

Eating Disorder Screening: a Systematic Review and Meta-analysis of Diagnostic Test Characteristics of the SCOFF



Amanda M. Kutz, PhD¹ , Alison G. Marsh, BA², Craig G. Gunderson, MD^{2,3}, Shira Maguen, PhD^{4,5}, and Robin M. Masheb, PhD^{2,6}

¹VA Salt Lake City Health Care System, Salt Lake City, UT, USA; ²VA Connecticut Healthcare System, West Haven, CT, USA; ³Department of Internal Medicine, Section of General Internal Medicine, Yale University School of Medicine, New Haven, CT, USA; ⁴San Francisco VA Healthcare System, San Francisco, CA, USA; ⁵Department of Psychiatry, University of California San Francisco, School of Medicine, San Francisco, CA, USA; ⁶Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA.

BACKGROUND: Eating disorders affect upwards of 30 million people worldwide and often go undertreated and underdiagnosed. The purpose of this systematic review and meta-analysis was to evaluate the diagnostic accuracy of the Sick, Control, One, Fat and Food (SCOFF) questionnaire for DSM-5 eating disorders in the general population.

METHOD: The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) were followed. A PubMed search was conducted among peer-reviewed articles. Information regarding validation of the SCOFF was required for inclusion. Study quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.

RESULTS: The final analysis included 25 studies. The validity of the SCOFF was high across samples with a pooled sensitivity of 0.86 (95% CI, 0.78–0.91) and specificity of 0.83 (95% CI, 0.77–0.88). Subgroup analyses were conducted to examine the impact of methodology, study quality, and clinical characteristics on diagnostic accuracy. Studies with the highest sensitivity tended to be case-control studies of young women with anorexia nervosa (AN) and bulimia nervosa (BN). Studies which included more men, included those diagnosed with binge eating disorder, and recruited from large community samples tended to have lower sensitivity. Few studies reported on BMI and race/ethnicity; thus, subgroups for these factors could not be examined. No studies used reference standards which assessed all DSM-5 eating disorders.

CONCLUSION: This meta-analysis of 25 validation studies demonstrates that the SCOFF is a simple and useful screening tool for young women at risk for AN and BN. However, there is not enough evidence to support utilizing the SCOFF for screening for the range of DSM-5 eating disorders in primary care and community-based settings.

Further examination of the validity of the SCOFF or development of a new screening tool, or multiple tools, to screen for the range of DSM-5 eating disorders heterogeneous populations is warranted.

TRIAL REGISTRATION: This study is registered online with PROSPERO (CRD42018089906).

KEY WORDS: eating disorders; SCOFF; screening; systematic review; diagnostic test accuracy.

J Gen Intern Med 35(3):885–93

DOI: 10.1007/s11606-019-05478-6

© Society of General Internal Medicine (This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply) 2019

INTRODUCTION

Eating disorders affect upwards of 30 million people and carry with them significant morbidity and mortality.¹ Effective screening for eating disorders is critical as these disorders are commonly underdiagnosed and undertreated.^{1–3} The 5-item SCOFF (Sick, Control, One, Fat and Food; see Fig. 1) questionnaire, developed in 1999 by Morgan and colleagues, is the most widely used screening measure for eating disorders. With the inclusion of binge eating disorder and other specified eating disorders (i.e., atypical anorexia, low frequency or limited duration bulimia nervosa and binge eating disorder, purging disorder, night eating syndrome) in DSM-5,⁴ it has become increasingly important to expand awareness of various types of eating pathology. Of particular importance, these new categories of eating disorders had not yet been defined at the time that the SCOFF was developed.

The changing landscape of diagnostic eating disorder categories since the publication of DSM-5 highlights the importance of ensuring screening tools are appropriate for detecting the full range of eating disorders in the general population. To date, the SCOFF has been the recommended screening tool across numerous validation studies; however, these recommendations have not been systematically assessed. The purpose of this systematic review and meta-analysis is to evaluate whether the SCOFF can appropriately screen patients in the general population for

Prior Presentations This project was presented at the annual meeting for the Society for Behavioral Medicine (SBM) in April 2018 and at the International Conference on Eating Disorders (ICED) in March 2019.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11606-019-05478-6>) contains supplementary material, which is available to authorized users.

Received July 15, 2019

Accepted October 7, 2019

Published online November 8, 2019

SCOFF Questionnaire

- 1) Do you make yourself **Sick** because you feel uncomfortably full?
- 2) Do you worry you have lost **Control** over how much you eat?
- 3) Have you recently lost **One** stone in a 3-month period?
- 4) Do you believe yourself to be **Fat** when others say you are too thin?
- 5) Would you say that **Food** dominates your life?

Figure 1 The SCOFF questionnaire.

the full range of eating pathology currently represented in DSM-5. To accomplish this, the literature was reviewed for studies that report the diagnostic test characteristics of the SCOFF.

METHOD

The Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guideline was followed in preparing this systematic review.⁵ This study is registered online with PROSPERO (CRD42018089906) and all search strategies and methods were determined before the onset of the study.

Search Strategy and Study Selection

We conducted a systematic literature search using PubMed from database inception through March 13, 2018. The search terms were “SCOFF” and “Feeding and Eating Disorders/Diagnosis” AND (“Psychometrics” OR “Sensitivity and Specificity”). Other search terms, such as “SCOFF questionnaire,” “Eating disorders screening,” and “Feeding and Eating Disorder/Diagnosis” AND “screening,” were attempted but revealed overlapping or extraneous results. The word “SCOFF” was searched for in all text of the Cochrane database in addition to these searches as this included the most comprehensive results in that database. Two reviewers (AMK and AGM) independently screened all abstracts generated from the subject search. Inclusion criteria specified that studies were published in English or were available in translation to English. To be included, it was required that validation information for the SCOFF could be derived from articles and included some specific demographic information (i.e., age range or standard deviation, gender, eating disorder diagnosis). The two independent reviewers (AMK and AGM) had high inter-rater agreement ($K = 1$) for exclusion of articles.

