Research article

# Comparison of DASS-21, PHQ-8, and GAD-7 in a virtual behavioral health care setting

Lila Peters, Aimee Peters, Evie Andreopoulos, Naomi Pollock, Reena L. Pande, Heidi Mochari-Greenberger *

*AbleTo Clinical Research, New York, NY, United States*

A B S T R A C T

*Background:* Validated depression and anxiety symptom screeners are commonly used in clinical settings. How results from different brief depression and anxiety symptom assessment tools compare to each other is not well established, especially in real world healthcare settings. This study aimed to compare the Depression Anxiety Stress Scales 21 Depression scale (DASS-Depression) and Anxiety (DASS-Anxiety) scale to the Patient Health Questionnaire 8 (PHQ-8) and Generalized Anxiety Disorder 7 (GAD-7) respectively, in a real-world virtual behavioral healthcare setting.
*Methods:* This was a retrospective comparison study of clinical data from a population of adults who completed a consultation via telephone or secure video with a licensed therapist as part of a standardized, evidence-based, virtual behavioral therapy program for individuals with comorbid medical and behavioral health conditions. The joint distributions and correlations between scores yielded by each depression and anxiety scale were assessed using descriptive and Spearman correlation statistics.
*Results:* The DASS-Depression and PHQ-8 were highly correlated (r = .71; p=<.001); the DASS-Anxiety and GAD-7 correlation was also high (r = .61; p=<.001). The PHQ-8 categorized more individuals as having above-threshold depression scores versus the DASS-Depression (71.5% vs. 43.5%; p < .001). The GAD-7 categorized more individuals as having above-threshold anxiety scores versus the DASS-Anxiety (59.0% vs. 45.0%; p < .001).
*Limitations:* This study compared results yielded by validated screeners, precluding conclusions related to the validity of screener results.
*Conclusions:* The DASS-Depression and PHQ-8 and the DASS-Anxiety and GAD-7 similarly ranked symptom severity. The PHQ-8 and GAD-7 were more likely than the DASS-21 Depression or Anxiety scales to classify individuals as having above-threshold symptom severity.

## Introduction

As the prevalence of comorbid general medical and mental health conditions is more broadly recognized, screening for depression and anxiety symptoms has become a more standard practice. Evidence-based treatment guidelines for several common medical conditions recommend screening for psychiatric comorbidity with validated brief assessment tools (American Diabetes Association, 2018; Andersen et al., 2014; Jha et al., 2019; Siu and US Preventive Services Task Force, 2016).

Which depression or anxiety assessment tool to use may depend on several factors including the clinical setting, practitioner type, patient characteristics, and how the results will be utilized. For example, results yielded from a tool that is used in the psychiatric setting to support

accurate diagnosis may be interpreted differently than those from a tool used by non-mental health providers to screen for risk.

Most recently, increased availability of behavioral telehealth services has broadened mental health care access and treatment opportunities (FAIR Health, 2019; Centers for Medicare and Medicaid Services, 2019). With this shift to virtually delivered care, psychiatric symptom assessment tools are being routinely administered over telephone or video conference in both medical and behavioral health care settings (National Committee for Quality Assurance, 2019).

The COVID-19 pandemic has further amplified screening and referral into tele-delivered mental health care, increasing the utility of information that distinguishes and eases interpretation of results produced by different scales. While several commonly used depression and anxiety

screeners such as the Patient Health Questionnaire 9 (PHQ-9) and the Generalized Anxiety Disorder 7 (GAD-7) have been used in telehealth clinical trials (Mohr et al., 2012; Salisbury et al., 2016; Titov et al., 2011), little is known about whether rates of identification of symptomatic individuals differ between these tools and those produced by other brief assessment tools used in the field, especially under real-world virtual care conditions. For example, the Depression Anxiety Stress Scales 21 (DASS-21) may be used to assess depression and anxiety symptom severity in clinical settings; but few data have compared the DASS-Depression scale to the PHQ screeners, and fewer have compared anxiety symptom severity as assessed by the DASS-21 Anxiety and Stress scales to those yielded by GAD-7 (Lambert et al., 2015; Orta et al., 2015; Sakakibara et al., 2009). This information is important, as it would inform interpretation and comparison of results yielded by different tools in the same population.

To address this gap in knowledge, we sought to compare the results from two commonly used depression symptom assessment tools (the DASS-21 Depression Scale (DASS-Depression) and the Patient Health Questionnaire-8 (PHQ-8)) and two common anxiety symptom assessment tools (the DASS-21 Anxiety and Stress scales (DASS-Anxiety and DASS-Stress) and Generalized Anxiety Disorder-7 (GAD-7)) in an adult population with medical and behavioral health issues being treated in a virtual behavioral health care setting. Specifically, we aimed to (1) determine the correlation and compare symptom severity classification between the PHQ-8 and DASS-Depression, and (2) determine the correlation and compare symptom severity classification between the GAD-7 and DASS-Anxiety and DASS-Stress Scales.

## Methods

### Study population

The present investigation was a cross-sectional comparison study of previously collected clinical data from a virtually delivered behavioral health program. The program comprised protocolized cognitive behavior therapy delivered in a fully remote telehealth setting by telephone or video to adults with comorbid medical (e.g. cardiovascular disease, diabetes, chronic pain) and mental health (e.g. depression, anxiety) conditions (Dent et al., 2018; Mochari-Greenberger et al., 2016, 2017; Pande et al., 2015). The clinical protocol consisted of sixteen clinical sessions delivered over approximately 8 weeks, starting with a comprehensive initial consultation conducted by a licensed clinical social worker (LCSW) and subsequent sessions alternating between psychotherapy and behavioral coaching (Dent et al., 2018). Participants were 18 years old or older and access to a telephone was required to enroll. The presence of acute suicidal risk, recent psychiatric hospitalization, or cognitive impairment precluded participation.

