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

Julio C. Nunes, MD\textsuperscript{a}, Megan K. Carroll, MS\textsuperscript{b}, Kenneth W. Mahaffey, MD\textsuperscript{a}, Robert M. Califf, MD\textsuperscript{b}, P. Murali Doraiswamy, MBBS\textsuperscript{c,d}, Sarah Short, MPH\textsuperscript{b}, Svati H. Shah, MD, MHS\textsuperscript{e,f}, Susan Swope, RN, MS\textsuperscript{a}, Donna Williams, RN, MSN, MPH\textsuperscript{a}, Adrian F. Hernandez, MD, MHS\textsuperscript{e,g}, David S. Hong, MD\textsuperscript{h}

\textsuperscript{a}Stanford Center for Clinical Research, Stanford University School of Medicine, Stanford, CA, USA; \textsuperscript{b}Verily Life Sciences (Alphabet), San Francisco, CA, USA; \textsuperscript{c}Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA; \textsuperscript{d}Duke Institute for Brain Sciences, Duke University School of Medicine, Durham, NC, USA; \textsuperscript{e}Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC, USA; \textsuperscript{f}Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, NC, USA; \textsuperscript{g}Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, USA; \textsuperscript{h}Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA

**Corresponding author:**
Julio C. Nunes, MD
Stanford Center for Clinical Research
Stanford University
3180 Porter Dr.
Palo Alto, CA 94304
jcnunes@stanford.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
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. Even less is known about how GAD-7 relates to race, ethnicity, and socioeconomic status (SES). 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 ≤$25,000, and ≤12 years of education. After adjustment for 166 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.

Key words: Generalized Anxiety Disorder-7; measures of health and disease; effective clinical intervention; social determinants of health
INTRODUCTION

Anxiety disorders are among the most common psychiatric diagnoses, with a lifetime prevalence of 32% in the United States. Worldwide, the World Health Organization estimates that about 300 million people are affected. In the United States, anxiety disorders may contribute $110 million to $600 million per million inhabitants in indirect costs to the health care system. 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. Anxiety is closely related to multiple chronic disorders, such as cardiovascular disease, diabetes mellitus type 2, hypertension, arthritis, and chronic obstructive pulmonary disease.

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.

Spitzer et al. 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. 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. 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. Vismara et al. recently reviewed and summarized the relationship between anxiety disorders and these biomarkers. However, despite these significant advances in understanding the physiology of anxiety, little has been published about how it correlates with wide-ranging population health markers, demographic characteristics, and social determinants of health.

The Baseline Health Study (BHS) 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.
Nunes et al. GAD-7 as a Marker of Inequity

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 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.

METHODS

The Baseline Health Study

The Baseline Health Study is an ongoing multicenter longitudinal observational study. Multimodal data, including 166 different demographic, clinical, imaging, and laboratory measurements, were collected for 2502 adult participants. The BHS design has been previously described, 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.

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.

SES Risk Score

Socioeconomic status variables were combined into a single “SES risk score” variable that was used in subsequent LASSO regressions and on Figure 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.

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. The process generally consists of 4 steps: 1) a simple imputation (i.e., mean) is done for each missing value and used as placeholder, 2) placeholder imputations are set back to missing for a single “target” variable \( X_i \), 3) \( X_i \) becomes the dependent variable and is regressed on all other variables in the imputation model, and 4) missing values in \( X_i \) are replaced with predicted values from the model in step 3, and are then used in following regression models on variables with missing data. Steps 3-4 employ predictive mean matching in order to calculate the missing values of both continuous and categorical target variables; this ensures that missing data are assigned a value that has already been observed in the data. Steps 2-4 are repeated for each variable with missing data until all data are complete, a process that constitutes 1 “cycle.” A preset number of cycles were completed, at which point a final imputed dataset was computed. This process was then repeated for a prespecified number of imputed datasets.\(^{25}\)

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 \( \ell_1 \)-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 (\( \hat{\beta}_{1,j}, \ldots, \hat{\beta}_{D,j} \), 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 (\( \lambda \)), which is optimized by selecting the model that minimizes the Bayesian Information Criterion (BIC).\(^{26}\)

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.\(^{27}\) MI-LASSO regressions were run using the MI.LASSO function developed by Chen and Wang.\(^{26}\) Figures were created using the ‘ggplot2’ package v3.3.0.\(^{28}\)

