APPENDIX A. SEARCH STRATEGY

Step	Category	Terms	Result[a]
1	**Disorders**	(generalized AND anxiety AND disorder[tiab]) OR panic disorder[tiab] OR "generalized anxiety disorder" OR panic disorder[mesh] OR panic[title/abstract]	12293
2	**Measurement instruments**		
	GAD or PD	"gad7"[tiab] OR "generalized anxiety disorder 7"[tiab] OR "gad-7"[tiab] OR "beck anxiety"[tiab] OR "geriatric anxiety inventory"[tiab] OR "short anxiety screening test"[tiab] OR "hospital anxiety and depression scale"[tiab] OR PHQ[tiab] OR "patient health questionnaire"[tiab] OR "zung anxiety scale"[tiab] OR "penn state worry questionnaire"[tiab] OR "multicenter collaborative panic disorder severity scale"	3801
		OR	
	Broad terms for instruments	"Psychiatric Status Rating Scales"[Mesh] OR questionnaires[MeSH Terms] OR questionnaires[tiab] OR questionnaire[tiab] OR tools[tiab] OR tool[tiab] OR scale[tiab] OR scales[tiab] OR inventory[tiab] OR screening[tiab]	1,094,242
3	**Instrument characteristics**	medical history taking[mh] OR reproducibility of results[mh] OR observer variation[mh] OR sensitivity[tiab] OR specificity[tiab] OR "sensitivity and specificity"[mh] OR likelihood [tiab] OR accura*[tiab]	1,249,615
4	**Combine results and apply limits**	#1 AND #2 AND #3 English and Human and Adult	850

[a]Numbers reflect the result of the PubMed search only.

APPENDIX B. STUDY SELECTION FORM

INCLUSION CRITERIA:

- Sample population is adults age ≥18 years presenting with a somatic symptom or presenting to a medical clinic for a scheduled appointment.
- Setting is primary care (general internal medicine, family medicine, geriatrics) or general medical (emergency department, women's health clinic).
- Intervention is a self-report instrument (index test) designed to screen for or facilitate diagnosis of GAD, PD, or anxiety disorders. The instrument must be feasible in a clinical setting without requiring special equipment and may be performed by a nonexpert.
- Reference standard diagnosis of GAD or PD is made using acceptable criteria (e.g., DSM-III or later, ICD-9 or later) and administered by a trained clinician.
- Study reports a measure of reliability or sensitivity/specificity or the data to calculate at least one of these performance characteristics.
- Study design is prospective comparison of an anxiety questionnaire to a reference standard; reference standard must be applied to all subjects or to a randomly selected subsample that allows correction for verification bias.
- Study must be published in a peer-reviewed publication.

EXCLUSION CRITERIA:

- Study is a non-English language publication. English language articles that address Spanish version of instruments will be included.
- Study is conducted outside of North America, Western Europe, New Zealand or Australia.
- Study populations are patients with current mental illness (e.g., substance abuse disorder), and screening is for comorbid anxiety disorder.
- Anxiety measure and reference standard are performed by the same individual.

APPENDIX C. EXCLUDED STUDIES

All studies listed below were reviewed in their full-text version and excluded for the reason indicated. An alphabetical reference list follows the table.

Reference	Population not of interest	Setting not PC, clinic, or ER	No self-reported index test at screening	Reference standard not acceptable	No instrument characteristics data	Design not prospective	Reference standard not applied correctly	Publication not English	Screening tool not English/ Spanish
Andersson, 2004 (422)		X							
Andjreu, 2008 (1551)	X								
Andreescu, 2008 (124)						X			
Apfeldorf, 1994 (1690)	X								
Argyropoulos, 2007 (247)		X							
Austin, 2006 (321)		X							
Baughman, 1994 (2675)			X						
Beck, 1996 (801)		X							
Behar, 2003 (505)		X							
Berrocal, 2006 (316)		X							
Berrocal, 2006 (362)		X							
Bieling, 1998 (721)		X							
Bobes, 2006 (315)								X	
Bucholz, 1991 (2532)	X								
Bystritsky, 1996 (810)		X							
Clum, 1990 (3010)	X								
Connor, 2001 (2399)	X								
Dammen, 1999 (674)								X	
Eack, 2006 (1478)		X							
Eack, 2008 (149)		X							
Epstein, 2001 (2417)		X							
Farvolden, 2003 (486)		X							
Fleet, 1997 (759)								X	
Gladstone, 2005 (345)	X								
Gloster, 2008 (174)	X								
Jackson, 2007 (243)					X				
Kobak, 1997 (751)	X								
Kuijpers, 2003 (497)							X		
Lowe, 2003 (477)									X
Lykouras, 1996 (2256)		X							