As part of a quality improvement initiative, a group of 9 LCSWs were uniformly trained to administer the PHQ-8 and GAD-7, in addition to the DASS-21 which was standard as part of the protocolized comprehensive clinical interview. The order in which the three tools were utilized was neither randomized nor prescribed. The initial consultation also included collection of sociodemographic characteristics (age, sex) and clinical information (past psychiatric and medical histories), and a comprehensive psychiatric risk assessment. De-identified data from a sample of 202 participants were abstracted for this analysis. The present study was approved by the Sterling Institutional Review Board.

### Measures

#### DASS-21

The DASS-21, a brief version of the 42-item Depression Anxiety Stress Scales (DASS-21), is a 21-item clinical assessment with subscales for depression (DASS-Depression), anxiety (DASS-Anxiety), and stress (DASS-Stress) (Lovibond and Lovibond, 1995; Psychology Foundation of Australia, 2018a). Clinical severity categories and cut points of the

DASS-21 are described in Supplemental Figure 1. The depression subscale assesses the loss of motivation and self-esteem. The anxiety subscale primarily measures symptoms of persistent anxiety and fear. The stress subscale addresses symptoms of persistent arousal and irritability (Lovibond and Lovibond, 1995). The DASS-21 has been validated in clinical and community populations against the gold-standard DSM-IV structured clinical interview (Antony et al., 1998). It has also been demonstrated to have excellent internal consistency in clinical populations (Cronbach's α = .94 (DASS-Depression), .87 (DASS-Anxiety), and .91 (DASS-Stress)) (Antony et al., 1998).

#### PHQ-8

The PHQ-8 is a version of the Primary Care Evaluation of Mental Disorders (PRIME-MD) depression measure (PHQ-9), typically utilized to assess depressive symptom severity and to assess for the presence of major depressive disorder (Kroenke et al., 2001, 2008; Kroenke and Spitzer, 2002). Clinical symptom severity categories and cut points for the PHQ-8 are described in Supplemental Figure 1. A PHQ-8 score ≥5 is considered above minimal threshold and a PHQ-8 score ≥10 points is consistent with moderate or greater depressive symptoms. The PHQ-8 scale has been shown to have good internal reliability in medical populations (Cronbach's α = 0.82) (Pressler et al., 2010).

A diagnostic algorithm is offered by the authors of the PHQ-8 as an alternative to using cut points to evaluate current depression (Kroenke et al., 2008). Using the PHQ-8 algorithm, the presence of "major depression" is indicated when: 1) 5 to 8 symptoms on the PHQ-8 are endorsed as having been present "more than half the days" (score ≥2 points for each item) in the past two weeks, *and* 2) one or more of the 5 to 8 symptoms endorsed is/are anhedonia and/or depressed mood (PHQ-8 questions 1 and/or 2). The presence of "other depression" is indicated when: 1) 2 to 4 of the eight symptoms measured by the PHQ-8 were present on "more than half the days" in the past two weeks (score ≥2 points for each item) *and* 2) one or more of the 2 to 4 symptoms endorsed is/are anhedonia and/or depressed mood (PHQ-8 questions 1 and/or 2). The PHQ-8 algorithm indicates *any* depression if criteria for *major depression* or *other depression* are met.

The PHQ-8 differs from the more commonly-used PHQ-9 in that it excludes the final question addressing suicidal ideation (Kroenke et al., 2008). In this study, the final question was not included because suicide risk was comprehensively assessed separately using a standardized risk assessment during the same session in which the PHQ-8 was administered. The intention was to reduce redundancy and number of questions to optimize the patient and clinician experience. The PHQ-8 has been shown to be similar to the PHQ-9 in identifying depression (Kroenke et al., 2008); a meta-analysis of PHQ-9 validation studies found that the PHQ-9 had a pooled sensitivity of 92% and a specificity of 80% compared to the DSM clinical interview, and has diagnostic properties comparable with longer diagnostic instruments (Gilbody et al., 2007). To assess the potential impact that utilizing the PHQ-8 versus the PHQ-9 could have had on study results, and allow for comparability of these results to those from past and future investigations which may utilize the PHQ-9, a supplemental analysis (described below under Data Analysis) was conducted to look at the data from the suicide risk screening alongside PHQ-8 data.

#### GAD-7

The GAD-7 is a seven item scale used to assess symptoms of generalized anxiety disorder (GAD). The GAD-7 was developed to reflect symptom criteria of GAD as defined by the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) criteria and other validated anxiety measures (Spitzer et al., 2006). Clinical severity categories and cut points of the GAD-7 are described in Supplemental Figure 1. The GAD-7 has been shown to have high internal reliability (Cronbach's α = .92) and validity to identify the presence of GAD when compared to structured psychological interview by a mental health professional (DSM-IV criteria) in medical care settings (sensitivity =

89%, specificity 82% to identify GAD when GAD-7 score is $\geq$10 points) (Spitzer et al., 2006).

### Data analysis

The distribution of baseline demographic and clinical characteristics were calculated using means and proportions. The distribution of scores on the DASS-Depression, DASS-Anxiety, PHQ-8 and GAD-7 scales were evaluated against standard cut-points to identify above threshold scores (DASS-Depression $\geq$ 10, DASS-Anxiety $\geq$ 8, PHQ-8 $\geq$ 5 and GAD-7 $\geq$ 5) and scores of moderate or higher symptom severity (Kroenke et al., 2008; Lovibond and Lovibond, 1995; Spitzer et al., 2006). Internal consistency was quantified for each scale using Cronbach's alpha. The relation between DASS-Depression and PHQ-8 and between DASS-Anxiety and GAD-7 were assessed by Spearman rank correlation. Categorical distributions of DASS-Depression and PHQ-8 severity categories were compared using proportions (McNemar's test statistic for paired nominal data) and illustrated using a bubble plot; DASS-Anxiety and GAD-7 were similarly compared.

