HHS Public Access

Author manuscript

J Affect Disord. Author manuscript; available in PMC 2022 March 01.

Published in final edited form as:

J Affect Disord. 2021 March 01; 282: 1021–1029. doi:10.1016/j.jad.2021.01.005.

Validation of Brief Screening Measures for Depression and Anxiety in Young People with Substance Use Disorders

Kate H. Bentley[1,2,*], **Hitoshi Sakurai**[1,2,3,*], **Kelsey L. Lowman**[1], **Lisa Rines-Toth**[1], **James McKowen**[1,2], **Paola Pedrelli**[1,2], **A. Eden Evins**[1,2], **Amy M. Yule**[1,2,4,5]

[1]Department of Psychiatry, Massachusetts General Hospital

[2]Harvard Medical School

[3]Department of Neuropsychiatry, Keio University School of Medicine

[4]Department of Psychiatry, Boston Medical Center

[5]Boston University School of Medicine

Abstract

Background: It is critical to promptly identify and monitor mood and anxiety symptoms in young people with SUD. The primary aim of this study was to conduct a psychometric validation of the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) for depression and anxiety screening in young people seeking outpatient treatment for SUD. Our secondary aim was to compare the performance of the PHQ-9 and GAD-7 to their briefer two-item versions (PHQ-2 and GAD-2) in terms of detecting probable mood and anxiety disorders.

Method: Data were extracted from the electronic health records of patients (ages 14 to 26) who received a diagnostic evaluation following clinical implementation of the PHQ-9 and GAD-7 at a hospital-based outpatient SUD treatment program (N=121, average age 19.1 ± 3.1 years).

Results: The PHQ-9 and GAD-7 showed excellent internal consistency. A PHQ-9 cut score of 7 or 8 (PHQ-2 cut score: 2) and GAD-7 cut score of 6 (GAD-2 cut score: 2) had the best balance of

Corresponding Author: Kate H. Bentley, PhD, Assistant Professor, Harvard Medical School, Psychologist, Depression Clinical and Research Program, Massachusetts General Hospital, One Bowdoin Square, 6th Floor, Boston MA 02114, KBENTLEY@mgh.harvard.edu, Phone: 617-724-7741.

*These authors contributed equally to this work.

Contributors:

Drs. Bentley, Yule, and McKowen designed the study. Dr. Bentley, Dr. Yule, Ms. Lowman, and Ms. Rines-Toth wrote the protocol. Dr. Bentley, Dr. Sakurai, Ms. Lowman, and Ms. Rines-Toth managed data collection. Drs. Bentley, Sakurai, Yule, Pedrelli, and Evins contributed to the statistical approach, and Drs. Bentley and Sakurai conducted all statistical analyses. Dr. Bentley, Dr. Sakurai, Dr. Yule and Ms. Lowman contributed to the first draft of the manuscript, and all authors significantly contributed to and have approved the final manuscript.

Conflict of Interest Declaration

Dr. Sakurai has received manuscript or speaker's honoraria from Dainippon Sumitomo, Meiji-Seika Pharma, Mochida Pharmaceutical, and Yoshitomi Yakuhin, and grants from the Japanese Society of Clinical Neuropsychopharmacology and the Uehara Memorial Foundation within the past three years. All other authors declare that they have no conflicts of interest.

sensitivity, specificity, and positive and negative predictive power in these data. These measures also showed good convergent and acceptable discriminant validity.

Limitations: The sample was predominantly White and non-Hispanic, and a validated (semi-)structured diagnostic interview was not used to establish mood and anxiety disorder diagnoses.

Conclusions: Results suggest the PHQ-9 and GAD-7 are reliable and potentially clinically useful screening tools for depression and anxiety in young people with SUD, and that the two-item versions may have similar clinical utility as the full measures.

Keywords

Screening; depression; anxiety; substance use disorder; adolescents; young adults

Substance use disorders (SUD) peak in prevalence in young adults ages 18 to 25 years, with recent twelve-month prevalence rates ranging from 15% (Substance Abuse and Mental Health Services Administration, 2019) to 44% (Arterberry et al., 2019) in this age group. Both young adults and adolescents with SUD commonly have co-occurring psychiatric conditions (Chan et al., 2008; Couwenbergh et al., 2006; Kandel et al., 1997), often mood and anxiety disorders (Armstrong & Costello, 2002; Grella et al., 2001; Lai et al., 2015; Lubman et al., 2007). Co-occurring mood and anxiety disorders in individuals with SUD are associated with more substance-related problems (e.g., Lubman et al., 2007), greater treatment attrition (e.g., Krawczyk et al., 2017), and poorer treatment outcomes (e.g., Boden & Moos, 2009). Adolescents and young adults with SUD and mood and anxiety disorders are also at increased risk for adverse and lethal sequelae of both conditions including overdose and suicide (Kelly et al., 2004; Lyons et al., 2019; Yule et al., 2018). Thus, there is a need for prompt, reliable identification of depressive and anxiety symptoms within young people with SUDs to inform assessment and treatment planning.

The 9-item Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001) and the 7-item Generalized Anxiety Disorder scale (GAD-7; Spitzer et al., 2006) are brief, widely-used self-report screening tools for depression and anxiety (Dear et al., 2011; Titov et al., 2011; Toussaint et al., 2020; Zimmerman, 2019). Both measures were initially developed and validated in adult primary care, and have since been extended to general samples of adolescents (Allgaier et al., 2012a; Richardson et al., 2010; Tiirikainen et al., 2019) and young adults (Keum et al., 2018; Parkerson et al., 2015), as well as adults in psychiatric (e.g., Beard et al., 2016; Beard & Björgvinsson, 2014; Johnson et al., 2019; Kertz et al., 2013; Kung et al., 2013; Rutter & Brown, 2017) and addiction treatment settings (e.g., Delgadillo et al., 2011, 2012; Dum et al., 2008; Hepner et al., 2009). Across samples and settings, the PHQ-9 and GAD-7 have shown robust internal consistency ($\alpha \approx 0.82–0.90$ for the PHQ-9 and $\alpha \approx 0.85–0.92$ for the GAD-7), as well as convergent validity with other self-report measures of depression and anxiety (Beard & Björgvinsson, 2014b; Dum et al., 2008b).

Recommended cut scores for the PHQ-9 and GAD-7 (indicating a probable diagnosis of major depressive disorder [MDD] or anxiety disorder, respectively) vary across the

literature. Whereas a cut score of ≥10 on the PHQ-9 is the most widely implemented cut point for depression screening in medical settings (Kroenke et al., 2001; Gilbody et al., 2007), other suggested cut scores on the PHQ-9 have ranged from ≥ 8 in adolescents (Allgaier et al., 2012b) to ≥ 12 (Delgadillo et al., 2011b) and ≥ 16 (Levitt et al., 2021) for adults in outpatient and inpatient SUD treatment, respectively. Likewise, whereas the original GAD-7 validation study indicated a cut score of ≥ 10 (Spitzer et al., 2006b), other research proposes using ≥ 9 for adults in SUD treatment (Delgadillo et al., 2012b; Levitt et al., 2021), and results have been inconclusive for adults in psychiatric treatment (Kertz et al., 2013b; Rutter & Brown, 2017b).

There has also been recent interest in briefer versions of these measures – namely, the PHQ-2 (comprising the first two items of the PHQ-9; Kroenke et al., 2003) and the GAD-2 (first two items of the GAD-7; Kroenke et al., 2007) – either as standalone screening tools or to identify patients (based on a cut score of typically 2 or 3; e.g., Levis et al., 202) to administer the full measure in a two-step screening process (Levis et al., 2020; Staples et al., 2019). Indeed, a recent meta-analysis suggests that the PHQ-2 followed by the PHQ-9 may be the optimal approach, as it has higher sensitivity or specificity than the PHQ-2 alone, and reduces the number of patients needing to complete the full PHQ-9 (Levis et al., 2020). Regarding anxiety screening, a previous meta-analysis indicated that both the GAD-2 and GAD-7 had acceptable properties for screening for heterogeneous anxiety disorders (Plummer et al., 2016).