Reference	Population not of interest	Setting not PC, clinic, or ER	No self-reported index test at screening	Reference standard not acceptable	No instrument characteristics data	Design not prospective	Reference standard not applied correctly	Publication not English	Screening tool not English/ Spanish
McQuaid, 2000 (633)							X		
Means-Christensen 2005 (343)				X					
Means-Christensen, 2006 (319)	X								
Meyer, 1990 (946)		X							
Mori, 2003 (3846)				X					
Morlock, 2008 (190)	X								
Mowry, 1990 (2735)		X							
Mussell, 2008 (150)	X								
Newman, 2006 (310)		X							
Novy, 2001 (587)	X								
Olssøn, 2005 (1625)				X					
Parker, 1997 (747)	X								
Parkerson, 1997 (767)				X					
Robinson, 2010 (1021)		X							
Rollman, 2005 (371)							X		
Sandin, 1996 (800)	X								
Senior, 2007 (3868)	X								
Stein, 1999 (2268)							X		
Svanborg, 1994 (872)	X								
Vujanovic, 2007 (227)				X					
Webb, 2008 (114)	X								
Weissman, 1998 (735)			X						
Wetherell, 2007 (271)		X							
Yingling, 1993 (886)					X				

LIST OF EXCLUDED STUDIES

Andersson G, Carlbring P, Kaldo V, et al. Screening of psychiatric disorders via the Internet. A pilot study with tinnitus patients. *Nord J Psychiatry*. 2004;58(4):287-91.

Andjreu Y, Galdón MJ, Dura E, et al. Psychometric properties of the Brief Symptoms Inventory-18 (BSI-18) in a Spanish sample of outpatients with psychiatric disorders. *Psicothema*. 2008;20(4):844-850.

Andreescu C, Belnap BH, Rollman BL, et al. Generalized anxiety disorder severity scale validation in older adults. *Am J Geriatr Psychiatry*. 2008;16(10):813-8.

Apfeldorf WJ, Shear MK, Leon AC, et al. A brief screen for panic disorder. *Journal of Anxiety Disorders*. 1994;8(1):71-78.

Argyropoulos SV, Ploubidis GB, Wright TS, et al. Development and validation of the Generalized Anxiety Disorder Inventory (GADI). *J Psychopharmacol*. 2007;21(2):145-52.

Austin DW, Richards JC, Klein B. Modification of the Body Sensations Interpretation Questionnaire (BSIQ-M): validity and reliability. *Journal of Anxiety Disorders*. 2006;20(2):237-51.

Baughman OL. Rapid diagnosis and treatment of anxiety and depression in primary care: The somatizing patient. *The Journal of Family Practice*. 1994;39(4):373-378.

Beck JG, Stanley MA, Zebb BJ. Characteristics of generalized anxiety disorder in older adults: a descriptive study. *Behav Res Ther*. 1996;34(3):225-34.

Behar E, Alcaine O, Zuellig AR, et al. Screening for generalized anxiety disorder using the Penn State Worry Questionnaire: a receiver operating characteristic analysis. *J Behav Ther Exp Psychiatry*. 2003;34(1):25-43.

Berrocal C, Ruiz Moreno M, Merchan P, et al. The Mood Spectrum Self-Report: validation and adaptation into Spanish. *Depress Anxiety*. 2006;23(4):220-35.

Berrocal C, Ruiz Moreno MA, Gil Villa M, et al. Multidimensional assessment of the Panic-Agoraphobic Spectrum: reliability and validity of the Spanish version of the PAS-SR. *Journal of Anxiety Disorders*. 2006;20(5):562-79.

Bieling PJ, Antony MM, Swinson RP. The State-Trait Anxiety Inventory, Trait version: structure and content re-examined. *Behav Res Ther*. 1998;36(7-8):777-88.

Bobes J, Garcia-Calvo C, Prieto R, et al. Psychometric properties of the Spanish version of the screening scale for DSM-IV Generalized Anxiety Disorder of Carroll and Davidson. *Actas Esp Psiquiatr*. 2006;34(2):83-93.

Bucholz KK, Robins LN, Shayka JJ, et al. Performance of two forms of a computer psychiatric screening interview: Version I of the DISSI. *J Psychiatr Res*. 1991;25(3):117-129.

Bystritsky A, Waikar SV, Vapnik T. Four-dimensional Anxiety and Depression Scale: a preliminary psychometric report. *Anxiety*. 1996;2(1):47-50.

Clum GA, Broyles S, Borden J, et al. Validity and reliability of the panic attack symptoms and cognitions questionnaires. *Journal of Psychopathology and Behavioral Assessment*. 1990;12(3):233-245.

Connor KM, Kobak KA, Churchill LE, et al. Mini-SPIN: A brief screening assessment for generalized social anxiety disorder. *Depress Anxiety*. 2001;14(2):137-140.

Dammen T, Ekeberg O, Arnesen H, et al. The detection of panic disorder in chest pain patients. *Gen Hosp Psychiatry*. 1999;21(5):323-32.

Eack SM, Greeno CG, Lee B-J. Limitations of the Patient Health Questionnaire in Identifying Anxiety and Depression in Community Mental Health: Many Cases are Undetected. *Research on Social Work Practice*. 2006;16(6):625-631.

Eack SM, Singer JB, Greeno CG. Screening for anxiety and depression in community mental health: the beck anxiety and depression inventories. *Community Ment Health J*. 2008;44(6):465-74.

