ORIGINAL ARTICLE

Neurogastroenterology & Motility N G M WILEY

Examining the optimal cutoff values of HADS, PHQ-9 and GAD-7 as screening instruments for depression and anxiety in irritable bowel syndrome

Johanna T. W. Snijkers[1,2] | Wendy van den Oever[1,2] | Zsa Zsa R. M. Weerts[1,2] | Lisa Vork[1,2] | Zlatan Mujagic[1,2] | Carsten Leue[3,4] | Martine A. M. Hesselink[1,2] | Joanna W. Kruimel[1,2] | Jean W. M. Muris[5] | Roel M. M. Bogie[1,6] | Ad A. M. Masclee[1,2] | Daisy M. A. E. Jonkers[1,2] | Daniel Keszthelyi[1,2]

[1]Division of Gastroenterology-Hepatology, Department of Internal Medicine, Maastricht University Medical Center+, Maastricht, The Netherlands

[2]NUTRIM, School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, The Netherlands

[3]Department of Psychiatry and Psychology, Maastricht University Medical Center+, Maastricht, The Netherlands

[4]MHeNS, School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands

[5]Department of Family Medicine, CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, The Netherlands

[6]GROW, School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands

Correspondence
Johanna Theodora Wilhelmina Snijkers, Department of Internal Medicine, Division of Gastroenterology-Hepatology, Maastricht University, NUTRIM, PO Box 5800, 6202 AZ Maastricht, The Netherlands.
Email: j.snijkers@maastrichtuniversity.nl

Abstract

Background: Self-rating scales are frequently used to screen for anxiety and depression in patients with irritable bowel syndrome (IBS). Different cutoff values are recommended in literature, and guidelines have suggested the use of other screening instruments over time. The aim of this study was to assess the correlation between the most commonly used psychological screening instruments for anxiety and depression in IBS and to compare custom cutoff scores for these instruments.

Methods: Irritable bowel syndrome patients (*n* = 192) completed several questionnaires including the Hospital Anxiety and Depression Scale (HADS, HADS-A and HADS-D subscale), Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7). Agreement at different cutoff points, for depressive and anxiety disorder, was assessed by use of the Gwet AC1 coefficient.

Key Results: Hospital Anxiety and Depression Scale (HADS)-D and PHQ-9 scores, and HADS-A and GAD-7 scores showed high correlations (r_s = 0.735 and r_s = 0.805, respectively). For depressive disorder, a Gwet AC1 value of 0.829 was found when recommended cutoff points from literature were compared (PHQ-9 cutoff ≥10, HADS-D cutoff ≥8). For anxiety disorder, a Gwet AC1 value of 0.806 was found when recommended cutoff points from literature were compared (GAD-7 cutoff ≥10, HADS-A cutoff ≥8). Even higher agreements were found when higher HADS cutoff values were chosen, with impact on sensitivity and specificity.

Conclusions & Inferences: Custom cutoff values deem the HADS subscales (HADS-D and HADS-A) concordant to PHQ-9 and GAD-7 scores. The choice of a cutoff value has substantial impact on sensitivity/specificity and is dependent on patient population, setting, and the purpose of use.

Key points

- There is a very good agreement between recommended cutoffs of self-ratings scales used to screen for anxiety and depression in patients with Irritable Bowel Syndrome (IBS): Hospital Anxiety and Depression Scale-subscale Depression and HADS-subscale Anxiety cutoff ≥8, Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7 cutoff ≥10.
- The choice of a cutoff value has substantial impact on sensitivity/specificity and should be dependent on patient population, setting, and the purpose of use.

1 | INTRODUCTION

Irritable bowel syndrome (IBS) is a disorder involving the gut-brain interaction, characterized by recurrent chronic abdominal pain and altered bowel habits. Its prevalence is 5–15% in the Western population.[1,2] The pathophysiology is incompletely understood, but its underlying mechanism is multifactorial and includes brain-gut dysfunction, visceral hypersensitivity, microbial imbalance, genetic susceptibility, immune activation, and increased intestinal permeability.[3,4]

Psychological comorbidities are often reported in patients with IBS, specifically anxiety or depressive disorders. Psychological co-morbidity is associated with increased symptom severity and persistence, health-seeking behavior, and therapy resistance.[5,6] IBS patients with a comorbid anxiety or depressive disorder may benefit from a different therapeutic approach than those without any psychological comorbidity.[7] It is therefore important to screen for the presence of psychological comorbidities in daily clinical practice, to gain insight into factors that contribute to symptoms.

Self-rating scales are frequently used to screen for anxiety and depression in clinical practice. Validated tools include the Hospital Anxiety and Depression Scale (HADS),[8] Patient Health Questionnaire-9 (PHQ-9),[9] and the Generalized Anxiety Disorder-7 assessment (GAD-7).[10] Historically in IBS, the HADS was most commonly used to screen for both anxiety (HADS-A subscale) and depression (HADS-D subscale). However, recent guidelines have suggested the use of GAD-7 (for anxiety) and PHQ-9 (for depression).[11] The GAD-7 and PHQ-9 both measure a heterogeneous spectrum of symptoms in anxiety and depressive disorders, respectively. In contrast, the HADS focuses more on the emotional aspects of anxiety and depression and does not contain items that measure somatic symptoms.[12]

With regard to depression, studies comparing HADS and PHQ-9 draw incongruent conclusions about their diagnostic accuracy.[13–15] As for anxiety, HADS and GAD-7 both appeared adequate and of comparable diagnostic accuracy, but their diagnostic performance largely depended on the cutoff chosen.[16–18] Importantly, optimal cutoff values can differ for patient populations and setting.[19] Limited number of studies have directly examined the performance of these psychological self-rating scales for agreement with regard to cutoff points. Only a single study[20] performed in Swedish primary care and psychiatric outpatients, examined how a modified cutoff score influences the diagnostic performance of different screening tools. They found a low agreement between HADS and PHQ-9 when using the officially recommended cutoffs. This implies that recommended cutoffs do not necessarily apply universally in different settings as this might diminish their performance in detecting psychopathology.