Despite the body of literature on the psychometric properties of the PHQ and GAD scales, no research to date has focused on adolescents or young adults with SUD. It is important to establish that these widely used tools are valid, reliable, and clinically useful screening tools in this specific and unique patient population. For one, measurement invariance has been observed for psychological measures used across in different populations and settings (e.g., Bach et al., 2018; Meredith, 1993), thus impacting clinical interpretation and utility. Indeed, the psychometric performance and clinical utility of screening measures for depression and anxiety have been shown to vary across age groups (Balsis & Cully, 2008; Morin et al., 1999; Trainor et al., 2013). Converging evidence also indicates that depression and anxiety may have different presentations in, for example, adolescent versus adult populations (e.g., Dickstein, 2011; Rice et al., 2019), further underscoring the value of explicitly testing these screening measures for depression and anxiety screening in younger people with SUD (e.g., Stockings et al., 2015).

The current study aims to address this gap using electronic health record (EHR) data from intake evaluations at an urban hospital-based outpatient SUD treatment program for “young people” (here defined as ages 14 to 26, the age range served by the treatment program and thus including adolescents and young adults) after clinical implementation of the PHQ-9 and GAD-7 in 2018. To accomplish an overview of the validity, reliability, and clinical utility of these screening tools, we analyzed the internal consistency, classification accuracy, and convergent and discriminant validity of the PHQ-9 and GAD-7. We also compared the classification accuracy and convergent and discriminant validity of the PHQ-9 and GAD-7 to their briefer, and thus potentially easier to implement in routine SUD care, two-item versions.

Method

Participants

The sample included young people aged 14 to 26 years who underwent an intake evaluation (and as part of this completed the PHQ-9 and GAD-7) between January and September 2018 at an outpatient SUD treatment program for young people at an urban academic medical center in the Northeastern United States.

Measures

PHQ-9/2.—The 9-item PHQ-9 assesses the frequency of depressive symptoms in the past two weeks (Kroenke et al., 2001a). Items are rated on a Likert scale from 0 (not at all) to 3 (nearly every day) and summed for a composite score ranging from 0 to 27. The PHQ-2 consists of the first two items of the PHQ-9 (depressed mood and anhedonia), with total scores ranging from 0 to 6.

BDI-II.—The Beck Depression Inventory-II (BDI-II) is a 21-item self-report measure of depressive symptoms in the past two weeks (Beck et al., 1996). Patients respond to items on a Likert scale from 0 to 3, with total scores ranging from 0 to 63.

GAD-7/2.—The GAD-7 scale includes seven items that assess the frequency of generalized anxiety disorder (GAD) symptoms in the past two weeks (Spitzer et al., 2006a). Participants rate items on the same Likert scale as the PHQ-9/2 from 0 to 3, yielding a total score ranging from 0 to 21. The GAD-2 comprises the first two items of the GAD-7 (nervousness and uncontrollable worry), summed for a total score ranging from 0 to 6.

STAI.—The State-Trait Anxiety Inventory (STAI; Spielberger, 1983) measures anxiety symptoms via two 20-item subscales: the STAI-State (STAI-S) and STAI-Trait (STAI-T). Response choices are on a Likert scale from 1 (not at all/almost never) to 4 (very much so/almost always), yielding a total score of 20 to 80 per subscale.

TAS.—The Trait Anger Scale (TAS) is a 10-item subscale of the State-Trait Anger Expression Inventory (STAXI-2; Spielberger, 1988, 1999) that evaluates overall proneness to feeling and reacting to anger. Items are rated on a Likert scale ranging from 1 (almost never) to 4 (almost always) for a composite score range of 10 to 40.

LDQ.—The 10-item Leeds Dependence Questionnaire (LDQ; Raistrick et al., 1994) measures psychological symptoms of substance dependence. Patients rate how frequently they experienced symptoms in the past week, using a Likert scale from 0 (never) to 3 (nearly always). Total scores thus range from 0 to 30, with higher values indicating greater dependence severity.

Mood and anxiety disorders.—Presence versus absence of current and lifetime Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychiatric Association, 2013) non-substance induced unipolar depressive, bipolar, and anxiety disorders (i.e., GAD, social anxiety, panic disorder, specific phobias, post-traumatic

stress disorder (PTSD), obsessive-compulsive and related disorders) was extracted from the intake evaluation note.

Procedures

Evaluation.—Participants completed a self-report questionnaire battery on paper followed by a standardized biopsychosocial evaluation with a MA- or PhD-level provider with advanced training in addiction care. Providers had access to the self-report questionnaire data prior to conducting the evaluation, during which they used a structured note template in the EHR that included DSM-5 criteria for lifetime and current SUD and other psychiatric diagnoses. Evaluations were reviewed in multidisciplinary rounds with doctoral (PhD and MD) providers with formal advanced training in child and addiction psychiatry. Self-report questionnaire data were ultimately scanned into the EHR by administrative staff.

Data extraction.—Data were collected through a systematic, retrospective chart review study (with a waiver of informed consent) that was approved by the institutional human subjects committee. Clinic staff provided a medical record number and date of intake evaluation for each new patient who initiated care between January and September 2018. Eligible records (i.e., those with item-level PHQ-9 or GAD-7 data) were assigned a study ID in a password-protected file, and clinical information was recorded in a deidentified REDCap database hosted by Partners HealthCare (Harris et al., 2009). Study staff followed a standard operating procedure to code the following information from the intake evaluation: demographic characteristics, diagnoses, and self-report questionnaire data. If these variables were not included or clearly identified in the initial visit record, study staff consulted data from all clinic encounters within 30 days of the initial visit to clarify such information. Two BA- or MA-level study staff members (KLL and LRT) extracted diagnostic and behavioral data from the clinical assessment; all ambiguous cases or disagreements were reviewed and resolved by co-lead author and licensed clinical psychologist KHB.

Statistical Analysis

First, we compared total questionnaire scores as a function of age, gender, race, and ethnicity, and calculated internal consistencies (both alpha and omega; Dunn et al., 2014; McDonald, 1999) of the PHQ-9 and GAD-7. Second, to investigate the degree to which the PHQ-9 and PHQ-2 discriminated between individuals with and without a mood disorder diagnosis, we computed Hedge's g effect size estimates of differences in means for patients with/without current (and then lifetime) unipolar or bipolar depression.[1] For the GAD-7 and GAD-2, the same analysis was conducted for current and lifetime anxiety disorder diagnosis.

Third, to assess classification accuracy, we generated receiver operating characteristics (ROC) curves with PHQ-9, PHQ-2, GAD-7, and GAD-2 score as the continuous variable and the presence versus absence of a current mood or anxiety disorder as the categorical

[1]Whether patients with bipolar disorder ($n = 15$) were currently (or most recently) in a depressed, manic, or mixed episode at the time of the intake evaluation (which we expect to impact PHQ scores) was not consistently documented in the EHR. Thus, though our primary analyses collapsed across current unipolar and bipolar disorders, we also re-ran the diagnostic analyses excluding patients with current bipolar disorder (see footnotes in Results).

outcome. A ROC curve is a plot of the true positive rate (sensitivity) against the false positive rate (1 – specificity) for different possible cut points. Thus, ROC curves offered visual representations of the trade-off between sensitivity and specificity of cut scores when using the questionnaire as a diagnostic measure. Chi-square tests, sensitivity (e.g., percentage of those who were diagnosed with a mood disorder correctly identified by the PHQ as depressed), and specificity (e.g., percentage of those who were not diagnosed with a mood disorder correctly identified by the PHQ as nondepressed) were used to determine the percentage of patients correctly classified. Estimates of positive predictive power (PPP; e.g., percentage of those classified by the PHQ as depressed who were actually diagnosed with a mood disorder) and negative predictive power (NPP; e.g., percentage of those classified by the PHQ as nondepressed who were not diagnosed with depression) at each potential cut score were calculated. Given that PPP and NPP are affected by the prevalence of individuals classified as depressed or anxious, the base rates of those classified by the PHQ-9/2 as depressed and the GAD-7/2 as anxious at each cut score were determined to facilitate interpretation of these values. Kappa coefficients indicating agreement between the questionnaire score and diagnostic classification after correcting for chance (Cohen, 1960) were computed to provide further indication of diagnostic utility after correcting for base rates. These various indices were used to select an appropriate cut score on the PHQ-9/2 for identifying patients with a probable current mood disorder and the GAD-7/2 for those with probable current anxiety.

