Screening for Anxiety in Adolescent and Adult Women

Systematic Review for the Women's Preventive Services Initiative

Heidi D. Nelson, MD, MPH
Amy Cantor, MD, MPH
Miranda Pappas, MA
Chandler Weeks, MPH

Pacific Northwest Evidence-based Practice Center
Oregon Health & Science University

April 30, 2020

INTRODUCTION

Anxiety disorders include several related conditions characterized by excessive, uncontrollable worry.[1] These include generalized anxiety disorder, panic disorder, social or school anxiety disorder, and other specific types (**Table 1**).[1] Anxiety disorders cause significant impairment in daily activities, health, and function, including work and school responsibilities, and adversely impact well-being and social relationships.[2] Anxiety increases risk for major depression over the following year,[3] and is associated with unhealthy behaviors[4] and higher medical utilization.[5] Over 30 million Americans have anxiety during their lifetimes, and its economic impact has been estimated as $42 billion dollars per year.[6]

Table 1. Anxiety Disorders[1]

Generalized Anxiety Disorder	Excessive anxiety and worry (apprehensive expectation) occurring more days than not for at least 6 months about a number of events or activities (such as work or school performance); the person finds it difficult to control the worry.
Separation Anxiety Disorder	Developing inappropriate and excessive fear or anxiety concerning separation from those to whom the individual is attached.
Social Phobia or Anxiety	Marked fear or anxiety about one or more social situations in which the individual is exposed to possible scrutiny by others.
Specific Phobia	Marked fear or anxiety about a specific object or situation (e.g., flying, heights, animals, receiving an injection, seeing blood).
Panic Disorder	Recurrent unexpected panic attacks. A panic attack is an abrupt surge of intense fear or intense discomfort that reaches a peak within minutes.
Selective Mutism	A childhood disorder typified by an inability to speak in certain circumstances. Specifically, it is a consistent failure to speak in certain social situations where there is a natural expectation of speaking**.**
Agoraphobia	A disproportionate fear of public places, often perceiving such environments as too open, crowded, or dangerous.

Anxiety disorders are the most frequent mental health disorders in the general population,[7] with approximately 31% of adults in the United States experiencing anxiety disorders during their lifetimes[8] and 19% over the past year.[9] These estimates are likely inaccurate because anxiety disorders are often undiagnosed.[2] Prevalence is higher among women compared with men (23% versus 14%).[9] The prevalence of anxiety disorders among U.S adolescents age 13 to 18 years is

32%, with higher rates for girls than boys (38% versus 26%).[7] Among adolescents with anxiety disorders, 8% meet Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for severe impairment.[7] In children and adolescents, anxiety disorders are frequently associated with other conditions including depression, eating disorders, and attention-deficit/hyperactivity disorder.[10]

Although research on anxiety disorders in women is limited, it suggests sex-specific features. Studies of anxiety during pregnancy describe the effects of elevated maternal cortisol on the developing fetus.[11] These include effects on sex-specific neonatal amygdala connectivity that manifests in behavioral problems of female offspring at age 2 years.[12] A longitudinal study of young girls indicated that early behaviors and emotional symptoms predicted anxiety diagnosis in adulthood.[13] Previous studies have shown associations of anxiety with environmental causes or triggers, particularly in teenage females. These include worries about school performance, concerns about appearance, earlier sexualization, changing media and consumer culture, and poor self-esteem.[14] In addition, females are more attentive to social and emotional experiences that increase stress.

Several brief screening instruments have been validated for identification of anxiety in primary care clinical settings. The diagnosis of an anxiety disorder is established by a clinical diagnostic interview using DSM-V criteria[1,15] (see example in **Table 2**). Importantly, when evaluating a patient for suspected anxiety disorders, other potential medical conditions must be ruled out (e.g. endocrine, cardiopulmonary, neurologic diseases). Other psychiatric disorders including depression and bipolar disorder must be considered, in addition to the use of caffeine, medications (e.g., decongestants, albuterol, levothyroxine), addictive substances, or substance withdrawal.

Table 2. DSM-V Criteria for Generalized Anxiety Disorder[1]

A.	Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months, about a number of events or activities (such as work or school performance).
B.	The individual finds it difficult to control the worry.
C.	The anxiety and worry are associated with three or more of the following six symptoms (with at least some symptoms having been present for more days than not for the past 6 months): Note: Only one item is required in children 1. Restlessness or feeling keyed up or on edge 2. Easily fatigued 3. Difficulty concentrating or mind going blank 4. Irritability 5. Muscle tension 6. Sleep disturbance (difficulty falling or staying asleep, or restless, unsatisfying sleep)
D.	Anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
E.	The disturbance is not attributable to the physiological effects of a substance (e.g. a drug of abuse, a medication) or another medical condition (e.g. hyperthyroidism).
F.	The disturbance is not better explained by another mental disorder, such as panic disorder.

Cognitive behavioral therapy or other forms of psychotherapy[16] are first-line therapy for most patients, while medications are second-line.[17] These include selective serotonin reuptake inhibitors (SSRI), serotonin-norepinephrine reuptake inhibitors (SNRI), and azapirone

(buspirone). Tricyclic antidepressants and calcium modulators (pregabalin) are sometimes used, while benzodiazepines are not recommended for treatment other than during an acute crisis.

Screening for anxiety has not been addressed by the U.S. Preventive Services Task Force (USPSTF), although screening for depression is recommended and has become standard practice in primary care.[18,19] Anxiety disorders are often missed by clinicians because patients may be reluctant to discuss their distress, symptoms may be attributed to other causes, or anxiety may co-exist with other conditions, such as depression and substance use. The purpose of screening is to identify individuals for further evaluation of the whole spectrum of anxiety disorders and related conditions. As with other disorders, such as depression, screening itself is not diagnostic. Screening has the potential to identify previously unrecognized anxiety and related disorders, initiate individualized treatment, and prevent progression and impairment. An example of a clinical approach to screening is described below (**Table 3**).

Table 3. Clinical Approach to Screening for Anxiety[20]

Screening		Interpretation & Diagnosis	Proposed Treatment Actions*
GAD-7 Score	**Anxiety Severity**		
0 - 4	None	Diagnostic criteria not met	None
5 - 9	Mild		Watchful waiting, repeat at follow up
10 - 14	Moderate	Diagnostic criteria met	Initiate cognitive behavioral therapy and consider pharmacotherapy
15 - 21	Severe		Initiate cognitive behavioral therapy and pharmacotherapy; consider referral to mental health specialist

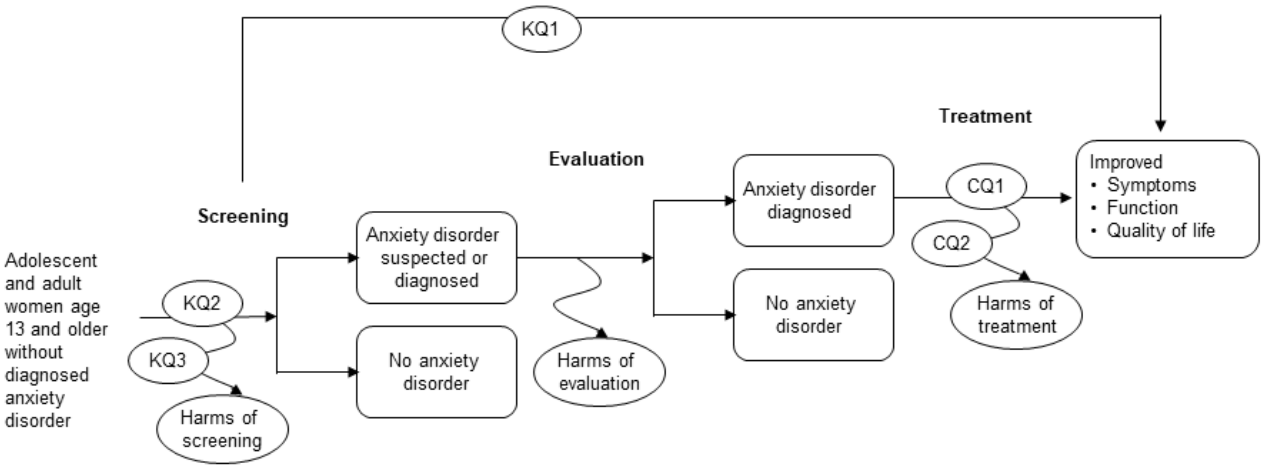
GAD=Generalized Anxiety Disorder Scale
*Examples only, treatment requires a patient-specific approach.

The purpose of this systematic review is to evaluate evidence on the effectiveness and harms of screening for anxiety disorders in adolescent and adult women, including those pregnant or postpartum, in improving symptoms, function, and quality of life; the accuracy of screening instruments; and the effectiveness and harms of treatment to inform new practice recommendations from the Women's Preventive Services Initiative (WPSI).

METHODS

The WPSI Advisory Panel determined the scope and key questions for this review to inform the development of new screening recommendations. The protocol was developed according to established methods[21,22] with input from experts and the public. Investigators created an analytic framework outlining the key questions and patient populations, interventions, and outcomes (**Figure 1**). The target population includes women and adolescent girls age 13 and older without known current anxiety disorders, including those pregnant and postpartum.

Figure 1. Analytic Framework



Key Questions (KQ)

1. In women and adolescent girls age 13 and older without currently diagnosed anxiety disorders, what is the effectiveness of screening and evaluation for anxiety to improve symptoms, function, and quality of life?
2. What is the accuracy of methods to screen for anxiety? How does accuracy vary between age, pregnancy status, social-demographic, and cultural groups; and among women with comorbid conditions or who use additional medications?
3. What are potential harms of screening for anxiety?

Contextual Questions (CQ)

Two contextual questions were also included to provide additional information that could support the chain of evidence for screening. Contextual questions were addressed by reviewing recently published systematic reviews of randomized controlled trials (RCTS).

1. What is the effectiveness of treatments for anxiety in improving symptoms, function, and quality of life?
2. What are the potential harms of treatments for anxiety?

Literature Searches
A research librarian conducted electronic database searches in Ovid MEDLINE, the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews from January 1, 1996 to November 4, 2019 (**Appendix 1**). Investigators also manually reviewed reference lists of relevant systematic reviews and articles.

Study Selection
Investigators reviewed all titles and abstracts identified through searches and secondary referencing and determined inclusion based on pre-specified criteria defined by PICOTS components (population, intervention, comparator, outcome, timing, study design) (**Appendix 2**). Studies meeting eligibility criteria for possible inclusion by a reviewer at the abstract level subsequently underwent full-text review. Each full-text article was independently reviewed by two investigators based on the pre-specified eligibility criteria. All results were tracked in an EndNote® database (Thomson Reuters, New York, NY).

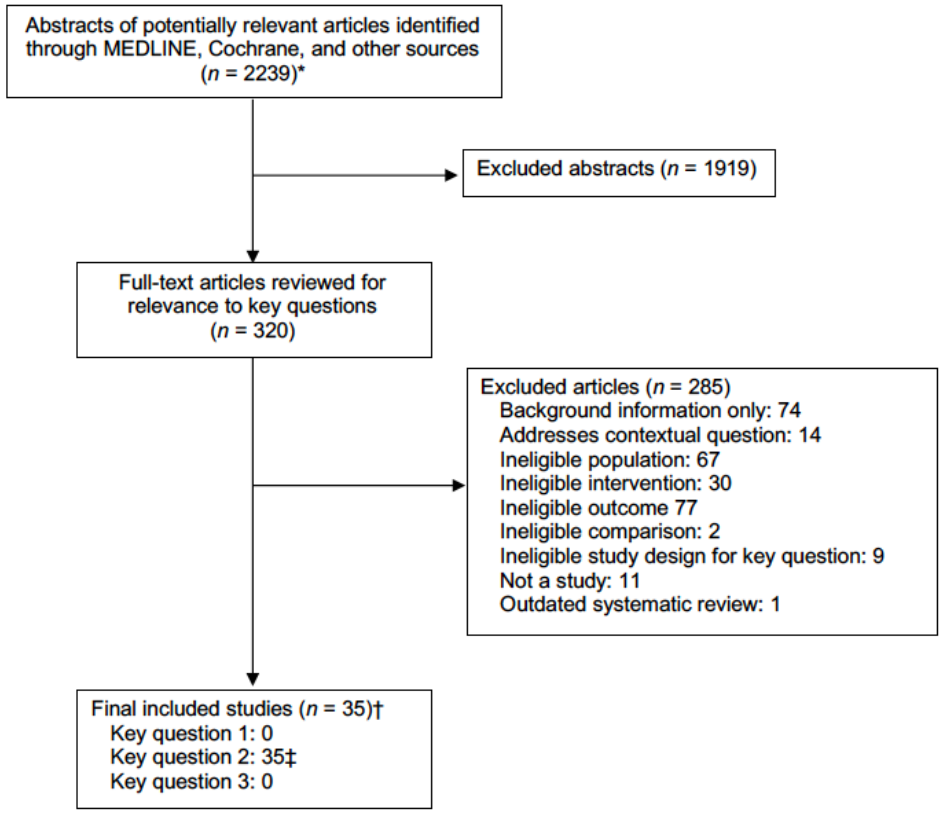
Investigators applied a best evidence approach when reviewing abstracts and selecting studies to include for this review that involves using the most relevant studies with the strongest methodology.[21,23,24] Disagreements regarding inclusion of studies were resolved by discussion and consensus involving a third reviewer. Results of the full text review were tracked in the EndNote® database, including reasons for exclusion. Results of searches and study selection are described in **Figure 2.**

Studies were included that enrolled predominantly adolescent girls or adult women (>50% female participants) and were applicable to clinical practice in the United States. Findings related to specific populations were included when available. Randomized controlled trials, large (>100) prospective cohort studies, diagnostic accuracy studies, and systematic reviews meeting eligibility criteria were included. Other study designs, such as case-control and modeling studies, were included when evidence from other study designs was lacking.