RESULTS
A total of 2,453 participants completed baseline GAD-7 assessments, of whom 1,343 (55%) were female and 1,110 (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
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.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Overall</th>
<th>Female GAD-7</th>
<th>Male GAD-7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 2453</td>
<td>N = 531</td>
<td>N = 1822</td>
</tr>
<tr>
<td>Age, median (IQR)</td>
<td>50.0 (35.3-64.1)</td>
<td>57.2 (35.3-62.4)</td>
<td>47.5 (32.7-65.0)</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, n (%)</td>
<td>1561 (63.6)</td>
<td>1034 (61.9)</td>
<td>1280 (59.7)</td>
</tr>
<tr>
<td>Black, n (%)</td>
<td>390 (15.9)</td>
<td>240 (16.0)</td>
<td>600 (26.6)</td>
</tr>
<tr>
<td>Asian, n (%)</td>
<td>254 (10.4)</td>
<td>120 (7.5)</td>
<td>334 (15.3)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander, n (%)</td>
<td>27 (1.1)</td>
<td>6 (0.4)</td>
<td>21 (1.0)</td>
</tr>
<tr>
<td>American Indian or Alaska Native, n (%)</td>
<td>28 (1.1)</td>
<td>4 (0.3)</td>
<td>24 (1.1)</td>
</tr>
<tr>
<td>Other race, n (%)</td>
<td>193 (7.9)</td>
<td>117 (7.1)</td>
<td>76 (3.4)</td>
</tr>
<tr>
<td>Hispanic ethnicity, n (%)</td>
<td>279 (11.3)</td>
<td>190 (11.3)</td>
<td>89 (4.0)</td>
</tr>
<tr>
<td>Site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Lake, n (%)</td>
<td>468 (19.1)</td>
<td>303 (18.1)</td>
<td>165 (7.4)</td>
</tr>
<tr>
<td>Durham, n (%)</td>
<td>479 (19.5)</td>
<td>284 (16.5)</td>
<td>195 (8.7)</td>
</tr>
<tr>
<td>Kannapolis, n (%)</td>
<td>511 (20.8)</td>
<td>313 (19.1)</td>
<td>198 (8.8)</td>
</tr>
<tr>
<td>Palo Alto, n (%)</td>
<td>995 (40.6)</td>
<td>613 (37.9)</td>
<td>382 (16.9)</td>
</tr>
</tbody>
</table>

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.
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 18.5% 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.
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.