Two supplemental analyses were also conducted. First, because the DASS-Stress scale also assesses symptoms similar to the DSM-IV diagnosis of Generalized Anxiety Disorder (Lovibond and Lovibond, 1995; Psychology Foundation of Australia, 2018a, 2018b, 2018c), we conducted an exploratory analysis to compare symptom severity classifications yielded by 1) the DASS-Stress Scale versus the GAD-7, and 2) the DASS-Anxiety and the DASS-Stress Scale combined versus the GAD-7. Kappa statistics were used to describe agreement between scales. Second, because the difference in score between the PHQ-8 and the PHQ-9 could have affected results we conducted a supplemental analysis to estimate how results might have differed if the PHQ-9 was utilized. To do this, we imputed 3 points (highest possible item score) into the PHQ-8 score for each participant that endorsed any frequency or intensity of suicidal ideation in the past month during their initial consultation with their therapist to create a pseudo PHQ-9 score. The analyses to compare PHQ-8 and DASS-Depression described above were repeated replacing the PHQ-8 score with the pseudo PHQ-9 score. All data were analyzed using SAS statistical software (version 9.4). Statistical significance was set at p < 0.05.

### Results

The baseline characteristics of the study population are described in Table 1. The 202 participants were 68.3% female; mean age was 51 years. More than half had been previously diagnosed with depression (55.9%) and almost half had been diagnosed with anxiety (45.0%). Almost two in five had a history of diabetes, 45.5% had a history of hypertension, and 16.8% had a history of heart disease.

### Comparison of DASS-Depression and PHQ-8

The DASS-Depression (Cronbach's $\alpha$ = .90) and PHQ-8 (Cronbach's $\alpha$ = .85) were highly correlated (Spearman r = .71; p < .001); however, the distribution of score severity varied between the two measures (Figure 1). The PHQ-8 identified more patients as having above threshold (PHQ-8 $\geq$5) depression symptom scores versus the DASS-Depression ($\geq$10 points; 71.5% vs. 43.5%; p < .001). The PHQ-8 also classified more individuals as having moderate or higher depressive symptoms (PHQ-8 $\geq$10 points) versus the DASS-Depression ($\geq$14 points; 44% vs. 31.5; p = 0.0002).

In order to better understand these differences in classification, we explored the specific items within each scale that may have contributed. Among the subset of participants with above threshold scores on the PHQ-8 but not the DASS-Depression (n = 60; 41.9% of those with above threshold PHQ-8 scores), PHQ-8 scale items related to somatic symptoms not expressly measured by the DASS-Depression were frequently endorsed (Table 2). Additionally, these participants reported the somatic

**Table 1.** Baseline characteristics of a consecutive sample of virtual behavioral healthcare program participants (N = 202).

| Baseline Characteristics | Distribution |
|---|---|
| *Age* | |
| Years (mean [±SD]) | 51.0 [±10.6] |
| Range (years) | 24–68 |
| *Female (n [%])* | 138 [68.3] |
| *Race/Ethnicity* | |
| Black or African American (n [%]) | 36 [17.8] |
| Hispanic or Latino (n [%]) | 18 [8.9] |
| White (n [%]) | 129 [63.9] |
| Other/Declined to Answer (n [%]) | 19 [9.4] |
| *Employment Status* | |
| Currently Employed/Working for Pay (n [%]) | 119 [58.9] |
| *Past Psychiatric History* | |
| Depression (n [%]) | 113 [55.9] |
| Anxiety (n [%]) | 91 [45.0] |
| Panic Attacks (n [%]) | 35 [17.3] |
| *Past Medical History* | |
| Diabetes (n [%]) | 80 [39.6] |
| Heart Disease (n [%]) | 34 [16.8] |
| Hypertension (n [%]) | 92 [45.5] |
| Kidney Disease (n [%]) | 15 [7.4] |
| *Symptom Severity* | |
| Above Threshold DASS-Depression ($\geq$10) (n [%]) | 88 [43.6] |
| Above Threshold PHQ-8 ($\geq$5) (n [%])* | 143 [71.5] |
| Above Threshold DASS-Anxiety ($\geq$8) (n [%]) | 91 [45.0] |
| Above Threshold GAD-7 ($\geq$5) (n [%])** | 118 [59.0] |

* N = 200; n missing = 2.
** N = 200; n missing = 2.

symptoms with increased severity (indicated by scores $\geq$2) relative to non-somatic symptoms addressed by both the DASS-Depression and PHQ-8 such as anhedonia, dysphoria, and self-deprecation. Other symptoms measured by the PHQ-8 but not the DASS-Depression, such as difficulty concentrating and psychomotor retardation, were not endorsed with as great severity or frequency as the symptoms described above.

When the PHQ-8 algorithm was applied to the raw PHQ-8 data, almost half of participants with raw PHQ-8 scores $\geq$5 points were classified as having *any depression,* and the proportion with *any depression* rose to nearly three out of four when the sample was restricted to those with a raw PHQ-8 score $\geq$10 points (Table 3).

