

# Diagnostic Accuracy of Spanish Language Depression-Screening Instruments

Daniel S. Reuland, MD, MPH<sup>1</sup>

Andrea Cherrington MD, MPH<sup>2</sup>

Garth S. Watkins, MD, MPH<sup>3</sup>

Daniel W. Bradford, MD, MPH<sup>4</sup>

Roberto A. Blanco, MD<sup>3</sup>

Bradley N. Gaynes, MD, MPH<sup>3</sup>

<sup>1</sup>Department of Medicine, Division of General Medicine and Clinical Epidemiology, and the Cecil G. Sheps Center for Health Services Research, University of North Carolina, Chapel Hill, North Carolina

<sup>2</sup>Department of Medicine, Division of General Internal Medicine, University of Alabama-Birmingham, Birmingham, Alabama

<sup>3</sup>Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

<sup>4</sup>Psychosocial Rehabilitation and Recovery Center, Durham VA Medical Center, Durham, and Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, North Carolina

*Conflict of interest: Dr Blanco was awarded the American Psychiatric Association/Bristol-Myers Squibb Fellowship in Public Psychiatry for 2008-2010 prior to submission. Within the past 12 months, Dr Gaynes has received grants and research support from the National Institute of Mental Health, Agency for Healthcare Research and Quality, the M-3 Corporation, Bristol-Myers Squibb Company, and Novartis. He has performed as an advisor or consultant for Bristol-Myers Squibb Company.*

## CORRESPONDING AUTHOR

Daniel S. Reuland MD, MPH  
Division of General Internal Medicine and Clinical Epidemiology  
University of North Carolina at Chapel Hill  
Campus Box 7110  
Chapel Hill, NC 27599 – 7110  
dreuland@med.unc.edu

## ABSTRACT

**PURPOSE** To make decisions about implementing systematic depression screening, primary care physicians who serve Spanish-speaking populations need to know whether Spanish language depression-screening instruments are accurate. We aimed to review systematically the evidence regarding diagnostic accuracy of depression-screening instruments in Spanish-speaking primary care populations.

**METHODS** We searched PubMed, PsycINFO, CINAHL, EMBASE, and Cochrane Libraries from inception to May 28, 2008, for studies examining the diagnostic accuracy of Spanish language depression case-finding instrument(s) administered to primary-care outpatients. Two authors independently assessed studies for inclusion and quality.

**RESULTS** Twelve studies met inclusion criteria. In general primary care screening, the Spanish language version of the Center for Epidemiologic Studies-Depression scale (CES-D) had sensitivities ranging from 76% to 92% and specificities ranging from 70% to 74%. We found no US study reporting the accuracy of the Primary Care Evaluation of Mental Disorders (PRIME-MD-9) or the Patient Health Questionnaire (PHQ-9) depression module in Spanish-speakers. One fair-quality European study and 1 poor-quality study conducted in Honduras found the 9-item PRIME-MD had sensitivities ranging from 72% to 77% and specificities ranging from 86% to 100%. The 2-item PRIME-MD was 92% sensitive, but only 44% specific for depression in 1 US study. In geriatric outpatients, the 15-item Spanish language version of the Geriatric Depression Scale (GDS) had sensitivities ranging from 76% to 82%, and specificities ranging from 64% to 98%. In postpartum women, the Spanish language version of the Edinburgh Postnatal Depression Scale (EPDS) was 72% to 89% sensitive and 86% to 95% specific for major depression (2 non-US studies). The Spanish language version of the Postpartum Depression Screening Scale (PDSS) was 78% sensitive and 85% specific for combined major/minor depression (1 US study).

**CONCLUSIONS** For depression screening in Spanish-speaking outpatients, fair evidence supports the diagnostic accuracy of the CES-D and PRIME-MD-9 in general primary care, the GDS-15-Spanish for geriatric patients, and the Spanish language versions of the EPDS or PDSS for postpartum patients. The ultrashort 2-item version of PRIME-MD may lack specificity in US Spanish-speakers.

*Ann Fam Med* 2009;7:455-462. doi:10.1370/afm.981.

## INTRODUCTION

Depression is common and costly, and it causes considerable suffering for patients and their families. Recent advances in clinical detection and treatment of depression led to a 2002 US Preventive Services Task Force (USPSTF) recommendation for depression screening in primary care settings.<sup>1</sup> Despite progress in this area, however, evidence also suggests disparities exist in detection and treatment of depression and other mental health disorders by ethnic and racial minorities.<sup>2,3</sup> These disparities are particularly apparent in US Latinos,<sup>4-7</sup> who now constitute the nation's largest minority group.

Although legal and economic barriers contribute to these disparities, language and cultural barriers also play an important role.<sup>8</sup> In recent years, the number of US residents who are Spanish speakers has risen dramatically. More than 31 million US residents now speak Spanish at home.<sup>9</sup> Within this changing socio-cultural context, those who provide primary care to Spanish-speaking populations must make decisions about implementing systematic depression screening in their clinics and health centers. Among the key pieces of information clinicians and clinical leaders need in order to make these decisions is how the various available depression-screening instruments perform in Spanish-speaking primary care populations.