<table>
<thead>
<tr>
<th>Socioeconomic Status</th>
<th>Overall</th>
<th>0 (N = 404)</th>
<th>1-4 (N = 531)</th>
<th>5-9 (N = 257)</th>
<th>10-14 (N = 103)</th>
<th>15+ (N = 48)</th>
<th>p-value</th>
<th>Female GAD-7</th>
<th>Overall</th>
<th>0 (N = 418)</th>
<th>1-4 (N = 479)</th>
<th>5-9 (N = 137)</th>
<th>10-14 (N = 54)</th>
<th>15+ (N = 22)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>2453</td>
<td>176 (7.2)</td>
<td>28 (6.9)</td>
<td>32 (6.0)</td>
<td>17 (6.6)</td>
<td>15 (14.6)</td>
<td>9 (18.8)</td>
<td>0.0024</td>
<td>25 (6.0)</td>
<td>20 (4.2)</td>
<td>15 (10.9)</td>
<td>11 (20.4)</td>
<td>4 (18.2)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>High school or less, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college, n (%)</td>
<td>455</td>
<td>70 (17.3)</td>
<td>103 (19.4)</td>
<td>71 (27.6)</td>
<td>25 (24.3)</td>
<td>11 (22.9)</td>
<td>0.0076</td>
<td>65 (15.6)</td>
<td>70 (14.6)</td>
<td>24 (17.5)</td>
<td>10 (18.5)</td>
<td>6 (27.3)</td>
<td>0.2192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College, n (%)</td>
<td>634</td>
<td>120 (29.7)</td>
<td>178 (33.5)</td>
<td>68 (26.5)</td>
<td>24 (23.3)</td>
<td>11 (22.9)</td>
<td>0.0721</td>
<td>95 (22.7)</td>
<td>105 (21.9)</td>
<td>24 (17.5)</td>
<td>8 (14.8)</td>
<td>1 (4.5)</td>
<td>0.0192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduate degree or higher, n (%)</td>
<td>670</td>
<td>123 (30.4)</td>
<td>140 (26.4)</td>
<td>58 (22.6)</td>
<td>13 (12.6)</td>
<td>4 (8.3)</td>
<td>&lt;0.0001</td>
<td>144 (34.4)</td>
<td>144 (30.1)</td>
<td>37 (27.0)</td>
<td>6 (11.1)</td>
<td>1 (4.5)</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; $25,000, n (%)</td>
<td>191</td>
<td>24 (5.9)</td>
<td>40 (7.5)</td>
<td>34 (13.2)</td>
<td>20 (19.4)</td>
<td>9 (18.8)</td>
<td>&lt;0.0001</td>
<td>20 (4.8)</td>
<td>24 (5.0)</td>
<td>6 (4.4)</td>
<td>10 (18.5)</td>
<td>6 (27.3)</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$25,000-50,000, n (%)</td>
<td>256</td>
<td>43 (10.6)</td>
<td>52 (9.8)</td>
<td>39 (15.2)</td>
<td>15 (14.6)</td>
<td>9 (18.8)</td>
<td>0.0186</td>
<td>34 (8.1)</td>
<td>44 (9.2)</td>
<td>14 (10.2)</td>
<td>6 (11.1)</td>
<td>0 (0.0)</td>
<td>0.7462</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$50,000-100,000, n (%)</td>
<td>493</td>
<td>86 (21.3)</td>
<td>142 (26.7)</td>
<td>53 (20.6)</td>
<td>26 (25.2)</td>
<td>6 (12.5)</td>
<td>0.5255</td>
<td>70 (16.7)</td>
<td>77 (16.1)</td>
<td>24 (17.5)</td>
<td>7 (13.0)</td>
<td>2 (9.1)</td>
<td>0.4488</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$100,000-150,000, n (%)</td>
<td>308</td>
<td>57 (14.1)</td>
<td>61 (11.5)</td>
<td>28 (10.9)</td>
<td>2 (1.9)</td>
<td>2 (4.2)</td>
<td>0.0066</td>
<td>64 (15.3)</td>
<td>71 (14.8)</td>
<td>19 (13.9)</td>
<td>2 (3.7)</td>
<td>2 (9.1)</td>
<td>0.0678</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$150,000-200,000, n (%)</td>
<td>197</td>
<td>35 (8.7)</td>
<td>50 (9.4)</td>
<td>16 (6.2)</td>
<td>5 (4.9)</td>
<td>2 (4.2)</td>
<td>0.0643</td>
<td>43 (10.3)</td>
<td>30 (6.3)</td>
<td>12 (8.8)</td>
<td>3 (5.6)</td>
<td>1 (4.5)</td>
<td>0.1106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; $200,000, n (%)</td>
<td>358</td>
<td>64 (15.8)</td>
<td>77 (14.5)</td>
<td>27 (10.5)</td>
<td>8 (7.8)</td>
<td>3 (6.2)</td>
<td>0.0028</td>
<td>79 (18.9)</td>
<td>75 (15.7)</td>
<td>20 (14.6)</td>
<td>5 (9.3)</td>
<td>0 (0.0)</td>
<td>0.0077</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married, n (%)</td>
<td>1054</td>
<td>179 (44.3)</td>
<td>242 (45.6)</td>
<td>97 (37.7)</td>
<td>31 (30.1)</td>
<td>11 (22.9)</td>
<td>0.0002</td>
<td>232 (55.5)</td>
<td>200 (41.8)</td>
<td>44 (32.1)</td>
<td>13 (24.1)</td>
<td>5 (22.7)</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divorced, n (%)</td>
<td>180</td>
<td>42 (10.4)</td>
<td>49 (9.2)</td>
<td>30 (11.7)</td>
<td>12 (11.7)</td>
<td>4 (8.3)</td>
<td>0.7771</td>
<td>20 (4.8)</td>
<td>15 (3.1)</td>
<td>4 (2.9)</td>
<td>2 (3.7)</td>
<td>2 (9.1)</td>
<td>0.7563</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formerly in long term relationship, n</td>
<td>97</td>
<td>12 (3.0)</td>
<td>22 (4.1)</td>
<td>11 (4.3)</td>
<td>3 (2.9)</td>
<td>0 (0.0)</td>
<td>0.9021</td>
<td>16 (3.8)</td>
<td>24 (5.0)</td>
<td>5 (3.6)</td>
<td>3 (5.6)</td>
<td>1 (4.5)</td>
<td>0.6686</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living together, n (%)</td>
<td>205</td>
<td>26 (6.4)</td>
<td>46 (8.7)</td>
<td>31 (12.1)</td>
<td>9 (8.7)</td>
<td>8 (16.7)</td>
<td>0.0084</td>
<td>23 (5.5)</td>
<td>34 (7.1)</td>
<td>19 (13.9)</td>
<td>8 (14.8)</td>
<td>1 (4.5)</td>
<td>0.0049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never in long term relationship, n (%)</td>
<td>255</td>
<td>38 (9.4)</td>
<td>60 (11.3)</td>
<td>22 (8.6)</td>
<td>13 (12.6)</td>
<td>9 (18.8)</td>
<td>0.1884</td>
<td>26 (6.2)</td>
<td>54 (11.3)</td>
<td>21 (15.3)</td>
<td>9 (16.7)</td>
<td>3 (13.6)</td>
<td>0.0004</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Socioeconomic Status</th>
<th>Overall</th>
<th>Female GAD-7</th>
<th>Male GAD-7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 2453</td>
<td>0 (N = 404)</td>
<td>1-4 (N = 531)</td>
</tr>
<tr>
<td>Separated, n (%)</td>
<td>48 (2.0)</td>
<td>9 (2.2)</td>
<td>6 (1.1)</td>
</tr>
<tr>
<td>Widowed, n (%)</td>
<td>75 (3.1)</td>
<td>29 (7.2)</td>
<td>23 (4.3)</td>
</tr>
</tbody>
</table>

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 |
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 339 total predictors and was stratified by sex, are presented in Figure 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 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 more than 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.