Last, to investigate convergent and discriminant validity, correlations between the PHQ-9/2 and GAD-7/2 and other well-established measures of depression (BDI-II), anxiety (STAI-T/S), anger (TAS), and substance dependence (LDQ) were computed. The magnitudes of correlations were interpreted in accordance with commonly used benchmarks suggested by Cohen (1988); effect sizes between .10 and .30 were considered small; between .30 and .50 were considered medium, and .50 or above were considered large. Large positive correlations of the PHQ/2) with the BDI-II and the GAD-7/2) with the STAI-T/S were interpreted as evidence for convergent validity, whereas correlations with the TAS and LDQ of small to moderate magnitudes were considered supportive of discriminant validity. A two-tailed p value of < 0.05 was considered statistically significant for all tests. All statistical analyses were conducted using the Statistical Package for Social Science (SPSS) version 24.0 (IBM Corp, 2016).

Results

Sample Demographics

The mean age of the sample ($N = 121$) was 19.6 years (SD = 3.1) (see Table 1). The sample identified as 68.6% male (including .01% transgender male) and 31.4% female. The sample primarily identified as White (75.2%), followed by 5.8% Asian, 2.5% Black/African American, and 4.1% more than one race; information about race was unavailable for 12.4% of the sample. Regarding ethnicity, 82.6% identified as not Hispanic or Latino, 9.1% identified as Hispanic or Latino; information about ethnicity was not available for 8.3% of the sample.

The mean number of SUD diagnoses was 1.44 (SD = 0.97, range 0–5). The most common SUD was cannabis use disorder (63.6%), followed by alcohol use disorder (32.2%), opioid use disorder (17.4%), stimulant use disorder (13.2%), sedative, hypnotic, or anxiolytic use disorder (11.6%), and inhalant use disorder (0.8%). Almost half (46.3%) of the sample met criteria for current unipolar depressive disorder (71.9% lifetime) and 12.4% for bipolar disorder (20.7% lifetime). Just over half (51.2%) met criteria for a current anxiety disorder (74.4% lifetime).

PHQ-9 and GAD-7 Scores and Internal Consistency

There was very little missing item-level data (0.6% for the PHQ-9 and 0.2% for the GAD-7). The mean PHQ-9 and GAD-7 scores (N = 121) were 10.6 (SD = 7.9, range = 0 to 27) and 8.5 (SD = 6.6, range = 0 to 21). The mean PHQ-2 and GAD-2 scores were 2.6 (SD = 2.2, range = 0 to 6) and 2.7 (SD = 2.1, range = 0 to 6). See Supplemental Material for descriptive statistics and response frequencies for individual items (Table S1). Females scored higher than males on the PHQ-9 (M = 13.5, SD = 8.6 versus M = 9.2, SD = 7.2; $t_{(119)} = -2.66$, $p < .05$), GAD-7 (M = 10.4, SD = 7.0 versus M = 7.6, SD = 6.2; $t_{(119)} = -2.22$, $p < .05$), and GAD-2 (M = 3.3, SD = 2.3 versus M = 2.5, SD = 2.0; $t_{(119)} = -2.04$, $p < .05$). GAD-7/2 scores also differed by age, in that older age was correlated with higher scores ($r = .21$, $p < .05$ and $r = .22$, $p < .05$). Although there was an overall effect of race on PHQ-2 scores ($F_{(4, 115)} = 2.6$, $p < .05$), no significant between-group differences were observed in post-hoc tests. Hispanic patients ($n = 10$) scored higher on the PHQ-2 than non-Hispanic patients (M = 4.3, SD = 1.8 versus M = 2.5, SD = 2.2; $t_{(108)} = -2.60$, $p < .05$). Cronbach's alphas for the nine PHQ-9 and seven GAD-7 items were .92 and .93, and coefficient omegas were also .92 (95% CI: .89, .94) and .93 (95% CI: .91, .95), indicating excellent internal consistency.

Effects of Mood and Anxiety Disorder Diagnoses

Patients with a current mood disorder had higher PHQ-9 and PHQ-2 scores than those without (Hedge's $g = 0.97$ and 1.06) (see Table 2). Patients with a lifetime mood disorder also had higher PHQ-9 and PHQ-2 scores than those without (Hedge's $g = 1.03$ and 0.86).[2] Patients with a current anxiety disorder had higher GAD-7 and GAD-2 scores than those without (Hedge's $g = 0.97$ and 0.99). Patients with a lifetime anxiety disorder also had higher GAD-7 and GAD-2 scores than those without (Hedge's $g = 1.06$ and 1.13). The magnitudes of these effect sizes indicate that the presence of a mood disorder was strongly associated with PHQ-9/2 scores, and the presence of an anxiety disorder strongly associated with GAD-7/2 scores.

Sensitivity, Specificity, PPP, NPP, Agreement, and Correct Classification

ROC curves with PHQ-9 and PHQ-2 scores and presence versus absence of current mood disorder are shown in Figure 1. The chi-square, sensitivity and specificity, PPP and NPP, base rates, kappa coefficients, and percentages of patients correctly classified are presented in Table 3. Cut scores of ≥ 7 or ≥ 8 on the PHQ-9 correctly classified 71% of the sample and evidenced the most favorable balances of sensitivity and specificity (≥ 7: .77 and .62; ≥

[2]As a sensitivity analysis, we also excluded those patients with current bipolar disorder ($n = 15$) and the interpretability of results (i.e., large effect sizes) did not change.

8: .74 and .66) and PPP and NPP (≥ 7: .74 and .66; ≥ 8: .75 and .65) in these data. These cut scores also had the highest rates of agreement between clinical mood disorder diagnosis and PHQ-9 classification (κ =.39 and .40). Of note, a PHQ-9 cut score of ≥ 6 also correctly classified 71% of patients, but when we excluded patientsn with current bipolar disorder, ≥ 7 or ≥ 8 clearly outperformed ≥ 6.[3] A cut score of ≥ 2 on the PHQ-2 successfully classified 74% of the sample, evidenced the most favorable balance of sensitivity (80%) and specificity (65%), and PPP (77%) and NPP (70%), and showed the highest agreement between mood disorder diagnosis and PHQ-2 classification (κ =.46).[4]

ROC curves with GAD-7 and GAD-2 scores and presence versus absence of current anxiety disorder are presented in Figure 1, and chi-square, sensitivity and specificity, PPP and NPP, base rates, kappa coefficients, and percentages of patients correctly classified in Table 4. A GAD-7 cut score of ≥ 6 was optimal as it successfully classified 73% of patients, evidenced the most favorable balance of sensitivity (81%) and specificity (64%), and PPP (70%) and NPP (76%), and showed the highest agreement between anxiety disorder diagnosis and GAD-7 classification (κ=.45). A GAD-2 cut score of ≥ 2 successfully classified 74% of the sample, evidenced the most favorable balance of both sensitivity (87%) and specificity (60%), and PPP (70%) and NPP (81%), and showed the highest agreement between anxiety disorder diagnosis and GAD-2 classification (κ=.48).

Convergent and Discriminant Validity

Associations between the PHQ-9 and PHQ-2 and other measures are displayed in Table 5. The PHQ-9 demonstrated convergent validity with the BDI-II (large, statistically significant correlation of .88). The PHQ-9 also evidenced positive, significant correlations of moderate to large magnitudes with the TAS and LDQ scores (.35 and .56). The PHQ-2 showed convergent validity with the BDI-II (large, significant correlation of .78), and positive, significant correlations of small to medium magnitudes with the TAS and LDQ scores (.21 and .49). There was also evidence of convergent validity of the GAD-7 with the STAI-S and STAI-T (large, significant correlations of .72 and .77). The GAD-7 also had positive, significant correlations with the TAS and LDQ (.52 and .58). The GAD-2 demonstrated convergent validity with the STAI-S and STAI-T scores (large, significant correlations of .63 and .70). Additionally, the GAD-2 evidenced positive, significant correlations of moderate to large magnitudes with the TAS and LDQ (.45 and .53).