Despite their large and growing numbers, US Latinos with limited English proficiency are poorly represented in large clinical studies of depression screening. In a large systematic review and meta-analysis<sup>10</sup> on which current USPSTF recommendations are based, only 6 of the 41 included studies reported enrolling more than 5% Hispanic participants. Moreover, even when Hispanics have been enrolled in trials of depression screening, studies typically exclude non-English speakers,<sup>11,12</sup> do not report the proportions of enrolled participants who are Spanish speakers,<sup>13-16</sup> or do not report the diagnostic accuracy of the screening instrument used for Spanish speakers.<sup>17-20</sup>

Because clinical manifestations of mental health disorders are partly dependent upon linguistically and culturally determined conceptions of illness, evidence based mainly on studies in English-speaking populations may not be generalizable to non-English-speaking populations. Language and culture can affect patterns of endorsement of depression scale items and dimensions, even when a well-translated instrument is used.<sup>17,21-23</sup> Hence, the accuracy of a depression-screening instrument that is translated from English into Spanish may be different from the accuracy of the original version.

We undertook a systematic review of the literature aimed at summarizing the evidence regarding diagnostic accuracy of depression-screening instruments for Spanish-speaking populations studied in primary care settings.

## METHODS

### Search Strategy

We searched PubMed for articles through May 28, 2008, using medical subject headings (MeSH) terms "depressive disorder" or "depression," and combined this result with the MeSH terms "mass screening" and with the terms "Hispanic Americans" (includes Latino) or "Mexican Americans." We then searched the PubMed database using the same publication date

range for the MeSH term "Spanish" combined with MeSH terms "screening instrument" or "tool" or "questionnaire" and "depressive disorder" or "depression." We used analogous search strategies to search for additional articles in PsycINFO, CINAHL, and EMBASE. Results from all 3 searches were combined, and duplicate abstracts were eliminated. We supplemented these sources by searching the Cochrane collection database for articles on "depression," "neurosis," and "anxiety disorders," searching further by hand the bibliographies from the articles yielded from the literature search, and searching evidence tables and bibliographies from pertinent reviews of depression screening in primary care. A doctoral-level research librarian assisted with all electronic searching.

### Eligibility and Exclusion Criteria

To be eligible, studies had to report data describing the diagnostic accuracy of 1 or more Spanish language versions of depression-screening instruments compared with a reference standard. Acceptable reference standards included either a diagnosis by a mental health professional or a structured diagnostic interview, such as the Structured Clinical Interview for the *DSM-IV* (SCID),<sup>24</sup> Schedule for Clinical Assessment in Neuropsychiatry (SCAN),<sup>25</sup> or the Composite International Diagnostic Interview (CIDI).<sup>26</sup> The diagnostic standard had to be applied both to screen-positive and screen-negative cases, or a randomly selected subsample thereof. Eligible studies had to provide both the sensitivities and specificities of the screening instrument compared with the reference standard, or data that would allow calculation of these values. We included studies that examined the detection of related mood disorders, such as dysthymia, minor (subsyndromal) depression, and anxiety disorders, provided that they also examined the detection of major depression.

Studies had to enroll unselected patients in an outpatient primary care setting, defined as family medicine, internal medicine, general practice, pediatrics, geriatrics, and general obstetrics-gynecology, including prenatal and postpartum care. Because we sought evidence generalizable to primary care practice, we excluded studies if they were performed in nonclinical, community-based settings, or in disease-specific or referral populations. We excluded studies from inpatient, institutional, or psychiatric settings. As other reviews of depression screening have done,<sup>10,27</sup> we included studies in primary care settings outside the United States, though we abstracted them into a separate table. We did not limit our search by language. Two authors independently reviewed abstracts of identified articles. Studies were excluded at this stage if both reviewers agreed that eligibility criteria were

clearly not met. If either reviewer could not exclude the study based on the abstract, the full article was reviewed independently by 2 authors. If these authors were discordant in their assessment, a third author independently reviewed the article, and consensus was reached by discussion among all 3 authors.

### Data Abstraction

Data from articles published in English or Spanish were abstracted by 1 author directly into evidence tables. A second author confirmed the accuracy of the data abstraction. We abstracted sensitivity and specificity data that corresponded to "standard" (as reported by study authors) cut points. When the standard cut points for a given screening instrument varied between studies, however, we chose a common cut point to allow for comparisons among studies. When needed, we calculated 95% confidence intervals (CIs) for sensitivity and specificity using data from the articles. We also abstracted area under the curve (AUC) as an overall measure of an instrument's diagnostic accuracy, when available.

### Quality Appraisal

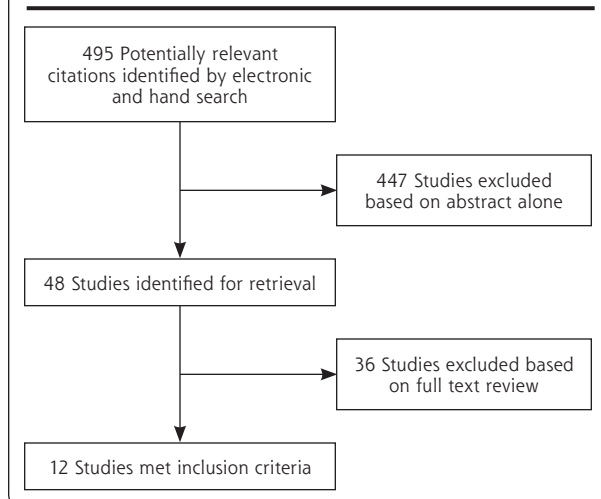
Studies meeting inclusion criteria were independently rated as good, fair, or poor by 2 authors. Differences were resolved by consensus. We used the QUADAS (a tool for the quality assessment of studies of diagnostic accuracy) checklist<sup>28</sup> to guide the appraisals, and we emphasized independent and blind administration of the screening test and the diagnostic reference standard, as well as the potential for bias in selection of subjects for screening or for administration of the diagnostic assessment.