Figure 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. There was no clear trend between SES risk score and GAD-7 severity for those who reported other races, but those participants tended to have lower SES risk scores overall.

**DISCUSSION**

Our study enrolled a large, deeply phenotyped cohort that allowed us to assess a broad array of anxiety correlates. Our findings confirm and highlight previous literature data regarding GAD-7 and different measurements of SES. We emphasize the importance of using GAD-7 with psychosocial contextual knowledge, as it raises concerns not only about psychological symptoms, but also about broader social determinants of health. Although some of the individual correlations are expected, their assembly highlights 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. 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. (2020)
Nunes et al. GAD-7 as a Marker of Inequity

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. reported similar findings in an analysis of 13,775 medical records.

When designing interventions, clinicians and policy makers must consider socioeconomic issues, access to opportunities, and social context to better reduce anxiety. Different treatment modalities may be more effective if employed in a holistic approach that is sensitive to more than just biochemical matters or cognitive function. Furthermore, SES may limit the types of treatment and therapy patients may receive and thus, clinicians and health systems should be focused on innovative methods to enable people with anxiety to access the standard of care.

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 healthcare, distance from family, use of English as a second language, and systemic racism and prejudice. Additionally, concerns about migration and visa status could lead certain ethnic groups to not enroll, which is reflected on the current underrepresentation of Latinx populations in our sample. We did not ask participants about their visa status and that is a limitation of this study. This aspect is important, as insecurity around residency status may also contribute to higher degrees of psychosocial distress among Hispanic populations.

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. However, when focusing on the relationships between race, SES risk score, and GAD-7 score (Figure 2), Black persons with high SES risk mostly have higher scores, suggesting that poor SES may be a significant associated factor in the development of anxiety for that 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.

The BHS is longitudinally assessing a large cohort of patients for markers of social, biological, clinical, and behavioral function; once it is completed, we will have 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 how social determinants of health shape the profile of anxiety in specific populations. The representation of each demographic group is a priority for the BHS study and its final population pool will reflect every race/ethnicity per the US census. The BHS is also an opportunity to understand the validity of behavioral data collection in the context of digital health.
Limitations
A number of limitations to this study should be noted. Cross-sectional studies cannot determine causality. Therefore, this study is capable of demonstrating 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.

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.
Nunes et al. GAD-7 as a Marker of Inequity

Contributors
JCN drafted and wrote the paper and performed the analysis. MKC contributed data or analysis tools, performed the analysis, wrote the paper. KWM conceived and designed the analysis, reviewed the paper. RMC conceived and designed the analysis, reviewed the paper. PMD conceived and designed the analysis, reviewed the paper. S Short contributed data or analysis tools, performed the analysis, reviewed the paper. SHS contributed data or analysis tools, reviewed the paper. S Swope collected the data, reviewed the paper. DW collected the data, reviewed the paper. AFH conceived and designed the analysis, reviewed the paper. DSH conceived and designed the analysis, wrote and reviewed the paper, coordinated writing and analysis efforts.

Declaration of Competing Interest
None.

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 Medixi 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, Relypsia, and Verily. Dr. Doraiswamy notes external relationships with Advera Health, Apollo Hospitals, Danone, Evidation Health, Live Love Laugh, Luminos Labs, MarvelBiome, Neuroggle, 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.

Acknowledgments
Nunes et al. GAD-7 as a Marker of Inequity


Data availability
The Baseline Study data will be available to qualified researchers for exploratory analysis after the data are adequately curated and initial planned primary manuscripts are written. Qualified external researchers will be able to apply through applications reviewed by the Proposal Review and Publications Committee and Scientific Executive Committee.
Funding
The Baseline Health Study and this analysis were funded by Verily Life Sciences, South San Francisco, CA. This analysis was funded by Verily Life Sciences (South 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.
REFERENCES


Nunes et al. GAD-7 as a Marker of Inequity


Figure Legends

Figure 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.

Figure 2. Associations between socioeconomic status (SES) and GAD-7 score, stratified by race. Each panel describes the interaction of GAD-7 severity and SES risk score within each of 4 racial groupings. 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 Exact Tests for count data with simulated p-value were performed within each race; significant overall associations between SES risk score and GAD-7 severity for all races (White: p < 0.001; Black or African American: p = 0.012; Asian: p = 0.010; Other: p = 0.02