To our knowledge, the performance of psychological self-rating scales has never been formally evaluated in IBS patients and no comparative research has been reported to ascertain the differences in performance of HADS, PHQ and GAD-7. The aim of this study therefore was given as:

I) to assess the correlation between the HADS-D and PHQ-9, and HADS-A and GAD-7; and II) to determine custom cutoff scores for the HADS-D vs. the PHQ-9 and HADS-A vs. the GAD-7 in IBS.

2 | METHODS

2.1 | Study population

Data were obtained via the Maastricht Irritable Bowel Syndrome (MIBS) cohort study. The MIBS cohort was designed to establish a large cohort of IBS patients and to identify etiological and patho-physiological factors, and different disease characteristics in sub-groups of IBS patients according to phenotypical and genotypical patterns. Data collection started prospectively in 2009. In 2016, the PHQ-9 and GAD-7 questionnaires were added in addition to the HADS questionnaire, as these became the general standard according to a European guideline for standardized phenotyping of IBS patients for research purposes.[11] Patients who had been included previously were then invited for a follow-up assessment including the old and new questionnaires. The study protocol of the MIBS cohort had been approved by the MUMC+Committee of Ethics (METC08-2–066). All study procedures were performed in compliance with Good Clinical Practice Guidelines and according to the revised Declaration of Helsinki. Written informed consent was given by each participant prior to participation.

Participants were recruited via the gastroenterology outpatient clinic of Maastricht University Medical Center+ (MUMC+, a secondary/tertiary referral hospital) and general practitioners in South-Limburg, the Netherlands. Patients with IBS diagnosed according to the Rome III criteria and aged between 18 and 75 years were included. A history of abdominal surgery was reason for exclusion, except for cases of uncomplicated appendectomy, cholecystectomy, or hysterectomy. (In- and exclusion criteria shown in Table S1). Patients participated in the follow-up assessment of the MIBS cohort between September 2016 and March 2017. A trained clinical investigator contacted all patients who did respond and confirmed the Rome III diagnostic criteria during a telephonic interview. Only those patients that completed both the HADS and the PHQ-9 and GAD-7 questionnaires were included in the current study.

2.2 | Data collection

Data were collected on demographics (eg, gender, age, and educational attainment), BMI (in kg/m^2), IBS subtype (Rome III criteria), treatment center (general practitioner or secondary/tertiary care), experienced GI symptoms (Gastrointestinal Symptom Rating Scale), Quality of Life (SF-36), and psychological comorbidities (HADS, PHQ-9, and GAD-7).

Information on demographics, BMI, and treatment center was collected via a self-report questionnaire. Educational attainment was based on the Dutch educational system and then converted to the International Standard Classification of Education (ISCED) prior to analysis. ISCED categories include lower education, intermediate education, and tertiary education. GI symptom score was determined using the Gastrointestinal Symptom Rating Scale (GSRS, scale 1–7), a fifteen-item questionnaire evaluating five GI symptom clusters including abdominal pain, bloating, diarrhea, satiety, and constipation.[21] Quality of life was based on the rand 36-item Short-Form Health Survey (SF-36, scale 1–6) which generates a physical and a mental health summary score.[22,23]

Additionally, all participants completed the HADS, PHQ-9, and GAD-7, to screen for a depressive and/or anxiety disorder. The HADS is a fourteen-item combined questionnaire assessing symptomatology of both anxiety (HADS-A subscale) and depression (HADS-D subscale).[12] Items were scored on a four-point scale with total scores per subscale ranging from zero to twenty-one. The recommended screening cutoff score of ≥8 was used as indication of anxiety or depression.[8] The PHQ-9 is a nine-item questionnaire for screening on the presence of depressive symptoms and monitoring depression severity.[9] Items were scored on a four-point scale with total scores ranging from zero to twenty-seven. Scores were defined as: ≥5 mild, ≥10 moderate, and ≥15 severe level of depression.[24] The recommended screening cutoff was ≥10, corresponding with at least a moderate level of depression.[11,19] The GAD-7 is a seven-item questionnaire for screening on the presence of generalized anxiety disorder and assessing its severity.[10]

Items were scored on a four-point scale with total scores ranging from zero to twenty-one. Scores were defined as: ≥5 mild, ≥10 moderate, and ≥15 severe anxiety. The recommended screening cutoff was ≥10, corresponding with at least a moderate level of anxiety.[11,25]

2.3 | Statistical analysis

Statistical analyses were performed using IBM SPSS statistics for Windows, version 26.0 (IBM Corp., Armonk, N.Y., USA) and R Statistical Software version 3.6.3.[26] Baseline characteristics were analyzed using an independent t-test for normally distributed continuous variables, Mann-Whitney U test for not normally distributed continuous variables, and Chi-square test for dichotomous variables. A two-sided p-value of ≤0.05 was considered statistically significant.