Discussion

Results suggest that the PHQ-9 and GAD-7 are valid, reliable, and potentially clinically useful screening tools for depression and anxiety in young people with SUD. The briefer two-item versions performed similarly in terms of identifying patients with probable mood or anxiety disorders. The high rate of mood and anxiety disorders in this sample highlights

[3]When excluding bipolar patients (n = 15) in a sensitivity analysis: a PHQ-9 cut-score of ≥ 7 had sensitivity = .80, specificity = .62, PPP = .70, NPP = .74, and κ = .42, and ≥ 8 had sensitivity = .76, specificity = .66, PPP = .71, NPP = .72, and κ = .43. Both ≥ 7 and ≥ 8 correctly classified 71% of the sample.

[4]When excluding bipolar patients (n = 15), in a sensitivity analysis, a PHQ-2 cut-score of ≥ 2 was optimal: sensitivity = .80, specificity = .65, PPP = .72, NPP = .74, κ = .46, and 73% correctly classified.

the importance of assessing and monitoring depressive and anxiety symptoms in young people with SUD.

Regarding clinical utility as screening measures, cut-scores of 7 or 8 for the PHQ-9, 6 for the GAD-7, and 2 for the PHQ-2 and GAD-2 appeared optimal in these data for identifying probable mood or anxiety disorders in this population. There is substantial variability across the literature in terms of recommended cut scores for the PHQ-9 (e.g., Moriarty et al., 2015), with one meta-analysis showing no significant differences for scores from 8 to 11 across clinical settings (Manea et al., 2012), and for the GAD-7, one meta-analysis recommending a cutoff of 8 or 9 for screening adult patients (Plummer et al., 2016). Thus, we found evidence of slightly lower PHQ-9 and GAD-7 cut scores in young people with SUD.

Overall, accuracy of the PHQ-9 (e.g., sensitivity and specificity of .77 and .62 for the optimal cut-score) and GAD-7 (e.g., .81 and .64) for mood and anxiety disorder screening were lower than would be ideal and relative to other recent work. For example, the Manea et al. (2012) PHQ-9 meta-analysis reported pooled sensitivity and specificity estimates from .83 to .89 and .73 to .89 (for cut scores 7 to 11) for detecting major depressive disorder across different clinical settings, and Plummer et al. (2016) found pooled sensitivity and specificity estimates of .83 and .84 for their recommended GAD-7 cut score of 8. Our results suggest that about 23% of patients with depression and 20% with anxiety will be missed when using the PHQ-9 and GAD-7 to screen young people with SUD. Thus, providers working with this high-risk population are cautioned against assuming a “low” PHQ-9 or GAD-7 score means no additional assessment is needed and should consider inquiring verbally about depression and anxiety if time permits. It is also noteworthy that sensitivity was consistently higher than specificity for candidate PHQ-9 and GAD-7 cut scores. Given the purpose of mood and anxiety disorder screening, prioritizing sensitivity over specificity (i.e., reducing the number of false negatives versus false positives) may not be inherently problematic, as the benefit of identifying more patients with depression or anxiety (and thus potentially changing the course of mental illness, especially for young people) may outweigh the additional time burden associated with following up on false positives.

The PHQ-2 and GAD-2 had similar (and based on absolute values, even slightly better) performance in terms of detecting probable mood and anxiety disorders than the 9-item and 7-item versions. These findings are in line with previous work indicating that both the GAD-2 and the GAD-7 have acceptable properties for anxiety disorder screening (e.g., Plummer et al., 2016), but less consistent with a recent large meta-analysis showing significantly lower specificity or sensitivity (depending on the cut-score) for the PHQ-2 versus the PHQ-9 alone (Levis et al., 2020). Given the obvious advantages of using screening tools that are as brief as possible, future research that replicates these findings among young people with SUD would be valuable. Given that young people with SUD are at particularly high risk for suicidal behavior (Yule et al., 2018), using the PHQ-9 (that contains suicidal/self-injury ideation item) may alert providers from the outset to clinically meaningful information about suicidal thoughts, and thus be worth the additional time investment of a longer screening tool.

The PHQ-9, PHQ-2, GAD-7, and GAD-2 all demonstrated strong convergent validity. Only the PHQ-2, however, met our *a priori* definition of discriminant validity. That said, it is not unexpected that depressive and anxiety symptoms would track with anger (e.g., Cassiello-Robbins & Barlow, 2016) and substance dependence (e.g., Conner et al., 2009; Lai et al., 2015), and the magnitudes of these correlations were consistently smaller than those with measures of depression and anxiety. Future research that incorporates measures of constructs less strongly linked to depression and anxiety (e.g., personality traits such as openness and agreeableness) (e.g., Kotov et al., 2010) may be optimally relevant to establishing discriminant validity.

Results from this study must be considered in light of its limitations. The sample was predominantly White and non-Hispanic; whether findings on the psychometric properties of these measures extend to young people with SUD from minority groups is unclear. Second, the diagnostic information used in this study was from routine clinical evaluations, not a "gold-standard" structured or semi-structured diagnostic interview. Classification accuracy might have been increased with use of a validated or structured interview as the reference standard (versus clinical interviewing); indeed, results from a recent meta-analysis indicate that the type of clinical interview used to diagnose disorders impacts classification accuracy for the PHQ-2 (Levis et al., 2020). Third, the relatively small sample size and ordinal (and in some cases, not normally distributed) item-level PHQ-9 and GAD-7 data rendered us unable to conduct adequately powered factor analyses to examine the latent structure of these scales. Given controversy about whether for example the PHQ-9 has a single- or two-factor structure (e.g., (Boothroyd et al., 2019), future studies on these measures in young people with SUD should evaluate this. Fourth, the small number of patients aged 17 or younger (n = 38) in this sample rendered us unable to formally test for measurement invariance between adolescents (under 18) and for example young adults (ages 18 to 25), which may be an important step for future research especially it is perhaps more typical for clinics to either serve patients under 18 or ages 18+. Last, and as noted earlier, the assessment of discriminant validity was limited.

In summary, our findings indicate the PHQ-9 and GAD-7 are valid, reliable, and potentially clinically useful screening tools for depression and anxiety in young people with SUDs. The briefer PHQ-2 and GAD-2 also appear to discriminate just as well between patients with and without mood and anxiety disorders. Providers and settings that treat young people with SUD may consider using these brief, relatively low-burden tools to screen for depression and anxiety. Several key directions for future research remain, such as examining the sensitivity to change of these measures for monitoring progress of dually diagnosed young people over time (e.g., Zimmerman, 2019) and the performance and effectiveness of step-wise combined screening approaches (e.g., Levis et al., 2020) in this unique, high-risk population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Role of the Funding Source:

Research reported in this publication was supported by the National Institute on Drug Abuse (grants K12DA000357 [Dr. Yule] and K12DA043490 [Dr. Evins]) and the National Institute of Mental Health (grant K23MH120436 [Dr. Bentley]) of the National Institutes of Health.

References

Allgaier A, Pietsch K, Frühe B, Sigl-Glöckner J, & Schulte-Körne G (2012a). Screening for depression in adolescents: Validity of the patient health questionnaire in pediatric care. Depression and Anxiety, 29(10), 906–913. 10.1002/da.21971 [PubMed: 22753313]

Allgaier A, Pietsch K, Frühe B, Sigl-Glöckner J, & Schulte-Körne G (2012b). Screening for depression in adolescents: Validity of the patient health questionnaire in pediatric care. Depression and Anxiety, 29(10), 906–913. 10.1002/da.21971 [PubMed: 22753313]

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders, 5th Edition.

