The Validity and Reliability of Screening Measures for Depression and Anxiety Disorders in Inflammatory Bowel Disease

Charles N. Bernstein, MD,* Lixia Zhang MSc,[†] Lisa M. Lix, PhD,[†] Lesley A. Graff, PhD,[‡]
John R. Walker, PhD,[‡] John D. Fisk, PhD,[§] Scott B. Patten, MD, PhD,[¶] Carol A. Hitchon, MD, MSc,*
James M. Bolton, MD,[∥] Jitender Sareen, MD,[∥] Renée El-Gabalawy, PhD,[‡,**]
James Marriott, MD, MSc, MD* and Ruth Ann Marrie, MD, PhD*,[†];
for the CIHR Team in Defining the Burden and Managing the Effects of Immune-mediated Inflammatory Disease

Background: We evaluated the validity and reliability of multiple symptom scales for depression and anxiety for persons with inflammatory bowel disease (IBD).

Methods: IBD participants in a cohort study completed a Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID) and completed the Patient Health Questionnaire (PHQ-9), Hospital Anxiety and Depression Scale (HADS), Kessler-6 Distress Scale, PROMIS Emotional Distress Depression Short-Form 8a (PROMIS Depression) and Anxiety Short-Form 8a (PROMIS Anxiety), Generalized Anxiety Disorder 7-item Scale, and Overall Anxiety and Severity Impairment Scale. We computed sensitivity, specificity, and positive and negative predictive values for the screening measures with the SCID diagnoses as the reference standard, conducted receiver operating curve (ROC) analysis, and assessed internal consistency and test-retest reliability.

Results: Of 242 participants, the SCID classified 8.7% as having major depression and 17.8% as having anxiety disorders. Among the depression scales, the PHQ-9 had the highest sensitivity (95%). Specificity was generally higher than sensitivity and was highest for the HADS-D (cut-point of 11; 97%). The area under the ROC curve (AUC) did not differ significantly among depression scales. Among the anxiety scales, sensitivity was highest for the PROMIS (79%). Specificity ranged from 82% to 88% for all tools except the HADS-A (cut-point of 8; 65%). The AUC did not differ between depression and anxiety tools.

Conclusions: Overall, the symptom scales for depression and anxiety were similar in their psychometric properties. The anxiety scales did not perform as well as the depression scales. Alternate cut-points may be more relevant when these scales are used in an IBD sample.

Key Words: depression, anxiety, inflammatory bowel disease, psychometric properties.

INTRODUCTION

Inflammatory bowel disease (IBD) is associated with high morbidity from intestinal and extra-intestinal symptoms and fatigue.[1] It is also highly associated with psychiatric comorbidity. The incidence and prevalence of depression and anxiety are increased in IBD as compared with populations without IBD.[2] Depression may affect more than 25% and anxiety more than 30% of persons with IBD, a burden 2 to 3 times

higher than in the general population.[2] In a population-based study, psychiatric comorbidity antedated the diagnosis of IBD by at least 5 years.[3] Hence, many persons with newly diagnosed IBD may already have depression or anxiety. The incidence of psychiatric diagnoses is also increased after the diagnosis of IBD,[4] and the periods of increased incidence exist around major disease events such as in the first year postsurgery.[5]

Psychiatric disorders are associated with reduced quality of life in general,[6, 7] and also specifically in IBD.[8, 9] Psychiatric comorbidity can be associated with increased health care utilization in persons with IBD,[10, 11] and conversely, improvements in mental health can be associated with reduced health care utilization.[12] Concomitant depression or anxiety is also associated with loss of work productivity in persons with IBD.[13] Hence, identifying mental health disorders in persons with IBD either at the time of diagnosis or during the course of the disease is highly relevant to the care of the patient, to ensure the appropriate treatment can be initiated promptly, with the aim of improving outcomes across many domains.

Screening and monitoring treatment of these disorders in persons with IBD is important in both primary care and specialist settings. Screening in primary care for depression and anxiety specifically has been endorsed by the US Preventive Service Task Force (USPSTF).[14, 15] Several symptom scales have been developed for the common disorders of depression and anxiety, and it can be challenging for health care providers to determine the optimal tool to use. Moreover, as psychiatric comorbidity is increasingly recognized as an important contributor to health outcomes in IBD, various symptom scales have been included in IBD research studies. Self-administered validated symptom scales can take much less time than clinical interviews and do not require trained personnel. They can measure psychological issues in a severity continuum, and scoring can be standardized. Further, scales may be useful to help identify cases when used judiciously and could have a big role in monitoring progress during treatment. However, most were developed for the general population, and validation is needed to determine their applicability and interpretation in particular chronic disease populations. To our knowledge, a rigorous assessment of the psychometric performance of the various psychological symptom scales in IBD has yet to be undertaken.

We aimed to compare the reliability and diagnostic performance of several brief, self-administered symptom scales of anxiety and depression against a reference standard, the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV-TR) Axis I Disorders–Research version (SCID) interview in a cohort of persons with IBD.

METHODS

As detailed elsewhere, over a 20-month period, 247 individuals with definite diagnoses of IBD, aged 18 years or older, were enrolled in a longitudinal cohort study exploring psychiatric comorbidity in IBD and 2 other chronic immune-mediated inflammatory diseases (multiple sclerosis and rheumatoid arthritis).[16] They were recruited from self-referrals, tertiary care, and community clinics. For referrals originating outside our institution, diagnoses were confirmed by medical record review (C.B.). The study was approved by the University of Manitoba Health Research Ethics Board.