Spearman's rank correlation coefficient was used to determine the correlation between scores. Agreement at different cutoff points, for the depression and anxiety self-rating scales, was measured with the AC1 coefficient introduced by Gwet as paradox-resistant alternative agreement coefficient, as a low Cohen's Kappa value was found despite the high agreement between the questionnaires.[27-29] This low Cohen's kappa was caused by imbalanced data distribution across the study groups. Gwet's AC1 is less affected by prevalence and provides a more stable inter-rater reliability coefficient than Cohen's Kappa.[28] Sensitivity and specificity for the different HADS-D cutoff values were calculated when compared to the PHQ-9 cutoffs as standard, and for the different HADS-A cutoff values to the GAD-7 cutoffs as standard.

3 | RESULTS

3.1 | Study population

In total, 379 patients were invited to participate of whom 192 participants completed the HADS as well as the PHQ-9 and GAD-7 questionnaires (shown in Figure 1). Of these, 160 patients were reached by telephone to evaluate Rome III criteria for IBS. A total of 111 patients (69.4%) still met these criteria at follow-up. The group who still fulfilled the Rome III criteria did not differ significantly compared to the group who did not fulfilled the Rome III criteria at the follow-up measurement for all variables in the characteristics table, except the gastrointestinal symptom scores and physical quality of life (data not shown). The scores for abdominal pain, diarrhea syndrome, indigestion syndrome, and constipation syndrome were significantly higher, and the physical quality of life score was significantly lower for the patients who still fulfilled the Rome III criteria.

Participants' characteristics off all (previously confirmed) IBS patients (n = 192) are described for depressive and anxiety disorder determined by PHQ-9 and GAD-7 cutoffs ≥10 (shown in Table 1).

In total, 21.9% (n = 42, 69.0% female, median age 50 years) of the participants showed indications for a depressive disorder using a PHQ-9 score ≥10 and 14.1% (n = 27, 63.0% female, median age 48 years) showed indications for an anxiety disorder using a GAD-7 score ≥10. Of the patients having an indication for either a depressive or anxiety disorder, 38% (n = 19) had an overlapping depressive and anxiety disorder. Based on the patients who still fulfilled the Rome III criteria for IBS at follow-up (n = 111), the prevalence was 25.2% for depressive disorder based on PHQ-9 score ≥10 and 17.1% for anxiety disorder based on GAD-7 score ≥10.

3.2 | Depressive disorder

Hospital Anxiety and Depression Scale (HADS)-D and PHQ-9 scores showed a significant positive correlation (r_s = 0.735, p < 0.001). (shown in Figure S1).

Agreement for HADS-D and PHQ-9 were measured at different cutoff points for the presence of a depressive disorder (shown in Table 2). A Gwet AC1 value of 0.829 (CI 0.76–0.90) was found when recommended cutoff points from literature were compared (PHQ-9 cutoff ≥10, HADS-D cutoff ≥8). Corresponding sensitivity and specificity were 69.0% and 94.0%, respectively. Prevalence of depression at this cutoff was 19.8% when using HADS-D vs. 21.9% when using PHQ-9. Although higher agreements were found for higher HADS-D cutoffs, this gives rather low sensitivity values (<55%).

3.3 | Generalized anxiety disorder

Hospital Anxiety and Depression Scale (HADS)-A and GAD-7 scores showed a significant positive correlation (r_s = 0.805, p < 0.001). (shown in Figure S2).

Agreement at different cutoff points for HADS-A and GAD-7 for the presence of an anxiety disorder are presented in Table 3. A Gwet AC1 value of 0.806 (CI 0.73–0.88) was found when recommended cutoff points from literature were compared (GAD-7 cutoff ≥10, HADS-A cutoff ≥8). Corresponding sensitivity and specificity were 96.3% and 85.3%, respectively. Prevalence of generalized anxiety disorder at this cutoff was 26.3% when using HADS-A vs. 14.2% when using GAD-7. This underscores the substantial difference between the prevalences using the different screening tools. Higher agreement was found for HADS-A ≥9 (Gwet AC1 value 0.843, sensitivity 88.9%, and specificity 90.0%) and HADS-A ≥10 (Gwet AC1 value 0.886, sensitivity 77.8%, and specificity 93.9%). Corresponding prevalences for HADS-A ≥9 and HADS-A ≥10 were 22.10% and 16.3%, respectively.

4 | DISCUSSION

To the best of our knowledge, this is the first study to compare the HADS depression and anxiety subscales (HADS-D and HADS-A) with the PHQ-9 and GAD-7, respectively, as screening tools for depression and anxiety in an IBS population. Significant positive correlations were found between both depression and anxiety scales. There was a very good agreement between recommended cutoffs (HADS-D and -A cutoff ≥8, PHQ-9 and GAD-7 cutoff ≥10). This agreement increased when higher HADS cutoff values were chosen, but this affects sensitivity and specificity.

The positive correlations found between HADS-D and PHQ-9, and HADS-A and GAD-7 are in line with findings of previous studies in patient populations other than IBS. The strength of the correlations found previously varied according to the disease population examined.[13,14,30–32] Overall, studies (including our study) have found a moderate to high correlation between HADS-D and PHQ-9 scores, and HADS-A and GAD-7 scores. Therefore, the convergent validity, indicating the degree to which two scales are related, was good.