For diagnostic accuracy of screening instruments, studies that used screening methods applicable to primary care settings in the United States were included, such as brief self-report or clinician-administered questionnaires. While only primary care relevant methods were included, they may have been developed in other settings. Included studies reported measures of test performance, such as areas under the receiver-operating characteristic curve (AUC) (also known as the c-statistic), sensitivity and specificity, or likelihood ratios as reported by the studies. In general, AUC levels above 0.80 indicate high diagnostic accuracy, 0.70 to 0.8 good, 0.60 to 0.70 sufficient, and levels less than 0.60 may not be clinically useful.[25] Potential harms of screening included false-positive or false-negative results, anxiety, distress, and other adverse events affecting quality of life.

For contextual questions on treatment (CQ 1, 2), studies were included that compared treatment against a placebo, no treatment, waitlist control, or usual care group. Treatment effectiveness outcomes included clinical response, reduction in anxiety symptoms or improvement in scores on validated scales, and quality of life measures. Multiple adverse effects outcomes were included as reported in studies.

Figure 2. Literature Flow Diagram



*Cochrane databases include the Cochrane Central Register of Controlled Trials and the Cochrane Database of Systematic Reviews. Other sources include reference lists and hand-searching.
†Studies that provided data for the key questions.
‡Includes 2 systematic reviews.

Data Extraction and Synthesis

For studies meeting inclusion criteria, data were abstracted into tables to summarize relevant information including characteristics of study populations, interventions, comparators, outcomes, study designs, settings, methods, and results. All data abstractions were reviewed for completeness and accuracy by a second investigator.

Predefined criteria were used to assess the quality of individual controlled trials, observational studies, systematic reviews, and diagnostic accuracy studies rating them as "good," "fair," or "poor."[22] Critical appraisal criteria for the diagnostic accuracy of screening tests were based on USPSTF methods,[22] which are similar to other established methods, including QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2).[26,27] Each study was independently rated for quality by two investigators and disagreements were resolved by consensus involving a third reviewer. Studies were synthesized qualitatively. No statistical meta-analyses were conducted because of methodological and clinical heterogeneity.

Assessing Applicability

Applicability is defined as the extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under "real-world" conditions.[21] It is an indicator of the extent to which research included in a systematic review might be useful for informing clinical decisions. Factors important for understanding applicability were considered for each study including differences in the interventions, comparators, populations, and settings. Based on these factors, applicability was rated "high" or "low."

Establishing the Strength of Evidence

The strength of evidence for each key question was assessed by using the approach described in the AHRQ Methods Guide.[21] Grades were based on study limitations, consistency, directness, precision, and reporting bias (**Appendix 3**).

RESULTS

Key Questions 1 and 3. Effectiveness and Harms of Screening for Anxiety

A total of 2239 abstracts and 320 full-text articles were reviewed. Of these, no studies directly evaluated the overall effectiveness or harms of screening.

Key Question 2. Accuracy of Screening Methods

Thirty-three studies and 2 systematic reviews that included 171 studies evaluated the accuracy of 27 clinician or self-administered screening instruments and their multiple variations (**Appendix 4**). Studies were conducted in either the general adult population[20,28-42] or among specific populations including adolescents,[43-47] pregnant and postpartum women,[48-54] and older adults.[55-57] All studies included at least 50% female participants. Most studies met criteria for good or fair quality (**Appendix 5**).

Screening methods included various clinician or self-administered questionnaires addressing symptoms of anxiety designed for use in clinical practice (**Table 4**). Responses were typically scored using a Likert scale or other point system. Diagnostic accuracy measures were determined by comparing scores against reference standards that generally included clinical diagnosis using DSM criteria. Additional reference standards included clinical interviews, more comprehensive instruments, or combinations. Results were expressed as AUC c-statistics, sensitivity and specificity values, or positive and negative likelihood ratios.

Table 4. Instruments Included in Studies

Abbreviation	Screening Instrument	Study (author, year)
None	Anxiety Disorders-13	Fairbrother, 2019[58]
BAI	Beck Anxiety Inventory	Leyfer, 2006[37]; Dennis, 2007[59]; O'Hara, 2012[51]
BSI-A	Brief Symptom Inventory-18	Wetherell, 2007[57]
CES-D	Center for Epidemiologic Studies- Depression Scale	Dozeman, 2011[60]; Dierker, 2001[46]
CIDI	Composite International Diagnostic Interview	Austin, 2010[61]; Rowe, 2008[62]
DASS-21	Depression, Anxiety, and Stress Scale 21	Somerville, 2014[53]
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders	Newman, 2002[40]; Houston, 2011[34]
DUKE-AD	Duke Anxiety-Depression Scale	Parkerson, 1997[41]
EK10	Extended Kessler-10	Donker, 2010[30]
EPDS	Edinburgh Postnatal Depression Scale	Austin, 2010[61]; Fairbrother, 2019[58]; Matthey, 2013[48]; McDonald, 2012[49]; Meades, 2011[50]; O'Hara, 2012[51]; Petrozzi, 2013 [63]; Rowe, 2008[62]; Simpson, 2014[52]; Somerville, 2014[53] ; Tendais, 2014[54]
FEAR	Frequency of anxiety; enduring nature of anxiety; alcohol or sedative use; restlessness or fidgeting	Krasucki, 1999[55]
GAD-2	Generalized Anxiety Disorder Scale-2 items	Donker, 2011[31]; Garcia-Campayo, 2012[33]; Fairbrother, 2019[58]
GAD-7	Generalized Anxiety Disorder Scale-7 items	Batterham, 2013[64]; Donker, 2011[31]; Fairbrother, 2019[58]; Munoz-Navarro, 2017[39]; Spitzer, 2006[20]; Vasiliadis, 2015[56]; Simpson, 2014[52];
GAD-Q-IV	Generalized Anxiety Disorder Questionnaire	Moore, 2014[38]; Newman, 2002[40]; Roemer, 1995[65]
GAD-SI	Single item from the GAD-7	Donker, 2011[31]
GADSS	Generalized Anxiety Severity Disorder	Weiss, 2009[66]
GAS/GAD	Goldberg Anxiety and Depression Scales	Kiely, 2015[35]
GHQ	General Health Questionnaire	Meades, 201[50]
HADS	Hospital Anxiety and Depression Scale	Dennis, 2007[59]; Jomeen, 2003[67]; Wetherell, 2007[57]; Matthey, 2013[48]; Meades, 2011[50]
K10	Kessler Psychological Distress Scale	Donker, 2010[30];Vasiliadis, 2015[56]
MASC	Multidimensional Anxiety Scale for Children	Dierker, 2001[46]
MCS-12	Mental Health Component Summary Scale Web-based depression and anxiety test	Kiely, 2015[35]
MGMQ[48]	Matthey Generic Mood Question	Matthey, 2013[48]; Matthey, 2019[68]; Fairbrother, 2019[58]
MINI	Mini-International Neuropsychiatric Interview	Grant, 2008[69]
PASS	Perinatal Anxiety Screening Scale	Somerville, 2014[53]
PDI-4	Provisional Diagnostic Instrument	Houston, 2011[34]
PDSQ	Psychiatric Diagnostics Screening Questionnaire	Leung, 2017[70]
PHQ-4	Patient Health Questionnaire for depression and anxiety (4 items)	Kroenke, 2009[36]; Cano-Vindel, 2018[71]
PHQ-9	Patient Health Questionnaire for depression and anxiety (9 items)	Kiely, 2015[35]
PSS	Cohen Perceived Stress Scale	McDonald, 2012[49]
PSWQ	Penn State Worry Questionnaire	Behar, 2003[28]; Schroder, 2019[72]
RCADS	Revised Child Anxiety and Depression Scale	Piqueras, 2017[47]
RCMAS	Revised Children's Manifest Anxiety Scale	Dierker, 2001[46]

Abbreviation	Screening Instrument	Study (author, year)
SCARED	Screen for Child Anxiety Related Emotional Disorders	Birmaher, 1997[43]; Birmaher, 1999 [44]; Bodden, 2009[45]; Crocetti, 2009[73]; Hale, 2013[74]
SCID	Structured Clinical Interview for DSM-IV	Farvolden, 2003[32]; O'Hara, 2012[51]
STAI	State Trait Anxiety Inventory	Dennis, 2007[59]; Grant, 2008[69]; McDonald, 2012[49]; Meades, 2011[50]; Somerville, 2014[53]; Tendais, 2014[54]
VAS	Visual Analogue Scale	Dennis, 2007[59]
WB-DAT	Web-based Depression and Anxiety Test	Farvolden, 2003[32]
WSQ	Web Screening Questionnaire	Donker, 2009[29]

Adolescents

Four studies[43-46] and one systematic review[47] of screening methods for adolescents met inclusion criteria (**Table 5**). Screening methods included four variations of Screen for Child Anxiety Related Emotional Disorders (SCARED);[43-45] the Revised Children's Manifest Anxiety Scale (RCMAS);[46] the Multidimensional Anxiety Scale for Children (MASC);[46] and the Revised Child Anxiety and Depression Scale (RCADS).[47]

The original SCARED instrument includes 38-items with five subscales specific for panic disorder, generalized anxiety disorder, separation anxiety disorder, social anxiety, and school anxiety. Several variations exist including 41-item, 71-item, and 5-item versions that were evaluated in the included diagnostic accuracy studies. The original SCARED instrument demonstrated sensitivity of 72% and specificity 64% in a study of adolescents age 9 to 18 years in a mood/anxiety disorders clinic.[43] A study comparing two versions of SCARED in adolescents age 9 to 19 years reported sensitivity and specificity of 71% and 67% for the 41-item version, and 74% and 73% for the 5-item version.[44] In another study, the 71-item version demonstrated sensitivity 64% and specificity 69% in adolescents age 8 to 18 years.[45]

The RCMAS and MASC instruments were evaluated in a study of 632 ninth graders enrolled in five high schools across the United States. The RCMAS is a 37-item self-report measure using yes or no responses to calculate a composite anxiety score; while the 39-item MASC rates the frequency of symptoms on a four-point scale. Results indicated AUC values specifically for girls of 0.62 for RCMAS and 0.82 for MASC.[46]

A systematic review of 146 studies evaluated RCADS, a 47-item instrument with six subscales for separation anxiety disorder, social phobia, generalized anxiety disorder, panic disorder, obsessive compulsive disorder, and major depressive disorder.[47] Studies included children and adolescents 6 to 18 years old, and combined results were reported as a reliability measure (0.91; 95% CI 0.90 to 0.92).