Epstein JF, Barker PR, Kroutil LA. Mode effects in self-reported mental health data. *Public Opin Q*. 2001;65(4):529-549.

Farvolden P, McBride C, Bagby RM, et al. A Web-based screening instrument for depression and anxiety disorders in primary care. *J Med Internet Res*. 2003;5(3):e23.

Fleet RP, Dupuis G, Marchand A, et al. Detecting panic disorder in emergency department chest pain patients: a validated model to improve recognition. *Ann Behav Med*. 1997;19(2):124-31.

Gladstone GL, Parker GB, Mitchell PB, et al. A Brief Measure of Worry Severity (BMWS): personality and clinical correlates of severe worriers. *Journal of Anxiety Disorders*. 2005;19(8):877-92.

Gloster AT, Rhoades HM, Novy D, et al. Psychometric properties of the Depression Anxiety and Stress Scale-21 in older primary care patients. *J Affect Disord*. 2008;110(3):248-59.

Jackson JL, Passamonti M, Kroenke K. Outcome and impact of mental disorders in primary care at 5 years. *Psychosom Med*. 2007;69(3):270-6.

Kobak KA, Taylor LH, Dottl SL, et al. Computerized screening for psychiatric disorders in an outpatient community mental health clinic. *Psychiatr Serv*. 1997;48(8):1048-57.

Kuijpers PM, Denollet J, Lousberg R, et al. Validity of the hospital anxiety and depression scale for use with patients with noncardiac chest pain. *Psychosomatics*. 2003;44(4):329-35.

Lowe B, Grafe K, Zipfel S, et al. Detecting panic disorder in medical and psychosomatic outpatients: comparative validation of the Hospital Anxiety and Depression Scale, the Patient Health Questionnaire, a screening question, and physicians' diagnosis. *J Psychosom Res*. 2003;55(6):515-9.

Lykouras L, Adrachta D, Kalfakis N, et al. GHQ-28 as an aid to detect mental disorders in neurological inpatients. *Acta Psychiatr Scand*. 1996;93(3):212-216.

McQuaid JR, Stein MB, McCahill M, et al. Use of brief psychiatric screening measures in a primary care sample. *Depress Anxiety*. 2000;12(1):21-9.

Means-Christensen AJ, Arnau RC, Tonidandel AM, et al. An efficient method of identifying major depression and panic disorder in primary care. *J Behav Med*. 2005;28(6):565-72.

Means-Christensen AJ, Sherbourne CD, Roy-Byrne PP, et al. Using five questions to screen for five common mental disorders in primary care: diagnostic accuracy of the Anxiety and Depression Detector. *Gen Hosp Psychiatry*. 2006;28(2):108-18.

Meyer TJ, Miller ML, Metzger RL, et al. Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther*. 1990;28(6):487-95.

Mori DL, Lambert JF, Niles BL, et al. The BAI–PC as a Screen for Anxiety, Depression, and PTSD in Primary Care. *Journal of Clinical Psychology in Medical Settings*. 2003;10(3):187-192.

Morlock RJ, Williams VS, Cappelleri JC, et al. Development and evaluation of the Daily Assessment of Symptoms - Anxiety (DAS-A) scale to evaluate onset of symptom relief in patients with generalized anxiety disorder. *J Psychiatr Res*. 2008;42(12):1024-36.

Mowry BJ, Burvill PW. Screening the elderly in the community for psychiatric disorder. *Aust N Z J Psychiatry*. 1990;24(2):203-206.

Mussell M, Kroenke K, Spitzer RL, et al. Gastrointestinal symptoms in primary care: prevalence and association with depression and anxiety. *J Psychosom Res*. 2008;64(6):605-12.

Newman MG, Holmes M, Zuellig AR, et al. The reliability and validity of the panic disorder self-report: a new diagnostic screening measure of panic disorder. *Psychol Assess*. 2006;18(1):49-61.

Novy DM, Stanley MA, Averill P, et al. Psychometric comparability of English- and Spanish-language measures of anxiety and related affective symptoms. *Psychol Assess*. 2001;13(3):347-55.

Olssøn I, Mykletun A, Dahl AA. The hospital anxiety and depression rating scale: A cross-sectional study of psychometrics and case finding abilities in general practice. *BMC Psychiatry*. 2005;5.

Parker G, Roussos J, Hadzi-Pavlovic D, et al. The development of a refined measure of dysfunctional parenting and assessment of its relevance in patients with affective disorders. *Psychol Med*. 1997;27(5):1193-203.

Parkerson GR, Jr., Broadhead WE. Screening for anxiety and depression in primary care with the Duke Anxiety-Depression Scale. *Fam Med*. 1997;29(3):177-81.

Robinson CM, Klenck SC, Norton PJ. Psychometric properties of the Generalized Anxiety Disorder Questionnaire for DSM-IV among four racial groups. *Cognitive Behaviour Therapy*. 2010;39(4):251-261.