The recommended PHQ-9 and GAD-7 cutoffs were based on validation studies that had compared the accuracy of PHQ-9 and GAD-7 to a diagnosis of depressive or anxiety disorder by the reference standard, which is a clinical diagnostic interview.[19,25] The optimal screening cutoff value that we found for HADS-D for IBS seems to be in line with the recommended HADS cutoff value in literature, HADS cutoff ≥8.[8] To date, Hansson et al. was the only study that compared the HADS-D with the PHQ-9 and investigated their agreement for depression at different cutoff points, in a primary care and psychiatric outpatient population in Sweden. They showed a moderate agreement when the recommended cutoff values (HADS-D cutoff ≥8 with PHQ-9 cutoff ≥10) were compared, and the highest though still moderate agreement for HADS-D cutoff ≥8 and PHQ-9 cutoff ≥12.[20] In contrast, our study showed a high agreement when recommended cutoffs from literature were compared (HADS-D ≥8 and PHQ-9 ≥10, AC1 0.829). The difference in agreement between the Hansson et al. study and our study could be due to the difference in the inclusion criteria, as Hansson et al. only included patients with depressive symptoms and an unknown number of the included patients were already being treated for their depressive symptoms, whereas in our study the prevalence of affective disorders was considerably lower. This highlights that specific population characteristics have profound influence on the performance of these screening instruments. In addition, we used the Gwet AC1 instead of Cohen's Kappa which provides a more stable inter-rater reliability coefficient.[28]

As for anxiety, the comparison between HADS-A and GAD-7 appears less clear-cut. The selection of the cutoffs has substantial impact on the presumed anxiety prevalence. In addition, it is important to realize that the highest AC1 values do not necessarily reflect the most suitable choice for a particular research or clinical question. In this respect, sensitivity and specificity should also be considered because they describe the performance of the test in medical diagnostics. A high specificity is preferred if the purpose is to diagnose the condition, while a high sensitivity is preferred if the purpose is to screen for the condition. Understanding of the performance of these commonly used screening instruments is also important as this allows intelligibility of findings across different studies. The findings from our study can therefore facilitate pooling data from different cohorts using divergent instruments to assess affective symptoms.
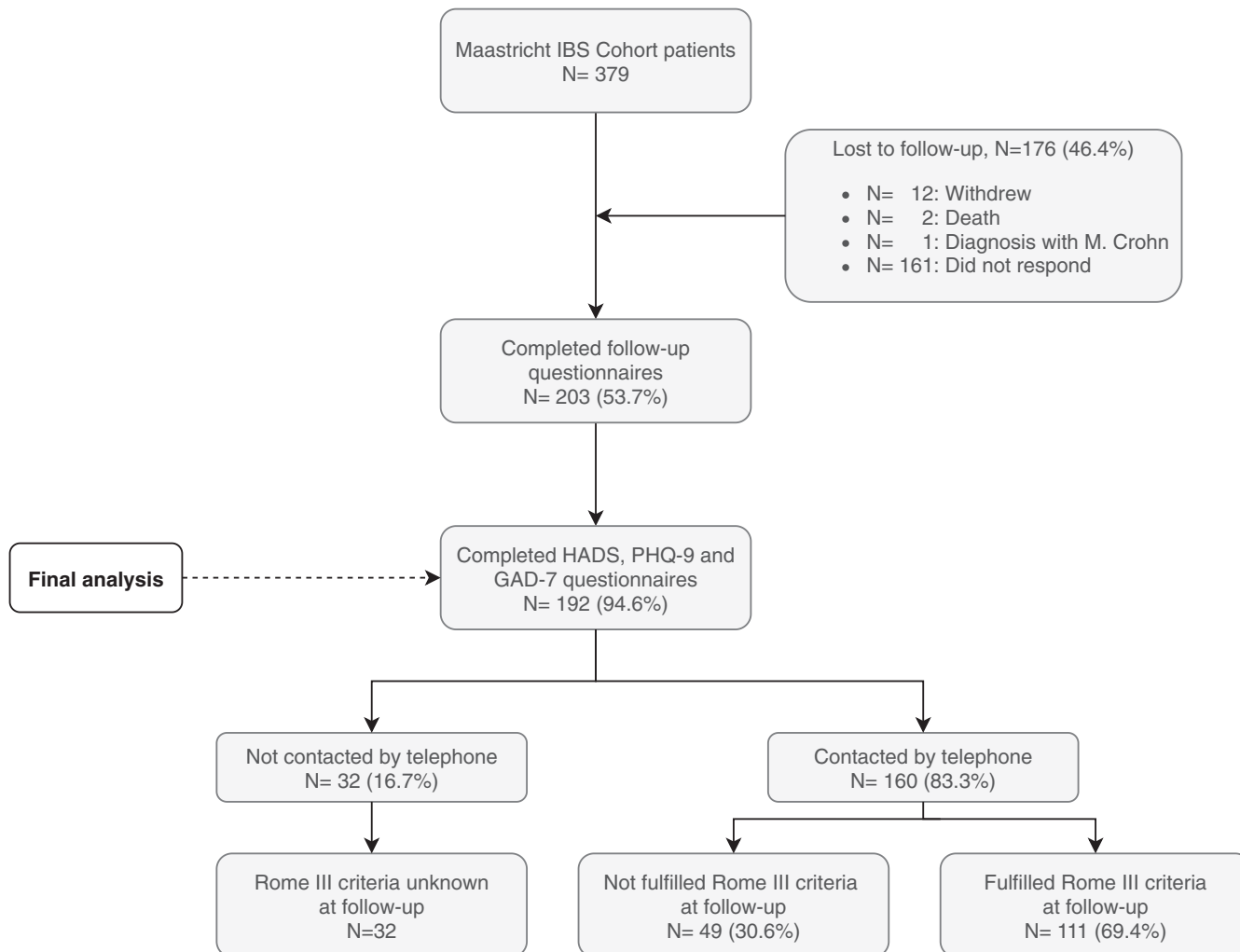
FIGURE 1 Flowchart participant selection