Table 5. Studies of Screening Instruments Developed for Children and Adolescents

Screening Instrument	Description	Study (author, year)	Participants	Reference standard	Performance characteristics (95% CI)	Quality Rating
SCARED[43]	38-items in 5 subscales: panic disorder, generalized anxiety disorder, separation anxiety disorder, social anxiety, school anxiety.	Birmaher, 1997[43]	341 adolescents 9 to 18 years old in a mood/anxiety disorders clinic	Either clinical interview using DSM-IV diagnosis criteria or K-SADS-P diagnosis	Sensitivity: 72% Specificity: 64%	Fair
SCARED-41[44]	41-item scale; addition of 3 items to the social phobia subscale of the SCARED scale.	Birmaher, 1999[44]	190 adolescents 9 to 19 years old in a mood/anxiety disorders clinic	Comprehensive symptom checklist for DSM-IV diagnostic criteria	Sensitivity: 71% Specificity: 67%	Fair
SCARED-71[45]	71-item scale; adds 3 additional subscales to the SCARED scale: specific phobia, obsessive-compulsive disorder, and post-traumatic stress disorder.	Bodden, 2009[45]	176 adolescents 8 to 18 years old; clinically anxious cases and controls	ADIS-C and ADIS-P	Sensitivity: 64% Specificity: 69%	Fair
5-item SCARED[44]	A shorter version of the SCARED-41; includes 1 item from each subscale that best discriminates between anxious and non-anxious respondents.	Birmaher, 1999[44]	190 adolescents 9 to 19 years old in a mood/anxiety disorders clinic	Comprehensive symptom checklist for DSM-IV diagnosis criteria	Sensitivity: 74% Specificity: 73%	Fair
RCMAS[75]	37-item self-report measure using yes or no responses to each statement. A composite anxiety score is calculated by summing the number of yes responses (range 0–28).	Dierker, 2001[46]	632 9th graders enrolled in 5 high schools across the United States	Diagnostic interview modules selected from the DSM-IV diagnosis criteria	AUC for girls: 0.62	Poor
MASC[76]	39-item rating scale; respondents rate the frequency of symptoms on a four-point scale: 0=never; 1=rarely; 2=sometimes; and 3=often.	Dierker, 2001[46]	632 9th graders enrolled in 5 high schools across the United States	Diagnostic interview modules selected from the DSM-IV diagnosis criteria	AUC for girls: 0.82	Poor
RCADS[77]	47 items in 6 subscales: separation anxiety disorder, social phobia, generalized anxiety disorder, panic disorder, obsessive compulsive disorder, and major depressive disorder.	Piqueras, 2017;[47] 146 studies in systematic review	88,648 children and adolescents 6 to 18 years old	Not reported, multiple studies	Reliability: 0.91 (0.90 to 0.92)*	Moderate

ADIS-C=Anxiety Disorder Interview Schedule-Child scale; ADIS-P=Anxiety Disorder Interview Schedule-Parent scale ; AUC=area under the receiver operating characteristic curve; CI=confidence interval; DSM=Diagnostic and Statistical Manual of Mental Disorders; GAD=generalized anxiety disorder; K-SADS-P= Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present Episode; MASC=Multidimensional Anxiety Scale for Children; RCADS=Revised Child Anxiety and Depression Scale; RCMAS=Revised Children's Manifest Anxiety Scale; SCARED=Screen for Child Anxiety Related Emotional Disorders.

*Determined by Cronbach's alpha measure to estimate internal consistency of RCADS.

General Adult Population

Seventeen studies of 10 screening methods for adults in the general population met inclusion criteria (**Table 6**). Screening methods included four variations of the Generalized Anxiety Disorder (GAD) scale;[20,31,33,36,38-40,71] the Penn State Worry Questionnaire (PSWQ);[28,72] Web Screening Questionnaire (WSQ);[29] Kessler-10 (K10) and extended version (EK-10);[30] Web-Based Depression and Anxiety Test (WB-DAT);[32] Provisional Diagnostic Instrument (PDI-4);[34] Goldberg Anxiety Scale (GAS);[35] Beck Anxiety Inventory (BAI);[37] Duke Anxiety-Depression Scale (DUKE-AD);[41] and 2 screening questions.[42]

The GAD-7 is the anxiety module of the Patient Health Questionnaire (PHQ) that assesses anxiety symptoms over the last 2 weeks. Three studies of the GAD-7 demonstrated sensitivities of 87% to 89%, specificities of 50% to 82%, and AUC 0.77 using a cut-point of 10.[20,31,39] The GAD-2, a shorter version, was evaluated in four studies indicating sensitivity from 70% to 91.5%, specificity 61% to 86%, and AUC 0.78 to 0.94.[31,33,36,71] A study of 2149 primary care patients reported an AUC value for GAD-2 of 0.908.[36] Studies of the GAD-Q-IV, an updated 9-item version, indicated sensitivity 97%, specificity 86%, and AUC 0.85 when using a DSM based algorithm[38]; and sensitivity ranging from 83% to 89% and specificity from 72% to 89% when using a cut-point of 7.6.[38,40]

The diagnostic accuracies of additional methods were reported in single studies. Of these, the PSWQ, WSQ, Kessler-10 and EK-10, PDI-4, GDS, GAS, BAI, DUKE-AD, and 2 screening questions demonstrated moderate to high performance measures, while the WB-DAT indicated lower performance.

Table 6. Studies of Screening Instruments in Adults

Screening Instrument	Description	Study (author, year)	Participants	Reference Standard	Performance Characteristics (95% CI)	Quality Rating
GAD-2	GAD-2 is derived from the anxiety module of the Patient Health Questionnaire (PHQ). Assesses anxiety symptoms over the last 2 weeks: • Feeling nervous, anxious, or on edge • Not being able to stop or control worrying • Feeling down, depressed, or hopeless • Little interest or pleasure in doing things	Donker, 2011[31]	502 adults age 18 to 80 compared with 20 psychology students; web-based (57% female)	DSM-IV CIDI GAD	Cutoff 3 Sensitivity: 70% Specificity: 76% AUC: 0.78 (0.69 to 0.86)	Fair
		García-Campayo, 2012[33]	220 adults age >18 (72% female)	HAM-A, HADS, and WHODAS II	Cutoff 3 Sensitivity: 91.5% Specificity: 85.8% AUC: 0.937	Fair
		Kroenke, 2009[36]	2149 primary care patients (66% female)	Structured interview using DSM-IV criteria	AUC: 0.908 (0.876 to 0.940)	Good
		Cano-Vindel, 2018[71]	1052 primary care patients (77% female)	SCID-I	Cutoff 3 Sensitivity: 88% Specificity: 61% AUC >0.85	Fair
GAD-7	GAD-7 is the anxiety module of the Patient Health Questionnaire (PHQ). Assesses anxiety symptoms over the last 2 weeks: • Feeling nervous, anxious or on edge • Not being able to stop or control worrying • Worrying too much about different things • Trouble relaxing • Being so restless that it's hard to sit still • Becoming easily annoyed or irritable • Feeling afraid as if something awful might happen	Donker, 2011[31]	502 adults age 18 to 80 compared with 20 undergraduate psychology students; web-based (57% female)	DSM-IV CIDI GAD	Cutoff 10 Sensitivity: 87 to 89% Specificity: 50 to 82% AUC: 0.77 (0.68 to 0.85)	Fair
		Munoz-Navarro, 2017[39]	178 adults age 18 to 65 in primary care (71% female)	CIDI for DSM-IV	Cutoff 10 Sensitivity: 87% Specificity: 78%	Fair
		Spitzer, 2006[20]	2740 adults in primary care clinics; mean age 47 (18-95 years) (65% female)	Structured interviews for 965	Cutoff 10: Sensitivity: 89% Specificity: 82%	Good
GAD-Q-IV	The fourth edition of the Generalized Anxiety Disorder Questionnaire (GAD-Q-IV) is a 9-item self-report measure.	Moore, 2014[38]	104 adults in primary care (69% female)	SCID-IV	AUC: 0.85 (0.76 to 0.93) DSM-based algorithm Sensitivity: 97% Specificity: 86% Cutoff 7.6 Sensitivity: 89% Specificity: 72%	Fair

Screening Instrument	Description	Study (author, year)	Participants	Reference Standard	Performance Characteristics (95% CI)	Quality Rating
		Newman, 2002[40]	143 undergraduates (80% female)	DSM structured interview	Sensitivity: 83% Specificity: 89%	Fair
PSWQ	A 16-item measure designed to assess the pathological worry characteristic of GAD. Including the generality, excessiveness, and uncontrollability of worry without focusing on particular domains of worry.	Behar, 2003[28]	2449 young adults (71% female)	GAD-Q-IV	Cutoff 62 Sensitivity: 75% Specificity: 86%	Fair
PSWQ brief versions	3-item and single-item variations focus on specific questions of the Penn State Worry Questionnaire.	Schroder, 2019[72]	1191 undergraduates and from community (73% female)	GAD-7	1-item version Sensitivity: 64% Specificity: 92% 3-item version Sensitivity: 68% Specificity: 92%	Fair
WSQ	Includes 15 items to screen for depression, GAD, panic disorder with and without agoraphobia, social phobia, specific phobia, OCD, PTSD, and alcohol abuse/dependence.	Donker, 2009[29]	502 adults age 18 to 80 recruited from the internet (57% female)	CIDI diagnosis with live phone interviews	Cutoff ≥10 Sensitivity: 89% Specificity: 82%	Fair
Kessler-10 (K10), EK-10 (extended version)	K-10: 10 questions; screens broadly for psychological distress. EK-10: Extended with five additional questions focusing on anxiety symptoms.	Donker, 2010[30]	1607 adults in primary care age 18 to 65 years (69% female)	CIDI interview, DSM-IV diagnosis	K10-20 Sensitivity: 94% Specificity: 67% EK10-20 Sensitivity: 95% Specificity: 61%	Fair
WB-DAT	11 broad preliminary questions; final report based on algorithm response to specific questions.	Farvolden, 2003[32]	32 adults (59% female)	SCID-I/P interview	Sensitivity: 63% Specificity: 94%	Poor
PDI-4	17-item instrument for provisional differential diagnosis with 4 items specific for anxiety.	Houston, 2011[34]	24 adults in (>60% female)	SCID/ACDS assessment, DSM-IV	Sensitivity: 83% Specificity: 75% Follow up with GAD-7 Sensitivity: 89% Specificity: 82%	Poor
Goldberg Anxiety Scales (GAS)	Asks respondents whether they experienced 9 anxiety symptoms over the past 4 weeks. Scores are summed to give a maximum total of 9 on each scale.	Kiely, 2015[35]	1015 community adults (ages 32 to 36 and 52 to 58) (59% female)	CIDI	Cutoff 7 Sensitivity: 84% Specificity: 86% AUC: 0.8957	Good

Screening Instrument	Description	Study (author, year)	Participants	Reference Standard	Performance Characteristics (95% CI)	Quality Rating
Beck Anxiety Inventory (BAI)	21-item self-report questionnaire that lists symptoms of anxiety. The respondent is asked to rate how much each symptom has bothered them in the past week.	Leyfer, 2005[37]	193 adults in the general population (76% female)	ADIS-IV	Cutoff 3.5 Sensitivity: 75% Specificity: 73%	Fair
DUKE-AD	Includes two items for negative affect (feeling depressed or sad, nervousness), two for somatic symptoms (trouble sleeping, getting tired easily), two for self-esteem (give up easily, uncomfortable being around people), and one for cognition (difficulty concentrating).	Parkerson, 1997[41]	481 adults in primary care age 18 to 64 (72% female)	DSM	Sensitivity: 71.4% Specificity: 59.2% AUC: 0.723	Poor
2 screening questions	Screening question for anxiety: "During the past month have you been worrying a lot about everyday problems?" If patients answered yes, then asked to complete a second question: "Is this something with which you would like help?"	Puddifoot, 2007[42]	982 adults (72% female)	HADS anxiety score >11	Two screening questions Sensitivity: 58% Specificity: 87% Worry question alone Sensitivity: 76% Specificity: 82%	Good

ADIS-IV=Anxiety Disorders Interview Schedule; AUC=area under the receiver operating characteristic curve; CI=confidence interval; CIDI=Composite International Diagnostic Interview; DSM=Diagnostic and Statistical Manual of Mental Disorders; DUKE-AD=Duke Anxiety-Depression Scale; GAD=generalized anxiety disorder; GAS=Goldberg Anxiety Scale; HADS=Hospital Anxiety and Depression Scale; HAM-A=Hamilton Anxiety Scale; OCD=obsessive compulsive disorder; PDI-4= Provisional Diagnostic Instrument-4; PHQ=Patient Health Questionnaire; PSWQ=Penn State Worry Questionnaire; PTSD=posttraumatic stress disorder; SCID/ACDS=Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, fourth edition axis I disorders and the Adult ADHD Clinician Diagnostic Scale version 1.2; SCID-IV=Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, fourth edition axis I disorders; WB-DAT=Web-Based Depression and Anxiety Test; WHODAS II=World Health's Organization Disability Assessment Scale; WSQ=Web Screening Questionnaire.