Participants completed several symptom scales of anxiety and depression and provided demographic information including sex, date of birth, ethnicity, and highest level of education, as delineated further below. Disease activity was also assessed. Participants were subsequently administered the SCID by trained personnel under the supervision of a clinical psychologist (J.R.W.) within 2 weeks of study enrollment.

Symptom Scales: Depression and Anxiety

Each participant completed the Patient Health Questionnaire (PHQ-9), from which we derived scores for the PHQ-9 and PHQ-2,[17,18] Hospital Anxiety and Depression Scale (HADS),[19] Kessler-6 Distress Scale,[20] Patient-Reported Outcomes Measurement Information System (PROMIS) Emotional Distress Depression Short-Form 8a (PROMIS Depression) and Anxiety Short-Form 8a (PROMIS Anxiety),[21] Generalized Anxiety Disorder 7-item Scale (GAD-7),[22] and the Overall Anxiety and Severity Impairment Scale (OASIS).[23] The HADS subscales of depression and anxiety have recommended cut-points of 8 to identify possible anxiety or depression[24] and 11 to identify probable anxiety or depression. All of the scales listed here were selected considering their brevity, prior use in IBD research, and open access for clinical use. See Table 1 for a summary of the characteristics of the scales. For all scales, higher scores indicated more pronounced symptoms.

Symptom Scales: Fatigue and Pain

Each participant completed the Fatigue Impact Scale for Daily Use (D-FIS), a validated instrument that includes 8 items scored ordinally from 0 to 4.[25] They also completed a pain scale, the MOS-Modified Pain Effects Scale, a valid and reliable instrument with scores ranging from 6 to 30.[26] Higher scores on both the D-FIS and the MOS-Modified Pain Effects Scale indicated more severe symptoms.

Disease Activity

Trained personnel assessed disease activity using validated clinical indices for IBD: the Powell-Tuck Activity Index for ulcerative colitis (UC)[27] and the Harvey-Bradshaw Disease Activity Index for Crohn's disease.[28] A score of ≥5 on either index indicates active disease.

Diagnostic Interview

The SCID is a semistructured clinical interview used to identify psychiatric disorders using DSM-IV criteria. For this

TABLE 1: Self-Administered Symptom Scales for Depression and Anxiety

Scale	Reporting Period	No. Items	Response Range/Item	Total Score[a]
PHQ-9	Last 2 wk	9	0–3	0–27
PHQ-2[22]	Last 2 wk	2	0–3	0–6
HADS	Past week (7 d)	Depression: 7	Depression: 0–3	Depression: 0–21
		Anxiety: 7	Anxiety: 0–3	Anxiety: 0–21
Kessler-6	Past 30 d	6 (5/6 for depression)	1–5	6–30
PROMIS Depression	Past week (7 d)	8	1–5	38.2–81.3 (transformed scores into T scores)
GAD-7	Past 2 wk	7	0–3	0–21
OASIS	Past wk (7 d)	5	0–4	0–20
PROMIS Anxiety	Past wk (7 d)	8	1–5	37.1–83.1 (transformed scores into T scores)

[a]For all measures used, higher scores indicate more severe symptoms.

study, we used SCID-based diagnoses of current major depression and generalized anxiety disorder as the reference standard in analyses of criterion validity. A clinical psychologist (J.R.W.) trained graduate students in clinical psychology, nurses, and research coordinators to conduct the interviews. The training process involved reviewing the SCID users' guide, a thorough review of the modules, watching video examples of interviews, observing an interview, and being monitored while conducting an interview. The interviewers met regularly to review their interviews and consulted the clinical psychologist periodically. Trained interviewers who were blinded to the results of the symptom scales administered the SCID. These SCID-based diagnoses served as the reference standard for analyses of criterion validity.

Analysis

We assessed criterion validity of the symptom scales and their construct validity (through hypothesis testing), content validity, internal consistency reliability, and test-retest reliability, as suggested by the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN).[29] For each symptom scale, we determined the median (interquartile range) and the percentage of participants scoring the minimum (floor) and maximum (ceiling) possible scores.

Criterion validity was evaluated by comparing depressive and anxiety disorder status based on the (i) SCID (criterion standard) and (ii) self-reported symptom scales. For each symptom scale, we computed sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) as compared with the criterion standard. In addition, we used receiver operating curve (ROC) analysis to identify the best cutpoint in this IBD sample for predicting the presence of current depressive or anxiety disorder by maximizing Youden's J index (sensitivity + specificity – 1),[30] in which sensitivity and specificity are balanced. We compared the area under the ROC

between symptom scales using binary logistic regression separately for depression and anxiety symptom scales.

Also, to assess construct validity, we used Spearman rank correlations (with 95% confidence intervals) to evaluate the associations between the depressive and anxiety symptom scales and measures of pain and fatigue, disease activity, and age. We expected the correlations with pain and fatigue to be positive and moderate as they represent different constructs assessed via similar methods (ie, all self-report symptom scales). We also expected correlations with disease activity to be positive. We anticipated that the relations with age would be weaker.[31]

We used Cronbach's alpha[32] to assess internal consistency for each symptom scale. We used an intraclass correlation coefficient (ICC) to assess test-retest reliability for a subgroup that completed the symptom scales 2 weeks apart; the SCID was not repeated. This provides information on the reproducibility of the scores over time for respondents who have not changed. The values of the ICC were classified as follows: <0.5 as poor, 0.5–0.75 as moderate, 0.75–0.90 as good, and >0.90 as excellent reliability.[33]

Sample Size

Assuming a lower bounds sensitivity = 0.75 and specificity ≥0.85, precision = 0.15, and α = 0.05, the required sample size was 247. For assessment of test-retest reliability, if the ICC is ≥0.6 (0.6 being lowest acceptable), precision is 0.1, and α = 0.05, then the required sample size was 158.