## RESULTS

Of 495 studies identified by the initial electronic and hand searches, we eliminated 447 based on abstract alone (Figure 1). The most common reasons for exclusion at this level were that studies were not conducted in a primary care setting and/or that there was no comparison of a screening instrument against a diagnostic standard. We reviewed the remaining 48 potentially eligible studies in full. Of these, 12 studies met inclusion criteria.

Included studies were heterogeneous with respect to care setting, demographic subpopulations studied, screening instrument used, and quality. Three studies were conducted in the United States: 1 in Puerto Rican outpatients,<sup>29</sup> 1 in postpartum women in Texas and Connecticut,<sup>30</sup> and 1 in immigrants from Mexico and Central America attending an urban general medical clinic in California.<sup>31</sup> Nine additional studies

**Figure 1. Flow diagram showing selection of studies addressing Spanish language depression-screening instruments in primary care.**



conducted in non-US settings met our inclusion criteria.<sup>32-40</sup> The most common, important methodologic limitation was lack of documentation that administration of the screening test and reference standard were blinded and independent of each other.

### Detailed Assessment, US Studies

Details of the US studies assessing depression-screening instruments are displayed in Table 1. Robison et al<sup>29</sup> determined diagnostic accuracy of 4 different depression-screening instruments: the Yale 1 Question, the 2-item version of the Primary Care Evaluation of Mental Disorders (PRIME-MD-2), both the 30-item and 15-item versions of the Geriatric Depression Scale (GDS), and the Center for Epidemiologic Studies-Depression scale (CES-D, both the 20-item and 10-item versions) in a sample of 303 middle-aged and older Puerto Ricans living in the northeastern United States. They found the 20-item and 10-item versions of the CES-D performed best, with sensitivities of 73% and 76% and specificities of 72% and 70%, respectively. These figures correspond to a positive likelihood ratio (+LR) of about 2.5 for both of these versions. The 15-item GDS had reasonable sensitivity (76%) but only fair specificity (64%) for major depression in this population, yielding +LR = 2.0. The 2 ultra-short screening instruments, the Yale 1 Question and the PRIME-MD-2, were sensitive for depression (86% and 92%, respectively) but had poor specificity (42% and 44%), yielding +LR = 1.5 and 1.6 only.

Beck and Gable assessed diagnostic accuracy of the Spanish version of the Postpartum Depression Screening Scale (PDSS),<sup>30</sup> an instrument they had previously

developed,<sup>41,42</sup> in 150 postpartum women. The full instrument contains 35 items and was administered in written form, effectively excluding subjects with low

literacy. The full PDSS was 84% sensitive and 84% specific for the presence of either major or minor postpartum depression (+LR = 5.3). The 7-item version per-

**Table 1. Studies of Diagnostic Accuracy of Spanish Language Depression-Screening Instruments in US Primary Care Settings**

Study	Sample Characteristics, Recruitment Setting (Country, %)	Reference Standard, Depressive Disorder (Prevalence)	Screening Instrument (No. of Items)	Score Range	Cut Point (≥)
Robison, 2002 <sup>29</sup>	303 Middle-aged and older patients from urban primary/acute care clinics in Connecticut (Puerto Rico, 100%)	CIDI Major depression (12%)	Yale 1 Question (1)	0-1	1
			PRIME MD-2 (2)	0-2	1
			GDS-30 (30)	0-30	9 <sup>a</sup>
			GDS-15 (15)	0-15	6
			CES-D (20)	0-60	21
Beck, 2005 <sup>30</sup>	150 Postpartum women aged 16-44 y in Texas and Connecticut (Puerto Rico, 43%; Mexico, 43%; other, 14%)	SCID Combined major and minor postpartum depression <sup>b</sup> (37%)	PDSS-Spanish, Full (35)	35-175	60
			Short (7)	7-35	14
Ring, 1991 <sup>31</sup>	48 Adult outpatients at 1 medical clinic in San Francisco, California (El Salvador, 40%; Nicaragua, 32%; Mexico, 16%)	SCID, Major depression (28%)	CES-D (20)	0-60	21

AUC = area under curve; CES-D = Center for Epidemiologic Studies-Depression scale; CI = confidence interval; CIDI = Composite International Diagnostic Interview; GDS-15 and GDS-30 = 15- and 30-item version of the Geriatric Depression Scale, respectively; NR = not reported; PDSS-Spanish = Postpartum Depression Screening Scale-Spanish version; PRIME-MD-2 = 2-item version of the Primary Care Evaluation of Mental Disorders; SCID = Structured Clinical Interview for the *Diagnostic and Statistical Manual of Mental Disorders (III or IV)*.

<sup>a</sup> Although some experts cite cut points of 10 and 5 as standard for the GDS-30 and GDS-15, these investigators reported results for cut points of 9 and 6, respectively, for these 2 instruments.

