvt	Self-reported PE or DVT at baseline	N/A
cc_epilepsy	Self-reported epilepsy	N/A
cc_hashimotos	Self-reported Hashimoto's disease	N/A
cc_pvd	Self-reported peripheral vascular disease	N/A
cc_prostate_cancer	Self-reported prostate cancer	N/A
cc_nonalcoholfattyliverdx	Self-reported non-alcoholic fatty liver disease	N/A
cc_thyroidgoiter	Self-reported goiter	N/A
cc_hepc	Self-reported Hepatitis C	N/A
cc_dm1	Self-reported Diabetes type 1	N/A
cc_macular_degeneration	Self-reported macular degeneration	N/A
cc_stroke	Self-reported stroke	N/A
cc_afib	Self-reported atrial fibrillation	N/A
cc_tia	Self-reported transient ischemic attack	N/A
cc_hepb	Self-reported Hepatitis B	N/A
Physical health-related symptoms		
Stiffness	Self-reported stiffness at baseline	N/A
Muscle_or_joint_pain	Self-reported muscle or joint pain at baseline	N/A
Nasal_stuffiness	Self-reported nasal stuffiness at baseline	N/A
Runny_nose	Self-reported runny nose at baseline	N/A
Urination_at_night	Self-reported urination at night at baseline	N/A
Floaters	Self-reported floaters at baseline	N/A
Joint_pain_swelling	Self-reported joint pain swelling at baseline	N/A
Itching_skin	Self-reported itching at baseline	N/A
Cough	Self-reported cough at baseline	N/A
Dryness	Self-reported dryness (skin) at baseline	N/A
Easy_bruising_or_bleeding	Self-reported easy bruising or bleeding at baseline	N/A
Tingling_or_numbness_in_extremities	Self-reported tingling or numbness in extremities at baseline	N/A
Tingling_or_pins_and_needles	Self-reported tingling or pins and needles at baseline	N/A
Heartburn	Self-reported heartburn at baseline	N/A
Frequency_of_urination	Self-reported frequency of urination at baseline	N/A
Constipation	Self-reported constipation at baseline	N/A
Leg_cramps	Self-reported leg cramps at baseline	N/A
Diarrhea	Self-reported diarrhea at baseline	N/A
Ear_ringing	Self-reported ear ringing at baseline	N/A
Heat_or_cold_intolerance	Self-reported heat or cold intolerance at baseline	N/A
Night_sweats	Self-reported night sweats at baseline	N/A
Dry_mouth	Self-reported dry mouth at baseline	N/A
Excessive_belching_or_passing_of_gas	Self-reported excessive belching or Passing of gas at baseline	N/A
Shortness_of_breath_with_exercise	Self-reported shortness of breath with exercise at baseline	N/A
Memory_change	Self-reported memory change at baseline	N/A
Lightheadedness	Self-reported lightheadedness at baseline	N/A
Sinus_pain	Self-reported sinus pain at baseline	N/A
Shortness_of_breath	Self-reported shortness of breath at baseline	N/A

(*continued on next page*)

(*continued*)

Variable name	Definition	Description of derivation
Swelling_in_calves_or_feet	Self-reported swelling in calves or feet at baseline	N/A
Coughing_up_sputum	Self-reported coughing up sputum at baseline	N/A
Urgency	Self-reported urgency at baseline	N/A
Hay_fever	Self-reported hay fever at baseline	N/A
Discharge	Self-reported discharge (nose and sinuses) at baseline	N/A
Hemorrhoids	Self-reported hemorrhoids at baseline	N/A
Cramping	Self-reported cramping at baseline	N/A
Numbness_or_loss_of_sensation	Self-reported numbness or loss of sensation at baseline	N/A
fa_ind	Self-reported food allergies (any vs. none, ignoring additional details about which allergen)	N/A
sa_ind	Self-reported seasonal allergies (any vs. none, ignoring additional details about which allergen)	N/A
nsa_ind	Self-reported non-seasonal allergies (any vs. none, ignoring additional details about which allergen)	N/A
ma_ind	Self-reported medication allergies (any vs. none, ignoring additional details about which allergen)	N/A

A.4. On site assessments - mental health surveys

Variable name	Definition	Description of derivation
gad7_total_score	Generalized Anxiety Disorder-7 total score (range 0, 21) at baseline	Sum of 7 individual questions For analysis: = log(gad7_total_score + 1)