Of the participants with an above threshold DASS-Depression score ($\geq$10 points), 65.5% were also classified with *any depression* by PHQ-8 algorithm. Among those with moderate or higher DASS-Depression score ($\geq$14 points), 76.2% were categorized with *any depression* by the PHQ-8 diagnostic algorithm. When the PHQ-8 diagnostic algorithm was applied to the 60 participants with above threshold scores on the PHQ-8 but not DASS-Depression, only 16.7% met the criteria for *any depression.*

Our supplemental analysis to assess the potential impact of using the PHQ-8 (versus the PHQ-9) on these observed results identified ten out of 202 participants (5%) with thoughts of suicide in the past month (2/10 daily in the past month, 7/10 one to two times in the past month, 1/10 none in the past month but conversely endorsed low intensity ideation in the past week). Agreement between the original PHQ-8 scores and the pseudo PHQ-9 (calculated by adding 3 points to the PHQ-8 score of those participants with thoughts of suicide) to identify individuals with scores $\geq$5 points (elevated) on both scales was perfect (Kappa = 1.0); agreement to identify individuals with $\geq$10 points (moderately elevated or more) on both scales was close to perfect (Kappa = .99), with the pseudo PHQ-9 categorizing one additional individual as having moderate or more severe depression symptoms, supporting the conclusion that study

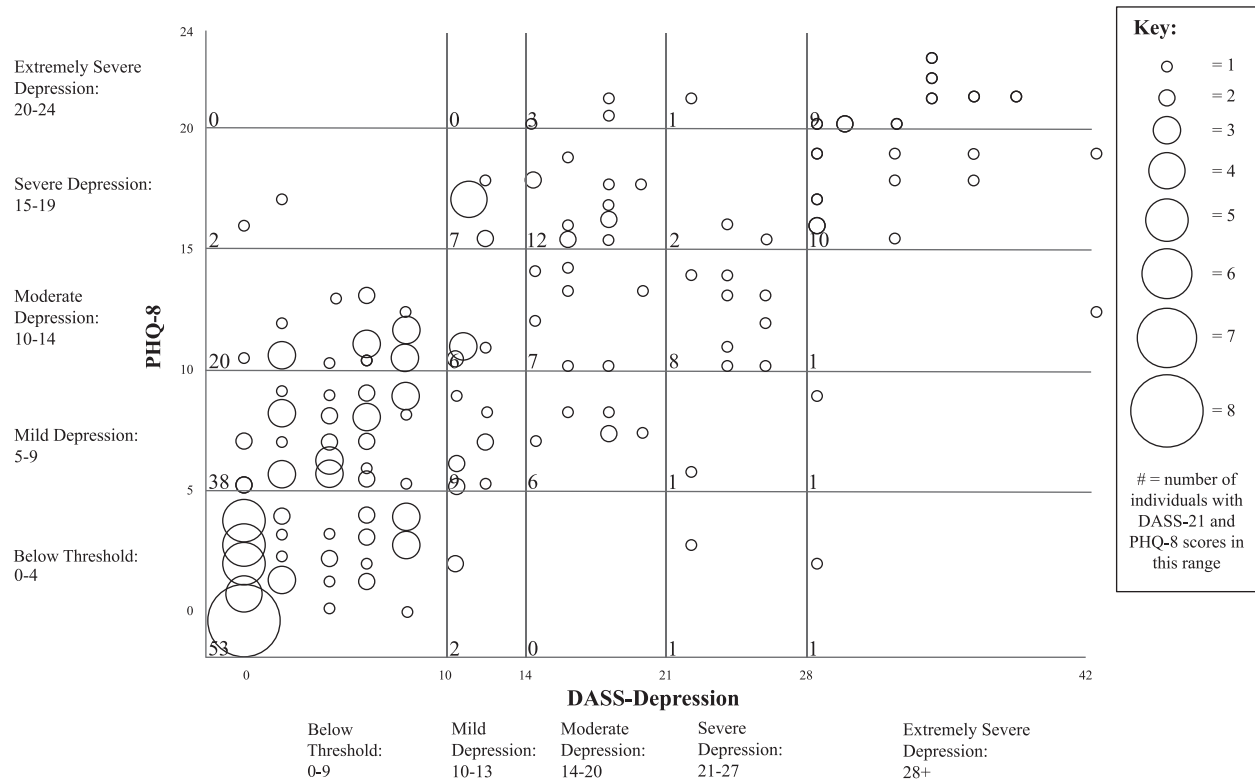**Comparison of Within-Person Depression Symptom Severity: DASS-21 Depression Scale Versus the PHQ-8**



**Figure 1.** Comparison of within-person depression symptom severity: DASS-21 depression scale versus the PHQ-8.

results would not have materially differed if the PHQ-9 was used instead of the PHQ-8 in this population.

*Comparison of DASS-Anxiety and GAD-7*

The DASS-Anxiety (Cronbach's α = .72) and GAD-7 (Cronbach's α = .87) were moderately correlated (Spearman r = .61; p=<.001).

Categorical data analysis of DASS-Anxiety and GAD-7 data documented a difference in the distribution of score severity between the measures (Figure 2). The GAD-7 classified significantly more individuals as having above threshold symptoms of anxiety than the DASS-Anxiety (59.0% vs. 45.0%; p < .001). Almost two-thirds (65.2%) of participants classified by the GAD-7 as having above threshold anxiety symptom scores were also classified as above

**Table 2.** Distribution of Within-Item PHQ Scores among Participants Who Screened Above Threshold on the PHQ-8 but not on the DASS-21 Depression Scale (n = 60).

| PHQ-8 Question | PHQ-8 Item Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | 3 |
| | n (%) | n (%) | n (%) | n (%) |
| 1. Little interest or pleasure in doing things | 27 (45) | 25 (42) | 8 (13) | 0 (0) |
| 2. Feeling down, depressed, or hopeless | 27 (45) | 29 (48) | 4 (7) | 0 (0) |
| 3. Trouble falling or staying asleep, or sleeping too much | 5 (8) | 12 (20) | 12 (20) | 31 (52) |
| 4. Feeling tired or having little energy | 5 (8) | 17 (29) | 18 (30) | 20 (33) |
| 5. Poor appetite or overeating | 16 (27) | 16 (27) | 14 (23) | 14 (23) |
| 6. Feeling bad about yourself – or that you are a failure or have let yourself or your family down | 35 (58) | 17 (29) | 5 (8) | 3 (5) |
| 7. Trouble concentrating on things, such as reading the newspaper or watching television | 28 (47) | 19 (32) | 6 (10) | 7 (11) |
| 8. Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual | 47 (79) | 8 (13) | 2 (3) | 3 (5) |