Armstrong TD, & Costello EJ (2002). Community studies on adolescent substance use, abuse, or dependence and psychiatric comorbidity. Journal of Consulting and Clinical Psychology, 70(6), 1224–1239. 10.1037/0022-006X.70.6.1224 [PubMed: 12472299]

Arterberry BJ, Boyd CJ, West BT, Schepis TS, & McCabe SE (2019). DSM-5 substance use disorders among college-age young adults in the United States: Prevalence, remission and treatment. Journal of American College Health, 1–8. 10.1080/07448481.2019.1590368

Bach B, Sellbom M, & Simonsen E (2018). Personality Inventory for DSM-5 (PID-5) in Clinical Versus Nonclinical Individuals: Generalizability of Psychometric Features. Assessment, 25(7), 815–825. 10.1177/1073191117709070 [PubMed: 28503935]

Balsis S, & Cully JA (2008). Comparing depression diagnostic symptoms across younger and older adults. Aging & Mental Health, 12(6), 800–806. 10.1080/13607860802428000 [PubMed: 19023732]

Beard C, & Björgvinsson T (2014). Beyond generalized anxiety disorder: Psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. Journal of Anxiety Disorders, 28(6), 547–552. 10.1016/j.janxdis.2014.06.002 [PubMed: 24983795]

Beard C, Hsu KJ, Rifkin LS, Busch AB, & Björgvinsson T (2016). Validation of the PHQ-9 in a psychiatric sample. Journal of Affective Disorders, 193, 267–273. 10.1016/j.jad.2015.12.075 [PubMed: 26774513]

Beck AT, Steer RA, & Brown BK (1996). Manual for the Beck Depression Inventory-II. Psychological Corporation.

Boden MT, & Moos R (2009). Dually diagnosed patients' responses to substance use disorder treatment. Journal of Substance Abuse Treatment, 37(4), 335–345. 10.1016/j.jsat.2009.03.012 [PubMed: 19540699]

Boothroyd L, Dagnan D, & Muncer S (2019). PHQ-9: One factor or two? Psychiatry Research, 271, 532–534. 10.1016/j.psychres.2018.12.048 [PubMed: 30553100]

Cassiello Robbins C, & Barlow DH. (2016). Anger: The Unrecognized Emotion in Emotional Disorders. Clinical Psychology: Science and Practice, 23(1), 66–85. 10.1111/cpsp.12139

Chan Y, Dennis ML, & Funk RR (2008). Prevalence and comorbidity of major internalizing and externalizing problems among adolescents and adults presenting to substance abuse treatment. Journal of Substance Abuse Treatment, 34(1), 14–24. 10.1016/j.jsat.2006.12.031 [PubMed: 17574804]

Cohen J (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37–46. 10.1177/001316446002000104

Conner KR, Pinquart M, & Gamble SA (2009). Meta-analysis of depression and substance use among individuals with alcohol use disorders. Journal of Substance Abuse Treatment, 37(2), 127–137. 10.1016/j.jsat.2008.11.007 [PubMed: 19150207]

Couwenbergh C, van den Brink W, Zwart K, Vreugdenhil C, van Wijngaarden-Cremers P, & van der Gaag RJ (2006). Comorbid psychopathology in adolescents and young adults treated for substance use disorders: A review. European Child & Adolescent Psychiatry, 15(6), 319–328. 10.1007/s00787-006-0535-6 [PubMed: 16648966]

Dear BF, Titov N, Sunderland M, McMillan D, Anderson T, Lorian C, & Robinson E (2011). Psychometric Comparison of the Generalized Anxiety Disorder Scale-7 and the Penn State Worry Questionnaire for Measuring Response during Treatment of Generalised Anxiety Disorder. Cognitive Behaviour Therapy, 40(3), 216–227. 10.1080/16506073.2011.582138 [PubMed: 21770844]

Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, & Dale V (2011a). How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. Journal of Affective Disorders, 134(1–3), 266–271. 10.1016/j.jad.2011.06.017 [PubMed: 21723619]

Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, & Dale V (2011b). How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. Journal of Affective Disorders, 134(1–3), 266–271. 10.1016/j.jad.2011.06.017 [PubMed: 21723619]

Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, & Dale V (2012a). Brief case finding tools for anxiety disorders: Validation of GAD-7 and GAD-2 in addictions treatment. Drug and Alcohol Dependence, 125(1–2), 37–42. 10.1016/j.drugalcdep.2012.03.011 [PubMed: 22480667]

Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, & Dale V (2012b). Brief case finding tools for anxiety disorders: Validation of GAD-7 and GAD-2 in addictions treatment. Drug and Alcohol Dependence, 125(1–2), 37–42. 10.1016/j.drugalcdep.2012.03.011 [PubMed: 22480667]

Dickstein D (2011). Anxiety in adolescents: Update on its diagnosis and treatment for primary care providers. Adolescent Health, Medicine and Therapeutics, 1 10.2147/AHMT.S7597

Dum M, Pickren J, Sobell LC, & Sobell MB (2008a). Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. Addictive Behaviors, 33(2), 381–387. 10.1016/j.addbeh.2007.09.017 [PubMed: 17964079]

Dum M, Pickren J, Sobell LC, & Sobell MB (2008b). Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. Addictive Behaviors, 33(2), 381–387. 10.1016/j.addbeh.2007.09.017 [PubMed: 17964079]

Dunn TJ, Baguley T, & Brunsden V (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105(3), 399–412. 10.1111/bjop.12046 [PubMed: 24844115]

Gilbody S, Richards D, Brealey S, & Hewitt C (2007). Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): A diagnostic meta-analysis. Journal of General Internal Medicine, 22(11), 1596–1602. 10.1007/s11606-007-0333-y [PubMed: 17874169]

Grella CE, Hser Y-I, Joshi V, & Rounds-Bryant J (2001). Drug Treatment Outcomes for Adolescents with Comorbid Mental and Substance Use Disorders: The Journal of Nervous and Mental Disease, 189(6), 384–392. 10.1097/00005053-200106000-00006 [PubMed: 11434639]

Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, & Conde JG (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics, 42(2), 377–381. 10.1016/j.jbi.2008.08.010 [PubMed: 18929686]

Hepner KA, Hunter SB, Edelen MO, Zhou AJ, & Watkins K (2009). A comparison of two depressive symptomatology measures in residential substance abuse treatment clients. Journal of Substance Abuse Treatment, 37(3), 318–325. 10.1016/j.jsat.2009.03.005 [PubMed: 19359127]

IBM Corp. (2016). IBM SPSS Statistics for Windows, Version 24.0. IBM Corp.

Johnson SU, Ulvenes PG, Øktedalen T, & Hoffart A (2019). Psychometric Properties of the General Anxiety Disorder 7-Item (GAD-7) Scale in a Heterogeneous Psychiatric Sample. Frontiers in Psychology, 10, 1713 10.3389/fpsyg.2019.01713 [PubMed: 31447721]

Kandel DB, Johnson JG, Bird HR, Canino G, Goodman SH, Lahey BB, Regier DA, & Schwab-Stone M (1997). Psychiatric disorders associated with substance use among children and adolescents: Findings from the Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) Study. Journal of Abnormal Child Psychology, 25(2), 121–132. 10.1023/a:1025779412167 [PubMed: 9109029]

Kelly TM, Cornelius JR, & Clark DB (2004). Psychiatric disorders and attempted suicide among adolescents with substance use disorders. Drug and Alcohol Dependence, 73(1), 87–97. 10.1016/j.drugalcdep.2003.10.004 [PubMed: 14687963]

Kertz S, Bigda-Peyton J, & Bjorgvinsson T (2013a). Validity of the Generalized Anxiety Disorder-7 scale in an acute psychiatric sample. Clinical Psychology & Psychotherapy, 20(5), 456–464. 10.1002/cpp.1802 [PubMed: 22593009]

Kertz S, Bigda-Peyton J, & Bjorgvinsson T (2013b). Validity of the Generalized Anxiety Disorder-7 scale in an acute psychiatric sample. Clinical Psychology & Psychotherapy, 20(5), 456–464. 10.1002/cpp.1802 [PubMed: 22593009]

Keum BT, Miller MJ, & Inkelas KK (2018). Testing the factor structure and measurement invariance of the PHQ-9 across racially diverse U.S. college students. Psychological Assessment, 30(8), 1096–1106. 10.1037/pas0000550 [PubMed: 29565614]