Pregnant and Postpartum Women

A total of 9 studies and 1 systematic review of 25 studies evaluated the diagnostic accuracy of 12 screening methods (**Table 7**). These include Anxiety Disorders-13;[58] two versions of the Edinburgh Postnatal Depression Scale (EPDS);[48,49,52,54,58] the Hospital Anxiety and Depression Scale-anxiety subscale (HADS-A);[48,50] Pregnancy-Related Thoughts (PRT)[48]; Pregnancy Related Anxiety Questionnaire-Revised (PRAQ-R);[48] Matthey Generic Mood Question (MGMQ);[48,58,68] McDonald Prenatal Screening Tool;[49] State Trait Anxiety Inventory (STAI);[50,54] General Health Questionnaire (GHQ);[50] Generalized Anxiety Disorder 7-item and 2-item scales (GAD-7; GAD-2);[52,58] Perinatal Anxiety Screening Scale (PASS);[53] and Beck Anxiety Inventory (BAI-Subj).[51]

The EPDS is a 10-item self-reported measure commonly used in the United States to assess pregnant and postpartum women for symptoms of emotional distress during the past 7 days. Results of three studies varied depending on cut-points and pregnancy status. The AUC value was 0.73 (95% CI 0.62 to 0.83) in a study of pregnant women in the community;[49] 0.75 (95% CI 0.66 to 0.82) in postpartum women in clinics;[58] and 0.62 in a study of pregnant women referred for psychiatric consultation.[52] Three studies of a 3-item variation of the EPDS indicated AUC values 0.69 to 0.76; sensitivity 54% to 68%; and specificity 63% to 76.0%.[48,52,58]

Several studies of pregnant and postpartum women evaluated instruments commonly used in general populations. In a study of pregnant and postpartum women using the GAD-7 with a cut-point of 10, sensitivity was 76.0% and specificity 51.5%,[52] while another study in postpartum women reported AUC values of 0.72 for both the GAD-7 and GAD-2.[58] Studies using variations of the GHQ indicated sensitivity from 75% to 83%, and specificity from 71% to 89%.[50] The STAI demonstrated sensitivity of 66% to 81%, and specificity 67% to 80% in two studies with various cut-points.[50,54] Two studies of the MGMQ in pregnant women attending their first prenatal visit indicated sensitivity ranging from 51% to 80% and specificity 94% to 96%.[48,68]

The diagnostic accuracies of additional methods were reported in single studies. Of these, the Anxiety Disorders-13,[58] McDonald Prenatal Screening Tool,[49] PASS,[53] and BAI-Subj[51] demonstrated moderate to high performance measures, while the PRT[48] and PRAQ-R,[48] indicated lower performance.

Table 7. Studies of Screening Instruments in Pregnant and Postpartum Women

Screening Instrument	Description	Study (author, year)	Participants	Reference Standard	Performance Characteristics (95% CI)	Quality Rating
Anxiety Disorders-13	13 items derived from the Generalized Anxiety Disorder Scale-7 (GAD-7) and other instruments specific to anxiety	Fairbrother, 2019[58]	115 postpartum women mean age 33.2 years in prenatal clinics, physician offices, midwifery clinics	SCID-IV diagnosis	Sensitivity: 86.5% Specificity: 68.2% PPV: 0.274; NPV: 0.973 AUC: 0.834 (0.776-0.893)	Fair
EPDS[78]	10-item self-report measure assessing pregnant and postpartum women for symptoms of emotional distress during the past 7 days.	Fairbrother, 2019[58]	115 postpartum women mean age 33.2 years in prenatal clinics, physician offices, midwifery clinics	SCID-IV diagnosis	Sensitivity: 73.7 Specificity: 63.7 PPV: 0.224; NPV: 0.944 AUC: 0.750 (0.663-0.824)	Fair
		McDonald, 2012[49]	567 pregnant women (<24 weeks) from the community	STAI-state anxiety scale	Sensitivity: 41% (27 to 61) Specificity: 88% (82 to 91) PPV: 0.34 (0.20 to 0.49) NPV: 0.91 (0.87 to 0.95) AUC: 0.73 (0.62 to 0.83)	Poor
		Simpson, 2014[52]	155 pregnant and 85 postpartum women mean age 30.5 years referred for psychiatric consultation	DSM-IV diagnosis	Cutoff 10 to 13 Sensitivity: 77.3 to 89.3% Specificity: 26.7 to 40.3% PPV: 0.36 to 0.38 NPV: 0.79 to 0.84 AUC for GAD: 0.62 AUC for GAD and MDD: 0.68	Poor
		Tendais, 2014[54]	35 pregnant and postpartum women mean age 28 years in obstetrics outpatient unit	SCID diagnosis	Cutoff >9 during pregnancy Sensitivity: 73.7% (56.9 to 86.6) Specificity: 70.0% (60.5 to 78.4) PPV: 0.46 (0.33 to 0.59) NPV: 0.89 (0.80 to 0.94) Cutoff >7 postpartum Sensitivity: 78.3% (56.3 to 92.5) Specificity: 81.6% (71.0 to 89.5) PPV: 0.56 (0.38 to 0.74) NPV: 0.93 (0.83 to 0.98)	Poor

Screening Instrument	Description	Study (author, year)	Participants	Reference Standard	Performance Characteristics (95% CI)	Quality Rating
EDS-3a[78,48]	3 items derived from the EPDS. Each item has 4 response options; total scores on the anxiety subscale range from 0 to 9; higher scores indicate increased anxiety.	Fairbrother, 2019[58]	115 postpartum women mean age 33.2 years in prenatal clinics, physician offices, midwifery clinics	SCID-IV diagnosis	Sensitivity: 65.8 Specificity: 76.0 PPV: 0.281; NPV: 0.940 AUC: 0.757 (0.678-0.836)	Fair
		Matthey, 2013[48]	391 pregnant women mean age 28.8 years attending first prenatal visit	MINI diagnosis	Sensitivity: 54%	Fair
		Simpson, 2014[52]	155 pregnant and 85 postpartum women mean age 30.5 years referred for psychiatric consultation	DSM-IV diagnosis	Sensitivity: 68.0% Specificity: 63.5% PPV: 0.46; NPV: 0.81 AUC for GAD: 0.69 AUC for GAD and MDD: 0.67	Poor
HADS-A[79]	7 items about general anxiety over the past 7 days. Total scores range from 0 to 21; higher scores indicate increased anxiety.	Matthey, 2013[48]	391 pregnant women mean age 28.8 years attending first prenatal visit	MINI diagnosis	Sensitivity: 35%	Fair
		Meades, 2011 (SR)[50]	441 pregnant women	MINI plus, semi-structured interview, or SCID diagnosis	Sensitivity: 92.9% Specificity: 90%	Low
PRT[80]	10 items about concerns regarding the health of the baby, labor and delivery, and caring for the baby over the past 7 days. Each item has 4 response options; total scores range from 10 to 40; higher scores indicate increased anxiety.	Matthey, 2013[48]	391 pregnant women mean age 28.8 years attending first prenatal visit	MINI diagnosis	Sensitivity: 20%	Fair
PRAQ-R[81]	10-item questionnaire with 3 domains using a 5-point scale. Total scores range from 10 to 50; higher scores indicate increased anxiety.	Matthey, 2013[48]	391 pregnant women mean age 28.8 years attending first prenatal visit	MINI diagnosis	Sensitivity: 33%	Fair

Screening Instrument	Description	Study (author, year)	Participants	Reference Standard	Performance Characteristics (95% CI)	Quality Rating
MGMQ[48]	1 question: In the last 2 weeks have you felt very stressed, anxious or unhappy, or found it difficult to cope, for some of the time? Follow-up question for those answering "Yes" or "Possible:" How bothered have you been by these feelings?	Matthey, 2013[48]	391 pregnant women mean age 28.8 years attending first prenatal visit	MINI diagnosis	Sensitivity: 80%	Fair
		Matthey, 2019[68]	252 pregnant women mean age 28.4 years at first prenatal visit	DSM-IV criteria	Sensitivity: 51-56% Specificity: 94-96% PPV: 0.67-0.73	Fair
McDonald Prenatal Screening Tool	Includes items relating to depression, stress, abuse history, and poor relationship quality.	McDonald, 2012[49]	567 pregnant women <24 weeks	STAI-state anxiety scale	Sensitivity: 44% (29 to 60) Specificity: 88% (82 to 91) PPV: 0.34 (0.20 to 0.49) NPV: 0.91 (0.87 to 0.95) AUC: 0.71 (0.61 to 0.82)	Poor
STAI [82]	Consists of 2 subscales with 20 items each. Measures anxiety at this moment or in general. Respondents endorse items on a 4-point scale.	Meades, 2011 (SR)[50]	100 pregnant women	MINI plus, semi-structured interview, or SCID diagnosis	Cutoff >40 Sensitivity: 80.95% Specificity: 79.75% PPV: 0.52; NPV: 0.94	Low
		Tendais, 2014[54]	35 pregnant women mean age 28 years	SCID diagnosis	Cutoff >40 during pregnancy Sensitivity: 65.7% (47.8 to 80.9) Specificity: 67.3% (57.8 to 75.8) PPV: 0.38 (0.26 to 0.52) NPV: 0.86 (0.77 to 0.93) Cutoff >34 postpartum Sensitivity: 71.4% (66.1 to 99.8) Specificity: 67.1% (56.0 to 76.9) PPV: 0.26 (0.13 to 0.43) NPV: 0.93 (0.84 to 0.98)	Poor
GHQ[83-85]	Asks how the respondent has felt recently using 4-point response scales; higher scores indicate increased likelihood of disorder. The GHQ has 4 different versions (60-item, 30- item; 28-item and 12-item versions) and can be scored 4 different ways.	Meades, 2011 (SR)[50]	2525 pregnant women	Clinical interview schedule, SCID, or ICD-o diagnosis; SADS, PAS, or ICD-9	GHQ-30 (3 studies) Sensitivity: 77 to 83% Specificity: 71 to 89% PPV: 0.37 to 0.53 NPV: 0.90 to 0.97 GHQ-28 (2 studies) Sensitivity: 75%; 82% Specificity: 83%; 85% PPV: 0.46; 0.53; NPV: 0.95; 0.96 GHQ-12 (2 studies) Sensitivity: 83%; 81% Specificity: 80%; 81%	Low

Screening Instrument	Description	Study (author, year)	Participants	Reference Standard	Performance Characteristics (95% CI)	Quality Rating
GAD-7[20]	7 self-rated items are each scored from 0 to 3; total score ranges from 0 to 21.	Simpson, 2014[52]	155 pregnant and 85 postpartum women mean age 30.5 years	DSM-IV diagnosis	Cutoff >10 Sensitivity: 76.0% Specificity: 51.5% PPV: 0.42; NPV: 0.83 Cutoff >13 Sensitivity: 61.3% Specificity: 72.7% PPV: 0.51; NPV: 0.81 AUC for GAD: 0.71 AUC for GAD and MDD: 0.74	Poor
		Fairbrother, 2019[58]	115 postpartum women mean age 33.2 years in prenatal clinics, physician offices, midwifery clinics	SCID-IV diagnosis	Sensitivity: 55.3% Specificity: 83.2% PPV: 0.318; NPV: 0.929 AUC: 0. 719 (0.619-0.818)	Fair
GAD-2	2-item variation of the GAD derived from the anxiety module of the PHQ.	Fairbrother, 2019[58]	115 postpartum women mean age 33.2 years in prenatal clinics, physician offices, midwifery clinics	SCID-IV diagnosis	Sensitivity: 81.6% Specificity: 50.9% PPV: 0.193; NPV: 0.951 AUC: 0.718 (0.675-0.829)	Fair
PASS	38-item self-report questionnaire with a 4-point Likert scale assessing the frequency of symptoms.	Somerville, 2014[53]	53 pregnant and postpartum women ≥18 years in prenatal clinic	ICD-10 diagnosis	Sensitivity: 70% Specificity: 30% AUC: 0.7 (SE 0.04)	Poor
BAI-Subj [86]	Assesses 21 affective and somatic symptoms of anxiety on a 4-point scale. The 6-item Subjective subscale (BAI-Subj) was used in the study because it most clearly represents general anxiety symptoms.	O'Hara, 2012[51]	353 postpartum women mean age 27 years; mean 21 weeks postpartum	SCID diagnosis	Cutoff >4 Sensitivity: 76% Specificity: 71% PPV: 0.31 AUC: 0.78 Cutoff >6 Sensitivity: 56% Specificity: 82% PPV: 0.35	Fair

AUC=area under the receiver operating characteristic curve; BAI=Beck Anxiety Inventory; CI=confidence interval; DSM=Diagnostic and Statistical Manual of Mental Disorders; EDS-3a=Edinburgh Depression Scale-anxiety subscale; EPDS=Edinburgh Postnatal Depression Scale; GAD=generalized anxiety disorder; GAD-7=Generalized Anxiety Disorder 7-item scale; GHQ=General Health Questionnaire; HADS-A=Hospital Anxiety and Depression Scale-anxiety subscale; MINI-Mini-International Neuropsychiatric Interview; MGMQ=Matthey Generic Mood Question; NPV=negative predictive value; PASS=Perinatal Anxiety Screening Scale; PPV=positive predictive value; PRAQ-R=Pregnancy Related Anxiety Questionnaire-Revised; PRT=Pregnancy-Related Thoughts; SCID=Structured Clinical Interview for DSM-IV; SE=standard error; SR=systematic review; STAI=State Trait Anxiety Inventory.