**Table 3.** Proportion of participants meeting PHQ-8 diagnostic criteria at various DASS-21 depression scale and PHQ-8 cut points.

| Above Threshold Scores On: | N | PHQ-8 Diagnostic Algorithm Classification Category | | |
|---|---|---|---|---|
| | | Any Depression[1] n (%) | Other Depression n (%) | Major Depression n (%) |
| PHQ-8 ≥ 5 | 143 | 67 (47%) | 18 (13%) | 49 (34%) |
| PHQ-8 ≥ 10 | 88 | 63 (72%) | 14 (16%) | 49 (56%) |
| DASS-Depression ≥ 10 | 87 | 57 (66%) | 10 (11%) | 47 (54%) |
| DASS-Depression ≥ 14 | 63 | 48 (76%) | 8 (13%) | 40 (64%) |
| PHQ-8 ≥ 5 & DASS-Depression ≥ 10 | 83 | 57 (69%) | 10 (12%) | 47 (57%) |
| PHQ-8 ≥ 5 & DASS-Depression < 10 | 60 | 10 (17%) | 8 (13%) | 2 (3%) |
| PHQ-8 < 5 & DASS-Depression ≥ 10 | 4 | 0 (0%) | 0 (0%) | 0 (0%) |

[1] Any Depression is equal to the total of Major Depression plus Other Depression.



**Comparison of Within-Person Anxiety Symptom Severity: DASS-21 Anxiety Scale Versus the GAD-7**
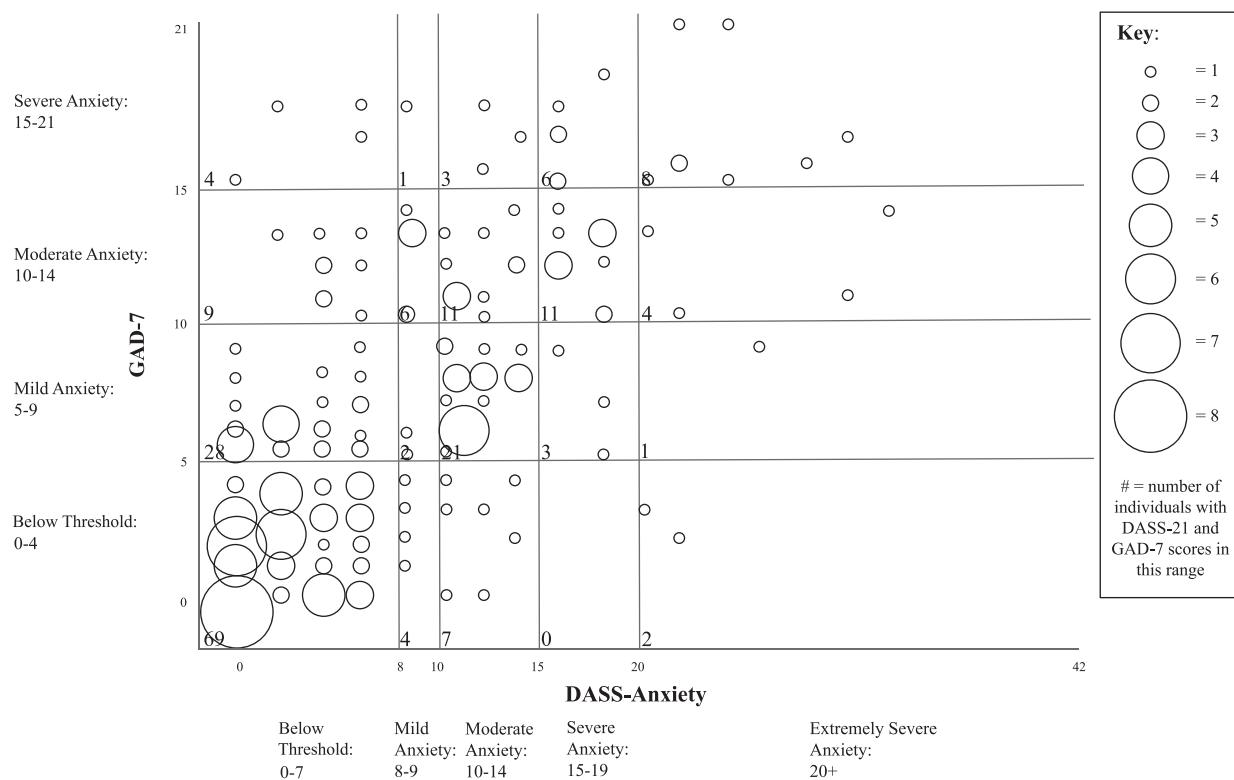
**Figure 2.** Comparison of within-person anxiety symptom severity: DASS-21 anxiety scale versus the GAD-7.

threshold by the DASS-Anxiety. Of the participants that were classified as below-threshold on the GAD-7, 84% of them were also classified as below threshold on the DASS-Anxiety. The GAD-7 classified a similar proportion of individuals as having moderate or more severe anxiety symptoms (GAD-7 ≥10) versus the DASS-Anxiety (32% vs. 39%; p = .06).

No striking clinical patterns or commonalities were observed in the GAD-7 item responses among participants with above threshold scores on the GAD-7 but not the DASS-Anxiety (n = 41; 34.7% of those with above threshold GAD-7 scores; Table 4). Review of the individual items on the GAD-7 scale and DASS-Anxiety scales suggest they assess different symptoms of anxiety. Participants who endorsed questions about worry on the GAD-7 also tended to endorse analogous questions on the DASS-Anxiety. However, scores on questions about worry on the GAD-7 were observed to be higher than scores about worry assessed by the DASS-Anxiety.