<sup>b</sup> The authors provided sensitivity and specificity data only for combined major and minor depression (not for major depression alone).

**Table 2. Studies of Diagnostic Accuracy of Spanish Language Depression-Screening Instruments in Non-US Primary Care Settings**

First Author, Publication Year (Language)	Sample Characteristics, Recruitment Setting (Country)	Reference Standard, Depressive Disorder (Prevalence)	Screening Instrument (No. of Items)	Score Range	Cut Point <sup>a</sup> (≥)
Aguilar-Navarro, 2007 (Spanish) <sup>32</sup>	199 Patients ages >64 y; outpatient geriatric clinic (Mexico)	SCID Depression <sup>b</sup> (56%)	MHAS (9)	0-9	5
Vega-Dienstmaier, 2002 (Spanish) <sup>33</sup>	321 Postpartum women (Peru)	SCID Major depression (7%)	EPDS (10)	0-30	13
Wulsin, 2002 (English) <sup>34</sup>	34 Mothers of young children; primary care clinics (Honduras)	SCID Major depression (38%)	PHQ-9 (9)	0-27	10
Aragónés Benaiges, 2001 (Spanish) <sup>35</sup>	350 Outpatients; primary care clinic (Spain)	SCID Major Depression (15%)	SDS (20)	25-100	50
Baca, 1999 (Spanish) <sup>36</sup>	312 Primary care patients (Spain)	SCAN Depression <sup>b</sup> (28%)	PRIME-MD-9 (9)	0-9	5
Fernández-San Martín, 2002 (English) <sup>39</sup>	192 Patients age 65 y or older; 3 health centers (Spain)	GMS Major or minor depression (31%)	GDS (30)	0-30	9
García-Esteve, 2003 (English) <sup>36</sup>	334 Women; routine postpartum care (Spain)	SCID Major depression (11%) Combined major or minor depression (30%)	EPDS (10)	0-30	13
Martínez de la Iglesia 2002 (Spanish) <sup>43</sup>	249 Outpatients aged 65 y or older; 1 health center (Spain)	MADRS Depression <sup>b</sup> (36%)	GDS-15	15	5
					6
Ortega-Orcos, 2007 (Spanish) <sup>37</sup>	301 Primary care patients aged >64 y (Spain)	DSM-IV-based clinical diagnosis Depression <sup>b</sup> (14.6%)	GDS-5 (5)	0-5	2
			GDS-15 (15)	0-15	5

AUC = area under curve; EPDS = Edinburgh Postnatal Depression Scale; GDS = Geriatric Depression Scale; GMS = Geriatric Mental State Schedule; MADRS = Montgomery-Asberg Depression Rating Scale; MHAS = Mexican Health and Age Study; NR = not reported; PHQ-9 = Patient Health Questionnaire; PRIME-MD-9 = Primary Care Evaluation of Mental Disorders; SCAN = Schedule for Clinical Assessment in Neuropsychiatry; SCID = Structured Clinical Interview for *DSM-IV*; SDS = Zung Self-rating Depression Scale.

<sup>a</sup> When sensitivity and specificity data were presented for multiple cut points, standard cut points (as reported by individual study authors) are presented. Where possible, we also include other cut points as needed to permit between-study comparisons for the same screening instrument.

<sup>b</sup> Depression severity (ie, major vs minor) not specified in article.

formed nearly as well, having a sensitivity of 78% and specificity of 85% (+LR = 5.2).

Ring et al examined the accuracy of the full 20-

item CES-D in 48 Hispanic immigrants at 1 urban outpatient clinic.<sup>31</sup> Using a cut point of 21 of 60, they found it to have a sensitivity of 92% and a specificity of 74% (+LR = 3.5).

### Detailed Assessment, Non-US Studies

Details of the non-US studies assessing depression-screening instruments are displayed in Table 2. Aguilar-Navarro et al examined the accuracy of the 9-item Mexican Health and Age Study (MHAS) in a sample of 199 geriatric outpatients in Mexico.<sup>32</sup> They found the MHAS to be 81% sensitive and 69% specific for "depression." Vega-Dienstmaier et al determined the accuracy of the Spanish language version of the Edinburgh Postnatal Depression Scale (EPDS) in 321 postpartum women in Peru and found a sensitivity of 89% and specificity of 72% for major postpartum depression.<sup>33</sup> Wulsin et al found the Patient Health Questionnaire depression module (PHQ-9) was 77% sensitive and 100% specific for major depression in a sample of Honduran mothers.<sup>34</sup> This study was rated as poor quality, however.

The remaining 6 non-US studies were from Spain. Aragonès Benaiges et al found the Spanish version of the Self-rating Depression Scale (SDS) to be very sensitive (95%) and moderately specific (74%) for major depression among 350 Spanish primary care patients.<sup>35</sup> Baca et al found the PRIME-MD-9 was 72% sensitive and 86% specific for major depression in a sample of 312 primary care patients in Spain.<sup>38</sup> Fernández-San Martín et al found that the 30-item GDS with a cut point of 9 yielded a sensitivity (87%) and specificity of (63%) for combined major and minor depression in 192 elderly Spanish primary care patients.<sup>39</sup> Garcia-Esteve et al<sup>36</sup> found that the Spanish version of the EPDS was highly accurate for detecting major depression (86% sensitivity, 95% specificity) and moderately accurate for detecting both major and minor depression (sensitivity 62%, specificity 98%) in a sample of 334 Spanish postpartum women. Martinez de la Iglesia et al administered the 15-item GDS to 249 geriatric patients, with a resultant sensitivity of 81% and a specificity of 77%.<sup>43</sup> Ortega-Orcos assessed the 15-item GDS and found a sensitivity of 82% and specificity of 98%.<sup>37</sup> A 5-item version was 86% sensitive and 85% specific.