In addition to the correlation between total scores, the agreement reveals more information about the concordance between the self-rating scales at different cutoff points for the presence of a depressive or anxiety disorder using PHQ-9 and GAD-7 cutoff ≥10 as the standard and how these cutoffs impact the prevalence of these disorders. In our study population, the prevalence of depression was 21.9% and of anxiety 14.1% based on a PHQ-9 and GAD-7 cutoff ≥10. Based on the patients who still fulfilled the Rome III criteria for IBS at follow-up (n = 111), the prevalence was 25.2% and 17.1%, respectively. These numbers appear lower than other reports from the literature. A systematic review with meta-analysis by Zamani et al,[33] found prevalence rates of 23.3% for a depressive disorder and 23% for an anxiety disorder in patients with IBS. Although the prevalences for depressive disorder are comparable to our study, there is an apparent difference in anxiety rates. This could be due to the different cutoffs used in the individual studies, and in this current paper, we have indeed pointed to the substantial impact on prevalence rates. Moreover, the Zamani paper indicated significant heterogeneity between studies. Another possible explanation for the lower anxiety prevalence could be the characteristics of the population examined

here as this was a follow-up measurement of IBS patients previously included. Recently, we have shown that anxiety symptoms tend to decrease with time in the same cohort, reflected by a significant difference between baseline and follow-up for the patients who did not fulfilled the Rome III criteria at follow-up, which was not the case for depressive symptoms.[34]

4.1 | Strengths and limitations

This study is the first to assess the performance of different cutoff points for the HADS subscales in relation to PHQ-9 and GAD-7 cutoff ≥10 in an IBS population. The current study population has been recruited from both primary care as well as secondary/tertiary care and provides a good reflection of the Dutch IBS population.

The current dataset is composed of 192 participants, who were included in the MIBS cohort with a mean follow-up time of ±4.6 years. As a consequence, 30.6% of the included participants no longer met the Rome III criteria for IBS. Given that these patients were still experiencing gastrointestinal symptoms and that fulfilling

TABLE 1 Characteristics table separated for depression and anxiety groups in a (previously confirmed) IBS population (n = 192), including n = 111 subjects who still fulfilled the Rome III criteria for IBS at follow-up. Depression and anxiety were characterized based on recommended cutoffs in literature

		PHQ–9 cutoff ≥10			GAD–7 cutoff ≥10		
	Total population n = 192[a]	No depression n = 150	Depression n = 42	p-value	No anxiety n = 165	Anxiety n = 27	p-value
Female gender, n (%)	144 (75.0)	115 (76.7)	29 (69.0)	0.313	127 (77.0)	17 (63.0)	0.119
Age, median (IQR)	53.0 (36–65)	53.5 (38–65)	49.5 (34–62)	0.346	53.0 (39–65)	48.0 (34–62)	0.327
BMI, median (IQR)	24.9 (21.8–28.4)	24.6 (21.8–28.0)	27.0 (21.8–30.7)	0.110	25.1 (21.8–28.4)	24.2 (21.5–29.2)	0.932
Fulfilled Rome III criteria, n (%)							
Yes	111	83 (66.4)	28 (80.0)	0.123	92 (67.2)	19 (82.6)	0.137
No	49	42 (33.6)	7 (20.0)		45 (32.8)	4 (17.4)	
IBS subtype, n (%)							
No IBS	49 (25.5)	42 (28.0)	7 (16.7)	0.452	45 (27.3)	4 (14.8)	0.613
IBS-D	51 (26.6)	40 (26.7)	11 (26.2)		43 (26.1)	8 (29.6)	
IBS-C	24 (12.5)	17 (11.3)	7 (16.7)		19 (11.5)	5 (18.5)	
IBS-M	22 (11.5)	15 (10.0)	7 (16.7)		18 (10.9)	4 (14.8)	
IBS-U	14 (7.3)	11 (7.3)	3 (7.1)		12 (7.2)	2 (7.4)	
Healthcare setting, n (%)							
General practitioner	59 (30.7)	50 (33.3)	9 (21.4)	0.221	52 (31.5)	7 (25.9)	0.791
Secondary/ tertiary care	127 (66.1)	96 (64.0)	31 (73.8)		108 (65.5)	19 (70.4)	
Other	6 (3.1)	4 (2.7)	2 (4.8)		5 (3.0)	1 (3.7)	
Educational attainment, n (%)							
Lower education	67 (34.9)	51 (34.0)	16 (38.1)	0.035 *	55 (33.3)	12 (44.4)	0.339
Intermediate education	68 (35.4)	48 (32.0)	20 (47.6)		58 (35.2)	10 (37.0)	
Tertiary education	57 (29.7)	51 (34.0)	6 (14.3)		52 (31.5)	5 (18.5)	
GSRS, median (IQR)							
Abdominal pain	3.0 (2.0–4.0)	2.8 (2.0–4.0)	3.7 (3.0–5.0)	<0.001 *	3.0 (2.0–4.0)	3.7 (3.0–5.0)	0.004 *
Reflux syndrome	1.5 (1.0–3.0)	1.5 (1.0–2.6)	2.5 (1.0–4.0)	0.006 *	1.5 (1.0–3.0)	2.0 (1.0–4.0)	0.307
Diarrhea syndrome	2.7 (1.7–4.0)	2.7 (1.3–4.0)	3.7 (1.9–5.3)	0.017 *	2.7 (1.7–4.0)	2.3 (1.3–5.3)	0.766
Indigestion syndrome	3.8 (2.8–4.5)	3.5 (2.5–4.3)	4.0 (3.4–4.6)	0.023 *	3.5 (2.5–4.3)	4.0 (3.5–4.8)	0.020 *
Constipation syndrome	2.7 (2.0–3.7)	2.7 (1.7–3.4)	3.3 (2.6–5.0)	0.003 *	2.7 (1.7–3.7)	3.3 (2.0–4.3)	0.147
Quality of life, median (IQR)							
PCS	45.6 (34.5–50.4)	46.1 (36.7–51.0)	41.0 (30.6–47.9)	0.006 *	45.3 (33.1–50.4)	47.9 (42.3–50.0)	0.278
MCS	52.0 (42.3–56.6)	54.7 (47.5–57.3)	34.5 (28.3–43.0)	<0.001 *	53.7 (46.1–57.1)	29.1 (23.5–37.0)	<0.001 *

Note: Not normally distributed continuous variables were analyzed with Mann-Whitney U test and reported as median and interquartile ranges. Dichotomous/categorical variables were analyzed with the Chi-square test and reported as frequencies.