Data Extraction and Quality Assessment

The reviewers (AMK and AGM) used a standardized data collection form to extract data on date of publication, country in which the study was conducted, recruitment method, reference measure utilized, sample size, age, gender of sample, and race/ethnicity of the sample, participants’ average BMI and weight category, and percentage of sample with eating disorder diagnoses.

The reviewers (AMK and AGM) independently assessed study quality of all included studies using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.⁶ Differences in assessment were resolved by consensus and inter-rater agreement was very high ($K = 1$). The original QUADAS-2 tool includes 4 domains: patient selection, index test(s), reference standard, and flow and timing. Given that there was little meaningful variation in the index test (i.e., the SCOFF) questions or administration, the index test domain was dropped from ratings. Specific rating criteria for each domain are presented in Appendix 1.

Data Synthesis and Analysis

Statistical measures of test performance (true positive, false positive, true negative, and false negative) were extracted from individual studies. Extracted test performance data can be found in Appendix 2. When sensitivity (true positive rate) and specificity (true negative rate) were reported separately for different eating disorder diagnoses or by gender, frequencies of the statistical measures for test performance were summed using data available in the manuscript. In two instances,^{7, 8} this data was not readily available or additional information from the authors was needed and corresponding authors on these manuscripts were contacted. Contacted authors provided data to calculate frequencies of the total sample (M. Tseng, personal communication, May 2018; S. Maguen, personal communication, November 2018). For all studies, statistical heterogeneity was estimated using the I^2 statistic. Statistical heterogeneity provides an estimate of the amount of variance that is attributable to variability between studies. To account for variability across studies, subgroup analyses were conducted. Subgroups were prespecified based on study methodology (case study vs non-case study; type of reference standard used—interview vs questionnaire), study quality based on QUADAS-2 ratings, and patient characteristics (gender, age, sample type, and location). Statistical analyses were performed using STATA version 14.2 (StataCorp, College Station, TX). The meta-analytical integration of diagnostic accuracy studies (MIDAS) command was used to obtain figures and descriptive summaries and conduct subgroup analyses.⁹

RESULTS

Literature Search and Study Selection

A total of 984 abstracts were identified through the included databases and three were identified through bibliographies (Fig. 2). After 47 duplicates were removed, all titles or abstracts were reviewed for relevance. Following initial review, 882 records were excluded leaving 58 full-text articles for full review. Of these 58 articles, 33 were excluded for the following reasons: no validation information available, no reference standard included, article and data not available in

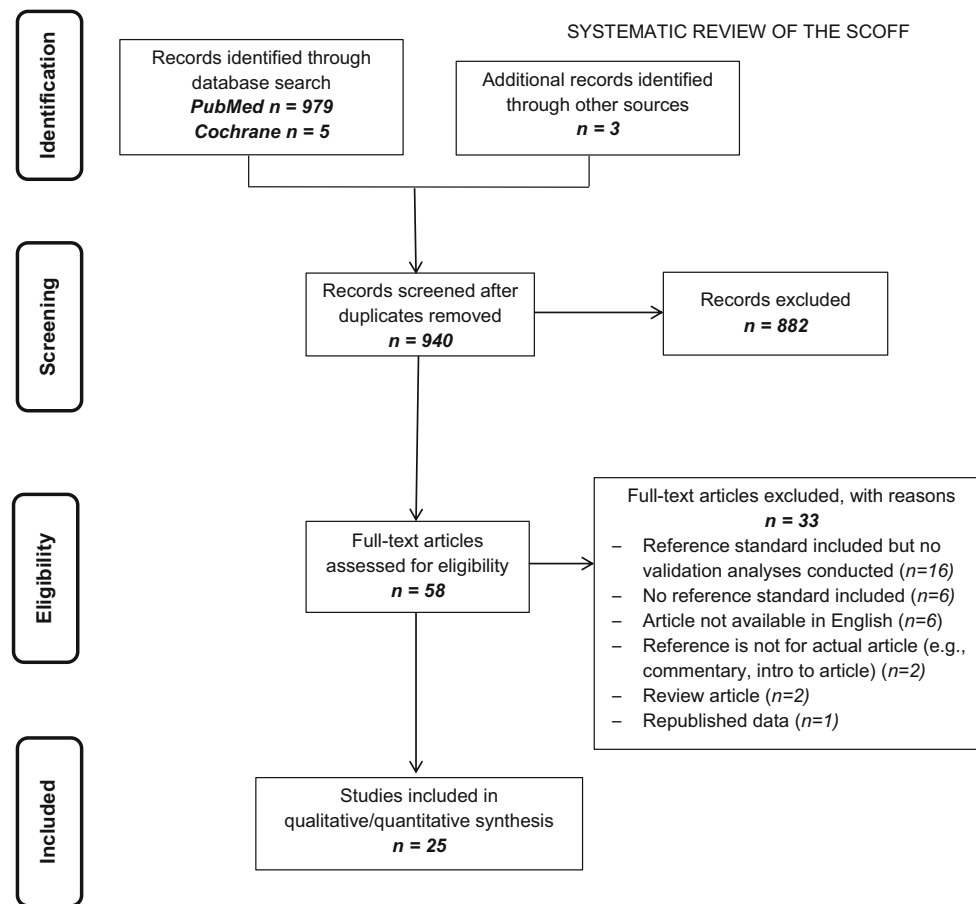


Figure 2 Study flow diagram of literature search.