Table 3 displays information for finding the Spanish language versions of the principal screening instruments.

## DISCUSSION

Our systematic review found limited evidence that directly guides primary care-based depression screening for Spanish speakers, a large and rapidly growing segment of the US population. Specifically, we found

Sensitivity % (95% CI)	Specificity % (95% CI)	AUC (SD or 95% CI)	Quality Rating
86 (70-95)	42 (40-44)	NR	Fair
92 (78-98)	44 (42-45)	0.68 (0.61-0.76)	
84 (68-93)	53 (51-55)	0.75 (0.67-0.82)	
76 (60-87)	64 (62-66)	0.76 (0.69-0.84)	
73 (57-85)	72 (70-74)	0.77 (0.70-0.85)	
76 (60-87)	70 (68-71)	0.77 (0.69-0.85)	
84 (71-92)	84 (75-91)	0.93 (0.02)	Fair
78 (65-88)	85 (76-91)	0.88 (0.03)	
92 (62-100)	74 (56-87)	NR	Fair

Sensitivity % (95% CI)	Specificity % (95% CI)	AUC (95%CI)	Quality Rating
81 (72-88)	69 (58-78)	0.79 (0.73-0.86)	Fair
89 (67-99)	72 (67-77)	NR	Fair
77 (46-94)	100 (81-100)	NR	Poor
94 (83-100)	70 (64-75)	0.93	Good
72 (62-81)	86 (80-90)	NR	Fair
87 (78-95)	63 (54-72)	0.85 (0.79-0.91)	Fair
86 (77-92)	95 (91-97)	0.98 (0.97-0.99)	Good
62 (44-79)	98 (96-99)	0.98 (0.97-0.98)	
81 (71-88)	77 (69-83)	0.84 (0.78-0.89)	Good
73 (63-82)	86 (80-91)	NR	
86 (72-94)	85 (80-89)	0.86 (0.80-0.92)	Good
82 (67-91)	98 (95-99)	0.90 (0.83-0.97)	

fair evidence suggesting that the 20-item and 10-item versions of the CES-D are accurate for depression screening in Spanish speakers seen general primary care settings. Although we found limited evidence from non-US studies suggesting that the PRIME-MD-9 accurately detects depression in Spanish speakers, we did not find studies that provided a direct assessment of diagnostic accuracy of the original PRIME-MD-9 or its newer relative, the PHQ-9,<sup>11</sup> in US Spanish speakers.

This review also summarizes evidence regarding the accuracy of instruments designed specifically for screening 2 specific Spanish-speaking primary care subpopulations, geriatric outpatients and postpartum women. For geriatric outpatients, we found evidence from both US and non-US studies that the 30-item and 15-item versions of the GDS-Spanish are accurate depression-screening instruments. For postpartum women, we found limited evidence suggesting that the PDSS-Spanish (1 fair quality US study) and probably the EPDS-Spanish (2 fair-good quality studies from Latin American and Europe) are accurate screening tests for postpartum depression. Because Hispanic women in the United States often have better access to primary health care during the perinatal period,<sup>44</sup> postpartum screening represents an opportunity to detect depression in this at-risk population.

To our knowledge this review of the evidence is the first regarding the diagnostic accuracy of depression-screening instruments in Spanish-speaking primary care populations. By specifically addressing the question of diagnostic accuracy of screening instruments

for Spanish-speaking Hispanic subpopulations, this study complements and extends findings from other published studies of primary care depression screening in US Latinos, especially those that found the PRIME-MD-9<sup>45</sup> and PHQ-9<sup>16,17</sup> are generally accurate across racially and ethnically diverse populations which included Latinos. Taken together, these findings suggest that accurate screening instruments are available for use in Spanish-speaking primary-care patients.

Our findings also point to important gaps in our knowledge that should be the subject of further research. Most notably, we found no evidence directly supporting the accuracy of ultrashort screening instruments in Spanish. Given the competing demands faced by primary care clinicians,<sup>46-48</sup> ultrashort depression-screening instruments are highly desirable. Unfortunately, the only study we identified that evaluated ultrashort screening instruments in a general primary care setting found that the "anhedonia" item in the PRIME-MD-2 lacked specificity compared with studies using the original PRIME-MD-2<sup>14</sup> or the newer PHQ-2 in English speakers.<sup>15</sup> This difference could represent true differential item functioning in US Spanish speakers compared with English speakers, an explanation supported by studies suggesting that US Latino patients with similar levels of depression are more likely to endorse anhedonia than non-Hispanic whites.<sup>17, 22</sup>

Another contributing factor could be that the original PRIME-MD-2 instrument studied by Robison et al<sup>29</sup> used dichotomous (yes/no) response options to questions about recent depressed mood and anhedonia,