Rollman BL, Belnap BH, Mazumdar S, et al. Symptomatic severity of PRIME-MD diagnosed episodes of panic and generalized anxiety disorder in primary care. *J Gen Intern Med*. 2005;20(7):623-8.

Senior, A.C, Kunik, M.E, Rhoades, H.M, et al. Utility of telephone assessments in an older adult population. *Psychol Aging.* 2007:22(2): p. 392-7.

Sandin B, Chorot P, McNally RJ. Validation of the Spanish version of the Anxiety Sensitivity Index in a clinical sample. *Behav Res Ther*. 1996;34(3):283-90.

Stein MB, Jang KL, Livesley WJ. Heritability of anxiety sensitivity: A twin study. *The American Journal of Psychiatry*. 1999;156(2):246-251.

Svanborg P, Asberg M. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. *Acta Psychiatr Scand*. 1994;89(1):21-8.

Vujanovic AA, Arrindell WA, Bernstein A, et al. Sixteen-item Anxiety Sensitivity Index: confirmatory factor analytic evidence, internal consistency, and construct validity in a young adult sample from the Netherlands. *Assessment*. 2007;14(2):129-43.

Webb SA, Diefenbach G, Wagener P, et al. Comparison of self-report measures for identifying late-life generalized anxiety in primary care. *J Geriatr Psychiatry Neurol*. 2008;21(4):223-31.

Weissman MM, Broadhead WE, Olfson M, et al. A diagnostic aid for detecting (DSM-IV) mental disorders in primary care. *Gen Hosp Psychiatry*. 1998;20(1):1-11.

Wetherell JL, Birchler GD, Ramsdell J, et al. Screening for generalized anxiety disorder in geriatric primary care patients. *Int J Geriatr Psychiatry*. 2007;22(2):115-23.

Yingling KW, Wulsin LR, Arnold LM, et al. Estimated prevalences of panic disorder and depression among consecutive patients seen in an emergency department with acute chest pain. *J Gen Intern Med*. 1993;8(5):231-5.

APPENDIX D. DATA EXTRACTION FORM

Data abstraction for anxiety screening in primary care

Reviewer initials Endnote ref #:
First Author: Year Published: Country:
Primary study: 1) Yes 2)No. Linked study: 1) Yes 2) No

I) Study setting
1)Outpatient primary care clinic 2) Specialty clinic (specify):
3)ER 4)OB/GYN or women's health

Comments:

II) Patient presentation: Did the patients present with a physical symptoms
1)Chest pain 2)Unselected 3)Other Symptom: 4)NR-99

III) Type of setting NR -99
1) Academic 2) Community 3) Mixed
4) Other (specify):

Comments:

IV) VA clinics NR -99 1) Only VA 2) Mixed 3) No VA

V) Selection of population for screening NR -99
1) Random 2)Consecutive 3)Convenience
4) Other (specify)

Comments:

VI) Selection of population for criterion standard NR -99
1) Random 2)Consecutive 3)Convenience
4) Other (specify)

Comments:

VII) Description of study population NR -99
Potentially eligible: N=
Met eligibility criteria: N=
Screened: N=
Completed criterion standard:

Comments:

VIII) Age NR-99 (Age is for results not for selection)

Mean age (SD)		Comments:
Age range:		

IX) Gender NR-99

Male (n)=		Comments:
Female (n)=		

X) Ethnicity NR-99

1) Caucasian N=		2) Black N=	
3) Hispanic N =		4) Asian N=	
5) Other N=			

Comments:

XI) Education NR-99

Mean years completed (±SD):		Comments:
Other measures:		

XII) Name of the screening instrument (specify version and number when applicable: eg GAD-7 OR GAD-2)

XIII) Methods of administration of screening test NR-99

1) Self-administered 2) Interviewer administered 3) Via telephone
4) Computer assisted 5) Other (specify):

XIV) What was the criterion standard NR-99

1) DSM IV 2) DSM IIIR 3) DSM III 4) ICD 9/10
5) Research diagnostic criteria (RDC) 6) Other (specify)

XV) Method used to determining standard NR-99

1) SCID 2)DIS 3)CIDI 4)DSM3/4 5)ADIS 6) Other (specify):

XVI) Medical comorbidity: specific diseases or average measures 1) Yes NR-99
List top 3 or measures:

XVII) Psychiatric comorbidity 1) Yes NR-99

	Excluded (1)	<10% (2)	10-25% (3)	>25% (4)	NR (-99)
Depression					
PTSD					
Substance abuse					
Social anxiety					
GAD					
PD					
Other ( )					
Other ( )					

XVIII) Other measures:

Responsiveness: NR-99

Test retest reliability: NR-99

RESULTS

Total sample/ Subgroup . If subgroup, specify:

Test used to detect 1)GAD 2)PD 3) Both Results for multiple cutoffs given: 1) Yes 2)No

Cutoff picked a priori? 1)Yes 2) No Same as traditional cutoff: 1) Yes 2) No

Gold standard ↓

	pos	neg	
pos co= ( )			
neg co= ( )			

Other measures, eg sensitivity, PPV, LR: (give 95% CI or NR-99)