English, and article representing re-publication of prior data or commentary on data.

Study Characteristics

Table 1 depicts the characteristics of included studies.^{7, 8, 10–32} The 25 studies reviewed included a total of 11,531 individuals. Thirteen unique countries were represented with 16 studies conducted in Europe, four in North America, three in Asia, and two in South America. Samples were recruited from three primary locations: medical settings (primary care clinics, specialty clinics), schools (grade school, high school, and universities), and the general community. Ages across studies ranged from 10 to 95, with the majority ($n = 18$) of studies including primarily adult samples and seven being conducted in entirely adolescent or young adult populations. Twelve studies included an entirely female sample and four additional studies included samples which were at least 70% female. The percentage of females included in the remaining eight studies ranged from 46.2 to 68%. Thirteen studies utilized interview format (SCID-I, CIDI interview, DSM-IV interview, EDE) for the reference standard. The remaining sixteen studies utilized various self-report measures including the EDI-3, the EDE-Q, the EAT-26, the Q-EDD, and an ICD-10 symptom rating scale. Eighteen studies reported on the percent of the sample

which was diagnosed with an eating disorder based on the criterion used. The range of any eating disorder diagnosis, which included anorexia nervosa, bulimia nervosa, binge eating disorder, and eating disorder not otherwise specified, was 1.2 to 64.3%. The range of the sample that had each eating disorder diagnosis was as follows: anorexia nervosa = 0 to 32.1%, bulimia nervosa = 0 to 23.1%, eating disorder not otherwise specified = 0.4 to 46.3%, binge eating disorder = 0.4 to 11.6%. Sixteen studies explicitly reported on the percent of individuals with anorexia and bulimia. Eleven studies reported on the percent of the sample meeting criteria for eating disorder not otherwise specified and six reported on binge eating disorder. Aside from binge eating disorder, no studies explicitly examined validity for any of the other newly included eating disorders in the DSM-5.

Other demographic characteristics were not frequently reported across all studies and thus are not included in Table 1. Only four studies reported any information about race or ethnicity, and samples tended to be primarily Caucasian (57.2 to 87.9%). Studies also did not frequently include information about BMI. The six studies reporting average BMI found that it ranged from 21.98 to 28.1. Of these, four studies included samples with an average BMI in the normal range (i.e., between 18.5 and 24.9).

Table 1 Characteristics of Included Studies

Article	Country	Setting	Age range	Gender (% female)	Reference standard	% sample AN/BN/EDNOS/BED	Sensitivity	Specificity
Aoun et al., 2015	Lebanon	Medical setting	15–55	100.0	MINI and DSM-IV interview	28.0 (0.8/11.4/16.3/–)	80.0	72.7
Berger et al., 2011	Germany	Community sample	12	52.7	EAT-26D	—	79.0	74.0
Caamano et al., 2002	Spain	Primary school	11–13	M/F included	EAT	—	64.1	87.2
Cotton, Ball, and Robinson 2003	UK	Medical setting and university	18–64	77.0	Q-EDD	12.0 (2.7/4.9/0.4/4.0)	78.0	88.0
Garcia et al., 2010	France	University	18–35	100.0	MINI and DSM-IV interview	9.3 (2.0/7.3/–/–)	94.6	94.8
Garcia et al., 2011	France	Medical setting	18–35	100.0	MINI and DSM-IV interview	49.5 (29.6/19.9/–/–)	94.0	94.0
Garcia-Campayo et al., 2005	Spain	Medical setting	15–53	100.0	SCAN interview	64.6 (14.2/23.1/27.0/–)	97.7	94.4
Lahteenmaki et al., 2009	Finland	Community sample	20–35	57.7	SCID interview	4.4 (2.1/2.3/–/–)	77.8	87.8
Leung et al., 2009	China	Primary school	12–25	46.2	EDE-Q	—	76.1	97.1
Lichtenstein, Hemmingsen, and Stoving, 2017	Denmark	Primary school and medical setting	11–20	85.9	EDI-3	42.9 (18.2/5.9/18.8/–)	77.1	71.9
Liu et al., 2015	Taiwan	Medical setting	18–45	60.7	SCID-I/P	8.8 (2.7/5.7/5.4/2.6)	85.7 (M) 94.5 (F)	74.2 (M) 66.1 (F)
Luck et al., 2002	UK	Medical setting	18–50	100.0	DSM-IV interview	3.8 (0.3/0.9/2.6/–)	84.6	89.6
Maguen et al., 2018	USA	Medical setting	18–70	100.0	EDE-Q	15.5 (2.0/2.2/–/11.5)	66.1	79.9
Mond et al., 2008	USA	Medical setting	18–40	100.0	EDE interview	17.0 (0/3.4/9.5/4.1)	72.0	73.0
Morgan, Reid, and Lacey, 1999	UK	Medical setting	18–40	100.0	DSM-IV interview	54.7 (32.1/22.6/–/–)	100.0	87.5
Muro-Sans, Amador-Campos, and Morgan, 2008	Spain	Primary school	10–17	49.9	EDI-2	—	73.1	77.7
Pannocchia et al., 2011	Italy	Medical setting	—	100.0	EDI-3	—	97.0	87.3
Parker, Lyons, and Bonner, 2005	USA	University	20–51	72.4	EDE-Q	20.3 (0/0.3/19.9/–)	54.3	93.2
Richter et al., 2017	Germany	Community sample	14–95	53.4	ICD-10 Symptom Rating Scale	—	25.7	97.0
Rueda, Diaz, Campo, et al., 2005	Colombia	University	17–35	100.0	CIDI interview	21.1 (0/3.3/17.8/–)	78.4	75.8
Rueda, Diaz, Ortiz, et al., 2005	Colombia	Primary school	10–19	100.0	CIDI interview	29.9 (1.7/5.8/22.4/–)	81.9	78.7
Sanchez-Armass et al., 2017	Mexico	University	17–56	68.0	EDE interview	25.0 (5.8/7.7/11.5/–)	78.0	84.0
Siervo et al., 2005	Italy	Medical setting	16–35	100.0	DSM-IV interview, EAT-26, and TFEQ	46.3 (0/0/38.9/7.4)	94.0	21.0
Solmi et al., 2015	UK	Community Sample	16–90	75.2	SCID interview	21.4 (–/–/–/–)	53.7	93.5
Wahida et al., 2017	Malaysia	University	18–22	64.7	EAT-26	43.5 (–/–/–/–)	77.4	60.5