Statistical analyses used SAS V9.4 (SAS Institute Inc., Cary, NC, USA).

RESULTS

We enrolled 247 participants, of whom 133 (53.8%) were recruited from clinic, 98 (39.7%) were recruited from the community or an IBD registry, and the remainder were self-referred.[16] Of those enrolled, we included 242 participants in the

analysis who completed the SCID within 2 weeks of enrollment. Specifically, 114 (46.2%) of the SCIDs were completed the day of study enrollment, when the symptom scales were completed, and the remainder were completed within 4 days of enrollment. Demographic characteristics and SCID results are reported in Table 2. Based on the SCID, the frequency of major depression did not differ between participants with CD (9.5%) or UC (7.4%, $P = 0.59$), nor did the frequency of generalized anxiety disorder (CD 6.7% vs UC 4.3%, $P = 0.42$).

Rates of depression and anxiety based on the different symptom scales used are reported in Table 3, and median, interquartile range, and minimum and maximum values for each scale are shown in Supplementary Table 1. Using a threshold of 15% for important effects,[34] no scale had ceiling effects, but the PHQ2, HADS-D, PROMIS Depression, and the GAD-7 and OASIS had floor effects.

Criterion Validity

Performance of the depression symptom scales based on the published cut-points, relative to the identification of generalized anxiety and major depressive disorders as derived via the SCID, is shown in Table 4. The scale with the highest sensitivity was the PHQ-9, whereas the scale with the lowest sensitivity was the HADS-D at a cut-point of 11. The scales with the highest specificity (97%) were the HADS-D at a cut-point of 11 and the Kessler-6; specificity of the PHQ-2, HADS-D at a cut-point of 8, and the PROMIS Depression were slightly lower, ranging from 88% to 90%.

Based on the ROC analysis, the optimal cut-points for some of the depression symptom scales differed from those routinely recommended, when balancing sensitivity and specificity (Table 5). The area under the ROC curve (AUC) did not differ among the HADS-D (AUC, 0.91; 95% confidence interval [CI], 0.86–0.95), PHQ-2 (AUC, 0.88; 95% CI, 0.80–0.97; $P = 0.36$), PHQ-9 (AUC, 0.93; 95% CI, 0.88–0.97; $P = 0.41$), PROMIS Depression (AUC, 0.90; 95% CI, 0.84–0.95; $P = 0.66$), and the Kessler-6 (AUC, 0.89; 95% CI, 0.83–0.95; $P = 0.55$) (Fig. 1).

Performance of the anxiety symptom scales based on the typically recommended cut-points is shown in Table 4. The symptom scale with the highest sensitivity was the PROMIS Anxiety, followed closely by the HADS-A at a cut-point of 8. The HADS-A with a cut-point of 11 and the GAD-7 had the lowest sensitivities. Based on the ROC analysis, the optimal cut-points for some of the anxiety screening measures differed from those routinely recommended (Table 5). The AUC did not differ among the HADS-A (AUC, 0.82; 95% CI, 0.72–0.92), GAD-7 (AUC, 0.87; 95% CI, 0.80–0.93; $P = 0.27$, PROMIS Anxiety (AUC, 0.86; 95% CI, 0.77–0.95; $P = 0.44$), and OASIS (AUC, 0.86; 95% CI, 0.78–0.94; $P = 0.52$) (Fig. 2).

Construct Validity

All of the symptom scales for depression and anxiety were moderately and positively associated with pain and fatigue and positively but more modestly associated with disease activity (Table 6). As expected, age was not correlated with these tools.

Reliability

All depression symptom scales and all anxiety symptom scales had acceptable internal consistency reliability, as measured by Cronbach's alpha (Table 7). Of the depression symptom scales, the PROMIS Depression tool had the highest internal consistency, whereas the HADS-D had the lowest. Of the anxiety scales, the PROMIS Anxiety had the highest internal consistency, whereas the HADS-A had the lowest.

TABLE 2: Demographic Characteristics of Participants

Characteristic	Value
Mean age (SD) at enrollment, y	47.5 (14.8)
Female, %	63
Race: white, %	85
Education, %	
Did not complete high school	3.3
Completed high school/GED	27.7
Some postsecondary college	20.7
Technical/trades	12.4
Completed postsecondary degree or higher	11.1
Crohn's disease	61.2
Ulcerative colitis	38.8
Disease Activity Index, mean (SD)	4.1 (3.7)
Active Disease (index ≥5), No. (%)	103 (41.7)

TABLE 3: Frequency (%) of Participants With Depression or Anxiety According to Each Scale Used, No. (%)

Scale	No. (%)
Depression	
SCID depressive disorder	21 (8.7)
PHQ-2	44 (18.6)
PHQ-9	56 (23.8)
HADS-D[8]	42 (17.4)
HADS-D[11]	15 (6.2)
PROMIS Depression	37 (15.3)
Kessler-6	16 (6.6)
Anxiety	
SCID anxiety disorder	43 (17.8)
SCID generalized anxiety disorder	14 (5.8)
GAD-7	36 (15.1)
OASIS	51 (21.1)
HADS-A[8]	89 (36.9)
HADS-A[11]	40 (16.6)
PROMIS Anxiety	45 (18.6)

TABLE 4: Test Characteristics for Previously Defined Cut-Points for Depression and Anxiety Symptom Scales