Older Adults

Three studies evaluated five screening methods in adults age 60 years and older (**Table 8**). These included the Anxiety Disorder Scale (ADS) and FEAR instruments specific to older adults[55]; and the GAD-7,[56] Hospital Anxiety and Depression Scale (HADS),[57] and Brief Symptom Inventory (BSI-18)[57] that are used in general adult populations.

The ADS was developed as a survey instrument for detecting anxiety disorders in individuals age 65 and older and includes 11 items in a generalized anxiety subscale. The FEAR instrument is a 4-item version of the ADS. A study of older adults in primary care settings indicated 85% sensitivity and 71% specificity for the ADS; and 74% sensitivity and 85% specificity for FEAR.[55] Additional studies of older patients in primary care clinics indicated AUC values of 0.695 for GAD-7;[56] 0.80 for HADS;[57] and 0.573 for BSI-18.[57]

Table 8. Studies of Screening Instruments in Older Adults

Screening Instrument	Description	Study (author, year)	Participants	Reference Standard	Performance Characteristics	Quality Rating
ADS FEAR	ADS: Developed as a survey instrument for detecting anxiety disorders in a community sample of individuals age 65 and over. Includes 11 items in a generalized anxiety subscale. FEAR: 4-item version of the ADS.	Krasucki, 1999[55]	88 adults age >65 in primary care settings (64% female)	Clinical Interview, ICD-10 diagnosis	ADS: Sensitivity: 85% Specificity: 71% FEAR: Sensitivity: 74% Specificity: 85%	Poor
GAD-7	Evaluates anxiety symptoms using a 4-point Likert scale; total scores range from 0 to 21. Higher scores indicate higher severity, while scores below 5 indicate minimal anxiety.	Vasiliadis, 2015[56]	1775 adults age ≥65 in primary care clinics (57% female)	DSM-IV diagnosis, in person interview	Cutoff 5: Sensitivity: 71% Specificity: 57% AUC 0.695	Good
HADS BSI-18	HADS: 14-item questionnaire to detect anxiety and depression in the general medical outpatient population. BSI-18: Includes 6 items scored on a 5-point Likert scale. Includes items assessing depression and anxiety.	Wetherell, 2007[57]	68 adults >60 in primary care clinics (67% female)	ADIS-IV interview, DSM diagnosis	HADS: Sensitivity: 97% Specificity: 67% AUC 0.80 BSI-18: AUC 0.573, SE 0.092	Poor

ADIS-IV=Anxiety Disorders Interview Schedule; ADS=Anxiety Disorder Scale; AUC=area under the receiver operating characteristic curve; BSI=Brief Symptom Inventory; DSM=Diagnostic and Statistical Manual of Mental Disorders; FEAR=Frequency of anxiety; Enduring nature of anxiety; Alcohol or sedative use; Restlessness or fidgeting; GAD=generalized anxiety disorder; HADS=Hospital Anxiety and Depression Scale; ICD=International Statistical Classification of Diseases; SE=standard error.

Contextual Question 1. Effectiveness of Treatments for Anxiety

The effectiveness of treatments for anxiety has been evaluated by studies that are summarized in systematic reviews of psychological[16,87-90] and pharmacological treatments[91-99] (**Appendix 6**).

Psychological Therapy

Cognitive behavioral therapy (CBT) is the first line psychological therapy for generalized anxiety disorder and usually includes a combination of psychoeducation, worry exposure, relaxation, applied relaxation, problem-solving, cognitive re-structuring, and interpersonal psychotherapy.[100] Therapy can be delivered individually or as a group, face-to-face, over the internet, or via the telephone.

Five systematic reviews evaluated the effectiveness of psychological therapies for anxiety disorders (**Table 9**). Two Cochrane reviews, one of therapies for children and adolescents,[89] and the other for adults,[88] compared psychological therapies with waitlist controls, active treatment arms (either another psychological therapy or pharmacotherapy), usual care, and psychological placebos. Three additional reviews of adults compared individual, group, computer or internet delivered CBT with face-to-face therapy, an alternate media-delivered intervention, waitlist control, psychological placebos, or usual care.[16,87,90]

Adolescents

A Cochrane review of psychological therapies in children and adolescents included 42 studies (41 in meta-analysis) enrolling 1806 participants.[89] Most studies enrolled children ages 7 to 14 years, although some included up to age 18 years. Studies were predominantly conducted in research settings, such as university outpatient clinics, although some were conducted in community clinics and inner-city schools. Symptoms of anxiety were assessed using the Revised Children's Manifest Anxiety Scale (RCMAS); Fear Survey for Children Revised (FSSC-R); Social Anxiety Scale for Adolescents (SAS-A); Social Phobia and Anxiety Inventory for Children (SPAI); Mood and Anxiety Symptom Scale (MASQ); Spence Child Anxiety Scale, child and parent versions (SCAS); Child Behavior Checklist (CBCL); and the Clinical Global Impressions Scale.

In a meta-analysis of studies, anxiety remission was improved for CBT compared with waitlist controls (odds ratio [OR] 7.85; 95% CI 5.31 to 11.60; 25 studies). Results were similar for individual and group CBT compared with waitlist controls (individual therapy OR 7.92; 95% CI 3.37 to 18.63; 7 studies; group therapy OR 7.86; 3.83 to 16.12; 13 studies). Differences were not statistically significant for CBT compared with active controls (i.e., another psychological therapy or pharmacotherapy) or usual care. Long-term remission was more likely for CBT compared with active therapy (OR 2.03; 95% CI 1.22 to 3.36; 2 studies), but not compared with waitlist controls.

The standard mean difference (SMD) score on standardized scales was used to measure changes in anxiety symptoms. Symptoms were significantly reduced for CBT versus waitlist controls (SMD 0.98; 95% CI -1.21 to -0.74; 30 studies), but not compared with active controls or usual care. Long-term remission was similar for CBT compared with waitlist controls and active controls.

Table 9. Systematic Reviews of Cognitive Behavior Therapy

Systematic review (author, year)	Intervention and Comparison	Measure	Outcome	Number of trials; number of participants	Summary of Main Findings (95 % CI)
Adolescents					
James, 2015[89]	CBT versus wait list control	Self-reported measures	Remission for all anxiety disorders	26; 1350	OR = 7.85 (5.31 to 11.60)
			Reduction in anxiety symptoms for all anxiety disorders	30; 1394	SMD = -0.98 (-1.21 to -0.74)
Adults					
Andrews, 2018[87]	Internet CBT versus wait list, placebo, or usual care	PSWQ or GAD-7	Treatment of anxiety for GAD	9; 1103	Hedge's g effect size = 0.70 (0.39 to 1.01)
Hunot, 2010[88]	All types of psychotherapy with components of CBT versus wait list or usual care	Most often a 20% reduction in anxiety symptoms from pre to post intervention (mainly the HAM-A or STAI-T)	Treatment response	8; 334	RR = 0.64 (0.55 to 0.74)
			Reduction in anxiety symptoms	12; 330	SMD = -1.00 (-1.24 to 0.77)
Mayo-Wilson, 2013[90]	Media-delivered therapy versus no intervention	Any self-reported measure	Treatment response for GAD	4; 342	RR = 4.60 (2.75 to 7.68)
			Reduction in anxiety symptoms for GAD	10; 649	SMD = 0.95 (0.44 to 1.45)
	Media-delivered therapy versus face-to-face intervention	Any self-reported measure	Treatment response for all anxiety disorders	10; 575	RR = 0.78 (0.56 to 1.09)
			Reduction in anxiety symptoms for all anxiety disorders	24; 1360	SMD = -0.23 (-0.36 to -0.09)
van Dis, 2019[16]	Individual, group, or internet CBT versus usual care, relaxation, psychoeducation, pill placebo, supportive therapy, or wait list	Structured diagnostic interview	Anxiety symptoms	Immediate: 14; 856; 12 months: 10; 657	Immediate: Hedge's g effect size = 0.39 (0.12 to 0.66); 12 months: Hedge's g effect size = 0.22 (0.02 to 0.42)

CI=confidence interval; CBT=cognitive behavioral therapy; HAM-A=Hamilton Anxiety Scale; GAD=generalized anxiety disorder; GAD-7=Generalized Anxiety Disorder-7; OR=odds ratio; PSWQ=Penn State Worry Questionnaire; SMD=standard mean difference; STAT-T=Spielberger State-Trait Anxiety Inventory--Trait subscale; RR=risk ratio.

Adults

CBT versus controls or active therapy groups. A Cochrane review of psychological therapies for adults included 25 studies (22 in the meta-analysis) enrolling 1305 participants.[88] Sample sizes ranged from 12 to 119, with an average of 54. Most studies were conducted in the United States and the United Kingdom, with others in Canada and European countries. Most studies were set in out-patient psychiatric or psychology department clinics or community mental health settings, while two were conducted in primary care clinics and one at a university campus. The majority of the participants were female (68.6%) and the mean age across studies was 47.2 years (mean age 38.1 years in studies of general adult populations and 61.1 years in studies of older populations). Twenty-three studies included participants with a primary diagnosis of generalized anxiety disorder.

The Hamilton Anxiety Scale (HAM-A) was the most frequently used clinician-rated outcome measure (13 studies), and the Trait subscale of the Spielberger State-Trait Inventory (STAI-I) was the most frequently used self-report measure (16 studies). The Penn State Worry Questionnaire (PSWQ) was used in 10 studies, the Beck Anxiety Inventory (BAI) in 9, and the Zung Anxiety Inventory (ZAI) in 8.

In a meta-analysis of studies, clinical response was improved with CBT compared with waitlist or usual care controls (relative risk [RR] 0.64; 95% CI 0.55 to 0.74; 8 studies), psychodynamic therapy (RR 0.77; 95% CI 0.65 to 0.92; 1 study), and behavioral therapy (RR 0.70; 95% CI 0.56 to 0.87; 5 studies). Results were not statistically significantly different between CBT and supportive therapy. At 6-months follow-up, differences were statistically significantly different for cognitive compared with behavioral therapy (RR 0.70; 95% CI 0.56 to 0.87; 5 studies), but not for CBT compared with psychodynamic therapy, or CBT compared with supportive therapy.

The standard mean difference (SMD) score on standardized scales was used to measure changes in anxiety symptoms. Symptoms were significantly reduced with CBT versus waitlist or usual care controls (SMD -1.00; 95% CI -1.24 to -0.77; 12 studies), psychodynamic therapy (SMD -6.85; 95% CI -11.20 to -2.50; 2 studies), and supportive therapy (SMD -0.40; 95% CI -0.66 to -0.14; 7 studies), but not behavioral therapy. Differences between groups were significant at 6-months follow-up for CBT compared with psychodynamic therapy (SMD -13.41; 95% CI -19.09 to -7.74; 2 studies) and supportive therapy (SMD -0.42; 95% CI -0.83 to -0.02; 3 studies), but not with behavioral therapy. Differences were not statistically significantly different at 12-months follow-up for CBT compared with supportive therapy, or cognitive therapy compared with behavioral therapy.

A second meta-analysis of trials compared individual, group, or internet CBT with usual care, relaxation, psychoeducation, pill placebo, supportive therapy, or wait list controls.[16] Overall, anxiety symptoms, determined by structured diagnostic interviews, improved for CBT at both immediate (Hedge's g effect size 0.39; 95% CI 0.12 to 0.66; 14 studies); and 12-month time points (Hedge's g effect size 0.22; 95% CI 0.02 to 0.42; 10 studies).

Media-delivered CBT versus controls or face-to-face CBT. A Cochrane review comparing media-delivered therapy of any type (print, audio or video recordings, or computers including the

internet) with no intervention and with face-to-face CBT or behavioral therapy in adults with anxiety disorders included 101 studies (91 studies in meta-analysis) enrolling 8043 participants.[90] Most participants were white (94%) and female (67%), with a mean age of 37 years. The review included 10 studies of generalized anxiety disorder, while other studies included other types of anxiety disorders.