MINI, Mini International Neuropsychiatric Interview; EAT, Eating Attitudes Test; Q-EDD, Questionnaire for Eating Disorder Diagnosis; SCAN, Schedules for Clinical Assessment in Neuropsychiatry; SCID, Structured Clinical Interview for DSM-IV; EDE-Q, Eating Disorder Examination-Questionnaire; EDE, Eating Disorder Examination; EDI, Eating Disorder Inventory; ICD, International Classification of Diseases; CIDI, Composite International Diagnostic Interview; TFEQ, Three-Factor Eating Questionnaire

Quality Assessment

Summary index scores for the QUADAS-2 are depicted in Table 2. Risk of bias and applicability concerns were rated as low risk/concern (depicted as “+” signs in the table), high risk/concern (represented as “-” signs in the table), and unknown risk/concern (depicted as “?” signs in the table). Only two studies

were rated as low across all risk of bias and applicability concern domains. Risk of bias was high within patient selection across five studies. Four of these studies used a case-control design while one recruited an at-risk sample as opposed to utilizing a random or consecutive sample. Risk of bias was high in four studies in the flow and timing domain due to the SCOFF and reference standard being administered at different times (i.e., not

Table 2 Study Quality Assessment

Reference	Risk of bias			Applicability concerns	
	Patient selection	Reference standard	Flow and timing	Patient selection	Reference standard
Aoun et al., 2015	+	+	+	+	+
Berger et al., 2011	?	+	?	-	+
Caamano et al., 2002	+	+	+	-	+
Cotton, Ball, and Robinson, 2003	+	+	+	+	?
Garcia et al., 2010	+	+	+	-	-
Garcia et al., 2011	-	?	+	-	-
Garcia-Campayo et al., 2005	?	+	?	-	+
Lahteenmaki et al., 2009	+	-	-	+	+
Leung et al., 2009	+	+	?	-	+
Lichtenstein, Hemmingson, and Stoving, 2017	-	+	-	-	-
Liu et al., 2015	+	+	+	+	+
Luck et al., 2002	+	+	+	-	+
Maguen et al., 2018	+	+	?	+	+
Mond et al., 2008	+	+	-	+	+
Morgan, Reid, and Lacey, 1999	-	?	?	-	+
Muro-Sans, Amador-Campos, and Morgan, 2008	+	+	+	-	-
Pannocchia et al., 2011	-	?	+	?	?
Parker, Lyons, and Bonner, 2005	+	+	+	+	+
Richter et al., 2017	+	?	+	+	-
Rueda et al., 2005a	?	+	?	?	+
Rueda et al., 2005b	?	+	?	-	+
Sanchez-Armass et al., 2017	+	?	?	+	+
Siervo et al., 2005	-	?	?	-	+
Solmi et al., 2015	+	?	-	-	+
Wahida et al., 2017	+	+	+	-	+

Risk of bias and applicability concerns were rated as low risk/concern (depicted as “+” signs in the table), high risk/concern (represented as “-” signs in the table), and unknown risk/concern (depicted as “?” signs in the table)

sequentially) or if there was any ambiguity about questionnaires being completed at the same time (such as would be present in surveys that were mailed). In general, risk of bias was low across studies in the reference standard domain. Applicability concerns were most prevalent in the patient selection domain, with 14 studies having high applicability concerns. This was often because the sample utilized was restricted demographically (e.g., only included females). Applicability concerns were high across five studies for the reference standard. In these studies, certain subgroups of patients were not given the reference standard or were excluded for unknown reasons.

Diagnostic Accuracy

Diagnostic accuracy rates for each study are depicted in the forest plot in Figure 3 and the receiver operating curve (SROC) in Figure 4. Pooled sensitivity was 0.86 (95% CI, 0.78–0.91) and specificity was 0.83 (95% CI, 0.77–0.88). The area under the curve (AUC) was 0.91 (95% CI, 0.88–0.93). Heterogeneity was statistically significant for sensitivity ($I^2 = 97.63$; 95% CI, 97.17–98.09) and specificity ($I^2 = 98.22$; 95% CI, 97.91–98.54). Differences in study methodology or clinical characteristics of the sample may result in elevated heterogeneity. Heterogeneity may also be elevated due to different thresholds utilized across studies to define cases. This was not the case for the SCOFF as the threshold effect was not significant ($r = -0.21$; $p = 0.32$).