Kotov R, Gamez W, Schmidt F, & Watson D (2010). Linking “big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. Psychological Bulletin, 136(5), 768–821. 10.1037/a0020327 [PubMed: 20804236]

Krawczyk N, Feder KA, Saloner B, Crum RM, Kealhofer M, & Mojtabai R (2017). The association of psychiatric comorbidity with treatment completion among clients admitted to substance use treatment programs in a U.S. national sample. Drug and Alcohol Dependence, 175, 157–163. 10.1016/j.drugalcdep.2017.02.006 [PubMed: 28432939]

Kroenke K, Spitzer RL, & Williams JB (2001a). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606–613. 10.1046/j.1525-1497.2001.016009606.x [PubMed: 11556941]

Kroenke K, Spitzer RL, & Williams JB (2001b). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606–613. 10.1046/j.1525-1497.2001.016009606.x [PubMed: 11556941]

Kroenke K, Spitzer RL, & Williams JB (2003). The Patient Health Questionnaire-2: Validity of a two-item depression screener. Medical Care, 41(11), 1284–1292. 10.1097/01.MLR.0000093487.78664.3C [PubMed: 14583691]

Kroenke K, Spitzer RL, Williams JB, Monahan PO, & Löwe B (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. Annals of Internal Medicine, 146(5), 317–325. 10.7326/0003-4819-146-5-200703060-00004 [PubMed: 17339617]

Kung S, Alarcon RD, Williams MD, Poppe KA, Jo Moore M, & Frye MA (2013). Comparing the Beck Depression Inventory-II (BDI-II) and Patient Health Questionnaire (PHQ-9) depression measures in an integrated mood disorders practice. Journal of Affective Disorders, 145(3), 341–343. 10.1016/j.jad.2012.08.017 [PubMed: 23021192]

Lai HMX, Cleary M, Sitharthan T, & Hunt GE (2015a). Prevalence of comorbid substance use, anxiety and mood disorders in epidemiological surveys, 1990–2014: A systematic review and meta-analysis. Drug and Alcohol Dependence, 154, 1–13. 10.1016/j.drugalcdep.2015.05.031 [PubMed: 26072219]

Lai HMX, Cleary M, Sitharthan T, & Hunt GE (2015b). Prevalence of comorbid substance use, anxiety and mood disorders in epidemiological surveys, 1990–2014: A systematic review and meta-analysis. Drug and Alcohol Dependence, 154, 1–13. 10.1016/j.drugalcdep.2015.05.031 [PubMed: 26072219]

Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, Neupane D, Imran M, Brehaut E, Negeri Z, Fischer FH, Benedetti A, Thombs BD, Depression Screening Data (DEPRESSD) PHQ Collaboration, Che L, Levis A, Riehm K, Saadat N, Azar M, … Zhang Y. (2020). Accuracy of the PHQ-2 Alone and in Combination With the PHQ-9 for Screening to Detect Major Depression: Systematic Review and Meta-analysis. JAMA, 323(22), 2290–2300. 10.1001/jama.2020.6504 [PubMed: 32515813]

Levitt EE, Syan SK, Sousa S, Costello MJ, Rush B, Samokhvalov AV, McCabe RE, Kelly J, & MacKillop J (2021). Optimizing screening for depression, anxiety disorders, and post-traumatic stress disorder in inpatient addiction treatment: A preliminary investigation. Addictive Behaviors, 112, 106649 10.1016/j.addbeh.2020.106649 [PubMed: 32979691]

Lubman DI, Allen NB, Rogers N, Cementon E, & Bonomo Y (2007). The impact of co-occurring mood and anxiety disorders among substance-abusing youth. Journal of Affective Disorders, 103(1–3), 105–112. 10.1016/j.jad.2007.01.011 [PubMed: 17291589]

Lyons RM, Yule AM, Schiff D, Bagley SM, & Wilens TE (2019). Risk Factors for Drug Overdose in Young People: A Systematic Review of the Literature. Journal of Child and Adolescent Psychopharmacology, 29(7), 487–497. 10.1089/cap.2019.0013 [PubMed: 31246496]

Manea L, Gilbody S, & McMillan D (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis. CMAJ: Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne, 184(3), E191–196. 10.1503/cmaj.110829

McDonald RP (1999). Test theory: A unified treatment. Lawrence Erlbaum.

Meredith W (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika, 58(4), 525–543. 10.1007/BF02294825

Moriarty AS, Gilbody S, McMillan D, & Manea L (2015). Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): A meta-analysis. General Hospital Psychiatry, 37(6), 567–576. 10.1016/j.genhosppsych.2015.06.012 [PubMed: 26195347]

Morin CM, Landreville P, Colecchi C, McDonald K, Stone J, & Ling W (1999). The Beck Anxiety Inventory: Psychometric Properties with Older Adults. Journal of Clinical Geropsychology, 5(1), 19–29. 10.1023/A:1022986728576

Parkerson HA, Thibodeau MA, Brandt CP, Zvolensky MJ, & Asmundson GJG (2015). Cultural-based biases of the GAD-7. Journal of Anxiety Disorders, 31, 38–42. 10.1016/j.janxdis.2015.01.005 [PubMed: 25725310]

Plummer F, Manea L, Trepel D, & McMillan D (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic metaanalysis. General Hospital Psychiatry, 39, 24–31. 10.1016/j.genhosppsych.2015.11.005 [PubMed: 26719105]

Raistrick D, Bradshaw J, Tober G, Weiner J, Allison J, & Healey C (1994). Development of the Leeds Dependence Questionnaire (LDQ): A questionnaire to measure alcohol and opiate dependence in the context of a treatment evaluation package. Addiction (Abingdon, England), 89(5), 563–572. 10.1111/j.1360-0443.1994.tb03332.x

Rice F, Riglin L, Lomax T, Souter E, Potter R, Smith DJ, Thapar AK, & Thapar A (2019). Adolescent and adult differences in major depression symptom profiles. Journal of Affective Disorders, 243, 175–181. 10.1016/j.jad.2018.09.015 [PubMed: 30243197]

Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, Rockhill C, & Katon W (2010). Evaluation of the Patient Health Questionnaire-9 Item for Detecting Major Depression Among Adolescents. PEDIATRICS, 126(6), 1117–1123. 10.1542/peds.2010-0852 [PubMed: 21041282]

Rutter LA, & Brown TA (2017a). Psychometric Properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in Outpatients with Anxiety and Mood Disorders. Journal of Psychopathology and Behavioral Assessment, 39(1), 140–146. 10.1007/s10862-016-9571-9 [PubMed: 28260835]

Rutter LA, & Brown TA (2017b). Psychometric Properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in Outpatients with Anxiety and Mood Disorders. Journal of Psychopathology and Behavioral Assessment, 39(1), 140–146. 10.1007/s10862-016-9571-9 [PubMed: 28260835]

Spielberger CD (1983). Manual for the State-Trait Anxiety Inventory (STAI). Consulting Psychologists Press.

Spielberger CD (1988). Manual for the State-Trait Anger Expression Inventory. Psychological Assessment Resources.

Spielberger CD (1999). STAXI-2: State-Trait Anger Expression Inventory-2, Professional Manual. Psychological Assessment Resources.