Instrument	Cut-Point ≥	Sens (95% CI)	Spec (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Depression						
PHQ-2	3	0.81 (0.58–0.95)	0.88 (0.82–0.92)	0.39 (0.24–0.55)	0.98 (0.95–0.99)	0.87 (0.82–0.91)
PHQ-9	10	0.95(0.76–0.99)	0.83 (0.77–0.88)	0.36 (0.23–0.50)	0.99 (0.97–1.0)	0.84 (0.79–0.89)
HADS-D	8	0.71 (0.48–0.89)	0.88 (0.83–0.92)	0.36 (0.22–0.52)	0.97 (0.94–0.99)	0.86 (0.81–0.90)
HADS-D	11	0.38 (0.18–0.62)	0.97 (0.94–0.99)	0.53 (0.27–0.79)	0.94 (0.90–0.97)	0.92 (0.88–0.95)
PROMIS Depression	T score 60	0.67 (0.43–0.85)	0.90 (0.85–0.93)	0.38 (0.22–0.55)	0.97 (0.93–0.99)	0.88 (0.83–0.91)
Kessler-6	19	0. 43 (0.22–0.66)	0.97 (0.94–0.99)	0.56 (0.30–0.80)	0.95 (0.91–0.97)	0.92 (0.88–0.95)
Anxiety						
GAD-7	10	0.64 (0.35–0.87)	0.88 (0.83–0.92)	0.25 (0.12–0.42)	0.98 (0.94–0.99)	0.87 (0.82–0.91)
OASIS	8	0.71 (0.42–0.92)	0.82 (0.76–0.87)	0.20 (0.10–0.33)	0.98 (0.95–0.99)	0.81 (0.76–0.86)
HADS-A	8	0.77 (0.46–0.95)	0.65 (0.59–0.72)	0.11 (0.05–0.20)	0.98 (0.94–1.00)	0.66 (0.60–0.72)
HADS-A	11	0.61 (0.31–0.86)	0.86 (0.81–0.90)	0.20 (0.09–0.36)	0.98 (0.94–0.99)	0.85 (0.79–0.89)
PROMIS Anxiety	T score 60	0.79 (0.49–0.95)	0.85 (0.80–0.89)	0.24 (0.13–0.40)	0.98 (0.96–1.0)	0.85 (0.80–0.89)

Test-retest reliability for the depression scales, as measured by an intraclass correlation coefficient, was good, ranging from 0.83 (PROMIS Depression) to 0.90 (HADS-D) (Table 7). On average, test-retest reliability for the anxiety scales was lower but considered good; values ranged from 0.79 (PROMIS Anxiety short) to 0.83 (HADS-A).

A comparison of the scales in terms of individual advantages and disadvantages is presented in Table 8.

DISCUSSION

We comprehensively evaluated the validity and reliability of multiple scales measuring symptoms of depression and anxiety in the IBD population. As compared with standardized clinical interviews, using self-administered symptom scales to identify depression and anxiety has several advantages, including that they are less time-consuming, which is important in a busy clinical practice, do not require trained personnel, can measure psychological issues in a severity continuum, and allow standardized scoring.[35] The scales might also be useful for tracking progress during the treatment or monitoring of depression or anxiety. Although the anxiety scales did not perform as well as the depression scales with respect to criterion validity, their performance was adequate. Within the domains of depression and anxiety, we did not find substantive differences in the individual psychometric properties of the different symptom scales used. The ROC analysis suggested criterion validity was adequate, although the optimal cut-points seem to differ slightly in the IBD population and the general population when the goal is balancing sensitivity and specificity. This should be kept in mind, for instance, if a practitioner chooses to use the HADS, where a cut-point of 11 has a lower sensitivity than other measures or than a cut-point of 8. The lower bounds estimates for test-retest reliability for the OASIS and GAD-7 were only 0.64–0.68. All of the GAD-7, OASIS, PHQ2,

TABLE 5: Test Characteristics for Optimal Cut-Points for Depression and Anxiety Symptom Scales

| Instrument	Cut-Point ≥	Sens (95% CI)	Spec (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)
Depression						
PHQ-9	11	0.86 (0.64–0.97)	0.87 (0.82–0.91)	0.39 (0.25–0.55)	0.98 (0.95–0.99)	0.87 (0.82–0.91)
HADS-D	7	0.81 (0.58–0.96)	0.83 (0.77–0.86)	0.31 (0.19–0.45)	0.99 (0.95–0.99)	0.83 (0.77–0.99)
PROMIS Depression	57.7 (T score)	0.76 (0.52–0.92)	0.81 (0.76–0.87)	0.29 (0.17–0.42)	0.97 (0.94–0.99)	0.81 (0.76–0.86)
Kessler-6	13	0.81 (0.58–0.94)	0.78 (0.72–0.83)	0.26 (0.16–0.38)	0.98 (0.94–0.99)	0.78 (0.73–0.83)
Anxiety						
GAD-7	8	0.71 (0.42–0.92)	0.80 (0.75–0.85)	0.18 (0.09–0.31)	0.98 (0.95–0.99)	0.79 (0.74–0.85)
HADS-A	9	0.77 (0.46–0.95)	0.76 (0.70–0.82)	0.16 (0.08–0.27)	0.98 (0.95–1.0)	0.76 (0.70–0.82)
PROMIS Anxiety	59.4 (T score)	0.86 (0.57–0.98)	0.82 (0.77–0.87)	0.23 (0.12–0.37)	0.99 (0.96–1.0)	0.83 (0.77–0.87)
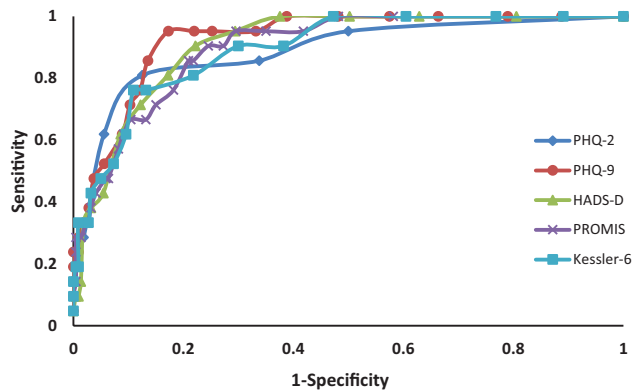
FIGURE 1. Receiver operating characteristic curves for depression screening measures as compared with the SCID depression.