In order to address the significant heterogeneity found across studies, subgroup analyses were conducted to examine the impact of methodological (i.e., case-control vs non-case-control;

interview vs questionnaire reference standard) and clinical characteristics (age, gender, location, and diagnosis) on diagnostic accuracy. Table 3 presents pooled sensitivity, specificity, and heterogeneity values for each subgroup. The diagnostic accuracy of the SCOFF was higher in case-control studies ($p < 0.01$), when an interview was used as a reference standard as opposed to a questionnaire ($p = 0.05$) and when the percentage of women in the sample was larger than the percentage of men ($p < 0.01$). Additionally, diagnostic accuracy was higher when risk of bias was high for patient selection ($p < 0.01$). Sensitivity and specificity were lower in studies which included individuals diagnosed with BED; however, this difference was not significant ($p = 0.22$). Of note, subgroup analysis did not explain the high overall heterogeneity of the included studies as all subgroups had an I^2 value of greater than 60%.

The likelihood ratio scattergram (Fig. 5) shows the distribution of positive and negative likelihood ratios. The pooled positive likelihood ratio of 5.0 (95% CI, 3.6–6.8) suggests that the SCOFF is moderately helpful in detecting eating disorders. The negative likelihood ratio of 0.17 (95% CI, 0.11–0.27) suggests that the SCOFF is moderately helpful in ruling out the presence of an eating disorder.^{33, 34}

DISCUSSION

We conducted a meta-analysis of 25 validation studies on the SCOFF to determine whether this screen is a valid tool for identifying eating disorders in diverse settings and populations.

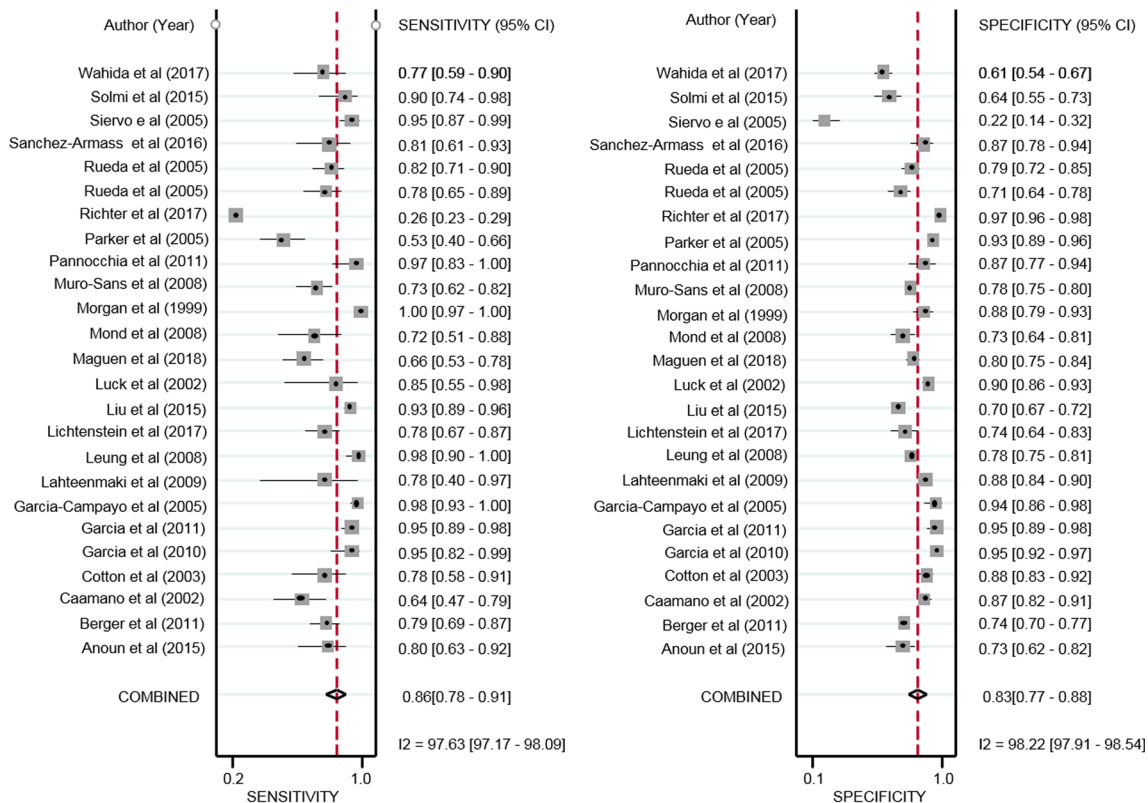


Figure 3 Forest plot of included studies.

Our in-depth examination of the SCOFF calls into question the effectiveness of this tool for eating disorder screening in primary care and community settings, with diverse populations, and with the full range of DSM-5 eating disorder diagnoses. This examination provides a critical context given that we also found in our study, as was found in a previous meta-analysis,³⁵ that the SCOFF is an effective tool for identifying the presence of particular eating disorders (i.e., AN and BN) in the population for which it was initially developed (i.e., young women with eating disorder symptoms).

The purpose of screening is to capture the range of pathological eating and identify cases that might not be identified by other means. The SCOFF was originally developed and subsequently validated several times using case-control study designs. Case-control studies dramatically limit samples to a specific target population (i.e., cases and matched controls) and do not capture the diversity and range of disorders in the general population. Additionally, validity data from case-control studies may artificially inflate the efficacy of screening measures and lead to erroneous conclusions about the utility of the measure in the general population.⁶ As expected, our analyses revealed that the highest levels of sensitivity were found in case-control studies including young women diagnosed with AN and BN. These findings are important as higher rates of sensitivity and specificity in case-control samples highlight that when patients are at risk for AN and BN, the SCOFF is a highly robust screening measure. Conversely, studies with lower sensitivity rates were primarily

recruited from community samples and included the highest reported rates of BED. Sensitivity was also lower in locations where rates of obesity tend to be higher (e.g., North America).