A.5. Blood draw - standard laboratory data

Variable name	Definition	N (%) imputed
Standard_labs_hemoglobin_gdl	Hemoglobin (g/dl) at baseline	144 (6 %)
Standard_labs_serum_creatinine_mgdl	Serum creatinine (mg/dl) at baseline	132 (5 %)
Standard_labs_hdl_mgdl	High density lipoprotein (mg/dl) at baseline	132 (5 %)
Standard_labs_ldl_mgdl	Low density lipoprotein (mg/dl) at baseline	183 (7 %)
Standard_labs_triglycerides_mgdl	Triglycerides (mg/dl) at baseline	132 (5 %)
Standard_labs_hba1c_pct_tl_hgb	Hemoglobin A1c (%) at baseline	131 (5 %)
Standard_labs_alt_ul	Alanine aminotransferase (U/L) at baseline	132 (5 %)
Standard_labs_ast_ul	Aspartate aminotransferase (U/L) at baseline	132 (5 %)
Standard_labs_vitamind_ngml	Vitamin D (ng/ml) at baseline	137 (5 %)
Standard_labs_crp_mgl	C-reactive protein (mg/l) at baseline	142 (%)
Standard_labs_blood_glucose_mgdl	Blood glucose (mg/dl) at baseline	132 (5 %)
Standard_labs_neutrophil_segs_pct_wbc	Neutrophil segments (% WBC) at baseline	164 (7 %)
Standard_labs_neutrophils_thou_per_mcl	Total neutrophils (k/mcL) at baseline	164 (7 %)
Standard_labs_lymphocytes_thou_per_mcl	Total lymphocytes (k/mcL) at baseline	164 (7 %)
Standard_labs_magnesium_meql	Magnesium (MEQ/L) at baseline	132 (5 %)
Standard_labs_monocytes_thou_per_mcl	Absolute monocytes (k/mcL) at baseline	164 (7 %)
Standard_labs_eosinophils_thou_per_mcl	Absolute eosinophils (k/mcL) at baseline	164 (7 %)
Standard_labs_basophils_thou_per_mcl	Absolute basophils (k/mcL) at baseline	163 (7 %)
Standard_labs_hematocrit_pct_rbc_blood	Hematocrit (% RBC to whole blood volume) at baseline	144 (6 %)
Standard_labs_mcv_fl	Mean corpuscular volume (fL) at baseline	144 (6 %)
Standard_labs_mch_pg	Mean corpuscular hemoglobin (pg) at baseline	144 (6 %)
Standard_labs_mpv_fl	Mean platelet volume (fL) at baseline	155 (6 %)
Standard_labs_platelets_per_cumm	Platelet count (cumm) at baseline	156 (6 %)
Standard_labs_rbc_mill_per_mcl	Red blood cell count (millions/mcL) at baseline	144 (6 %)
Standard_labs_wbc_thou_per_mcl	White blood cell count (millions/mcL) at baseline	164 (7 %)
Standard_labs_calcium_mgdl	Calcium (mg/dL) at baseline	132 (5 %)
Standard_labs_cholesterol_mgdl	Total cholesterol (mg/dL) at baseline	132 (5 %)
Standard_labs_chloride_meql	Chloride (MEQ/L) at baseline	132 (5 %)
Standard_labs_potassium_meql	Potassium (MEQ/L) at baseline	132 (5 %)
Standard_labs_sodium_meql	Sodium (MEQ/L) at baseline	132 (5 %)
Standard_labs_protein_serum_gdl	Protein in serum (g/dL) at baseline	132 (5 %)
Standard_labs_albumin_gdl	Albumin (g/L) at baseline	132 (5 %)
Standard_labs_uric_acid_mgdl	Uric acid (mg/dL) at baseline	133 (5 %)
Standard_labs_creatinine_random_urine_mgl	Creatinine in urine (mg/dL) at baseline	50 (2 %)
Standard_labs_gfr_mdrd_ml_min	Glomerular filtration rate (mL/min/1.73 m^2) based on Modification of Diet in Renal Disease Study equation at baseline	132 (5 %)
Standard_labs_reticulocytes_bill_per_liter	Absolute reticulocytes (billions/L) at baseline	144 (6 %)
Standard_labs_tsh_miu_per_liter	Thyroid stimulating hormone (mIU/L) at baseline	171 (5 %)
Standard_labs_specific_gravity	Urine specific gravity at baseline	50 (2 %)
Standard_labs_reaction_ph	Urine reaction pH at baseline	50 (2 %)