**Table 3. Depression-Screening and Diagnostic Instruments Assessed in the Studies**

Instrument	Abbreviation	Location
Center for Epidemiologic Studies-Depression scale	CES-D	<a href="http://patienteducation.stanford.edu/research/cesdesp.pdf">http://patienteducation.stanford.edu/research/cesdesp.pdf</a>
Composite International Diagnostic Interview	CIDI	<a href="http://www.hcp.med.harvard.edu/wmhcdi/instruments.php">http://www.hcp.med.harvard.edu/wmhcdi/instruments.php</a>
<i>Diagnostic and Statistical Manual of Mental Disorders- fourth edition</i>	DSM-IV	<a href="http://www.psychiatryonline.com/referral.aspx?gclid=CKuFsOaR7JoCFQ9JagodtRYpCQ">http://www.psychiatryonline.com/referral.aspx?gclid=CKuFsOaR7JoCFQ9JagodtRYpCQ</a>
Edinburgh Postnatal Depression Scale	EPDS	<a href="http://www.aap.org/practicingsafety/Toolkit_Resources/Module2/EPDS.pdf">http://www.aap.org/practicingsafety/Toolkit_Resources/Module2/EPDS.pdf</a>
Geriatric Depression Scale	GDS	<a href="http://www.stanford.edu/~yesavage/Spanish5.html">http://www.stanford.edu/~yesavage/Spanish5.html</a>
Geriatric Mental State Schedule	GMS	<a href="http://www.liv.ac.uk/gms/">http://www.liv.ac.uk/gms/</a>
Montgomery-Asberg Depression Rating Scale	MADRS	<a href="http://www.cnsforum.com/streamfile.aspx?filename=MADRS.pdf&amp;path=pdf">http://www.cnsforum.com/streamfile.aspx?filename=MADRS.pdf&amp;path=pdf</a>
Mexican Health and Age Study	MHAS	Aguilar-Navarro et al. <sup>32</sup>
Postpartum Depression Screening Scale	PDSS	Beck and Gable <sup>30</sup>
Primary Care Evaluation of Mental Disorders, 9-item version (2-item version)	PRIME-MD-9 (2)	<a href="http://health.utah.gov/rhp/pdf/PHQ-9.pdf">http://health.utah.gov/rhp/pdf/PHQ-9.pdf</a>
Patient Health Questionnaire depression module, 9-item version (2-item version)	PHQ-9 (2)	<a href="http://www.depression-primarycare.org/clinicians/toolkits/materials/forms/phq9">http://www.depression-primarycare.org/clinicians/toolkits/materials/forms/phq9</a> <a href="http://www.docsfortots.org/documents/phqscreeningtool.pdf">http://www.docsfortots.org/documents/phqscreeningtool.pdf</a>
Structured Clinical Interview for the DSM-IV	SCID	<a href="http://cpmcnet.columbia.edu/dept/scid/">http://cpmcnet.columbia.edu/dept/scid/</a>
<i>The Schedules for Clinical Assessment in Neuropsychiatry</i>	SCAN	<a href="http://gdp.ggz.edu/scandocs/">http://gdp.ggz.edu/scandocs/</a>
Zung Self-rating Depression Scale	SDS	Aragonès Benaiges et al <sup>35</sup>

Note: Spanish language versions of the principal screening instruments can be found at the Web sites shown or from corresponding authors for referenced studies.

whereas the newer version (ie, the PHQ-2) allows 4 graded response options that correspond to the number of days during the past 2 weeks that patients are bothered by these symptoms (not at all, several days, more than one-half the days, or nearly every day).<sup>11</sup> Studies suggest the PHQ, with these graded response options, has better sensitivity and specificity for depressive disorders than the original PRIME-MD.<sup>11,15,45</sup> Whether these improved operating characteristics apply to Spanish-speaking Latinos is unclear, particularly for the ultrashort 2-item instrument.

Subsequent research should assess the accuracy of the Spanish language PHQ-2 and/or other brief, practical screening instruments in Spanish-speaking primary care populations. To be most generalizable to US Hispanic patients, such studies should take place in settings where these patients commonly receive care, such as community health centers. Failure to identify brief, accurate screening instruments for use in US Spanish-speaking populations could result in missed diagnoses and wasted resources.

Our study has limitations. First, despite a vigorous attempt to identify relevant articles, it is possible that some were missed or that publication bias led to underreporting of data related to our question. Second, our ability to make strong recommendations for clinicians is hindered by the limited number and quality of studies, as well as by their heterogeneity in setting, patient demographics (including country of origin), screening instrument used, patient selection methods, and educational and literacy level of participants. Third, because the US Hispanic population is not homogeneous, evidence that a depression-screening instrument is accurate in one subpopulation may not generalize to other subpopulations.<sup>49</sup> Fourth, our focus on identifying direct evidence for accuracy among Spanish-speaking patients in general primary care settings led to the exclusion of high-quality studies that enrolled Latino participants but did not report how many of these participants were Spanish speakers. Finally, we caution that our findings apply only to the question of diagnostic accuracy of depression-screening instruments. They do not address other important issues relevant to decisions about implementing systematic screening of Spanish-speaking populations in a practice or health system (eg, availability of treatment). Benefits from depression screening are most likely to accrue when systems are also in place for treatment and follow-up of patients found to be depressed at screening.<sup>1</sup>

In summary, we found fair evidence supporting the diagnostic accuracy of Spanish language versions of the CES-D and PRIME-MD-9 in general primary care patients, the GDS in geriatric patients, and the EPDS and PDSS in postpartum patients. This evidence may

be remarkably limited compared with comparable evidence in English-speaking primary care populations. Available evidence suggests that the original PRIME-MD-2 may be inaccurate (nonspecific) in US Spanish-speaking populations. Our review points to the need for further studies addressing the diagnostic accuracy of ultrabrief, practical depression-screening instruments for use in US Spanish-speaking populations seen in primary care settings.