Comparing demographic variables across studies shows that while the SCOFF has been validated numerous times

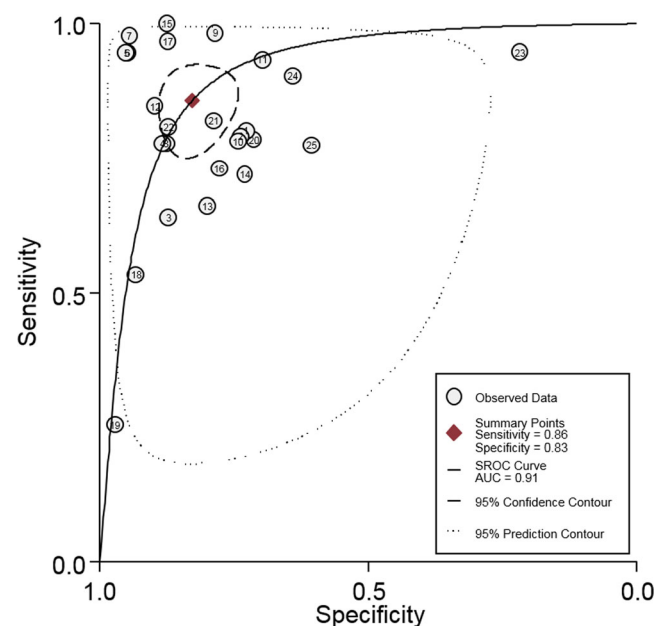


Figure 4 Receiver operating curve (SROC).

Table 3 Subgroup Analyses

Subgroup	N	Sensitivity (95% CI)	Specificity (95% CI)	Heterogeneity I ²
Methodologic				
Recruitment				
Case-control	5	0.96 (0.92–1.00)	0.89 (0.81–0.98)	85 (68–100)
Random/consecutive sample	20	0.80 (0.73–0.88)	0.81 (0.75–0.87)	
Reference standard type				
Interview	14	0.91 (0.86–0.96)	0.81 (0.74–0.89)	66 (24–100)
Questionnaire	11	0.75 (0.63–0.87)	0.84 (0.77–0.92)	
Quality				
Patient selection (bias)				
High bias	5	0.95 (0.91–1.00)	0.78 (0.63–0.94)	97 (95–99)
Low bias	16	0.79 (0.70–0.89)	0.84 (0.78–0.91)	
Patient selection (applicability)				
High applicability concerns	14	0.91 (0.85–0.96)	0.81 (0.73–0.89)	94 (89–99)
Low applicability concerns	9	0.72 (0.57–0.87)	0.86 (0.78–0.94)	
Flow and timing				
High bias	12	0.81 (0.63–0.99)	0.76 (0.61–0.91)	99 (98–99)
Low bias	4	0.81 (0.71–0.91)	0.87 (0.82–0.93)	
Demographics				
Diagnosis*				
BED included	6	0.77 (0.60–0.94)	0.78 (0.65–0.91)	33 (0–100)
BED not included	19	0.88 (0.82–0.94)	0.84 (0.78–0.90)	
Gender†				
Majority female	19	0.88 (0.81–0.93)	0.82 (0.74–0.88)	89 (78–100)
Equal female and male	5	0.75 (0.46–0.91)	0.86 (0.73–0.93)	
Sample				
Medical/research settings	12	0.90 (0.78–0.96)	0.86 (0.80–0.91)	—
Schools/universities	9	0.83 (0.74–0.89)	0.81 (0.73–0.87)	
Community	4	0.71 (0.45–0.96)	0.86 (0.74–0.97)	
Age‡				
Adults	12	0.84 (0.74–0.94)	0.86 (0.79–0.93)	83 (65–100)
Children, adolescents, young adult	12	0.86 (0.74–0.95)	0.79 (0.69–0.88)	
Location§				
Europe	17	0.88 (0.82–0.95)	0.84 (0.78–0.98)	97 (95–99)
North America	4	0.69 (0.43–0.95)	0.85 (0.72–0.98)	

*Diagnosis: Binge eating disorder assessed in study and rates reported in manuscript; binge eating disorder not assessed and/or rates not included in manuscript

† Majority female: sample is ≥ 60% female; Equal female and men: sample is < 60% female

‡ Adults: average age is ≥ 25; Children, adolescents, young adult: average age is < 25

§ There were too few studies conducted in Asia and South America to obtain pooled results

since its development, it is often validated in samples highly similar to the population in which it was initially validated

(i.e., young women with AN and BN). In fact, of the 25 studies reviewed, more than half utilized a predominately or entirely