In a meta-analysis of studies, compared with no intervention, clinical response was improved for media-delivered interventions for all anxiety disorders (RR 2.34; 95% CI 1.81 to 3.03; 21 studies) and for generalized anxiety disorder specifically (RR 4.60; 95% CI 2.75 to 7.68; 4 studies). Symptoms were also significantly reduced in studies comparing media-delivered interventions with no intervention for all anxiety disorders (RR 0.67; 95% CI 0.55 to 0.78, 76 studies), but not for generalized anxiety disorder specifically. Recovery, as determined by clinical assessment at post-treatment, was reduced for media-delivered interventions compared with no intervention for all anxiety disorders (RR 0.40; 95% CI 0.20 to 0.60; 9 studies). Compared with face-to-face interventions, symptoms were reduced for media-delivered interventions (SMD -0.23; 95% CI -0.36 to -0.09; 24 studies), while clinical response and recovery were not statistically significantly different.

Internet CBT versus controls or face-to-face CBT. A systematic review comparing internet CBT with face-to-face CBT, waitlist control, information control, care as usual, or placebo in adults with either depression or anxiety included nine studies of 1103 participants.[87] Among participants treated for generalized anxiety disorder, differences in symptoms between internet CBT and all other treatments combined were not statistically significantly different (Hedge's g effect size 0.70; 95% CI 0.39 to 1.01).

Pharmacological Therapy

Nine systematic reviews summarized RCTs of the effectiveness of pharmacological treatments.[91-99] Results of systematic reviews of first-line (SSRI, SNRI, buspirone) and second-line (tricyclic antidepressants, calcium modulators) agents are included in this report.[94,98,99]

Adolescents

A Cochrane review of SSRIs and SNRIs in children and adolescents included short-term (≤16 weeks) trials.[94] Studies ranged in size from 15 to 322 participants with a mean age of 12 years; more than half were female (52.1%). Most trials were conducted in the United States. In this review, two trials treated patients with generalized anxiety disorder and three trials included patients with either generalized anxiety disorder, social phobia, or separation anxiety disorder. Medications included fluoxetine, fluvoxamine, sertraline (with or without CBT), and venlafaxine-ER.

Treatment response for generalized anxiety disorder was improved for all medications compared with placebo for fluoxetine 10 to 20 mg/day (RR 1.74; 95% CI 1.04 to 2.89); fluvoxamine 50-300 mg/day (dose based on weight) (RR 2.61; 95% CI 1.74 to 3.90); sertraline 50 mg/day (RR 10.00; 95% CI 1.53 to 65.41); sertraline 25 to 200 mg/day (with or without CBT) (RR 2.32; 95% CI 1.50 to 3.57); and venlafaxine-ER 37.5 to 225 mg/day (dose based on weight) (RR 1.44; 95% CI 1.19 to 1.75).

A more recent systematic review and meta-analysis included additional medications.[98] Improvement in anxiety based on clinician evaluations was statistically significant for all SSRIs (citalopram, paroxetine, sertraline, fluoxetine) and SNRIs (duloxetine, venlafaxine, atomoxetine, fluvoxamine) evaluated compared with placebo (**Table 10**).

Table 10. Treatment Effects of Anti-Anxiety Medications versus Placebo for Children and Adolescents[98]

Class	Medication	Dose	Number of trials; number of participants	Effect on Anxiety (clinician report)* Standard Mean Difference (95% CI)
SSRI	Citalopram (Celexa)	Not approved	1; 272	-0.43 (-0.67 to -0.19)
	Paroxetine (Paxil)	Not approved	1; 137	-0.71 (-1.06 to -0.37)
	Sertraline (Zoloft)	50-200 mg	2; 231	-0.71 (-0.99 to -0.42)
	Fluoxetine (Prozac)	Not approved	2; 154	-0.40 (-0.72 to -0.01)
SNRI	Duloxetine (Cymbalta)	30-120 mg	1; 272	-0.43 (-0.67 to -0.19)
	Venlafaxine (Effexor)	Not approved	1; 153	-0.42 (-0.74 to -0.10)
	Atomoxetine (Strattera)	Not approved	2; 331	-0.56 (-0.78 to -0.34)
	Fluvoxamine (Luvox)	Not approved	2; 153	-0.97 (-1.31 to -0.63)

CI=confidence interval; SNRI=serotonin-norepinephrine reuptake inhibitor; SSRI=selective serotonin reuptake inhibitor.
*Compared with pill placebo

Adults

The efficacy of SSRIs and SNRIs for treating anxiety in adults has been well established in RCTs and they are generally FDA approved for this use. Information on older patients is limited and pregnant women were not included in trials, although these medications are widely used in these patient groups.

A systematic review and network meta-analysis published in 2019 included trials of anti-anxiety medications compared with pill placebo.[99] Treatment effects were measured using scores from the Hamilton Anxiety Rating Scale (HAM-A). Results of for first-line (SSRI, SNRI, buspirone) and second-line (tricyclic antidepressants, calcium modulators) pharmacologic therapies used in the United States are summarized in **Table 11** below. Medications with statistically significant reductions in mean anxiety scores compared with placebo included SSRIs (46 trials; 4,229 participants), SNRIs (22 trials; 3,652 participants), buspirone (6 trials; 311 participants), and pregabalin (11 trials; 1,957 participants).

Table 11. Treatment Effects of Anti-Anxiety Medications versus Placebo for Adults[99]

Class	Medication	Number of Trials; Number of Participants	Difference in Anxiety Score (HAM-A); Mean Difference (95% CrI)	Acceptability (discontinuation); Odds Ratio (95% CrI)
SSRI	Citalopram (Celexa)	2; 37	-2.22 (-4.28 to -0.19)	3.62 (0.74 to 20.27)
	Escitalopram (Cipralex)	13; 1581	-2.45 (-3.27 to -1.63)	0.96 (0.79 to 1.16)
	Fluoxetine (Prozac)	8; 264	-2.43 (-3.74 to -1.16)	1.36 (0.57 to 3.15)
	Paroxetine (Paxil)	17; 1862	-2.29 (-3.11 to -1.47)	1.24 (1.03 to 1.50)
	Sertraline (Zoloft)	6; 485	-2.88 (-4.17 to -1.59)	0.94 (0.65 to 1.35)
SNRI	Duloxetine (Cymbalta)	8; 1355	-3.13 (-4.13 to -2.13)	1.09 (0.89 to 1.32)
	Venlafaxine (Effexor)	14; 2275	-2.69 (-3.50 to -1.89)	0.98 (0.83 to 1.16)
Others	Buspirone	6; 311	-2.37 (-3.83 to -0.91)	0.76 (0.47 to 1.25)
	Imipramine (Tofranil)	1; 26	-0.59 (-3.85 to 2.70)	2.83 (0.74 to 12.10)
	Pregabalin (Lyrica)	11; 1957	-2.79 (-3.69 to -1.91)	0.80 (0.66 to 0.98)

CrI=credible interval; HAM-A=Hamilton Anxiety Rating Scale; SNRI=serotonin-norepinephrine reuptake inhibitor; SSRI=selective serotonin reuptake inhibitor.

Contextual Question 2. Adverse Effects of Treatments for Anxiety

Psychological Therapy

In the Cochrane review of 25 studies of psychological therapies for adults, attrition for any reason at post-treatment did not differ between CBT and waitlist or usual care controls (RR 1.00; 95% CI 0.65 to 1.54; 13 studies); or between CBT and various active treatment groups.[88] In a Cochrane review comparing media-delivered therapy with no intervention and with face-to-face CBT, attrition was lower for media-delivered interventions compared with no intervention (RR 0.96; 95% CI 0.94 to 0.99; 78 studies); but not for media-delivered versus face-to-face interventions (RR 0.99; 95% CI 0.95 to 1.03; 28 studies).[90] No other harms were reported.

Pharmacological Therapy

Adolescents

In a Cochrane review of SSRIs and SNRIs in children and adolescents, drop outs due to adverse effects did not differ between treatment and placebo groups for fluoxetine, fluvoxamine, sertraline with CBT, and venlafaxine-ER.[94] The most frequent treatment-emergent adverse effects were abdominal pain and nausea for fluoxetine; abdominal discomfort, increased motor activity, vomiting, tiredness/fatigue, muscle/joint pain, and decreased appetite for fluvoxamine; and anorexia for sertraline without CBT. None were reported for sertraline with CBT.

A more recent systematic review and meta-analysis of trials in children and adolescents included additional medications.[98] The review concluded that SSRIs and SNRIs were associated with increased risk of various short-term adverse effects that were overall not serious, similar to the Cochrane review. Studies were too small or too short to assess the effect on suicidal behavior,

although one study found that venlafaxine was associated with a statistically nonsignificant increase in the risk of suicidal ideation.

Adults

A systematic review and network meta-analysis published in 2019 included trials of anti-anxiety medications compared with placebo that reported discontinuation for any reason as an adverse effect.[99] Results indicated no differences in discontinuation between treatment and pill placebo groups (**Table 11**). In another systematic review of adverse effects of SSRIs and SNRIs for treating depression (not anxiety) in adults, 63% of patients experienced at least one adverse event, with diarrhea, dizziness, dry mouth, fatigue, headache, nausea, sexual dysfunction, sweating, tremor, and weight gain most commonly reported.[101] In general, trials were too small or too short to assess more serious adverse events, such as suicide, cardiovascular events, or others. Older patients were not specifically studied and pregnant women were not included in these trials.

CONCLUSIONS

A summary of evidence is described in **Table 12.** Results of this systematic review indicate that no studies have evaluated the overall effectiveness or harms of screening for anxiety in women and adolescent girls. The strength of evidence for the accuracy of screening methods to identify women with anxiety is high based on 33 studies and 2 systematic reviews evaluating 27 clinical screening instruments and their variations against a clinical diagnosis of anxiety disorder. Screening methods were similar across studies and included predominantly brief clinician or self-administered questionnaires describing symptoms that were easily scored and interpreted. Most studies enrolled predominantly women in community or primary care settings that are highly applicable to population screening and used DSM criteria as the reference standard.

Overall, most screening methods demonstrated moderate to high discriminatory accuracy in studies. For adolescents, studies of four versions of SCARED indicated sensitivity ranging from 64% to 74% and specificity from 64% to 73%. Notably, the 5-item version of SCARED performed similarly to longer versions. Other methods for adolescents demonstrated varying accuracy, with the highest accuracy reported for the MASC (0.82 AUC specifically for girls).

In adults, results of 17 studies evaluating 10 screening instruments and their variations indicated generally moderate to high discriminatory accuracy. Eight studies of the GAD (GAD-2, GAD-7, GAD-Q-IV) were evaluated in large primary care populations and indicated sensitivity ranging from 70% to 97% and specificity from 50% to 89%; AUCs from 0.77 to 0.94; with most results falling in the high end of these ranges. The GAD-2, with only two questions, performed as well as longer versions. Results were similar for the other methods studied, even when the method involved asking only one question.

For pregnant and postpartum women, results of studies of the EPDS, commonly used for depression screening, indicated varying results for anxiety (AUROC 0.62 to 0.73; sensitivity 41% to 89%; specificity 27% to 88%). Additional methods for general populations (BAI, GAD-7, GHQ, STAI) were more accurate than methods specific to pregnancy and postpartum (PRT, PRAQ-R, McDonald Prenatal Screening Tool). Use of the EPDS for screening in practice would

be an efficient method for both anxiety and depression in pregnant and postpartum women in clinical settings, although pairing the EPDS with the GAD-7, for example, could be more effective, although this has not been formally evaluated.

For older adults, methods specific to older patients (ADS, FEAR) had similar diagnostic accuracy as the HADS and GAD-7, while the BSI-18 was less accurate.

The strength of evidence for the effectiveness of treatment ranges from moderate to high, and for harms of treatment, is low for cognitive behavioral therapy and moderate for medications. Studies of treatment for anxiety indicate that CBT is effective in reducing symptoms and improving remission in adults and adolescents, and may be preferred for pregnant women and those intolerant of anti-anxiety medications. In addition, trials indicate effectiveness when CBT is delivered via the internet and by media in addition to face-to-face counseling, creating more opportunities for engagement.

SSRIs and SNRIs are the most common first-line pharmacologic treatments that have proven efficacy in RCTs. While these medications are generally approved by the U.S. Food and Drug Administration for treatment for anxiety in adults, few are approved for use in adolescents (e.g., sertraline, duloxetine). Information on older patients is limited and pregnant women were not included in trials, although these medications are widely used in these patient groups.

In conclusion, studies support a strong evidence base of moderate to highly accurate instruments for screening for anxiety that are applicable to clinical practices serving adolescent and adult women including those pregnant or postpartum. Brief instruments with as few as two questions are as accurate as longer instruments and are particularly useful for routine screening in primary care settings. Once identified, women with anxiety may benefit from CBT with or without pharmacologic therapies depending on severity of symptoms and preferences. CBT and anti-anxiety medications have proven effectiveness in randomized trials. Anti-anxiety medications, such as SSRIs and SNRIs, have proven effectiveness in RCTs, are widely used, generally well-tolerated, and are also effective for depression, which often accompanies anxiety or can develop subsequently. While trials of the overall effectiveness of screening for anxiety disorders are lacking, studies of the accuracy of screening methods and effectiveness and harms of treatment provide evidence supporting essential steps in the clinical pathway.