**To read or post commentaries in response to this article, see it online at <http://www.annfamned.org/cgi/content/full/7/5/455>.**

**Key words:** Depressive disorders; Hispanics; screening; language and cultural barriers

Submitted September 12, 2008, submitted, revised, January 9, 2009; accepted January 21, 2009.

A description of this project and preliminary data were presented at the Special Populations Workgroup of the North Carolina Community Health Center Association, October 3, 2007, Durham, North Carolina.

**Funding support:** Dr Reuland was supported by National Research Service Award (NRSA) Primary Care Research Fellowship, T32 HP14001. Dr Cherrington was supported by the Robert Wood Johnson Clinical Scholars' Program (047948) and the NRSA Primary Care Research Fellowship grant, T32HP14001.

**Acknowledgments:** We thank Lynn Whitener, DrPH, for invaluable assistance with electronic searches, and the University of North Carolina NRSA Primary Care Research fellows for their helpful input on this article.

## References

1. Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2002;136(10):765-776.
2. US Public Health Service. Reports of the Surgeon General. Mental health: culture, race, and ethnicity (supplement). August 2001. <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat5.chapter.971>. Accessed Sep 12, 2008.
3. Wells K, Klap R, Koike A, Sherbourne C. Ethnic disparities in unmet need for alcoholism, drug abuse, and mental health care. *Am J Psychiatry*. 2001;158(12):2027-2032.
4. Vega WA, Kolody B, Aguilar-Gaxiola S, Catalano R. Gaps in service utilization by Mexican Americans with mental health problems. *Am J Psychiatry*. 1999;156(6):928-934.
5. Wells K, Klap R, Koike A, Sherbourne C. Ethnic disparities in unmet need for alcoholism, drug abuse, and mental health care. *Am J Psychiatry*. 2001;158(12):2027-2032.
6. Alegría M, Canino G, Ríos R, et al. Inequalities in use of specialty mental health services among Latinos, African Americans, and non-Latino whites. *Psychiatr Serv*. 2002;53(12):1547-1555.
7. Simpson SM, Krishnan LL, Kunik ME, Ruiz P. Racial disparities in diagnosis and treatment of depression: a literature review. *Psychiatr Q*. 2007;78(1):3-14.
8. Sentell T, Shumway M, Snowden L. Access to mental health treatment by English language proficiency and race/ethnicity. *J Gen Intern Med*. 2007;22(0)(Suppl 2):289-293.
9. US Census Bureau. Facts for features; Hispanic heritage month 2005. September 8, 2005. Report CB05FF.14-3. <http://www.census.gov/Press-Release/www/releases/archives/cb05ff-14-3.pdf>.