Statistic:	Data	95% CI

Data validated? 1) Yes 2) No Data adjusted for sampling: 1) Yes 2) No

---

Total sample/ Subgroup . If subgroup, specify:

2) Test used to detect 1)GAD 2)PD 3) Both Results for multiple cutoffs given: 1) Yes 2)No

Cutoff picked a priori? 1)Yes 2) No Same as traditional cutoff: 1) Yes 2) No

Gold standard ↓

	pos	neg	
pos co= ( )			
neg co= ( )			

Other Measures., eg sensitivity, PPV, LR: (give 95% CI or NR-99)

Statistic:	Data	95% CI

Data validated? 1) Yes 2) No Data adjusted for sampling: 1) Yes 2) No

Total sample/ Subgroup . If subgroup, specify:

Test used to detect 1)GAD 2)PD 3) Both

Cutoff picked a priori? 1)Yes 2) No

Results for multiple cutoffs given: 1) Yes 2)No

Same as traditional cutoff: 1) Yes 2) No

Gold standard ↓

	pos	neg	
pos co= ( )			
neg co= ( )			

Other Measures., eg sensitivity, PPV, LR: (give 95% CI or NR-99)

Statistic:	Data	95% CI

Data validated? 1) Yes 2) No

Total sample/ Subgroup . If subgroup, specify:

Test used to detect 1)GAD 2)PD 3) Both

Cutoff picked a priori? 1)Yes 2) No

Results for multiple cutoffs given: 1) Yes 2)No

Same as traditional cutoff: 1) Yes 2) No

Gold standard ↓

	pos	neg	
pos co= ( )			
neg co= ( )			

Other Measures., eg sensitivity, PPV, LR: (give 95% CI or NR-99)

Statistic:	Data	95% CI

Data validated? 1) Yes 2) No

Inclusion and Exclusion Criteria:

APPENDIX E. CRITERIA USED IN QUALITY ASSESSMENT

QUADAS tool* with modified item 12.

Item		Yes	No	Unclear
1.	Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	( )	( )
2.	Were selection criteria clearly described?	( )	( )	( )
3.	Is the reference standard likely to correctly classify the target condition?	( )	( )	( )
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests? (Yes if one month or less)	( )	( )	( )
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	( )	( )	( )
6.	Did patients receive the same reference standard regardless of the index test result?	( )	( )	( )
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	( )	( )	( )
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?	( )	( )	( )
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	( )	( )
10.	Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	( )
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	( )
12.	Was the cut off point for the test chosen a priori?	( )	( )	( )
13.	Were uninterpretable/intermediate test results including missing data reported?	( )	( )	( )
14.	Were withdrawals from the study explained?	( )	( )	( )

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.

APPENDIX F. PEER REVIEW COMMENTS/AUTHOR RESPONSES

Reviewer	Comment	Response
Question 1: Are the objectives, scope, and methods for this review clearly described?		
1	Yes	Thank you.
2	Yes- The topic is important and is clearly justified in the introduction. The scope is clearly described. I was a bit disappointed that the scope did not include assessment of anxiety in the context of depression, given the high comorbidity. The authors did an exceptional job of writing methods that were easy for this reader to follow.	Thank you. Including studies that assess the performance of anxiety measures in patients with concurrent depression is an excellent idea. We did not encounter any such studies conducted in primary care settings. A future report could include a broader range of settings that might include this population.
3	Yes - Methodology is clearly described and appropriate to the question asked.	Thank you.
4	Yes- From these, we identified no recent systematic reviews and 12 observational reports on 9 unique studies that addressed one of the key questions. This sentence isn't clear to me; is it: 1) No systematic reviews; 2) 12 observational reports; 3) 9 unique studies?	We have changed this sentence to read "12 articles from 9 unique studies…" to clarify that there were nine studies, some of which had more than one resulting publication.
5	Yes	Thank you.
6	Yes- The objectives are clear-cut, and the review clarifies the potential and considerable limitations of prior research on screening tools for GAD and panic disorder. This report is timely and of great importance. The authors correctly point out that GAD and panic disorder are quite common mental illnesses in the VA population, with considerable impairment in quality of life and physical and cognitive health, and that treatments – SSRIs, other antidepressants, and CBT (all quite implementable within the VA health care system) – are effective for these common and typically undetected conditions. In my own opinion, the lack of detection of these anxiety disorders within the health care system is one of the "low-hanging fruit" in which to improve mental health treatment.	Thank you.
7	No- See my comment below re: Page 8, Table 1 inclusion and exclusion criteria and how they relate to KQ1.	Acknowledged
Question 2: Is there any indication of bias in our synthesis of the evidence?		
1	No	Acknowledged
2	No- There is no evidence of bias in the data synthesis.	Thank you.