FIGURE 2. Receiver operating characteristic curves for anxiety screening measures as compared with the SCID anxiety.

HADS-D, and PROMIS Depression had floor effects whereas there were no ceiling effects of any of the scales. Given the similarity of performance across scales, other factors such as time to completion and acceptability to patients may be considered in scale selection.

Studies assessing the merits of employing self-report symptom scales for anxiety and depression have mostly been conducted in primary care,[36] and not in persons with a chronic disease such as IBD, where the risk of depression and anxiety is particularly high and the yield may be higher. Yet, accurate identification rates for mental disorders in primary care can be less than 50%,[37, 38] and depression screening in primary care is estimated to be only 3.4%.[39] In a meta-analytic assessment of diagnosis of depression in primary care when assisted with symptom scales, the sensitivity for diagnosing depression by primary care providers was only 50.1%, and specificity was 80.1%.[40] In another meta-analytic assessment of the identification of anxiety disorders in primary care,[41] the sensitivity for diagnosing anxiety disorders was higher when diagnoses were assisted by the use of symptom scales (63.6%; 95% CI, 50.3%–75.1%) than when unassisted (30.5%; 95% CI, 20.7%–42.5%), with comparable specificity (87.9%; 95% CI, 81.3%–92.4%; vs 91.4%; 95% CI, 86.6–94.6%; respectively). All of this suggests that screening for depression and anxiety in primary care is suboptimal. Its occurrence and effectiveness in IBD specialty clinics has yet to be reported. However, symptom scales can potentially improve detection of depression and anxiety, even though some affected individuals will still be missed. Future studies will be required to determine to what extent consistent use of symptom scales in the IBD population will improve detection of depression and anxiety.

In terms of the comparability of screening tools in primary care, a recent systematic review concluded that all tools were comparable, although there was much more research published assessing the PHQ measures than other measures.[42] The PHQ-9 has the advantage of including a question on thoughts of death, which would then require active clinical follow-up if endorsed, but is important for clinical management. One
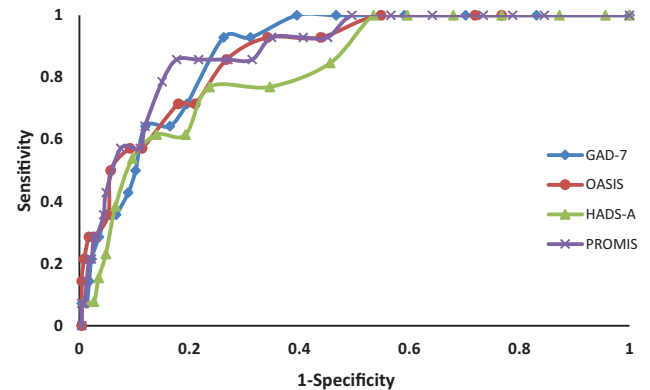
group, in their systematic review of depression screening tools in primary care, argued that because the PHQ-9 psychometrics has been studied much more often than other measures, it has greater credibility as a screening tool of choice.[43] The HADS has the advantage of evaluating both depression and anxiety using the same metric. The advantages of the PROMIS scales include development using contemporary psychometric standards (including item response theory), the availability of equivalent forms of varying lengths (4, 6, 8 items), normative data from a large community sample in the United States, and an item pool that may be used with computerized adaptive testing.[21] Cross-walks are being developed to relate the PROMIS scales to previously developed measures.

One would not expect complete correspondence between depression and anxiety scales and clinical interview for diagnostic evaluation. Symptoms of anxiety and depression are common even if they do not reach threshold criteria for a disorder, that is, are subsyndromal, and can present in the context of other mental disorders such as adjustment disorders. Some persons with depression or anxiety disorders experience periods of partial remission where their symptoms are at lower levels. An advantage of the symptom scales we examined is that they can identify persons experiencing high levels of emotional distress regardless of the reason. Many of these individuals may benefit from additional support even if they do not meet the full criteria for a psychiatric disorder, potentially facilitating earlier intervention.[37] Thus the clinician must pursue further evaluation to understand the reasons for elevated symptom scale scores. Moreover, the scales are not 100% sensitive; therefore, the use of any such scales does not obviate the need for clinical vigilance.

Our study had the strengths of a large sample size spanning clinic and community IBD participants with a well-established reference criterion and concurrent evaluation of an array of symptom scales. It is very important to study these scales specifically in an IBD setting and not simply exrapolate from their utility in primary care. For instance, in a postnatal setting, the Edinburgh Postnatal Depression Scale had a higher accuracy

TABLE 6: Construct Validity: Spearman Correlations (95% CIs) of Anxiety and Depression Symptom Scales With Pain, Fatigue, and Age (N = 227)