The DASS-Stress scale (Cronbach's α = .83) identified 38% of participants as above threshold (≥15 points) (versus the 59% originally identified by the GAD-7). However, when the DASS-Anxiety and the DASS-Stress scales were combined (i.e. if a participant was above threshold on either scale), 51% of participants were identified as above threshold; a proportion not statistically different from the 59% identified as above threshold by the GAD-7 (McNemar's p = 0.27). Agreement between the GAD-7 and the combined DASS-Anxiety and DASS-Stress scales to categorize individuals as above versus below threshold for anxiety symptoms was moderate (Kappa = 0.58) indicating the combined DASS scales were identifying many of the same individuals as above threshold as the GAD-7.

**Table 4.** Distribution of Within-item GAD-7 Scores Among Participants With an Above Threshold Score on the GAD-7 but not on the DASS-21 Anxiety Scale (n = 41).

| GAD-7 Question | GAD-7 Item Score | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | 3 |
| | n (%) | n (%) | n (%) | n (%) |
| 1. Feeling nervous, anxious, or on edge | 7 (17) | 23 (56) | 5 (12) | 6 (15) |
| 2. Not being able to control or stop worrying | 7 (17) | 20 (49) | 6 (15) | 8 (19) |
| 3. Worrying too much about different things | 8 (19) | 16 (39) | 7 (17) | 10 (25) |
| 4. Trouble relaxing | 6 (15) | 16 (39) | 7 (17) | 12 (29) |
| 5. Being so restless that it is hard to sit still | 20 (49) | 14 (34) | 5 (12) | 2 (5) |
| 6. Becoming easily annoyed or irritable | 3 (7) | 25 (61) | 7 (17) | 6 (15) |
| 7. Feeling afraid as if something awful might happen | 23 (56) | 10 (25) | 5 (12) | 3 (7) |

## Discussion

In this comparison study among adults with medical conditions who received behavioral therapy in a virtual behavioral health care setting, we documented significant correlation between commonly used screening tools for depression and anxiety as well as important differences in meeting thresholds and in grading severity. The PHQ-8 and the GAD-7 classified a higher proportion of individuals as above threshold versus the DASS-21. And, a significant proportion of individuals with above threshold PHQ-8 scores were not classified as above threshold for depression by PHQ-8 algorithm.

Few published studies have compared the PHQ to the DASS-Depression scale in medically comorbid populations (Lambert et al., 2015; Orta et al., 2015; Sakakibara et al., 2009). Those that have compared them directly have yielded patterns similar to those presented here. For example, in a study evaluating depression, anxiety, and stress comorbidity with the presence of migraines in a cohort of pregnant women, the PHQ-9 classified more participants as above threshold than the DASS-Depression (62.1% vs. 16.9%) (Orta et al., 2015). In another study evaluating the prevalence of depression in a cancer population, the PHQ-9 raw score also classified more participants as above threshold than the DASS-Depression, with authors concluding that the PHQ-9 raw score was identifying more individuals in the mild and moderate score severity categories (Lambert et al., 2015). This study adds to prior knowledge by corroborating these observations utilizing the PHQ-8, and in a virtual behavioral health care setting.

One possible explanation for why the PHQ-8 classified more participants as having above threshold symptoms versus the DASS-Depression is the difference in the depression symptoms the scales measure. In this study, participants with above threshold scores on the PHQ-8 but below threshold scores on the DASS-Depression were observed to endorse the PHQ-8 somatic symptoms (abnormal sleep behavior, lack of energy, and changes in appetite) more frequently and with higher scores than the PHQ-8 cognitive/affective symptoms. The DASS-Depression does not address these somatic depressive symptoms, which may in part explain the higher depression symptoms severity ratings observed on the raw PHQ-8 versus the DASS-Depression scale scores in this population.

Previous studies have also documented differences in the level of depression when assessed by the sum score of a standardized questionnaire (e.g. PHQ-8's raw symptom severity scores) versus clinical interview or applied algorithms (Hartung et al., 2019). This may also be attributable in part to the somatic symptoms associated with depression that medical populations may experience at a higher frequency than the general population in the presence or in the absence of cognitive/affective symptoms such as depressed mood or anhedonia (Ferrando et al., 2007; Grapp et al., 2019; Hartung et al., 2019). While the PHQ-8 diagnostic algorithm requires that depressed mood or anhedonia be present at least more than half the days for an individual to be classified with *any depression,* the standardized threshold scores for the same tool assigns all symptoms equal weight without this requirement. Our data had a much lower proportion of participants screening in as above

threshold for *any depression* when the PHQ-8 diagnostic algorithm was used instead of the raw cut-off scores (i.e. ≥5 points and ≥10 points). The results suggest assessing depression symptom severity using only the raw PHQ-8 cut-off scores, could yield a higher estimate for prevalence and severity of depression in medical populations with somatic symptoms than the algorithm would in the same population. However, it should be noted that other research has determined somatic items to be no less valid than cognitive/emotional items as indicators of depression in medically comorbid populations (Simon and Korff, 2006). Differences in response pattern between medical and non-medical populations have been observed; for example, a medically comorbid population may be more likely to report fatigue at a milder level of depression than a non-medical population control (Simon and Korff, 2006).

In this study we also observed the GAD-7 categorized a higher proportion of patients as having above threshold symptoms of anxiety compared to the DASS-Anxiety scale. This discordance could be related in part to the scales being designed to measure different types of anxiety symptoms. Based on the DSM-IV diagnosis of Generalized Anxiety Disorder, the GAD-7 assesses symptoms of worry, irritability, and restlessness (Spitzer et al., 2006). The DASS-Anxiety also assesses symptoms of worry along with symptoms of panic such as trembling of the hands, breathing difficulties, and increased heart rate (Psychology Foundation of Australia, 2018a).