A.6. Sensors data

Variable name	Definition	N (%) imputed
Mean_steps_first_30_days	Average daily number of steps in the first 30 days in study (measured with study watch)	263 (11 %)
Mean_days_10hrs_wear_first_30days	Number of days of 10 h of wear in the first 30 days in study (measured with study watch)	263 (11 %)

References

Adler, N.E., Stead, W.W., 2015. Patients in context-EHR capture of social and behavioral determinants of health. N. Engl. J. Med. 372 (8), 698–701. https://doi.org/10.1056/NEJMp1413945. PMID: 25693009.

Alif, A., Nelson, B., Stefancic, A., Ahmed, R., Okazaki, S., 2020. Documentation status and psychological distress among New York City community college students. Cult. Divers. Ethn. Minor. Psychol. 26 (1), 11–21. https://doi.org/10.1037/cdp0000290.

Alvarez, K., Fillbrunn, M., Green, J., Jackson, J., Kessler, R., McLaughlin, K., et al., 2018. Race/ethnicity, nativity, and lifetime risk of mental disorders in US adults. Soc. Psychiatry Psychiatr. Epidemiol. 54 (5), 553–565. https://doi.org/10.1007/s00127-018-1644-5.

Arges, K., Assimes, T., Bajaj, V., Balu, S., Bashir, M., Beskow, L., et al., 2020. The project baseline health study: a step towards a broader mission to map human health. NPJ Digit.Med. 3 (1) https://doi.org/10.1038/s41746-020-0290-y.

Assari, S., 2016. Number of chronic medical conditions fully mediates the effects of race on mortality; 25-year follow-up of a nationally representative sample of Americans. J. Racial Ethn. Health Disparities 4 (4), 623–631. https://doi.org/10.1007/s40615-016-0266-4.

Azur, M., Stuart, E., Frangakis, C., Leaf, P., 2011. Multiple imputation by chained equations: what is it and how does it work? Int. J. Methods Psychiatr. Res. 20 (1), 40–49. https://doi.org/10.1002/mpr.329.

Cameron, O., Smith, C., Lee, M., Hollingsworth, P., Hill, E., Curtis, G., 1990. Adrenergic status in anxiety disorders: platelet alpha2-adrenergic receptor binding, blood pressure, pulse, and plasma catecholamines in panic and generalized anxiety disorders patients and in normal subjects. Biol. Psychiatry 28 (1), 3–20. https://doi.org/10.1016/0006-3223(90)90427-4.

Chen, Q., Wang, S., 2013. Variable selection for multiply-imputed data with application to dioxin exposure study. Stat. Med. 32 (21), 3646–3659. https://doi.org/10.1002/sim.5783.

Chen, R., Kessler, R., Sadikova, E., NeMoyer, A., Sampson, N., Alvarez, K., et al., 2019. Racial and ethnic differences in individual-level and area-based socioeconomic status and 12-month DSM-IV mental disorders. J. Psychiatr. Res. 119, 48–59. https://doi.org/10.1016/j.jpsychires.2019.09.006.

Costello, H., Gould, R., Abrol, E., Howard, R., 2019. Systematic review and meta-analysis of the association between peripheral inflammatory cytokines and generalised anxiety disorder. BMJ Open 9 (7), e027925. https://doi.org/10.1136/bmjopen-2018-027925.

Emhan, A., Selek, S., Bayazıt, H., Fatih Karababa, İ., Katı, M., Aksoy, N., 2015. Evaluation of oxidative and antioxidative parameters in generalized anxiety disorder. Psychiatry Res. 230 (3), 806–810. https://doi.org/10.1016/j.psychres.2015.11.001.

Ercan, A., Bahceci, B., Polat, S., Cenker, O., Bahceci, I., Koroglu, A., et al., 2017. Oxidative status and prolidase activities in generalized anxiety disorder. Asian J. Psychiatr. 25, 118–122. https://doi.org/10.1016/j.ajp.2016.10.008.