Table 12. Summary of Evidence

Key Question	Studies; N	Summary of Findings	Limitations	Strength of evidence; applicability
KQ 1. Effectiveness of screening for anxiety	No studies	Not applicable	Not applicable	Insufficient; insufficient
KQ 2. Accuracy of screening methods	33 studies and 2 SRs with 171 studies of 27 instruments and their variations (*n*=112,574)	Accuracy varied by method; several methods have moderate to good discriminatory accuracy in identifying anxiety in adolescents, adults, pregnant and postpartum women, and older adults in primary care and maternity populations.	Studies varied in size, reference standards, and populations.	High; moderate
KQ 3. Harms of screening	No studies	Not applicable	Not applicable	Insufficient; insufficient
Contextual Question	**Studies; N**	**Summary of Findings**	**Limitations**	**Strength of evidence; applicability**
CQ 1. Effectiveness of treatment—cognitive behavioral therapy	5 systematic reviews of 246 RCTs (*n*=17,209)	Trials of CBT versus waitlist or usual care indicate improved remission/clinical response and reduced symptoms for various types of CBT including media and internet delivered.	Few trials for specific populations, such as adolescents, pregnant women, and older women; lack of long-term outcomes.	Moderate; moderate
CQ 1. Effectiveness of treatment—medication	3 systematic reviews of 126 RCTs (*n*=8,225)	SSRIs and SNRIs are effective first-line medication treatments for anxiety based on efficacy RCTs. Additional medications are effective for specific anxiety disorders or when SSRI/SNRIs are not effective or tolerated.	Few trials for specific populations; FDA approval for pediatric and pregnant patients is limited and some medications are used off label; lack of long-term outcomes.	Moderate to high; moderate
CQ 2. Harms of treatment— cognitive behavioral therapy	1 systematic review of 25 RCTs (*n*=1305)	Attrition for any reason at post-treatment did not differ between CBT and controls; no other harms were reported.	Other outcomes not specifically measured.	Low; low
CQ 2. Harms of treatment—medication	3 systematic reviews of 106 RCTs (*n*=8,225)	SSRIs and SNRIs are widely used and well-tolerated; adverse effects have been described and vary by medication. Discontinuation rates are similar between medications and pill placebos in trials.	Studies were too small or too brief to assess more serious adverse events, such as suicide and cardiovascular events. Older patients were not specifically studied and pregnant women were not included.	Moderate; moderate

CBT=cognitive behavioral therapy; CQ=contextual question; FDA=Food and Drug Administration; KQ=key question; RCT=randomized controlled trial; SNRI=serotonin-norepinephrine reuptake inhibitor; SR=systematic review; SSRI=selective serotonin reuptake inhibitor.
*Overall ranking of evidence: high=high confidence that the evidence reflects the true effect; further research is very unlikely to change confidence in the estimate of effect; moderate=moderate confidence that the evidence reflects the true effect; further research may change confidence in the estimate of effect and may change the estimate; low=low confidence that the evidence reflects the true effect; further research is likely to change confidence in the estimate of effect and is likely to change the estimate; insufficient=evidence either is unavailable or does not permit a conclusion.
†Applicability describes how well the overall body of evidence would apply to the U.S. population based on settings, populations, and intervention characteristics (high, moderate, low, insufficient).

REFERENCES

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). 5 ed. Washington, D.C.: American Psychiatric Pub; 2013.
2. Kroenke K, Spitzer RL, Williams JW, et al. Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. Ann Intern Med. 2007;146(5):317-25. doi: 10.7326/0003-4819-146-5-200703060-00004. PMID: 17339617.
3. Kessler RC, Nelson CB, McGonagle KA, et al. Comorbidity of DSM–III–R major depressive disorder in the general population: results from the US National Comorbidity Survey. Br J Psychiatry. 1996;168(S30):17-30. doi: 10.1192/S0007125000298371. PMID: 8864145.
4. Bonnet F, Irving K, Terra J-L, et al. Anxiety and depression are associated with unhealthy lifestyle in patients at risk of cardiovascular disease. Atherosclerosis. 2005;178(2):339-44. doi: 10.1016/j.atherosclerosis.2004.08.035. PMID: 15694943.
5. Stein MB, Roy-Byrne PP, Craske MG, et al. Functional impact and health utility of anxiety disorders in primary care outpatients. Med Care. 2005;43(12):1164-70. PMID: 16299426.
6. Greenberg PE, Sisitsky T, Kessler RC, et al. The economic burden of anxiety disorders in the 1990s. The Journal of Clinical Psychiatry. 1999;60(7):427-35. doi: 10.4088/JCP.v60n0702. PMID: 10453795.
7. Kessler RC, Chiu W, Demler O, et al. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry. 2005;62(6):617-27. doi: 10.1001/archpsyc.62.6.617. PMID: 15939839.
8. Kessler RC, Berglund P, Demler O, et al. Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication. Arch Gen Psychiatry. 2005;62(6):593-602. doi: 10.1001/archpsyc.62.6.593. PMID: 15939837.
9. Harvard Medical School. National Comorbidity Survey. Data Table 2. 2007.
10. National Institute of Mental Health. U.S. Department of Health and Human Services; 2017. https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder.shtml.
11. Rifkin-Graboi A, Meaney MJ, Chen H, et al. Antenatal maternal anxiety predicts variations in neural structures implicated in anxiety disorders in newborns. J Am Acad Child Adolesc Psychiatry. 2015;54(4):313-21.e2. doi: 10.1016/j.jaac.2015.01.013. PMID: 25791148.
12. Graham AM, Rasmussen JM, Rudolph MD, et al. Maternal systemic interleukin-6 during pregnancy is associated with newborn amygdala phenotypes and subsequent behavior at 2 years of age. Biol Psychiatry. 2018;83(2):109-19. doi: 10.1016/j.biopsych.2017.05.027. PMID: 2875415.
13. Rouquette A, Pingault J, Fried EI, et al. Emotional and behavioral symptom network structure in elementary school girls and association with anxiety disorders and depression in adolescence and early adulthood: a network analysis. JAMA Psychiatry. 2018doi: 10.1001/jamapsychiatry.2018.2119. PMID: 30128480.
14. Bor W, Dean AJ, Najman J, et al. Are child and adolescent mental health problems increasing in the 21st century? A systematic review. Aust N Z J Psychiatry. 2014;48(7):606-16. doi: 10.1177/0004867414533834. PMID: 24829198.
15. 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):122-39. doi: 10.1002/mpr.169. PMID: 15297907.
16. van Dis EAM, van Veen SC, Hagenaars MA, et al. Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders: A systematic review and meta-analysis. JAMA Psychiatry. 2019 Nov 23doi: 10.1001/jamapsychiatry.2019.3986. PMID: 31758858.
17. Bandelow B, Reitt M, Rover C, et al. Efficacy of treatments for anxiety disorders: a meta-analysis. Int Clin Psychopharmacol. 2015 Jul;30(4):183-92. doi: 10.1097/yic.0000000000000078. PMID: 25932596.
18. Siu AL. Screening for depression in children and adolescents: US Preventive Services Task Force recommendation statement. Ann Intern Med. 2016;164(5):360-6.
19. Siu AL, U.S. Preventive Services Task Force, Bibbins-Domingo K, et al. Screening for depression in adults: US Preventive Services Task Force recommendation statement. JAMA. 2016;315(4):380-7. doi: 10.1001/jama.2015.18392. PMID: 26813211.
20. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006 May;166(10):1092-7. doi: 10.1001/archinte.166.10.1092. PMID: 16717171.
21. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD; January 2014. https://effectivehealthcare.ahrq.gov/topics/cer-methods-guide/overview. Accessed May 23, 2019.
22. U.S. Preventive Services Task Force. Methods and Processes. 2016. https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes. Accessed March 6 2020.
23. Nelson HD. Systematic Reviews to Answer Health Care Questions: Lippincott Williams & Wilkins; 2014.
24. Institute of Medicine Committee on Standards for Systematic Reviews of Comparative Effectiveness R. In: Eden J, Levit L, Berg A, Morton S, eds. Finding What Works in Health Care: Standards for Systematic Reviews. Washington (DC): National Academies Press (US) Copyright 2011 by the National Academy of Sciences. All rights reserved.; 2011.
25. Šimundić A-M. Measures of diagnostic cccuracy: basic definitions. EJIFCC. 2009;19(4):203-11. PMID: 27683318.
26. Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EC017. Rockville, MD: Agency for Healthcare Research and Quality: June 2012.
27. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009. PMID: 22007046.
28. 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 Mar;34(1):25-43. doi: 10.1016/S0005-7916(03)00004-1. PMID: 12763391.
29. Donker T, van Straten A, Marks I, et al. A brief web-based screening questionnaire for common mental disorders: development and validation. J Med Internet Res. 2009 Jul 24;11(3):e19. doi: 10.2196/jmir.1134. PMID: 19632977.

30. Donker T, Comijs H, Cuijpers P, et al. The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. Psychiatry Res. 2010 Mar 30;176(1):45-50. doi: 10.1016/j.psychres.2009.01.012. PMID: 20071036.
31. Donker T, van Straten A, Marks I, et al. Quick and easy self-rating of generalized anxiety disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. Psychiatry Res. 2011 Jun;188(1):58-64. doi: 10.1016/j.psychres.2011.01.016. PMID: 21339006
32. Farvolden P, McBride C, Bagby R, et al. A web-based screening instrument for depression and anxiety disorders in primary care. J Med Internet Res. 2003 Nov;5(3):1-7. doi: 10.2196/jmir.5.3.e23. PMID: 14517114.
33. Garcia-Campayo J, Zamorano E, Ruiz MA, et al. The assessment of generalized anxiety disorder: psychometric validation of the Spanish version of the self-administered GAD-2 scale in daily medical practice. Health Qual Life Outcomes. 2012 Sep 19;10(114)doi: 10.1186/1477-7525-10-114. PMID: 22992432.
34. Houston JP, Kroenke K, Faries DE, et al. A provisional screening instrument for four common mental disorders in adult primary care patients. Psychosomatics. 2011 Jan;52(1):48-55. doi: 10.1016/j.psym.2010.11.011. PMID: 21300195.
35. Kiely KM, Butterworth P. Validation of four measures of mental health against depression and generalized anxiety in a community based sample. Psychiatry Res. 2015 Feb 28;225(3):291-8. doi: 10.1016/j.psychres.2014.12.023. PMID: 25578983.
36. Kroenke K, Spitzer RL, Williams JB, et al. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics. 2009 Nov;50(6):613-21. doi: 10.1176/appi.psy.50.6.613. PMID: 19996233.
37. Leyfer OT, Ruberg JL, Woodruff-Borden J. Examination of the utility of the Beck Anxiety Inventory and its factors as a screener for anxiety disorders. J Anxiety Disord. 2006;20(4):444-58. doi: 10.1016/j.janxdis.2005.05.004 PMID: 16005177.
38. Moore MT, Anderson NL, Barnes JM, et al. Using the GAD-Q-IV to identify generalized anxiety disorder in psychiatric treatment seeking and primary care medical samples. J Anxiety Disord. 2014 Jan;28(1):25-30. doi: 10.1016/j.janxdis.2013.10.009. PMID: 24334213.
39. Munoz-Navarro R, Cano-Vindel A, Moriana JA, et al. Screening for generalized anxiety disorder in Spanish primary care centers with the GAD-7. Psychiatry Res. 2017 Oct;256:312-7. doi: 10.1016/j.psychres.2017.06.023. PMID: 28666201.
40. Newman MG, Zuellig AR, Kachin KE, et al. Preliminary reliability and validity of the generalized anxiety disorder questionnaire-IV: a revised self-report diagnostic measure of generalized anxiety disorder. Behav Ther. 2002 Mar 01;33(2):215-33. doi: 10.1016/S0005-7894(02)80026-0.
41. Parkerson GR, Jr., Broadhead WE. Screening for anxiety and depression in primary care with the Duke Anxiety-Depression Scale. Fam Med. 1997 Mar 01;29(3):177-81. PMID: 9085098.
42. Puddifoot S, Arroll B, Goodyear-Smith FA, et al. A new case-finding tool for anxiety: a pragmatic diagnostic validity study in primary care. Int J Psychiatry Med. 2007 Dec 01;37(4):371-81. doi: 10.2190/PM.37.4.b. PMID: 18441626.
43. Birmaher B, Khetarpal S, Brent D, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry. 1997 Apr;36(4):545-53. doi: 10.1097/00004583-199704000-00018. PMID: 9100430