Spitzer RL, Kroenke K, Williams JB, & Löwe B (2006a). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166(10), 1092–1097. 10.1001/archinte.166.10.1092 [PubMed: 16717171]

Spitzer RL, Kroenke K, Williams JB, & Löwe B (2006b). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166(10), 1092–1097. 10.1001/archinte.166.10.1092 [PubMed: 16717171]

Staples LG, Dear BF, Gandy M, Fogliati V, Fogliati R, Karin E, Nielssen O, & Titov N (2019). Psychometric properties and clinical utility of brief measures of depression, anxiety, and general distress: The PHQ-2, GAD-2, and K-6. General Hospital Psychiatry, 56, 13–18. 10.1016/j.genhosppsych.2018.11.003 [PubMed: 30508772]

Stockings E, Degenhardt L, Lee YY, Mihalopoulos C, Liu A, Hobbs M, & Patton G (2015). Symptom screening scales for detecting major depressive disorder in children and adolescents: A systematic review and meta-analysis of reliability, validity and diagnostic utility. Journal of Affective Disorders, 174, 447–463. 10.1016/j.jad.2014.11.061 [PubMed: 25553406]

Substance Abuse and Mental Health Services Administration. (2019). Key substance use and mental health indicators in the United States: Results from the 2018 National Survey on Drug Use and Health (HHS Publication No. PEP19–5068; NSDUH Series H-54). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration https://www.samhsa.gov/data/

Tiirikainen K, Haravuori H, Ranta K, Kaltiala-Heino R, & Marttunen M (2019). Psychometric properties of the 7-item Generalized Anxiety Disorder Scale (GAD-7) in a large representative sample of Finnish adolescents. Psychiatry Research, 272, 30–35. 10.1016/j.psychres.2018.12.004 [PubMed: 30579178]

Titov N, Dear BF, McMillan D, Anderson T, Zou J, & Sunderland M (2011). Psychometric Comparison of the PHQ-9 and BDI-II for Measuring Response during Treatment of Depression. Cognitive Behaviour Therapy, 40(2), 126–136. 10.1080/16506073.2010.550059 [PubMed: 25155813]

Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, & Löwe B (2020). Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). Journal of Affective Disorders, 265, 395–401. 10.1016/j.jad.2020.01.032 [PubMed: 32090765]

Trainor K, Mallett J, & Rushe T (2013). Age related differences in mental health scale scores and depression diagnosis: Adult responses to the CIDI-SF and MHI-5. Journal of Affective Disorders, 151(2), 639–645. 10.1016/j.jad.2013.07.011 [PubMed: 23993442]

Yule AM, Carrellas NW, Fitzgerald M, McKowen JW, Nargiso JE, Bergman BG, Kelly JF, & Wilens TE (2018). Risk Factors for Overdose in Treatment-Seeking Youth With Substance Use Disorders. The Journal of Clinical Psychiatry, 79(3). 10.4088/JCP.17m11678

Zimmerman M (2019). Using the 9-Item Patient Health Questionnaire to Screen for and Monitor Depression. JAMA. 10.1001/jama.2019.15883

Highlights

- Depression and anxiety screening in young people with substance use is critical
- PHQ-9/2 and GAD-7/2 had good convergent and acceptable discriminant validity
- Classification accuracy was similar across brief (two-item) and full measures
- The two-item versions may have comparable clinical utility to the full measures

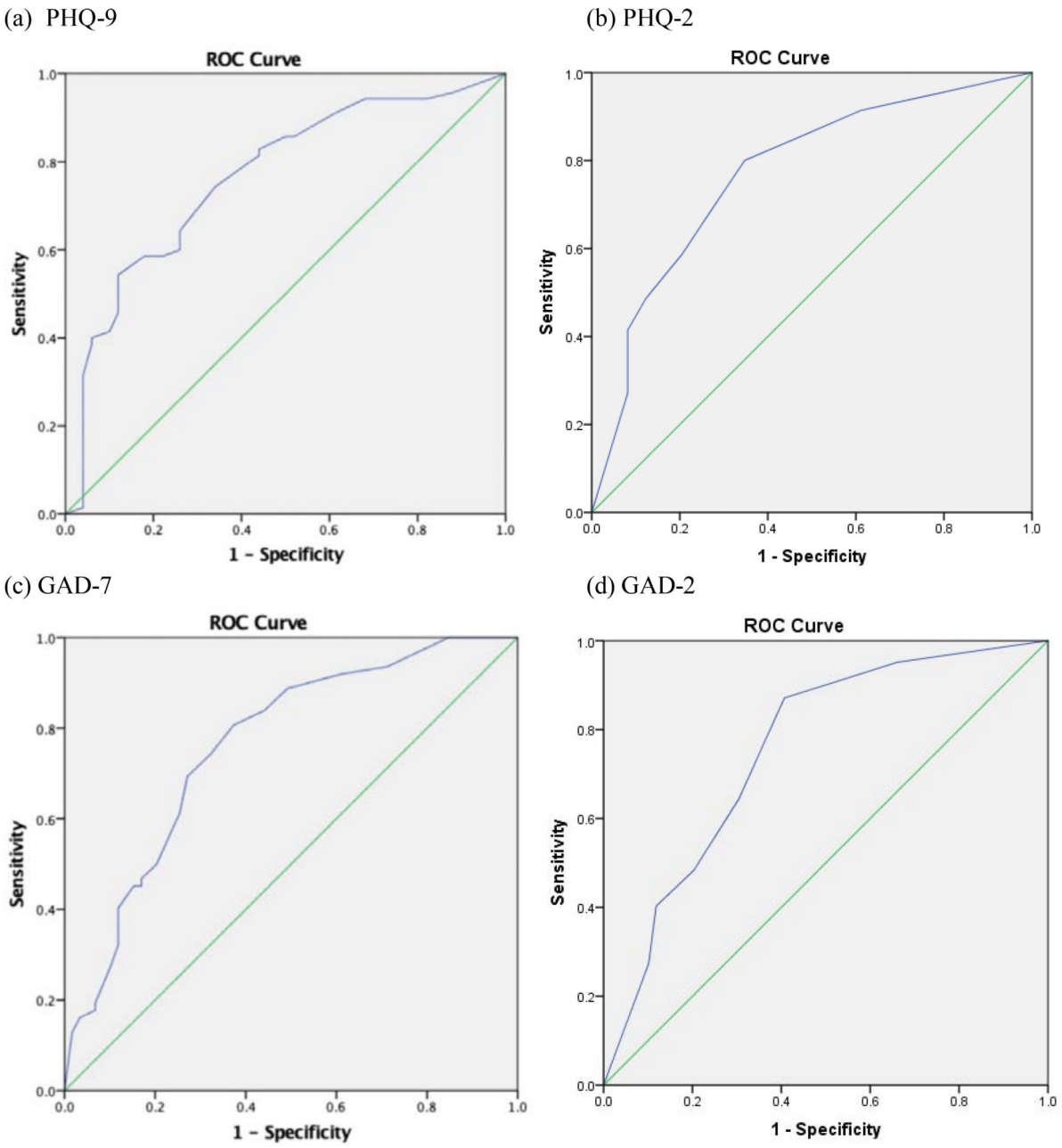

Figure 1.

Receiver operating characteristic (ROC) curves for (a) PHQ-9 and (b) PHQ-2 scores to predict presence of current mood disorder; (c) GAD-7 and (d) GAD-2 scores to predict presence of current anxiety disorder.

Table 1.

Demographic and Diagnostic Information

	n	%	Mean (SD)
Age	--	--	19.6 (3.1)
Gender			--
Female	38	31.4%	--
Male	83	68.6%	--
Race			--
White/Caucasian	91	75.2%	--
Asian	7	5.8%	--
Black/African American	3	2.5%	--
More than one race	5	4.1%	--
Ethnicity			--
Non-Hispanic	100	82.6%	--
Hispanic	11	9.1%	--
SUD diagnoses			
Alcohol use disorder	39	32.2%	--
Cannabis use disorder	77	63.6%	--
Opioid use disorder	21	17.4%	--
Stimulant use disorder	16	13.2%	--
Sedative use disorder	14	11.6%	--
Inhalant use disorder	1	0.8%	--
Mood and anxiety disorders			
Unipolar depressive disorder	56	46.3%	--
Bipolar disorder	15	12.4%	--
Anxiety disorder	62	51.2%	--

Note. SUD = substance use disorder. Only current (not lifetime) diagnoses presented here. SUD and mood and anxiety disorders could be primary or secondary, so percentages add to over 100%. Anxiety disorders include GAD, anxiety disorder unspecified, PTSD, social anxiety disorder, panic disorder, OCD, adjustment disorder, body dysmorphic disorder, and trichotillomania.

Table 2.