Evans-Lacko, S., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Benjet, C., Bruffaerts, R., et al., 2017. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO World Mental Health (WMH) surveys. Psychol. Med. 48 (9), 1560–1571. https://doi.org/10.1017/s0033291717003336.

Fındıklı, E., Camkurt, M., İzci, F., Karaaslan, M., Fındıklı, H., Sümer, P., Kurutaş, E., 2018. The diagnostic value of malondialdehyde, superoxide dismutase and catalase activity in drug naïve, first episode, non-smoker generalized anxiety disorder patients. Clin.Psychopharmacol.Neurosci. 16 (1), 88–94. https://doi.org/10.9758/cpn.2018.16.1.88.

Finegan, M., Firth, N., Delgadillo, J., 2019. Adverse impact of neighbourhood socioeconomic deprivation on psychological treatment outcomes: the role of area-level income and crime. Psychother. Res. 30 (4), 546–554. https://doi.org/10.1080/10503307.2019.1649500.

Georgiades, K., Paksarian, D., Rudolph, K., Merikangas, K., 2018. Prevalence of mental disorder and service use by immigrant generation and race/ethnicity among U.S. adolescents. J. Am. Acad. Child Adolesc. Psychiatry 57 (4), 280–287.e2. https://doi.org/10.1016/j.jaac.2018.01.020.

Giebel, C., Corcoran, R., Goodall, M., Campbell, N., Gabbay, M., Daras, K., et al., 2020. Do people living in disadvantaged circumstances receive different mental health treatments than those from less disadvantaged backgrounds? BMC Public Health 20 (1). https://doi.org/10.1186/s12889-020-08820-4.

Hernández, E., Lastra, S., Urbina, M., Carreira, I., Lima, L., 2002. Serotonin, 5-hydroxyindoleacetic acid and serotonin transporter in blood peripheral lymphocytes of patients with generalized anxiety disorder. Int. Immunopharmacol. 2 (7), 893–900. https://doi.org/10.1016/s1567-5769(02)00025-5.

Hinz, A., Klein, A., Brähler, E., Glaesmer, H., Luck, T., Riedel-Heller, S., et al., 2017. Psychometric evaluation of the Generalized Anxiety Disorder Screener GAD-7, based on a large German general population sample. J. Affect. Disord. 210, 338–344. https://doi.org/10.1016/j.jad.2016.12.012.

Iny, L., Pecknold, J., Suranyi-Cadotte, B., Bernier, B., Luthe, L., Nair, N., Meaney, M., 1994. Studies of a neurochemical link between depression, anxiety, and stress from [3H]imipramine and [3H]paroxetine binding on human platelets. Biol. Psychiatry 36 (5), 251–291. https://doi.org/10.1016/0006-3223(94)90625-4.

Johnson, S., Ulvenes, P., Øktedalen, T., Hoffart, A., 2019. Psychometric properties of the General Anxiety Disorder 7-Item (GAD-7) scale in a heterogeneous psychiatric sample. Front. Psychol. 10 https://doi.org/10.3389/fpsyg.2019.01713.

Kantor, B., Kantor, J., 2020. Mental health outcomes and associations during the COVID-19 pandemic: a cross-sectional population-based study in the United States. Front. Psychiatry 11. https://doi.org/10.3389/fpsyt.2020.569083.

Kessler, R., Berglund, P., Demler, O., Jin, R., Merikangas, K., Walters, E., 2005. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey replication. Arch. Gen. Psychiatry 62 (6), 593. https://doi.org/10.1001/archpsyc.62.6.593.

Kessler, R.C., Angermeyer, M., Anthony, J.C., De Graaf, R., Demyttenaere, K., Gasquet, I., et al., 2007. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 6 (3), 168–176.

Konnopka, A., Leichsenring, F., Leibing, E., König, H., 2009. Cost-of-illness studies and cost-effectiveness analyses in anxiety disorders: a systematic review. J. Affect. Disord. 114 (1–3), 14–31. https://doi.org/10.1016/j.jad.2008.07.014.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., Herzberg, P., 2008. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med. Care 46 (3), 266–274. https://doi.org/10.1097/mlr.0b013e318160d093.