When assessing symptoms related to worry, the DASS-Anxiety uses language that is specific to panic disorder and anticipatory anxiety for panic attacks more focused on panic (e.g. panic and worrying about panic), while the GAD-7 uses terminology inclusive of symptoms of generalized anxiety disorder more focused on worry (e.g. worrying too much and not being able to control worrying). Participant responses to items concerning worry tended to be higher on the GAD-7 than analogous questions on the DASS-Anxiety related to panic and this may have contributed to the GAD-7 assigning more above normal threshold scores versus the DASS-Anxiety. When the DASS-Stress scale, which measures symptoms of restlessness, difficulty relaxing and irritability associated with generalized anxiety disorder, was added to the analysis, we observed greater concordance between GAD-7 and the combined DASS-Anxiety and DASS-Stress scales. These observations underscore the value in pre-specifying what symptoms/condition you aim to screen for, before selecting the optimal brief scale(s) to utilize.

A final contributor to the observed discordance between the PHQ-8, GAD-7, and DASS-21 scales, could also be lookback time frame. While the DASS-21 has a look back period of one week, the PHQ-8 and GAD-7 have a lookback of two weeks (Psychology Foundation of Australia, 2018b; Kroenke et al., 2008; Spitzer et al., 2006). It is not possible given these data to know with certainty which direction time frame could have biased results. For example, one might hypothesize that a shorter timeframe would exclude individuals with symptoms that were present one to two weeks ago, but not in the past week. Alternatively over a longer timeframe, one might argue that the average level of symptom severity could be diluted for those whose symptoms have intensified over the past week.

The results from this comparison study add to our knowledge about how the DASS-21 Depression scale and the PHQ-8, and the DASS-21 Anxiety and Stress scales and the GAD-7, categorize symptom severity relative to each other in an adult population with medical and behavioral health issues. A primary limitation in this study was the small and heterogeneous convenience sample of participants with varying medical conditions which limits our ability to draw conclusions about the distribution of measures in specific medical populations. Additionally, the order in which the tools were utilized in each session was not randomized, which could have affected results; however the order was not prescribed either and the data represent results from a real-world clinical care setting. Though comparing results from these assessment tools administered in a virtual behavioral health setting is novel, this study was not designed to determine if results differ from those yielded by administration of the same tools in traditional health care settings. An important next research step will be to link these data to diagnosis yielded by clinical interview and to compare the criterion validity of the scales. The present study also aimed to compare different screeners to one another, not to a gold standard, which precludes us from drawing conclusions about validity. Nonetheless, this information may be informative to clinicians as they interpret results yielded by these tools in the populations they serve.

In this comparison study, we documented moderate to high correlation between the DASS-21 depression and anxiety scales with the PHQ-8 and GAD-7 respectively in a population with medical comorbidities receiving virtual behavioral health care. We also documented differences in the way the screeners categorized symptom severity, with the PHQ-8 and GAD-7 including more participants as above threshold than the corollary DASS-21 scales. The data underscore the importance of distinguishing what above-normal threshold scores indicate on each scale when selecting a brief tool to support assessment of depression and anxiety symptoms. Further research is needed to validate and examine the performance of these tools for screening and ongoing measurement in different clinical settings.

## Declarations

### Author contribution statement

Heidi Mochari Greenberger, Reena L Pande, Aimee Peters: Conceived and designed the analysis; Analyzed and interpreted the data; Contributed analysis tools or data; Wrote the paper.

Lila Peters, Evie Andreopoulos, Naomi Pollock: Conceived and designed the analysis; Analyzed and interpreted the data; Wrote the paper.

### Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Data availability statement

Data will be made available on request.

### Declaration of interests statement

L.P. is an employee of AbleTo. A.P. is an employee of and holds an equity interest in AbleTo. E.A. is an employee of AbleTo; N.P. is an employee of and holds an equity interest in AbleTo; R.L.P. is an employee of and holds an equity interest in AbleTo; she serves as AbleTo's Chief Medical Officer. H.M.G. is an employee of and holds an equity interest in AbleTo.

### Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2021.e06473.

## References

American Diabetes Association, 2018. Comprehensive medical evaluation and assessment of comorbidities. Diabetes Care 41 (Supplement 1), S28–S37.

Andersen, B.L., DeRubeis, R.J., Berman, B.S., Gruman, J., Champion, V.L., Massie, M.J., Holland, J.C., Partridge, A.H., Bak, K., Somerfield, M.R., Rowland, J.H., 2014. Screening, assessment, and care of anxiety and depressive symptoms in adults with cancer: an American Society of Clinical Oncology guideline adaptation. J. Clin. Oncol.: Off. J. Am. Soc. Clin. Oncol. 32 (15), 1605–1619.

Antony, M.M., Bieling, P.J., Cox, B.J., Enns, M.W., Swinson, R.P., 1998. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychol. Assess. 10 (2), 176–181.

Centers for Medicare & Medicaid Services (CMS), HHS, 2019. Medicare and Medicaid Programs; Policy and Technical Changes to the Medicare Advantage, Medicare Prescription Drug Benefit, Programs of All-Inclusive Care for the Elderly (PACE), Medicaid Fee-For-Service, and Medicaid Managed Care Programs for Years 2020 and 2021, 84 Fed. Reg. 73 (April 16, 2019). Federal Register: the Daily Journal of the United States. https://www.federalregister.gov/d/2019-06822. (Accessed 15 February 2021).

Dent, L., Peters, A., Kerr, P.L., Mochari-Greenberger, H., Pande, R.L., 2018. Using telehealth to implement cognitive-behavioral therapy. Psychiatr. Serv. 69 (4), 370–373.

FAIR Health, 2019. A Multilayered Analysis of Telehealth [White Paper]. 2019 from FAIR Health. https://www.fairhealth.org/article/fair-health-study-analyzes-telehealth. (Accessed 15 February 2021).