Reviewer	Comment	Response
3	No- Authors' disclosures indicate no overt bias. In selecting articles, they did exclude non-English-language measures and articles, possibly excluding high-quality studies, though it is true the excluded studies would likely have been less applicable to the VHA population. As the authors point out, there is some possibility of publication bias, as there is no trials register for diagnostic studies; inasmuch as possible, the search strategy was thorough and comprehensive.	Unfortunately, our resource limitations do not permit bilingual staff or translation services. Since foreign language publications often deal with foreign language questionnaires and this report was written to serve a Veteran population in the United States, we do not think we have missed many pertinent articles. We acknowledge the language limitation in the discussion.
4	No	Acknowledged
5	No	Acknowledged
6	Yes -None	Thank you.
7	No	Acknowledged
Question 3: Are there any studies of interest to the VA that we have overlooked?		
1	No	Acknowledged
2	No, I performed a separate search, particularly looking for anxiety assessment in the elderly, and I could find no studies that were not already included.	Thank you for checking our work! We are glad we did not miss key studies.
3	Yes- Non-English language articles (these studies were excluded).	Unfortunately, our resource limitations do not permit bilingual staff or translation services. Since foreign language publications often deal with foreign language questionnaires and this report was written to serve a Veteran population in the United States, we do not think we have missed many pertinent articles. We acknowledge the language limitation in the discussion.
4	No	Acknowledged
5	No- Search strategy documented in report appears thorough.	Thank you.
6	Yes. The DSM-V workgroup on late-life anxiety disorders has recently published a review of the difficulties of detecting anxiety disorders in older adults. Within this review are some potentially helpful recommendations for improving the characteristics of screening and diagnostic measures for this difficult to assess population (due to insight and memory problems). The citation is Mohlman et al, International Journal of Geriatric Psychiatry. If it is not yet available, you could get it directly from the 1st author, Jan Mohlman, Ph.D., jmohlman@rci.rutgers.edu.	Thank you for this suggestion. We have cited this article in our discussion.
7	No	Acknowledged

Reviewer	Comment	Response
Question 4: Are there any clinical performance measures, programs, quality improvement measures, patient care services, or conferences that will be directly affected by this report? If so, please provide detail.		
1	PACT and associated programs (e.g., primary care-mental health integration) are directly relevant to these results. Casefinding, identification of comorbid anxiety disorders, and tracking treatment progress (i.e., measurement-based care) are important activities for these programs.	We have revised the discussion to identify specific programs (e.g. PACT, primary care-mental health integration) that may want to the recommended instruments. As none of the instruments have been tested for response to change, we think it is too early to implement them for monitoring treatment response.
2	Given the review did not find one measure superior to others, it is not clear that this report will effect an immediate change in these areas. The report does highlight the need for future research on outcomes of anxiety screening.	Based on feasibility and performance characteristics, we identified and recommended the most promising instruments. We have noted the need for further research on the effects of routine screening for anxiety disorders.
3	No comment	Acknowledged
4	Not aware of any	Acknowledged
5	No. Report does not appear to recommend any additions to VA services at this time. However, report makes no practical recommendations so this question is hard to answer.	We have revised the report to make more explicit recommendations, including a summary table of recommendations.
6	I am insufficiently familiar with the VA programs to fully answer this, but it appears that the key conclusion from this report is that there is insufficient evidence regarding the value of existing screening methods for these disorders in VA settings (especially primary care). The logical conclusion would be to recommend to the VA HSR&D that a funding opportunity be made to create and test screening methods.	Thank you for your comment. We are assured that the report will be widely disseminated within the VA system. We have also included a specific recommendation for VA R&D to consider supporting studies on anxiety measures and anxiety screening.
7	No comment	Acknowledged
Question 5: Please provide any recommendations on how this report can be revised to more directly address or assist implementation needs.		
1	While the immediate and explicit aims of the report are specifically framed and very nicely accomplished, addition or further discussion of three issues could be made in several places (namely, framing the questions up front, recommending future research directions, and suggesting implementation needs) to further enhance the utility of this report or inform future work. Specifically, these three issues are: 1) the known or unknown science and the advisability on a practical level of using measures for following treatment progress in addition to casefinding; 2) the role of phone administration in future research; and 3) advice for implementation or research on the best clinical or population contexts for using these instruments for efficient and effective casefinding in general medical settings.	1) This is a very good point and idea. Our review did not specifically address the advisability of measurement-based care but we cite two anxiety care management studies that used this approach with positive results. 2) This is also a very good point, and we have amended the report to address it to a limited degree in the Recommendation for Future Research section as well as in the Summary of Recommendations. 3) We revised the discussion to comment on current recommended uses and the research on the performance of anxiety measures in specific populations.