Abbreviations: BMI, Body Mass Index (kg m^{-2}); GSRS, Gastrointestinal Symptom Rating Scale; IBS, Irritable Bowel Syndrome; IQR, interquartile range; MCS, Mental Composite Score; n, number of patients; PCS, Physical Composite Score.

[a]Numbers may not add up to total due to missing.

*p-value of <0.05 was considered statistically significant.

Difference between groups (no depression vs. depression, no anxiety vs. anxiety) were tested.

TABLE 2 Agreement at different cutoff points when comparing HADS-D cutoff values to PHQ-9 cutoff values in a (previously confirmed) IBS population (n = 192), including n = 111 subjects who still fulfilled the Rome III criteria for IBS at follow-up

	PHQ-9 at different cut-off values									
	Cut-off ≥8		Cut-off ≥9		Cut-off ≥10		Cut-off ≥11		Cut-off ≥12	
HADS-D at different cut-off values	Gwet AC1 (95% CI)	Sensitivity Specificity (%)	Gwet AC1 (95% CI)	Sensitivity Specificity (%)	Gwet AC1 (95% CI)	Sensitivity Specificity (%)	Gwet AC1 (95% CI)	Sensitivity Specificity (%)	Gwet AC1 (95% CI)	Sensitivity Specificity (%)
Cut-off ≥4	0.659 (0.56 - 0.76)	93.0 77.0	0.582 (0.47 - 0.69)	91.7 72.2	0.553 (0.44 - 0.66)	92.9 70.0	0.521 (0.41 - 0.63)	97.0 67.3	0.478 (0.37 - 0.59)	96.3 64.8
Cut-off ≥5	0.729 (0.64 - 0.82)	86.0 84.4	0.692 (0.6 - 0.79)	87.5 80.5	0.698 (0.61 - 0.79)	92.9 79.3	0.666 (0.57 - 0.76)	97.0 76.1	0.623 (0.53 - 0.72)	96.3 73.3
Cut-off ≥6	0.754 (0.67 - 0.84)	80.7 88.1	0.736 (0.65 - 0.82)	83.3 84.7	0.742 (0.66 - 0.83)	88.1 83.3	0.726 (0.64 - 0.81)	93.9 80.5	0.701 (0.61 - 0.79)	96.3 78.2
Cut-off ≥7	0.742 (0.66-0.83)	64.9 92.6	0.743 (0.66-0.83)	66.7 89.6	0.782 (0.71-0.86)	73.8 89.3	0.782 (0.71 - 0.86)	78.8 86.8	0.788 (0.72 - 0.86)	85.2 85.5
Cut-off ≥8	0.743 (0.66-0.83)	56.1 95.6	0.776 (0.70-0.85)	60.4 93.8	0.829 (0.76-0.90)	69.0 94.0	0.843 (0.78-0.91)	75.8 91.8	0.862 (0.80-0.92)	85.2 90.9
Cut-off ≥9	0.715 (0.63-0.8)	42.1 97.8	0.780 (0.71-0.85)	47.9 97.2	0.830 (0.76-0.90)	54.8 97.3	0.859 (0.80-0.92)	60.6 95.6	0.876 (0.82-0.93)	66.7 94.5
Cut-off ≥10	0.714 (0.63-0.8)	36.8 99.3	0.792 (0.72-0.86)	43.8 99.3	0.841 (0.78-0.90)	50.0 99.3	0.869 (0.81-0.92)	54.5 97.5	0.886 (0.83-0.94)	59.3 96.4
Cut-off ≥11	0.695 (0.61-0.78)	31.6 99.3	0.773 (0.70-0.85)	37.5 99.3	0.822 (0.76-0.89)	42.9 99.3	0.850 (0.79-0.91)	45.5 97.5	0.868 (0.81-0.92)	48.1 96.4
Cut-off ≥12	0.662 (0.58-0.75)	19.3 100	0.740 (0.66-0.82)	22.9 100	0.788 (0.72-0.86)	26.2 100	0.856 (0.80-0.91)	33.3 100	0.886 (0.84-0.94)	37.0 99.4

The red box represents the recommended cut-off ≥10 for PHQ-9.

Abbreviations: HADS-D, Hospital Anxiety and Depression Scale- subscale depression; PHQ-9, Patient Health Questionnaire-9; IBS, Irritable Bowel Syndrome; CI, Confidence Interval; n, number of patients; AC, Agreement Coefficient.