Maes, M., Bonifacio, K., Morelli, N., Vargas, H., Moreira, E., St. Stoyanov, D., et al., 2018. Generalized Anxiety Disorder (GAD) and comorbid major depression with GAD are characterized by enhanced nitro-oxidative stress, increased lipid peroxidation, and lowered lipid-associated antioxidant defenses. Neurotox.Res. 34 (3), 489–510. https://doi.org/10.1007/s12640-018-9906-2.

Olver, J., Hopwood, M., 2013. Depression and physical illness. Med. J. Aust. 199 (S6) https://doi.org/10.5694/mja12.10597.

Pantell, M., Rehkopf, D., Jutte, D., Syme, S.L., Balmes, J., Adler, N., 2013. Social isolation: a predictor of mortality comparable to traditional clinical risk factors. Am. J. Public Health 103 (11), 2056–2062. https://doi.org/10.2105/AJPH.2013.301261. Epub 2013 Sep 12. PMID: 24028260; PMCID: PMC3871270.

Ross, J., Hua, S., Perreira, K., Hanna, D., Castañeda, S., Gallo, L., et al., 2019. Association between immigration status and anxiety, depression, and use of anxiolytic and antidepressant medications in the Hispanic Community Health Study/Study of Latinos. Ann. Epidemiol. 37, 17–23.e3. https://doi.org/10.1016/j.annepidem.2019.07.007.

Ruiz, M., Zamorano, E., García-Campayo, J., Pardo, A., Freire, O., Rejas, J., 2011. Validity of the GAD-7 scale as an outcome measure of disability in patients with generalized anxiety disorders in primary care. J. Affect. Disord. 128 (3), 277–286. https://doi.org/10.1016/j.jad.2010.07.010.

Santabárbara, J., Lasheras, I., Lipnicki, D., Bueno-Notivol, J., Pérez-Moreno, M., López-Antón, R., et al., 2021. Prevalence of anxiety in the COVID-19 pandemic: an updated meta-analysis of community-based studies. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 109, 110207. https://doi.org/10.1016/j.pnpbp.2020.110207.

Sareen, J., Cox, B., Clara, I., Asmundson, G., 2005. The relationship between anxiety disorders and physical disorders in the U.S. National Comorbidity Survey. Depress. Anxiety 21 (4), 193–202. https://doi.org/10.1002/da.20072.

Schneider, L., Munjack, D., Severson, J., Palmer, R., 1987. Platelet [3H]Imipramine binding in generalized anxiety disorder, panic disorder, and agoraphobia with panic attacks. Biol. Psychiatry 22 (1), 59–66. https://doi.org/10.1016/0006-3223(87)90130-2.

Sevy, S., Papadimitriou, G., Surmont, D., Goldman, S., Mendiewicz, J., 1989. Noradrenergic function in generalized anxiety disorder, major depressive disorder, and healthy subjects. Biol. Psychiatry 25 (2), 141–152. https://doi.org/10.1016/0006-3223(89)90158-3.

Smyth, R., 2009. Depression in physical illness. J.R.Coll.Phys.Edinb. 39 (4), 337–342. https://doi.org/10.4997/jrcpe.2009.411.

Spitzer, R., Kroenke, K., Williams, J., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder. Arch. Intern. Med. 166 (10), 1092. https://doi.org/10.1001/archinte.166.10.1092.

Swinson, R.P., Antony, M.M., Bleau, P., 2006. Management of anxiety disorders. Can. J. Psychiatr. 51 (Suppl. 2), 1–92S.

van Buuren, S., Groothuis-Oudshoorn, K., 2011. MICE: Multivariate Imputation by Chained Equations in R. Available at: https://www.jstatsoft.org/v45/i03/ J. Stat. Softw. 45 (3), 1–67.

Vismara, M., Girone, N., Cirnigliaro, G., Fasciana, F., Vanzetto, S., Ferrara, L., et al., 2020. Peripheral biomarkers in DSM-5 anxiety disorders: an updated overview. Brain Sci. 10 (8), 564. https://doi.org/10.3390/brainsci10080564.

Weizman, R., Tanne, Z., Granek, M., Karp, L., Golomb, M., Tyano, S., Gavish, M., 1987. Peripheral benzodiazepine binding sites on platelet membranes are increased during diazepam treatment of anxious patients. Eur. J. Pharmacol. 138 (2), 289–292. https://doi.org/10.1016/0014-2999(87)90447-x.

Wickham, H., Sievert, C., 2016. Ggplot2. Springer, New York.