Reviewer	Comment	Response
2	As indicated above, the report might include some comment about screening for anxiety among depressives and comorbidity of these illnesses.	We discussed the potential for change in performance in individuals with depressive or medical illness. In addition, we commented on applicability to specific VA programs.
3	Given the important contribution of untreated mental illness to overall healthcare utilization and cost, reasonably effective and feasible diagnostic screening tools for patient self-administration in the medical setting could have an impact on overall health as well as healthcare expenditures. Use of screening tools for GAD and PD in primary care clinics may be an important first step; an algorithm for "what to do if the patient screens positive" might be helpful in encouraging implementation of a screening program.	We agree that such an algorithm would be highly useful if the policy implementation experts at the VA decide to start routine screening for anxiety disorders in primary care venues.
4	The report states: Patients referred to the integrated-care programs are also screened for comorbid conditions, including anxiety disorders. I'm not aware of screening for comorbid conditions including anxiety disorders in integrated care programs. If this were being done, it seems like we might have internal data to draw on or would have some information on what screening tools are being used.	The original call for proposals to establish mental health–primary care programs specified routine screening for anxiety disorders. However, these data are not being collected routinely at a national level. We will promote our report to the MH-PC program.
5	The methodology is sound and the evidence appears clear. The conclusions are theoretical and do not appear to provide any practical recommendations (e.g., that none of the measures examined should be implemented, that VA should devote funds to developing and researching new screening instruments, etc.). Also, it is unclear whether the overall VA policy will be to manage GAD and PD in primary care (hence necessitating a diagnostic instrument) or refer positive screens to Mental Health for more accurate diagnosis (which would necessitate only a brief screen, similar to the brief screenings VA uses now in primary care).	Thank you. We revised the report to offer more practical recommendations, including the need for research to inform the effects of screening in primary care. Making recommendations for VA policy—such as care for patients in primary care versus mental health settings—is beyond the scope of the report.

Reviewer	Comment	Response
6	I have several comments. My apologies if some of these go beyond the stated purpose of the expert review:	
	It is likely that a screening instrument will need to do more than simply detect anxiety. It will need to diagnose and track the severity of these disorders, as providers in the VA system (other than psychologists) will not have the time, ability, or inclination to do these.	We agree that tracking responsiveness to change is an important attribute of a good screening instrument. However, the instruments included in the review have not been evaluated for sensitivity to change. Therefore, we included this as a recommendation for future research and have highlighted it again in the Summary of Recommendations section.
	My understanding is that the VA health care system has a lot of older adults. A particular focus is needed on whether the screening instruments would have adequate ability to detect anxiety disorders in this age group. Older adults are notoriously difficult to screen for and diagnose anxiety disorders, given memory and insight issues, among others.	The reviewer is correct in that the VA does have a lot of older adults in whom detection of anxiety disorders is challenging. We have amended the Recommendations for Future Research section of the report to highlight this point. We agree that changes in the diagnostic criteria can affect the performance of an instrument that has been validated using a different version of the DSM. This has been addressed Summary and Discussion section.
	Another comment is that the report does not seem to be taking the changes in these disorders with DSM-V into account. For example, will the GAD-7 still be relevant once GAD is revised into a disorder that more reflects the core concept of worry (and less the associated symptoms)?	The potential changes in the diagnosis of anxiety disorders resulting from the current discussions about diagnostic boundaries are pertinent. We have addressed the specificity of scales under development in the Recommendations for Future Research section.
	Along this same line, there is increasingly a move to question the diagnostic boundaries of these disorders and instead focus on (in the case of anxiety disorders) core dimensions of pathological anxiety such as distress and avoidance. As a concrete example of this issue, wouldn't the VA be better off with an instrument that detected not only GAD but also "anxiety disorder NOS" in the context of substance abuse?	We have added a brief comment on the issue of developing and evaluating scales that detect general anxiety versus those that assess for specific disorders. There are tradeoffs for each decision.
	Finally, might the reviewers want to consider the PROMIS anxiety item banks in their review? To my knowledge, these have not been used in exactly the way the reviewers are examining, but they have been the most extensively psychometrically tested items for measuring anxiety symptoms. I've reviewed them in the past, and many of the items appear quite good – very effective at assessing both the presence and severity of pathological anxiety	Thank you for this suggestion. We contacted one of the investigators regarding the PROMIS scales and also reviewed their Web site. We also conducted a literature search for the PROMIS anxiety scale. It appears that the scale has not yet been validated in a primary care sample and, therefore, could not be included in this report.

Reviewer	Comment	Response
7	Discussion/conclusion sections: Include more of a discussion of implementation within VA settings. You provide a brief discussion of parallels with the PHQ-9 for depression, and expanding this discussion related to how the recommended screening tools could be implemented within VA settings could be helpful for policy makers and providers who will make use of the findings.	We have added our recommendations for current implementation, limited to case-finding and a recommendation for research to address systematic screening.
	Are there any recommendations for universal screening?	USPTF does not have a current recommendation on routine anxiety screening, and we have specifically noted the lack of a recommendation. We did not conduct a systematic review of the effect of anxiety screening; however, this is an important question for future research.
	Should certain tools be included in CPRS and administered to certain populations at certain intervals?	This is a good suggestion, and we have recommended that the most promising tools be added to the MH assistant.
Question 6: Please provide us with contact details of any additional individuals/stakeholders who should be made aware of this report.		
1	Andy Pomerantz	Thank you, we will make sure Dr. Pomerantz is aware of our report.
2	It may help to send this to Dr. Eric Lenze at Washington University, who is an expert on anxiety in the elderly. His email is lenzee@wustl.edu	Thank you. The report will be disseminated broadly.
3	As with all integrative (medical / mental health) work, this is important information for anyone involved in healthcare policy and reform.	Thank you.
4	No comment	Acknowledged
5	None that I can think of.	Acknowledged
6	The individuals cited above would be a good start.	Acknowledged
7	Use of the indicated screening tools should be implemented. This could be done at the national level through central office, or at the VISN or Chief of mental health level. The office(s) responsible for implementation should be made aware of this report.	Acknowledged. We are assured that the report will be widely circulated inside the VA system.
Question 7: Please write additional suggestions or comments below. If applicable, please indicate the page and line numbers from the draft report.		
1	No Comment	Acknowledged
2	A very well written report	Thank you.
3	No Comment	Acknowledged
4	No Comment	Acknowledged