Ferrando, S.J., Samton, J., Mor, N., Nicora, S., Findler, M., Apatoff, B., 2007. Patient Health Questionnaire-9 to screen for depression in outpatients with multiple sclerosis. Int. J. MS Care 9 (3), 99–103.

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. J. Gen. Intern. Med. 22 (11), 1596–1602.

Grapp, M., Terhoeven, V., Nikendei, C., Friederich, H.C., Maatouk, I., 2019. Screening for depression in cancer patients using the PHQ-9: the accuracy of somatic compared to non-somatic items. J. Affect. Disord. 254, 74–81.

Hartung, T.J., Fried, E.I., Mehnert, A., Hinz, A., Vehling, S., 2019. Frequency and network analysis of depressive symptoms in patients with cancer compared to the general population. J. Affect. Disord. 256, 295–301.

Jha, M.K., Qamar, A., Vaduganathan, M., Charney, D.S., Murrough, J.W., 2019. Screening and management of depression in patients with cardiovascular disease. JACC (J. Am. Coll. Cardiol.) 73 (14), 1827–1845.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613.

Kroenke, K., Spitzer, R.L., 2002. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr. Ann. 32 (9), 509–515.

Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B., Berry, J.T., Mokdad, A.H., 2008. The PHQ-8 as a measure of current depression in the general population. J. Affect. Disord. 114 (1-3), 163–173.

Lambert, S.D., Clover, K., Pallant, J.F., Britton, B., King, M.T., Mitchell, A.J., Carter, G., 2015. Making sense of variations in prevalence estimates of depression in cancer: a co-calibration of commonly used depression scales using Rasch Analysis. J. Natl. Compr. Canc. Netw. 13 (10), 1203–1211.

Lovibond, P.F., Lovibond, S.,H., 1995. The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the Beck depression and anxiety inventories. Behav. Res. Ther. 33 (3), 335–343.

Mochari-Greenberger, H., Vue, L., Luka, A., Peters, A., Pande, R.L., 2016. A tele-behavioral health intervention to reduce depression, anxiety, and stress and improve diabetes self-management. Telemed. E-Health 22 (8), 624–630.

Mochari-Greenberger, H., Peters, A., Vue, L., Pande, R.L., 2017. A nationally scaled telebehavioral health program for chronic pain: characteristics, goals, and psychological outcomes. Telemed. E-Health 23 (8), 640–648.

Mohr, D.C., Ho, J., Duffecy, J., Reifler, D., Sokol, L., Burns, M.N., Jin, L., Siddique, J., 2012. Effect of telephone-administered vs. face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients. JAMA 307 (21).

National Committee for Quality Assurance, 2019. HEDIS Summary Table of Measures, Product Lines, and Changes [Summary Table]. Retrieved Feb 15, 2021 from NCQA: https://www.ncqa.org/wp-content/uploads/2018/08/20190000_HEDIS_Measures_SummaryofChanges.pdf.

Orta, O.R., Gelaye, B., Qiu, C., Stoner, L., Williams, M.A., 2015. Depression, anxiety and stress among pregnant migraineurs in a pacific-northwest cohort. J. Affect. Disord. 172, 390–396.

Pande, R.L., Morris, M., Peters, A., Spettell, C.M., Feifer, R., Gillis, W., 2015. Leveraging remote behavioral health interventions to improve medical outcomes and reduce costs. Am. J. Manag. Care 21, e141–e151. https://pubmed.ncbi.nlm.nih.gov/26065105.

Pressler, S.J., Subramanian, U., Perkins, S.M., Gradus-Pizlo, I., Kareken, D., Kim, J., Ding, Y., Sauve, M.J., Sloan, R., 2010. Measuring depressive symptoms in heart failure: validity and reliability of the Patient Health Questionnaire-8. Am. J. Crit. Care 20 (2), 146–152.

Psychology Foundation of Australia, 2018a. Overview of the DASS and its Uses. http://www2.psy.unsw.edu.au/Groups/Dass/over.htm. (Accessed 15 February 2021).

Psychology Foundation of Australia, 2018b. Download DASS Forms. http://www2.psy.unsw.edu.au/dass/down.htm. (Accessed 15 February 2021).

Psychology Foundation of Australia, 2018c. DASS FAQ. http://www2.psy.unsw.edu.au/dass/DASSFAQ.htm#_14._What_does_the_Stress_scale_mea. (Accessed 15 February 2021).

Sakakibara, B.M., Miller, W.C., Orenczuk, S.G., Wolfe, D.L., 2009. A systematic review of depression and anxiety measures used with individuals with spinal cord injury. Spinal Cord 47 (12), 841–851.

Salisbury, C., O'Cathain, A., Edwards, L., Thomas, C., Gaunt, D., Hollinghurst, S., Nicholl, J., Large, S., Yardley, L., Lewis, G., Foster, A., Garner, K., Horspool, K., Man, M.S., Rogers, A., Pope, C., Dixon, P., Montgomery, A.A., 2016. Effectiveness of an integrated telehealth service for patients with depression: a pragmatic randomised controlled trial of a complex intervention. The Lancet Psychiatr. 3 (6), 515–525.

Simon, G.E., Korff, M.V., 2006. Medical co-morbidity and validity of DSM-IV depression criteria. Psychol. Med. 36 (1), 27–36.

Siu, A.L., The US Preventive Services Task Force (USPSTF), 2016. Screening for depression in adults: US Preventive Services Task Force recommendation statement. JAMA 315 (4), 380–387.

Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder. Arch. Intern. Med. 166 (10), 1092.

Titov, N., Dear, B.F., Schwencke, G., Andrews, G., Johnston, L., Craske, M.G., McEvoy, P., 2011. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav. Res. Ther. 49 (8), 441–452.