Interpretation Gwet AC1	Gwet AC1
Less than 0.2	Poor
0.2 to 0.4	Fair
0.4 to 0.6	Moderate
0.6 to 0.8	Good
0.8 to 1.0	Very Good

TABLE 3 Agreement at different cutoff points when comparing different HADS-A cutoff values to the GAD-7 cutoff values in a (previously confirmed) IBS population (n = 190), including n = 111 subjects who still fulfilled the Rome III criteria for IBS at follow-up

	GAD-7 at different cut-off values											
	Cut-off ≥7		Cut-off ≥8		Cut-off ≥9		Cut-off ≥10		Cut-off ≥11		Cut-off ≥12	
HADS-A at different cut-off values	Gwet AC1 (95% CI)	Sensitivity Specificity (%)	Gwet AC1 (95% CI)	Sensitivity Specificity (%)	Gwet AC1 (95% CI)	Sensitivity Specificity (%)	Gwet AC1 (95% CI)	Sensitivity Specificity (%)	Gwet AC1 (95% CI)	Sensitivity Specificity (%)	Gwet AC1 (95% CI)	Sensitivity Specificity (%)
Cut-off ≥7	0.708 (0.61-0.80)	89.4 81.1	0.636 (0.54-0.73)	90.9 75.2	0.632 (0.53-0.73)	93.3 74.4	0.628 (0.53-0.72)	96.3 73.6	0.604 (0.51-0.70)	100 71.6	0.585 (0.49-0.68)	100 70.3
Cut-off ≥8	0.839 (0.77- 0.91)	83.0 92.3	0.784 (0.71-0.86)	84.8 86.0	0.811 (0.74-0.88)	93.3 86.3	0.806 (0.73-0.88)	96.3 85.3	0.781 (0.71-0.85)	100 82.8	0.761 (0.69-0.84)	100 81.4
Cut-off ≥9	0.860 (0.80-0.92)	76.6 95.8	0.823 (0.75-0.89)	78.8 89.8	0.848 (0.78-0.91)	86.7 90.0	0.843 (0.78-0.91)	88.9 90.0	0.818 (0.75-0.88)	90.5 86.4	0.828 (0.76-0.89)	100 86.0
Cut-off ≥10	0.813 (0.74-0.88)	57.4 97.2	0.854 (0.79-0.91)	66.7 94.3	0.878 (0.82-0.93)	73.3 94.4	0.886 (0.83-0.94)	77.8 93.9	0.862 (0.80-0.92)	76.2 91.1	0.871 (0.82-0.93)	83.3 90.7
Cut-off ≥11	0.833 (0.77-0.90)	53.2 100	0.886 (0.83-0.94)	63.6 97.5	0.909 (0.86-0.96)	70.0 97.5	0.917 (0.87-0.96)	74.1 96.9	0.906 (0.86-0.95)	76.2 94.7	0.914 (0.87-0.96)	83.3 94.2
Cut-off ≥12	0.787 (0.72- 0.86)	38.3 100	0.883 (0.83-0.94)	51.5 99.4	0.905 (0.86-0.95)	56.7 99.4	0.927 (0.88-0.97)	63.0 99.4	0.929 (0.89-0.97)	66.7 97.6	0.936 (0.90-0.97)	72.3 97.1

The red box represents the recommended cut-off ≥10 for GAD-7.

Abbreviations: HADS-A, Hospital Anxiety and Depression Scale- subscale anxiety; GAD-7, Generalized Anxiety Disorder-7; IBS, Irritable Bowel Syndrome; CI, Confidence Interval; n, number of patients; AC, Agreement Coefficient.

Interpretation Gwet AC1	Gwet AC1
Less than 0.2	Poor
0.2 to 0.4	Fair
0.4 to 0.6	Moderate
0.6 to 0.8	Good
0.8 to 1.0	Very Good

the Rome criteria was not significantly associated with a depressive or anxiety disorder, we hypothesized that inclusion of these patients will not have a significant impact on the agreement between cutoff values. Indeed, in additional analysis the highest agreements were found for the same cutoff values in the subgroup of patients who met Rome III criteria (n = 111) compared to the total population (n = 192) (data not shown). Furthermore, we did not specifically assess medication use or psychotherapy. As a result, it is possible that some patients were on treatment for their psychological complaints and, therefore, their anxiety or depressive symptoms might be less

prominent. However, the mere prevalence of anxiety or depression was not the main purpose of our study. Lastly, in this study no data were available regarding a diagnosis of depressive and/or anxiety disorders based on the gold standard diagnostic method, that is, a psychiatric interview. Instead, sensitivity and specificity were calculated, based on the assumption that PHQ-9 and GAD-7 were the general standard to screen for depressive and anxiety disorders in this patient population. It was therefore not possible to define cutoffs for specific clinical or research settings. Given the excellent psychometric properties of the PHQ-9 and GAD-7 (PHQ-9 sensitivity 85% and specificity 89%[19] and GAD-7 sensitivity 74% and specificity 83%[25]), we think this assumption can be made safely without major impact on study findings.

5 | CONCLUSION

In conclusion, HADS-D vs. PHQ-9 scores and HADS-A vs. GAD-7 scores are highly correlated in a population of IBS patients. Custom cutoff values deem the HADS subscales (HADS-D and HADS-A) concordant to PHQ-9 and GAD-7 scores, the use of which are now recommended by current guidelines. The optimal cutoff value of ≥8 for HADS-D, was in line with the recommended cutoff from literature. However, modification of the HADS-A cutoff has substantial impact on sensitivity/specificity and this should be kept in mind when applying this instrument for screening, diagnostic or research purposes in IBS. It ought to be noted that the choice of an optimal cutoff value is dependent on patient population, setting, treatment monitoring, and the purpose of a study. Results from this study allow for easier comparison of results from different IBS populations regarding comorbid depression and anxiety.