Measure	Pain (95% CI)	Fatigue (95% CI)	Disease Activity (95% CI)	Age (95% CI)
Depression				
PHQ-2	0.62 (0.53–0.69)	0.53 (0.43–0.62)	0.39 (0.27–0.49)	–0.004 (–0.13 to 0.13)
PHQ-9	0.70 (0.63–0.76)	0.68 (0.60–0.74)	0.47 (0.36–0.56)	–0.06 (–0.19 to 0.067)
PROMIS Depression Short Form-8a	0.65 (0.57–0.72)	0.60 (0.51–0.68)	0.40 (0.29–0.50)	0.031 (–0.10 to 0.16)
HADS-D	0.63 (0.55–0.70)	0.63 (0.55–0.71)	0.39 (0.28–0.50)	0.024 (–0.0066 to 0.25)
Kessler-6	0.64 (0.56–0.71)	0.65 (0.57–0.72)	0.35 (0.23–0.46)	–0.12 (–0.37 to –0.13)
Anxiety				
OASIS	0.55 (0.46–0.64)	0.56 (0.46–0.64)	0.29 (0.17–0.40)	–0.052 (–0.18 to 0.079)
GAD-7	0.64 (0.56–0.71)	0.64 (0.56–0.71)	0.39 (0.27–0.49)	–0.11 (–0.23 to 0.025)
PROMIS Anxiety Short Form-8a	0.62 (0.53–0.69)	0.62 (0.53–0.69)	0.38 (0.27–0.48)	–0.035 (–0.16 to 0.096)
HADS-A	0.57 (0.47–0.65)	0.57 (0.48–0.65)	0.29 (0.17–0.40)	–0.087 (–0.21 to 0.044)

than generic screening tools.[44] However, in a meta-analysis of tools to diagnosis depression in persons with epilepsy, an epilepsy-specific tool, the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E), did not outperform the other general symptom scales for primary depression, including the HADS, PHQ-2, PHQ-9; they all performed reasonably well.[45] Our finding of different optimal cut-points among several of these commonly used scales in an IBD population (Table 5) suggests that studies like ours testing generic scales in a specific disease milieu is necessary. This was also shown in a stroke population where different cut-points on 4 common

self-screening mental health scales were identified.[46] Our study was limited by not having assessed all symptom scales, such as the Beck Depression Inventory or a brief disease-specific interview, the Luebeck Interview for Psychosocial Screening for IBD (LIPS-IBD).[47] However, we studied a broad array of symptom scales that are in the public domain (other than the HADS), which provides physicians with several choices for mental health symptom scales. Having drawn our IBD cohort broadly from community and referral practices suggests that these mental health symptom scales can be reliably used in varied IBD practices. We assessed test-retest reliability over a 2-week period; it is possible that symptoms may have changed over this interval, thus underestimating the performance of the scales evaluated. Finally, the rate of current depression (8.7%) as measured by the SCID, the study gold standard, was relatively low compared with what is expected from other studies in IBD, but this may reflect assessing a broad array of persons with IBD drawn from the community as opposed to what might be expected from a cohort of persons with IBD presenting to a gastroenterologist's clinic.

    There are many reasons why primary care providers and medical specialists may underdiagnose mental health disorders. These include the greater focus on physical symptoms or other disease presentation during an office encounter, the limited time available to explore mental health issues, the potential reluctance of patients to discuss mental health issues, lack of training, and the limited availability of resources to manage the uncovered mental health disorder (reviewed in 36). When left undiagnosed and subsequently untreated, both depression and anxiety disorders can be associated with suicidal behavior.[48] Persons with undetected anxiety disorders may undergo unnecessary, and potentially invasive, diagnostic investigations (eg, coronary angiographies) and increased emergency department

TABLE 7: Reliability of Anxiety and Depression Measures

Instrument	Internal Consistency Reliability Cronbach's Alpha (95% CI)	Test-Retest Reliability Intraclass Correlation Coefficient (95% CI)
Depression		
PHQ-2	0.82 (0.63–1.0)	0.86 (0.81–0.90)
PHQ-9	0.89 (0.82–0.95)	0.85 (0.80–0.89)
PROMIS Depression Short Form-8a	0.96 (0.89–1.0)	0.85 (0.80–0.89)
HADS-D	0.84 (0.76–0.92)	0.83 (0.77–0.87)
Kessler-6	0.87 (0.78–0.95)	0.87 (0.82–0.90)
Anxiety		
OASIS	0.91 (0.82–1.0)	0.73 (0.64–0.80)
GAD-7	0.91 (0.84–0.99)	0.76 (0.68–0.82)
PROMIS Anxiety Short Form-8a	0.94 (0.87–1.0)	0.79 (0.72–0.84)
HADS-A	0.87 (0.79–0.95)	0.83 (0.77–0.87)

TABLE 8: Advantages and Disadvantages of Depression and Anxiety Scales

Scale	Advantages	Disadvantages
PHQ-9	Brief; includes the key content areas for the formal diagnostic criteria Includes a question on suicidality supporting concurrent screening for depression and suicide risk	Some overlap with symptoms of inflammatory disorders (lacking energy, sleep, and appetite difficulties)
PHQ-2	Briefest of all depression measures	Substantial floor effects
HADS	Brief Evaluates depression and anxiety concurrently Designed for use in medically ill populations, so it minimizes overlapping physical symptoms	HADS-D and HADS-A had lowest sensitivity relative to other measures
Kessler-6	Brief Sensitive to any form of psychologic distress	Does not distinguish between depression or anxiety, which may be relevant to treatment
PROMIS Depression	Brief; 4- and 6-item versions also available Normative data from community samples available; other languages available Can be scored even with missing responses for some items	Floor effect Response pattern scoring for research purposes can be time-consuming, use of simple scale totals more feasible for clinical setting
GAD-7	Brief Focuses on a common anxiety disorder, generalized anxiety disorder	May not capture anxiety related to anxiety disorders other than generalized anxiety disorder Possibly only moderate test-retest reliability
OASIS	Briefest anxiety measure Captures severity of any anxiety disorder	Floor effect Response options are complex Possibly only moderate test-retest reliability
PROMIS Anxiety	Brief; 4- and 6-item versions also available Normative data from community samples available; other languages available Can be scored even with missing responses for some items	Response pattern scoring for research purposes can be time-consuming, use of simple scale totals more feasible for clinical setting

visits, all of which drive up the cost of care, and unnecessary referrals to specialists (eg, gastroenterologists, cardiologists) in an attempt to identify the cause of distressing physical symptoms that are often related to anxiety.[49, 50] Screening alone, however, does not improve health outcomes.[14] After deciding on an appropriate screening tool and implementing it in practice, the critical next step is ensuring that appropriate action is taken to improve mental health and reduce the impact of mental health disorders on quality of life and health in general.