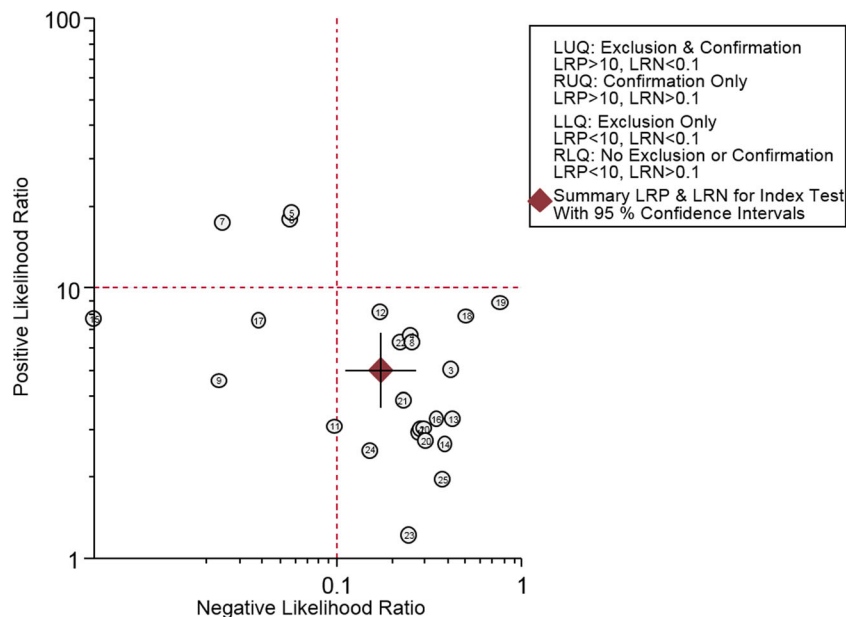


Figure 5 Likelihood ratio scattergram.

female sample. Of the studies that did include males, only three were conducted using adults. In addition, many studies did not report on important demographic and clinical characteristics including certain eating disorder diagnoses (e.g., BED), race, and BMI. Of those that did report on these characteristics, there was evidence that samples utilized in these validation studies often did not reflect the racial and weight diversity seen across DSM-5 eating disorders outside of AN and BN. In addition, only six studies explicitly examined the efficacy of the SCOFF for identifying BED and none examined efficacy for any of the other specified eating disorders in DSM-5. Reflecting the lack of demographic variability in the samples across the 25 studies, applicability concerns were high in many studies on the QUADAS-2 risk of bias tool in the patient selection domain. Given these high applicability concerns, it is difficult to make conclusions about the appropriateness of using the SCOFF to screen for eating disorders with the exception of young women at risk for AN and BN.

Compared with a prior systematic review on this topic conducted by Botella and colleagues,³⁵ the present systematic review provides a more in-depth and comprehensive analysis and, most importantly, includes an assessment of the quality of included studies. Additionally, ten new validation studies had been published and were included in our analysis for a total of 25 validation studies. As per PRISMA-DTA guidelines, this review also includes subgroup analyses. There were, however, several limitations to the current review. First, the literature search was limited to PubMed and Cochrane Library databases. Other databases were referenced in conducting initial searches; however, they were not included in the final search. Systematic reviews should include a range of databases as part of the final, systematic search strategy in general so that any possible articles are captured. The search was also limited to articles written or translated into English. These search limitations could have resulted in missing articles which might otherwise be included. With this being stated, there were no articles that the reviewers encountered that included validation of the SCOFF questionnaire and were inaccessible in English. This review was also limited to examining the validity of the SCOFF. A more comprehensive review of all eating disorder screening measures might have provided additional information regarding eating disorder screening; however, the SCOFF is the screening measure with the most extensive validity data and is frequently used in clinical practice. Another limitation was that we were unable to conduct subgroup analyses for other potentially relevant clinical characteristics (e.g., BMI, race, and ethnicity) as these variables were infrequently reported in the validation studies.

Conclusions

The current review was conducted to address concerns about the use of the SCOFF as a primary care screener for DSM-5 eating disorders, including BED and other specified eating disorders. Findings revealed that the psychometrics of the SCOFF are virtually unknown for the full range of DSM-5 eating disorder

diagnoses and for diverse populations. The present review suggests that the SCOFF is a highly sensitive screening measure for young women at risk for AN and BN but analyses and quality assessment of studies raised concerns about the generalizability and reliability of these results for other eating disorder diagnoses. Currently, there is insufficient evidence to recommend the use of the SCOFF for large-scale screening in primary care and diverse community settings. This review identifies the need for the development of a new screening tool, or multiple tools, for validation for the full range of DSM-5 eating disorder diagnoses in heterogeneous samples.

Corresponding Author: Amanda M. Kutz, PhD; VA Salt Lake City Health Care System, Salt Lake City, UT, USA (e-mail: Amanda.kutz@va.gov).

Author Contributions RMM, SM, and AMK were involved in conception of the review. AMK and AGM were involved in review of literature, extraction of data, and creation of data tables. AMK and CGG were involved in statistical analyses. CGG provided expert guidance on systematic review manuscript preparation. AMK, AGM, and RMM drafted the manuscript. All authors were involved in critical review of the manuscript.

Funding Information This project was supported in part by the VA's Health Services Research and Development (CIN 13-407) (HSR&D) Center of Innovation (COIN) Pain Research, Informatics, Multi-morbidities, and Education (PRIME) Center, West Haven, CT.

Compliance with Ethical Standards:

Conflict of Interest: The authors declare that they do not have a conflict of interest.

Disclaimer: The content of this research is solely the responsibility of the authors and does not necessarily represent the official views of the VA or the Veterans Health Administration.