ORCID
Johanna Theodora Wilhelmina Snijkers https://orcid.org/0000-0002-3582-2966
Zsa Zsa Regina Maria Weerts https://orcid.org/0000-0001-7522-9916
Lisa Vork https://orcid.org/0000-0001-7316-3191
Zlatan Mujagic https://orcid.org/0000-0001-5211-0381
Daniel Keszthelyi https://orcid.org/0000-0001-9856-9980

REFERENCES
1. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clin Gastroenterol Hepatol.* 2012;10(7):712-721.e714.
2. Hungin AP, Whorwell PJ, Tack J, Mearin F. The prevalence, patterns and impact of irritable bowel syndrome: an international survey of 40,000 subjects. *Aliment Pharmacol Ther.* 2003;17(5):643-650.
3. Holtmann GJ, Ford AC, Talley NJ. Pathophysiology of irritable bowel syndrome. *Lancet Gastroenterol Hepatol.* 2016;1(2):133-146.
4. Enck P, Aziz Q, Barbara G, et al. Irritable bowel syndrome. *Nat Rev Dis Primers.* 2016;2:16014.
5. Fond G, Loundou A, Hamdani N, et al. Anxiety and depression comorbidities in irritable bowel syndrome (IBS): a systematic review and meta-analysis. *Eur Arch Psychiatry Clin Neurosci.* 2014;264(8):651-660.
6. Drossman DA, Chang L, Bellamy N, et al. Severity in irritable bowel syndrome: a Rome foundation working team report. *Am J Gastroenterol.* 2011;106(10):1749-1759.quiz 1760.
7. Kruimel J, Leue C, Winkens B, et al. Integrated medical-psychiatric outpatient care in functional gastrointestinal disorders improves outcome: a pilot study. *Eur J Gastroenterol Hepatol.* 2015;27(6):721-727.
8. 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(2):69-77.
9. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606-613.
10. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* 2006;166(10):1092-1097.
11. Boeckxstaens GE, Drug V, Dumitrascu D, et al. Phenotyping of subjects for large scale studies on patients with IBS. *Neurogastroenterol Motil.* 2016;28(8):1134-1147.
12. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67(6):361-370.
13. Stafford L, Berk M, Jackson HJ. Validity of the hospital anxiety and depression scale and patient health questionnaire-9 to screen

for depression in patients with coronary artery disease. *Gen Hosp Psychiatry*. 2007;29(5):417-424.

14. Hartung TJ, Friedrich M, Johansen C, et al. The hospital anxiety and depression scale (HADS) and the 9-item patient health questionnaire (PHQ-9) as screening instruments for depression in patients with cancer. *Cancer*. 2017;123(21):4236-4243.

15. Lowe B, Grafe K, Zipfel S, Witte S, Loerch B, Herzog W. Diagnosing ICD-10 depressive episodes: superior criterion validity of the patient health questionnaire. *Psychother Psychosom*. 2004;73(6):386-390.

16. Esser P, Hartung TJ, Friedrich M, et al. The generalized anxiety disorder screener (GAD-7) and the anxiety module of the hospital and depression scale (HADS-A) as screening tools for generalized anxiety disorder among cancer patients. *Psycho-oncology*. 2018;27(6):1509-1516.

17. Baker AM, Holbrook JT, Yohannes AM, et al. Test performance characteristics of the AIR, GAD-7 and HADS-anxiety screening questionnaires for anxiety in chronic obstructive pulmonary disease. *Ann Am Thorac Soc*. 2018;15(8):926-934.

18. Marrie RA, Zhang L, Lix LM, et al. The validity and reliability of screening measures for depression and anxiety disorders in multiple sclerosis. *Mult Scler Relat Disord*. 2018;20:9-15.

19. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the patient health questionnaire (PHQ-9): a meta-analysis. *CMAJ*. 2012;184(3):E191-E196.

20. Hansson M, Chotai J, Nordstom A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. *Br J Gen Pract*. 2009;59(566):e283-288.

21. Svedlund J, Sjodin I, Dotevall G. GSRS–a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci*. 1988;33(2):129-134.

22. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993;31(3):247-263.

23. Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 health survey, V.I. *Health Qual Life Outcomes*. 2007;5:54.

24. Kroenke K, Spitzer RL, Williams JB, Lowe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *Gen Hosp Psychiatry*. 2010;32(4):345-359.

25. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. *Gen Hosp Psychiatry*. 2016;39:24-31.

26. R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020. https://www.R-project.org/

27. Gwet KL. Testing the difference of correlated agreement coefficients for statistical significance. *Educ Psychol Measur*. 2016;76(4):609-637.

28. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol*. 2013;13:61.

29. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008;61(Pt 1):29-48.

30. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *Br J Gen Pract*. 2008;58(546):32-36.

31. Omani-Samani R, Maroufizadeh S, Ghaheri A, Navid B. Generalized anxiety disorder-7 (GAD-7) in people with infertility: a reliability and validity study. *Middle East Fertil Soc J*. 2018;23(4):446-449.

32. Terrill AL, Hartoonian N, Beier M, Salem R, Alschuler K. The 7-item generalized anxiety disorder scale as a tool for measuring generalized anxiety in multiple sclerosis. *Int J MS Care*. 2015;17(2):49-56.

33. Zamani M, Alizadeh-Tabari S, Zamani V. Systematic review with meta-analysis: the prevalence of anxiety and depression in patients with irritable bowel syndrome. *Aliment Pharmacol Ther*. 2019;50(2):132-143.

34. Weerts Z, Vork L, Mujagic Z, et al. Reduction in IBS symptom severity is not paralleled by improvement in quality of life in patients with irritable bowel syndrome. *Neurogastroenterol Motil*. 2019;31(8):e13629.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.