10. Pignone M, Gaynes BN, Rushton JL, et al. Screening for depression. Systematic Evidence Review. In: *US Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Health Services/Technology Assessment Text (HSTAT)*. 2002;136:765-776. <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat3.chapter.1996>.
11. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA*. 1999;282(18):1737-1744.
12. Simon GE, Von Korff M. Recognition, management, and outcomes of depression in primary care. *Arch Fam Med*. 1995;4(2):99-105.
13. Whooley MA, Stone B, Soghikian K. Randomized trial of case-finding for depression in elderly primary care patients. *J Gen Intern Med*. 2000;15(5):293-300.
14. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med*. 1997;12(7):439-445.
15. Kroenke KMD, Spitzer RLMD, Williams JBWDSW. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;41(11):1284-1292.
16. 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.
17. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med*. 2006;21(6):547-552.
18. Williams JW Jr, Kerber CA, Mulrow CD, Medina A, Aguilar C. Depressive disorders in primary care: prevalence, functional disability, and identification. *J Gen Intern Med*. 1995;10(1):7-12.
19. Burnam MA, Wells KB, Leake B, Landsverk J. Development of a brief screening instrument for detecting depressive disorders. *Med Care*. 1988;26(8):775-789.
20. Williams JW Jr, Mulrow CD, Kroenke K, et al. Case-finding for depression in primary care: a randomized trial. *Am J Med*. 1999;106(1):36-43.
21. Carballeira Y, Dumont P, Borgacci S, et al. Criterion validity of the French version of Patient Health Questionnaire (PHQ) in a hospital department of internal medicine. *Psychol Psychother*. 2007;80(Pt 1):69-77.
22. Golding JM, Aneshensel CS, Hough RL. Responses to Depression Scale items among Mexican-Americans and non-Hispanic whites. *J Clin Psychol*. 1991;47(1):61-75.
23. Posner SF, Stewart AL, Marín G, Pérez-Stable EJ. Factor variability of the Center for Epidemiological Studies Depression Scale (CES-D) among urban Latinos. *Ethn Health*. 2001;6(2):137-144.
24. First M, Spitzer R, Gibbon M, Williams J. *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)*. Washington, DC: American Psychiatric Press, Inc; 1996.
25. Wing JK, Babor T, Brugha T, et al. Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Arch Gen Psychiatry*. 1990;47(6):589-593.
26. Kessler RC, Abelson J, Demler O, et al. Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI). *Int J Methods Psychiatr Res*. 2004;13(2):122-139.
27. Watson LC, Pignone MP. Screening accuracy for late-life depression in primary care: a systematic review. *J Fam Pract*. 2003;52(12):956-964.
28. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol*. 2006;6:9.
29. Robison J, Gruman C, Gaztambide S, Blank K. Screening for depression in middle-aged and older Puerto Rican primary care patients. *J Gerontol A Biol Sci Med Sci*. 2002;57(5):M308-M314.
30. Beck CT, Gable RK. Screening performance of the postpartum depression screening scale—Spanish version. *J Transcult Nurs*. 2005;16(4):331-338.
31. Ring JM, Marquis P. Depression in a Latino immigrant medical population: an exploratory screening and diagnosis. *Am J Orthopsychiatry*. 1991;61(2):298-302.
32. Aguilar-Navarro SG, Fuentes-Cantú A, Avila-Funes JA, García-Mayo EJ. [Validity and reliability of the screening questionnaire for geriatric depression used in the Mexican Health and Age Study]. *Salud Publica Mex*. 2007;49(4):256-262.
33. Vega-Dienstmaier JM, Mazzotti Suárez G, Campos Sánchez M. [Validation of a Spanish version of the Edinburgh Postnatal Depression Scale]. *Actas Esp Psiquiatr*. 2002;30(2):106-111.
34. Wulsin L, Somoza E, Heck J. The feasibility of using the Spanish PHQ-9 to screen for depression in primary care in Honduras. *Prim Care Companion J Clin Psychiatry*. 2002;4(5):191-195.
35. Aragonès Benaiges E, Masdèu Montalà RM, Cando Guasch G, Coll Borràs G. [Diagnostic validity of Zung's self-rating depression scale on primary care patients]. *Actas Esp Psiquiatr*. 2001;29(5):310-316.
36. García-Esteve L, Ascaso C, Ojuel J, Navarro P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. *J Affect Disord*. 2003;75(1):71-76.
37. Ortega Orcos R, Salinero Fort MA, Kazemzadeh Khajoui A, Vidal Aparicio S, de Dios del Valle R. [Validation of 5 and 15 items Spanish version of the geriatric depression scale in elderly subjects in primary health care setting]. *Rev Clin Esp*. 2007;207(11):559-562.
38. Baca E, Saiz J, Agüera L, et al. [Validation of the Spanish version of PRIME-MD: a procedure for diagnosing mental disorders in primary care]. *Actas Esp Psiquiatr*. 1999;27(6):375-383.
39. Fernández-San Martín MI, Andrade-Rosa C, Molina JD, et al. Validation of the Spanish version of the geriatric depression scale (GDS) in primary care. *Int J Geriatr Psychiatry*. 2002;17(3):279-287. Erratum in: *Int J Geriatr Psychiatry*. 2007;22(7):704.
40. Martínez de la Iglesia J, Onís Vilches MC, Dueñas Herrero R, Aguado Taberné C, Albert Colomer C, Arias Blanco MC. [Abbreviating the brief. Approach to ultra-short versions of the Yesavage questionnaire for the diagnosis of depression]. *Aten Primaria*. 2005;35(1):14-21.
41. Beck CT, Gable RK. Postpartum depression screening scale: Spanish version. *Nurs Res*. 2003;52(5):296-306.
42. Beck CT, Bernal H, Froman RD. Methods to document semantic equivalence of a translated scale. *Res Nurs Health*. 2003;26(1):64-73.
43. Martínez de la Iglesia J, Onís Vilches C, Duenas Herrero R, Albert Colomer C, Aguado Taberne C, Luque Luque R. Versión española del cuestionario de Yesavage abreviado (GDS) para el despistaje de depresión en mayores de 65 años: adaptación y validación. [Spanish version of the brief GDS for depression screening in patients over 65: adaptation and validation]. *Medifam*. 2002;12(10):620-630.
44. Fremstad SC. *Covering New Americans: A Review of Federal and State Policies Related to Immigrants' Eligibility and Access to Publicly Funded Health Insurance*. Henry J Kaiser Family Foundation; 2004.
45. Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *Am J Obstet Gynecol*. 2000;183(3):759-769.
46. Williams JW Jr. Competing demands: Does care for depression fit in primary care? *J Gen Intern Med*. 1998;13(2):137-139.
47. Klinkman MS. Competing demands in psychosocial care. A model for the identification and treatment of depressive disorders in primary care. *Gen Hosp Psychiatry*. 1997;19(2):98-111.
48. Kroenke K. Discovering depression in medical patients: reasonable expectations. *Ann Intern Med*. 1997;126(6):463-465.
49. Weinick RM, Jacobs EA, Stone LC, Ortega AN, Burstin H. Hispanic healthcare disparities: challenging the myth of a monolithic Hispanic population. *Med Care*. 2004;42(4):313-320.