Reviewer	Comment	Response
5	As noted in comments from item 5 above, some consideration should be given to the context in which this literature examination is taking place (i.e., VA setting versus community facility), and there should be some mention of possible ways that screening instruments could be used (e.g., whether positive screens will be assessed further and treated by PC personnel, whether they will be walked over to integrated MH in PC, whether they will be referred to MH), as this would affect the type of instrument that could be developed and researched.	We agree that the setting in which an instrument is administered is important. Though we would have liked to have included studies done in the VA, we did not identify any and have suggested this as a recommendation for future research. Subsequent treatment and referral of patients who screen positive is important; however, it was beyond the scope of this review.
6	Nice, well-written and well thought out report. I enjoyed reading it.	Thank you.
7	Page 1, paragraph 1: Provide citations for introductory paragraph.	We have added citations.
	Page 1, paragraph 1: Often is stated twice in the last sentence. Change last "often" to frequently.	We have made this change.
	Page 2, paragraph 3: should be "detailed review of" (not review on).	We have made this change.
	Page 8, Table 1: Inclusion/exclusion criteria related to population is unclear given KQ1 and the analytic framework described throughout the report. It is unclear whether "somatic symptoms" referred to in the KQ1 and analytic framework is the same population as is described in this table.	We have added text in the Methods section to clarify this further and have changed the wording in the table.
	Elsewhere (e.g., page 2, paragraph 5) you refer to patients in primary care settings. Clarify exactly which populations and settings were included and excluded from this report and use consistent terminology throughout the report (e.g., Included studies were all conducted in primary care settings with patients who (a) presented with somatic symptom(s) and (b) did not have a preexisting mental health diagnosis, hereafter referred to as "primary care patients with somatic symptoms"). In the introduction on page 6 you end with a description of you population (primary care settings), yet there should be clarification related to the presence of somatic symptoms and lack of preexisting mental health diagnosis. This is confusing because earlier in the paragraph you refer to the need for anxiety disorder screening tools and make reference to the likelihood that these disorders are present in populations with other mental illnesses—please clarify whether or not these populations are included in the scope of this report.	This has been clarified in the Results section.
	Also, the inclusion/exclusion criteria include non-primary care settings in the "setting" row—perhaps clarification that all these settings were included, however only primary care setting studies were found.	Thank you.

Reviewer	Comment	Response
7 (cont.)	Pages 10-11: Clear, concise description of quality assessment, data synthesis, and rating the body of evidence.	We have clarified these descriptions.
	Page 13, line 4: search of a relevant systematic review should be changed to search of relevant systematic reviews.	We have corrected this text.
	Page 14: The list of excluded articles includes 17 listed as "population not of interest" and 19 listed as "setting not PC/clinic/ER." Not sure if this needs more explanation, but it might be beneficial to describe these excluded studies in greater detail given the above comments re: population and setting description. I think it would be interesting to know more about these excluded studies and why they were irrelevant/beyond the scope of this report (if there are many studies conducted in MH clinics with populations who have a pre-existing MH diagnosis, for example, this would be an interesting future SR in and of itself, even if beyond the scope of this review).	We rechecked the 17 studies listed as "population not of interest." Fifteen studies were of subjects already diagnosed with an anxiety disorder; one was of Native Americans on a reservation; and the last was on inpatients. We also checked those listed as having "setting not PC/MH clinic/ER." Five were conducted at a university, five were recruited from MH clinics (and already diagnosed with an anxiety disorder), four were ads in the general community, two were specialty-based (neurology and geriatric), two were internet-based and one interviewed subjects in their homes.
	Pages 19-24: This is an excellent and concise description of measure characteristics. I'm a statistician, so it all made sense to me, however many readers likely don't have the stats background to understand the analyses. Try including a summary sentence for each type of analysis with a more "plain English" description of the analysis and what it means so that non-statsy folks can follow along, too.	We have included a section describing sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio in plain English.
	Page 25: Excellent figure!	Thank you.
	Pages 30 and 33: use either case finding or case-finding, not both.	We have made case-finding consistent.
	Page 30-31: This last/first paragraph on effective treatments for ADs seems a bit disjointed. Either just provide the citations or tie it in to the findings a little more.	We have clarified this text.
	Overall: Excellent, clear, and concise report. Very useful and well written. Will be very useful for implementing changes within the VA.	Thank you.