Publisher's Note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. **Hudson JI, et al.** The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biol Psychiatry*. 2007;61(3):348-358.
2. **Ogg EC, et al.** General practice consultation patterns preceding diagnosis of eating disorders. *Int J Eat Disord*. 1997;22(1): 89-93.
3. **Strother E, et al.** Eating disorders in men: underdiagnosed, undertreated, and misunderstood. *Eat Disord*. 2012;20(5):346-355.
4. Association, A.P Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub; 2013.
5. **McInnes MD, et al.** Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. 2018;319(4):p. 388-396.
6. **Whiting PF, et al.** QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
7. **Liu C-Y, et al.** Sex difference in using the SCOFF questionnaire to identify eating disorder patients at a psychiatric outpatient clinic. *Compr Psychiatry*. 2015;57:160-166.
8. **Maguen S, et al.** Screen for Disordered Eating: Improving the accuracy of eating disorder screening in primary care. *Gen Hosp Psychiatry*. 2018;50:20-25.
9. **Dwamena B.** MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies. 2009.

10. **Aoun A, et al.** Validation of the Arabic version of the SCOFF questionnaire for the screening of eating disorders. *East Mediterr Health J.* 2015;21(5).
11. **Berger U, et al.** Screening of disordered eating in 12-Year-old girls and boys: psychometric analysis of the German versions of SCOFF and EAT-26. *Psychother Psychosom Med Psychol.* 2011;61(7):311-318.
12. **Caamaño F, et al.** Validation of SCOFF questionnaire among preteenagers. *BMJ.* 2002.
13. **Cotton MA, Ball C, Robinson P.** Four simple questions can help screen for eating disorders. *J Gen Intern Med.* 2003;18(1):53-56.
14. **García FD, et al.** Validation of the French version of SCOFF questionnaire for screening of eating disorders among adults. *World J Biol Psychiatry.* 2010;11(7):888-893.
15. **García FD, et al.** Detection of eating disorders in patients: validity and reliability of the French version of the SCOFF questionnaire. *Clin Nutr.* 2011;30(2):178-181.
16. **García-Campayo J, et al.** Validation of the Spanish version of the SCOFF questionnaire for the screening of eating disorders in primary care. *J Psychosom Res.* 2005;59(2):51-55.
17. **Lähteenmäki S, et al.** Validation of the Finnish version of the SCOFF questionnaire among young adults aged 20 to 35 years. *BMC Psychiatry.* 2009;9(1):5.
18. **Leung SF, et al.** Psychometric properties of the SCOFF questionnaire (Chinese version) for screening eating disorders in Hong Kong secondary school students: a cross-sectional study. *Int J Nurs Stud.* 2009;46(2):239-247.
19. **Lichtenstein MB, Hemmingsen SD, Støving RK.** Identification of eating disorder symptoms in Danish adolescents with the SCOFF Questionnaire. *Nordic J Psychiatry.* 2017;71(5):340-347.
20. **Luck AJ, et al.** The SCOFF questionnaire and clinical interview for eating disorders in general practice: comparative study. *BMJ.* 2002;325(7367):755-756.
21. **Mond JM, et al.** Screening for eating disorders in primary care: EDE-Q versus SCOFF. *Behav Res Ther.* 2008;46(5):612-622.
22. **Morgan JF, Reid F, Lacey JH.** The SCOFF questionnaire: assessment of a new screening tool for eating disorders. *BMJ.* 1999;319(7223):1467-1468.
23. **Muro-Sans P, Amador-Campos JA, Morgan JF.** The SCOFF-c: Psychometric properties of the Catalan version in a Spanish adolescent sample. *J Psychosom Res.* 2008;64(1):81-86.
24. **Pannocchia L, et al.** A psychometric exploration of an Italian translation of the SCOFF questionnaire. *Eur Eat Disord Rev.* 2011;19(4):371-373.
25. **Parker SC, Lyons J, Bonner J.** Eating disorders in graduate students: exploring the SCOFF questionnaire as a simple screening tool. *J Am Coll Heal.* 2005;54(2):103-107.
26. **Richter F, et al.** Screening disordered eating in a representative sample of the German population: Usefulness and psychometric properties of the German SCOFF questionnaire. *Eat Behav.* 2017;25:81-88.
27. **Rueda GE, et al.** Validación de la encuesta SCOFF para tamizaje de trastornos de la conducta alimentaria en mujeres universitarias. *Biomedica.* 2005;25(2):196-202.
28. **Rueda GJ, et al.** Validation of the SCOFF questionnaire for screening the eating behaviour disorders of adolescents in school. *Aten Primaria.* 2005;35(2):89-94.
29. **Sanchez-Armass O, et al.** Validation of the SCOFF questionnaire for screening of eating disorders among Mexican university students. *Eat Weight Disord.* 2017;22(1):153-160.
30. **Servo M, et al.** Application of the SCOFF, Eating Attitude Test 26 (EAT 26) and Eating Inventory (TFEQ) Questionnaires in young women seeking diet-therapy. *Eat Weight Disord.* 2005;10(2):76-82.
31. **Solmi F, et al.** Validation of the SCOFF questionnaire for eating disorders in a multiethnic general population sample. *Int J Eat Disord.* 2015;48(3):312-316.
32. **Wahida WMZW, Lai PSM, Hadi HA.** Validity and reliability of the english version of the sick, control, one stone, fat, food (SCOFF) in Malaysia. *Clin Nutr ESPEN.* 2017;18:55-58.
33. **McGee S.** Simplifying likelihood ratios. *J Gen Intern Med.* 2002;17(8):647-650.
34. **Wilson MC, Henderson MC, Smetana GW.** Chapter 5. Evidence-Based Clinical Decision Making. In: Henderson MC, Tierney LM, Jr., Smetana GW. eds. *The Patient History: An Evidence-Based Approach to Differential Diagnosis.* New York: McGraw-Hill; 2012.
35. **Botella J, et al.** A meta-analysis of the diagnostic accuracy of the SCOFF. *Span J Psychol.* 2013;16:E92.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.