More controlled studies of pharmacological and psychological interventions for depression and anxiety in persons with IBD are needed. A systematic review found only 1 controlled drug intervention study (using lorazepam) for anxiety and no studies for depression in adults.[51] In a recent randomized controlled trial in adolescents, both cognitive behavioral therapy and supportive nondirective therapy significantly reduced depressive symptoms and improved global functioning, quality of life, and disease activity in adolescents with depression and IBD.[52] In the absence of robust data on treating depression and anxiety in persons with IBD, treatments are extrapolated from what is known about treating these diseases without concomitant chronic immune-mediated diseases. However, as the effect of antidepressant drugs may be attenuated when anti-inflammatory drugs are concurrently used, more studies on treating depression and anxiety specifically in persons with IBD are warranted.[53]

Clinicians assessing patients with IBD should feel comfortable incorporating a symptom scale into their practice, with the proviso that when scores exceed identified cut-points for this population, there is appropriate follow-up to clarify the context and manage the symptoms. Primary care practitioners and gastroenterologists caring for patients with IBD may be able to enhance their detection of depression and anxiety by implementing symptom scales. Table 8 includes a summary of all scales including specific advantages for each scale that may provide a rationale for a practitioner to choose 1 scale over the others.

SUPPLEMENTARY DATA

Supplementary data are available at *Inflammatory Bowel Diseases* online.

REFERENCES

1. Singh S, Blanchard A, Walker JR, et al. Common symptoms and stressors among individuals with inflammatory bowel diseases. *Clin Gastroenterol Hepatol.* 2011;9:769–75.
2. Walker JR, Ediger JP, Graff LA, et al. The Manitoba IBD Cohort Study: a population-based study of the prevalence of lifetime and 12-month anxiety and mood disorders. *Am J Gastroenterol.* 2008;103:1989–97.
3. Marrie RA, Walld R, Bolton JM, et al; CIHR Team in Defining the Burden and Managing the Effects of Psychiatric Comorbidity in Chronic Immunoinflammatory Disease. Increased incidence of psychiatric disorders in immune-mediated inflammatory disease. *J Psychosom Res.* 2017;101:17–23.
4. Bernstein CN, Hitchon CA, Walld R, et al. Increased burden of psychiatric disorders in inflammatory bowel disease. *Gastroenterology.* 2018; 152: S973.