PHQ-9, PHQ-2, GAD-7, and GAD-2 Scores in those with and without Mood and Anxiety Disorder Diagnoses

		Mean	SD	Range	n	Hedge's *g*
PHQ-9						
Current mood disorder	Yes	13.5	7.7	0–27	70	0.97
	No	6.5	6.3	0–27	50	
Lifetime mood disorder	Yes	12.5	7.8	0–27	90	1.03
	No	5.0	5.1	0–18	31	
PHQ-2						
Current mood disorder	Yes	3.5	2.1	0–6	70	1.06
	No	1.4	1.8	0–6	49	
Lifetime mood disorder	Yes	3.1	2.2	0–6	90	0.86
	No	1.3	1.7	0–6	30	
GAD-7						
Current anxiety disorder	Yes	11.2	6.2	1–21	62	0.97
	No	5.4	5.7	0–21	58	
Lifetime anxiety disorder	Yes	10.1	6.3	0–21	90	1.06
	No	3.7	4.9	0–19	31	
GAD-2						
Current anxiety disorder	Yes	3.6	1.9	0–6	62	0.99
	No	1.7	1.9	0–6	58	
Lifetime anxiety disorder	Yes	3.3	2.0	0–6	90	1.13
	No	1.1	1.8	0–6	31	

Note. Current and lifetime mood disorder included unipolar and bipolar depression. Anxiety disorder included prototypical anxiety disorders (e.g., GAD, panic disorder, social anxiety disorder) and related conditions (e.g., obsessive-compulsive and related disorders, PTSD).

Table 3.

Diagnostic Utility of the PHQ-9 and PHQ-2

	χ^2	Sensitivity	Specificity	PPP	NPP	Base rate	κ	% correctly classified
PHQ-9 score								
0	--	1.00	0.00	0.58	--	1.00	0.00	58
1	2.50	0.96	0.12	0.60	0.67	0.93	0.09	61
2	4.56*	0.94	0.18	0.62	0.69	0.89	0.14	63
3	14.51**	0.94	0.32	0.66	0.80	0.83	0.29	68
4	15.32**	0.91	0.38	0.67	0.76	0.79	0.32	69
5	18.01**	0.86	0.50	0.71	0.71	0.71	0.37	70
6	18.17**	0.81	0.56	0.72	0.68	0.66	0.38	71
7	18.76**	0.77	0.62	0.74	0.66	0.61	0.39	71
8	19.37**	0.74	0.66	0.75	0.65	0.58	0.40	71
9	17.12**	0.64	0.74	0.78	0.60	0.48	0.37	68
10	13.58**	0.60	0.74	0.76	0.57	0.46	0.33	66
11	19.75**	0.59	0.82	0.82	0.59	0.42	0.38	68
12	20.57**	0.57	0.84	0.83	0.58	0.40	0.39	68
13	22.46**	0.54	0.88	0.86	0.58	0.37	0.39	68
14	19.93**	0.51	0.88	0.86	0.56	0.35	0.37	67
15	15.32**	0.46	0.88	0.84	0.54	0.32	0.31	63
16	14.19**	0.41	0.90	0.85	0.52	0.28	0.28	62
17	17.60**	0.40	0.94	0.90	0.53	0.26	0.31	63
18	16.50**	0.39	0.94	0.90	0.52	0.25	0.29	62
19	13.71**	0.31	0.96	0.92	0.50	0.20	0.24	58
20	10.82*	0.27	0.96	0.90	0.48	0.18	0.20	56
21	6.46*	0.20	0.96	0.88	0.46	0.13	0.14	52
22	5.66*	0.19	0.96	0.87	0.46	0.13	0.13	51
23	3.43	0.14	0.96	0.83	0.44	0.10	0.09	48
24	2.75	0.13	0.96	0.82	0.44	0.09	0.08	48
25	0.98	0.09	0.96	0.75	0.43	0.07	0.04	45
26	0.18	0.06	0.96	0.67	0.42	0.05	0.01	43
27	0.79	0.01	0.96	0.33	0.41	0.03	−0.02	41
PHQ-2 score								
0	--	1.00	0.00	0.59	--	1.00	0.00	59
1	15.85**	0.91	0.39	0.68	0.76	0.79	0.33	70
2	24.95**	0.80	0.65	0.77	0.70	0.61	0.46	74

	χ^2	Sensitivity	Specificity	PPP	NPP	Base rate	κ	% correctly classified
3	17.14**	0.59	0.80	0.80	0.57	0.43	0.36	67
4	17.05**	0.49	0.88	0.85	0.54	0.34	0.33	65
5	15.92**	0.41	0.92	0.88	0.52	0.28	0.30	62
6	6.66*	0.27	0.92	0.83	0.47	0.19	0.17	54

Note. PPP = positive predictive power; NPP = negative predictive power; base rate = percentage scoring at or above cut score; κ = agreement between PHQ-9 and clinical diagnosis of current mood disorder after corrective for chance.

* = $p < .05$.

** $p < .001$.

Table 4.

Diagnostic Utility of the GAD-7 and GAD-2

	χ^2	Sensitivity	Specificity	PPP	NPP	Base rate	κ	% correctly classified
GAD-7 score								
0	--	1.00	0.00	0.52	--	1.00	0.00	52
1	10.40*	1.00	0.16	0.56	1.00	0.93	0.16	59
2	10.85**	0.94	0.29	0.59	0.81	0.83	0.23	63
3	16.72**	0.92	0.40	0.62	0.82	0.77	0.32	67
4	22.97**	0.89	0.52	0.66	0.81	0.69	0.41	71
5	21.66**	0.84	0.57	0.68	0.77	0.64	0.41	71
6	24.49**	0.81	0.64	0.70	0.76	0.59	0.45	73
7	22.43**	0.74	0.69	0.72	0.71	0.53	0.43	72
8	22.70**	0.69	0.74	0.74	0.69	0.48	0.43	72
9	16.84**	0.61	0.76	0.73	0.65	0.43	0.37	68
10	12.69**	0.50	0.81	0.74	0.60	0.35	0.31	65
11	13.53**	0.47	0.84	0.76	0.60	0.32	0.31	65
12	12.35**	0.45	0.84	0.76	0.59	0.31	0.29	64
13	14.04**	0.45	0.86	0.78	0.60	0.30	0.31	65
14	14.06**	0.40	0.90	0.81	0.58	0.26	0.29	64
15	8.48*	0.32	0.90	0.77	0.55	0.22	0.21	60
16	7.07*	0.27	0.91	0.77	0.54	0.18	0.18	58
18	4.02*	0.19	0.93	0.75	0.52	0.13	0.12	55
19	3.22	0.18	0.93	0.73	0.51	0.13	0.11	54
20	5.35*	0.16	0.97	0.83	0.52	0.10	0.12	55
21	5.40*	0.13	0.98	0.89	0.51	0.08	0.11	54
GAD-2 score								
0	--	1.00	0.00	0.52	--	1.00	0.00	52
1	17.00**	0.95	0.34	0.61	0.87	0.81	0.30	66
2	29.33**	0.87	0.60	0.70	0.81	0.64	0.48	74
3	14.89**	0.65	0.71	0.70	0.65	0.48	0.35	68
4	11.53**	0.48	0.81	0.73	0.59	0.34	0.29	64
5	14.06**	0.40	0.90	0.81	0.58	0.26	0.29	64
6	5.64*	0.27	0.90	0.74	0.54	0.19	0.17	58

Note. PPP = positive predictive power; NPP = negative predictive power; base rate = percentage scoring at or above cut score; κ = agreement between GAD-7 and clinical diagnosis of current anxiety disorder after corrective for chance.

* = $p < .05$.

** $p < .001$.

Table 5.

Correlations of PHQ-9, PHQ-2, GAD-7, and GAD-2 with Convergent and Discriminant Validity Measures

	BDI-II	TAS	LDQ	
PHQ-9	0.88**	0.35**	0.56**	
PHQ-2	0.78**	0.21*	0.49**	
	STAI-S	STAI-T	TAS	LDQ
GAD-7	0.72**	0.77**	0.52**	0.58**
GAD-2	0.63**	0.70**	0.45**	0.53**

Note. BDI-II = Beck Depression Inventory-II, LDQ = Leeds Dependence Questionnaire, STAI-S = State-Trait Anxiety Inventory-State subscale, STAI-T = State-Trait Anxiety Inventory-Trait subscale, TAS = Trait Anger Scale

** $p < 0.001$.