5. Ananthakrishnan AN, Gainer VS, Cai T, et al. Similar risk of depression and anxiety following surgery or hospitalization for Crohn's disease and ulcerative colitis. *Am J Gastroenterol*. 2013;108:594–601.
6. Rapaport MH, Clary C, Fayyad R, Endicott J. Quality-of-life impairment in depressive and anxiety disorders. *Am J Psychiatry*. 2005;162:1171–8.
7. Stein MB, Roy-Byrne PP, Craske MG, et al. Functional impact and health utility of anxiety disorders in primary care outpatients. *Med Care*. 2005;43:1164–70.
8. Guthrie E, Jackson J, Shaffer J, et al. Psychological disorder and severity of inflammatory bowel disease predict health-related quality of life in ulcerative colitis and Crohn's disease. *Am J Gastroenterol*. 2002;97:1994–9.
9. Kiebles JL, Doerfler B, Keefer L. Preliminary evidence supporting a framework of psychological adjustment to inflammatory bowel disease. *Inflamm Bowel Dis*. 2010;16:1685–95.
10. Click B, Ramos Rivers C, Koutroubakis IE, et al. Demographic and clinical predictors of high healthcare use in patients with inflammatory bowel disease. *Inflamm Bowel Dis*. 2016;22:1442–9.
11. Limsrivilai J, Stidham RW, Govani SM, et al. Factors that predict high health care utilization and costs for patients with inflammatory bowel diseases. *Clin Gastroenterol Hepatol*. 2017;15:385–92.e2.
12. Kiebles JL, Doerfler B, Keefer L. Preliminary evidence supporting a framework of psychological adjustment to inflammatory bowel disease. *Inflamm Bowel Dis*. 2010;16:1685–95.
13. Buist-Bouwman MA, de Graaf R, Vollebergh WA, Ormel J. Comorbidity of physical and mental disorders and the effect on work-loss days. *Acta Psychiatr Scand*. 2005;111:436–43.
14. O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Ann Intern Med*. 2009;151:793–803.
15. Siu AL, Bibbins-Domingo K, Grossman DC, et al; US Preventive Services Task Force (USPSTF). Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA*. 2016;315:380–7.
16. Marrie RA, Graff LA, Walker JR, et al. A prospective study of the effects of psychiatric comorbidity in immune-mediated inflammatory disease: rationale, protocol and participation. *JMIR Res Protoc*. In press.
17. Spitzer R, Kroenke K, Williams JW; Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA*. 1999;282:1737–44.
18. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;41:1284–92.
19. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand*. 1983;67:361–70.
20. Cairney J, Veldhuizen S, Wade TJ, et al. Evaluation of 2 measures of psychological distress as screeners for depression in the general population. *Can J Psychiatry*. 2007;52:111–20.
21. Pilkonis PA, Choi SW, Reise SP, et al; PROMIS Cooperative Group. Item banks for measuring emotional distress from the patient-reported outcomes measurement information system (PROMIS®): depression, anxiety, and anger. *Assessment*. 2011;18:263–83.
22. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166:1092–7.
23. Norman SB, Cissell SH, Means-Christensen AJ, Stein MB. Development and validation of an Overall Anxiety Severity and Impairment Scale (OASIS). *Depress Anxiety*. 2006;23:245–249.
24. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res*. 2002;52:69–77.
25. Fisk JD, Doble SE. Construction and validation of a fatigue impact scale for daily administration (D-FIS). *Qual Life Res*. 2002;11:263–72.
26. Ritvo PG, Fischer JS, Miller DM, et al. *Multiple Sclerosis Quality of Life Inventory: Technical Supplement*. New York: National Multiple Sclerosis Society; 1997.
27. Powell-Tuck J, Bown RL, Lennard-Jones JE. A comparison of oral prednisolone given as single or multiple daily doses for active proctocolitis. *Scand J Gastroenterol*. 1978;13:833–7.
28. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet*. 1980;1:514.
29. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539–49.
30. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–5.
31. Reynolds K, Pietrzak RH, El-Gabalawy R, et al. Prevalence of psychiatric disorders in U.S. older adults: findings from a nationally representative survey. *World Psychiatry*. 2015;14:74–81.
32. Bland JM, Altman DG. Cronbach's alpha. *BMJ*. 1997;314:572.
33. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–63.
34. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
35. Zimmerman M, Galione J. Psychiatrists' and nonpsychiatrist physicians' reported use of the DSM-IV criteria for major depressive disorder. *J Clin Psychiatry*. 2010;71:235–8.
36. Mitchell AJ. Clinical utility of screening for clinical depression and bipolar disorder. *Curr Opin Psychiatry*. 2012;25:24–31.
37. Sareen J, Stein MB, Campbell DW, et al. The relation between perceived need for mental health treatment, DSM diagnosis, and quality of life: a Canadian population-based survey. *Can J Psychiatry*. 2005;50:87–94.
38. Wittchen HU, Mühlig S, Beesdo K. Mental disorders in primary care. *Dialogues Clin Neurosci*. 2003;5:115–28.
39. Weiller E, Bisserbe JC, Maier W, et al. Prevalence and recognition of anxiety syndromes in five European primary care settings. A report from the WHO study on psychological problems in general health care. *Br J Psychiatry Suppl*. 1998;34:18–23.
40. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet*. 2009;374:609–19.
41. Olariu E, Forero CG, Castro-Rodriguez JI, et al. Detection of anxiety disorders in primary care: a meta-analysis of assisted and unassisted diagnoses. *Depress Anxiety*. 2015;32:471–84.
42. Mulvaney Day N, Marshall T, Downey Piscopo K, et al. Screening for behavioral health conditions in primary care settings: a systematic review of the literature. *J Gen Intern Med*. 2018;33(3):335–46.
43. El Den S, Chen TF, Gan YL, et al. The psychometric properties of depression screening tools in primary healthcare settings: a systematic review. *J Affect Dis*. 2017;225:503–22.
44. Chorwe-Sungani G, Chipps J. A systematic review of screening instruments for depression for use in antenatal services in low resource settings. *BMC Psych*. 2017;17:112.
45. Gill SJ, Lukmanji S, Fiest KM, et al. Depression screening tools in persons with epilepsy: a systematic review of validated tools. *Epilepsia*. 2017;58:695–705.
46. Prisnie JC, Fiest KM, Coutts SB, et al. Validating screening tools for depression in stroke and transient ischemic attack patients. *Int J Psychiatry Med*. 2016;51:262–77.
47. Kunzendorf S, Jantschek G, Straubinger K, et al. The luebeck interview for psychosocial screening in patients with inflammatory bowel disease. *Inflamm Bowel Dis*. 2007;13:33–41.
48. Sareen J, Cox BJ, Afifi TO, et al. Anxiety disorders and risk for suicidal ideation and suicide attempts: a population-based longitudinal study of adults. *Arch Gen Psychiatry*. 2005;62:1249–57.
49. Fleet RP, Beitman BD. Unexplained chest pain: when is it panic disorder? *Clin Cardiol*. 1997;20:187–94.
50. Logue MB, Thomas AM, Barbee JG, et al. Generalized anxiety disorder patients seek evaluation for cardiological symptoms at the same frequency as patients with panic disorder. *J Psychiatr Res*. 1993;27:55–9.
51. Fiest KM, Bernstein CN, Walker JR, et al; CIHR Team. Defining the burden and managing the effects of psychiatric comorbidity in chronic immunoinflammatory disease. Systematic review of interventions for depression and anxiety in persons with inflammatory bowel disease. *BMC Res Notes*. 2016;9:404.
52. Szigethy E, Bujoreanu SI, Youk AO, et al. Randomized efficacy trial of two psychotherapies for depression in youth with inflammatory bowel disease. *J Am Acad Child Adolesc Psychiatry*. 2014;53:726–35.
53. Warner-Schmidt JL, Vanover KE, Chen EY, et al. Antidepressant effects of selective serotonin reuptake inhibitors (SSRIS) are attenuated by antiinflammatory drugs in mice and humans. *Proc Natl Acad Sci U S A*. 2011;108:9262–7.