44. Birmaher B, Brent DA, Chiappetta L, et al. Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. J Am Acad Child Adolesc Psychiatry. 1999 Oct;38(10):1230-6. doi: 10.1097/00004583-199910000-00011. PMID: 10517055
45. Bodden DH, Bogels SM, Muris P. The diagnostic utility of the Screen for Child Anxiety Related Emotional Disorders-71 (SCARED-71). Behav Res Ther. 2009 May;47(5):418-25. doi: 10.1016/j.brat.2009.01.015. PMID: 19230863.
46. Dierker LC, Albano AM, Clarke GN, et al. Screening for anxiety and depression in early adolescence. J Am Acad Child Adolesc Psychiatry. 2001 Aug;40(8):929-36. doi: 10.1097/00004583-200108000-00015. PMID: 11501693.
47. Piqueras JA, Martin-Vivar M, Sandin B, et al. The revised child anxiety and depression scale: a systematic review and reliability generalization meta-analysis. J Affect Disord. 2017 Aug 15;218:153-69. doi: 10.1016/j.jad.2017.04.022. PMID: 28475961.
48. Matthey S, Valenti B, Souter K, et al. Comparison of four self-report measures and a generic mood question to screen for anxiety during pregnancy in English-speaking women.[Erratum appears in J Affect Disord. 2014 Feb;155:307]. J Affect Disord. 2013 Jun;148(2-3):347-51. doi: 10.1016/j.jad.2012.12.022. PMID: 23380518.
49. McDonald S, Wall J, Forbes K, et al. Development of a prenatal psychosocial screening tool for post-partum depression and anxiety. Paediatr Perinat Epidemiol. 2012 Jul;26(4):316-27. doi: 10.1111/j.1365-3016.2012.01286.x. PMID: 22686383.
50. Meades R, Ayers S. Anxiety measures validated in perinatal populations: a systematic review. J Affect Disord. 2011 Sep;133(1-2):1-15. doi: 10.1016/j.jad.2010.10.009. PMID: 21078523.
51. O'Hara MW, Stuart S, Watson D, et al. Brief scales to detect postpartum depression and anxiety symptoms. J Womens Health. 2012 Dec;21(12):1237-43. doi: 10.1089/jwh.2012.3612. PMID: 23130750.
52. Simpson W, Glazer M, Michalski N, et al. Comparative efficacy of the Generalized Anxiety Disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. Can J Psychiatry. 2014 Aug;59(8):434-40. doi: 10.1177/070674371405900806. PMID: 25161068.
53. Somerville S, Dedman K, Hagan R, et al. The perinatal anxiety screening scale: development and preliminary validation. Arch Womens Ment Health. 2014 Oct;17(5):443-54. doi: 10.1007/s00737-014-0425-8. PMID: 24699796.
54. Tendais I, Costa R, Conde A, et al. Screening for depression and anxiety disorders from pregnancy to postpartum with the EPDS and STAI. Span J Psychol. 2014 Mar 31;17(E7)doi: 10.1017/sjp.2014.7. PMID: 25012783.
55. Krasucki C, Ryan P, Ertan T, et al. The FEAR: a rapid screening instrument for generalized anxiety in elderly primary care attenders. Int J Geriatr Psychiatry. 1999 Jan;14(1):60-8. doi: 10.1002/(SICI)1099-1166(199901)14:1<60::AID-GPS893>3.0.CO;2-G. PMID: 10029937.
56. Vasiliadis HM, Chudzinski V, Gontijo-Guerra S, et al. Screening instruments for a population of older adults: the 10-item Kessler Psychological Distress Scale (K10) and the 7-item Generalized Anxiety Disorder Scale (GAD-7). Psychiatry Res. 2015 Jul 30;228(1):89-94. doi: 10.1016/j.psychres.2015.04.019. PMID: 25956759.

57. Wetherell JL, Birchler GD, Ramsdell J, et al. Screening for generalized anxiety disorder in geriatric primary care patients. Int J Geriatr Psychiatry. 2007 Feb;22(2):115-23. doi: 10.1002/gps.1701. PMID: 17096461.
58. Fairbrother N, Corbyn B, Thordarson DS, et al. Screening for perinatal anxiety disorders: Room to grow. J Affect Disord; 2019. p. 363-70.
59. Dennis R, Boddington S, Funnell N. Self-report measures of anxiety: are they suitable for older adults? Aging Ment Health. 2007 Nov;11(6):668-77. doi: 10.1080/13607860701529916. PMID: 18074254.
60. Dozeman E, van Schaik DJ, van Marwijk HW, et al. The Center for Epidemiological Studies Depression Scale (CES-D) is an adequate screening instrument for depressive and anxiety disorders in a very old population living in residential homes. Int J of Geriatr Psychiatry. 2011 Mar;26(3):239-46. doi: 10.1002/gps.2519. PMID: 20623777.
61. Austin MP, Hadzi-Pavlovic D, Priest SR, et al. Depressive and anxiety disorders in the postpartum period: how prevalent are they and can we improve their detection? Arch Women Ment Health. 2010 Oct;13(5):395-401. doi: 10.1007/s00737-010-0153-7. PMID: 20232218.
62. Rowe HJ, Fisher JR, Loh WM. The Edinburgh Postnatal Depression Scale detects but does not distinguish anxiety disorders from depression in mothers of infants. Arch Womens Ment Health. 2008 Jun;11(2):103-8. doi: 10.1007/s00737-008-0003-z. PMID: 18463939.
63. Petrozzi A, Gagliardi L. Anxious and depressive components of Edinburgh Postnatal Depression Scale in maternal postpartum psychological problems. J Perinatal Med. 2013 Jul;41(4):343-8. doi: 10.1515/jpm-2012-0258. PMID: 23426862.
64. Batterham PJ, Calear AL, Sunderland M, et al. Hierarchical screening for multiple mental disorders. J Affect Disord. 2013 Oct;151(1):229-36. doi: 10.1016/j.jad.2013.05.085. PMID: 23806587.
65. Roemer L, Borkovec M, Posa S, et al. A self-report diagnostic measure of generalized anxiety disorder. J Behav Ther Exp Psychiatry. 1995 Dec;26(4):345-50. doi: 10.1016/0005-7916(95)00040-2. PMID: 8675722.
66. Weiss BJ, Calleo J, Rhoades HM, et al. The utility of the Generalized Anxiety Disorder Severity Scale (GADSS) with older adults in primary care. Depress Anxiety. 2009 Jan 15;26(1):E10-5. doi: 10.1002/da.20520. PMID: 18839400.
67. Jomeen J, Martin C. Is the hospital anxiety and depression scale (HAPS) a reliable screening tool in early pregnancy? Psychol Health. 2004 Dec;19(6):787-800. doi: 10.1080/0887044042000272895.
68. Matthey S, Souter K, Valenti B, et al. Validation of the MGMQ in screening for emotional difficulties in women during pregnancy. J Affect Disord; 2019. p. 156-63.
69. Grant KA, McMahon C, Austin MP. Maternal anxiety during the transition to parenthood: a prospective study. J Affect Disord. 2008 May;108(1-2):101-11. doi: 10.1016/j.jad.2007.10.002. PMID: 18001841.
70. Leung B, Letourneau N, Bright K, et al. Appraisal of the psychiatric diagnostic screening questionnaire in a perinatal cohort: The APrON study. Scand J Public Health. 2017 Aug;45(6):658-65. doi: 10.1177/1403494817717835. PMID: 28707502.
71. Cano-Vindel A, Munoz-Navarro R, Medrano LA, et al. A computerized version of the Patient Health Questionnaire-4 as an ultra-brief screening tool to detect emotional

disorders in primary care. J Affect Disord. 2018;234:247-55. doi: 10.1016/j.jad.2018.01.030. PMID: 29549826.
72. Schroder HS, Clark D, Moser JS. Screening for problematic worry in adults with a single item from the Penn State Worry Questionnaire. Assessment; 2019. p. 336-46.
73. Crocetti E, Hale WW, III, Fermani A, et al. Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED) in the general Italian adolescent population: a validation and a comparison between Italy and the Netherlands. J Anxiety Disord. 2009 Aug;23(6):824-9. doi: 10.1016/j.janxdis.2009.04.003. PMID: 19427168.
74. Hale WW, Raaijmakers QA, Garcia-Lopez LJ, et al. Psychometric properties of the screen for child anxiety related emotional disorders for socially anxious and healthy Spanish adolescents. Span J Psychol. 2013 Jun 12;16:E25. doi: 10.1017/sjp.2013.36. PMID: 23866219.
75. Reynolds CR, Richmond BO. What I Think and Feel: A Revised Measure of Children's Manifest Anxiety. J Abnorm Child Psychol. 1997 February 01;25(1):15-20. doi: 10.1023/a:1025751206600. PMID: 9093896.
76. March JS, Parker JDA, Sullivan K, et al. The Multidimensional Anxiety Scale for Children (MASC): factor structure, reliability, and validity. J Am Acad Child Adolesc Psychiatry. 1997;36(4):554-65. doi: 10.1097/00004583-199704000-00019. PMID: 9100431.
77. Chorpita BF, Yim L, Moffitt C, et al. Assessment of symptoms of DSM-IV anxiety and depression in children: a revised child anxiety and depression scale. Behav Res Ther. 2000;38(8):835-55. doi: 10.1016/S0005-7967(99)00130-8. PMID: 10937431.
78. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry. 1987 Jun;150:782-6. doi: 10.1192/bjp.150.6.782. PMID: 3651732.
79. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983 Jun;67(6):361-70. doi: 10.1111/j.1600-0447.1983.tb09716.x PMID: 6880820.
80. Rini CK, Dunkel-Schetter C, Wadhwa PD, et al. Psychological adaptation and birth outcomes: the role of personal resources, stress, and sociocultural context in pregnancy. Health Psychol. 1999;18(4):333-45. doi: 10.1037//0278-6133.18.4.333. PMID: 10431934.
81. Huizink AC, Robles De Medina PG, Mulder EJH, et al. Psychological measures of prenatal stress as predictors of infant temperament. J Am Acad Child Adolesc Psychiatry. 2002;41(9):1078-85. doi: 10.1097/00004583-200209000-00008. PMID: 12218429.
82. Spielberger CD, Gorsuch RL, Lushene RE. Manual for the state-trait anxiety inventory; 1970.
83. Goldberg DP. The detection of psychiatric illness by questionnaire. Maudsley monograph. Oxford: Oxford University Press; 1972.
84. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychol Med. 1979;9(1):139-45. doi: 10.1017/S0033291700021644. PMID: 424481.
85. Ayers S. Assessing psychopathology in pregnancy and postpartum. J Psychosom Obstet Gynaecol. 2001 Jun;22(2):91-102. doi: 10.3109/01674820109049959. PMID: 11446159.
86. Beck AT, Steer RA. Beck Anxiety Inventory manual. San Antonio, TX: Psychological Corporation, Harcourt Brace; 1993.

87. Andrews G, Basu A, Cuijpers P, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord. 2018 Apr;55:70-8. doi: 10.1016/j.janxdis.2018.01.001. PMID: 29422409.
88. Hunot V, Churchill R, Teixeira V, et al. Psychological therapies for generalised anxiety disorder. Cochrane Database Syst Rev. 2010 Jan(6)doi: 10.1002/14651858.CD001848.pub4. PMID: 17253466.
89. James AC, James G, Cowdrey FA, et al. Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database Syst Rev. 2015 Feb 18(2):Cd004690. doi: 10.1002/14651858.CD004690.pub4. PMID: 25692403.
90. Mayo-Wilson E, Montgomery P. Media-delivered cognitive behavioural therapy and behavioural therapy (self-help) for anxiety disorders in adults. Cochrane Database Syst Rev. 2013 Sep 9(9)doi: 10.1002/14651858.CD005330.pub4. PMID: 24018460.
91. Chessick CA, Allen MH, Thase EM, et al. Azapirones for generalized anxiety disorder. Cochrane Database Syst Rev. 2006 Jul 19(3)doi: 10.1002/14651858.CD006115. PMID: 16856115.
92. Depping AM, Komossa K, Kissling W, et al. Second-generation antipsychotics for anxiety disorders. Cochrane Database Syst Rev. 2010 Dec 08(12):CD008120. doi: 10.1002/14651858.CD008120.pub2. PMID: 21154392.
93. Guaiana G, Barbui C, Cipriani A. Hydroxyzine for generalised anxiety disorder. Cochrane Database Syst Rev. 2010 Dec 8(12)doi: 10.1002/14651858.CD006815.pub2. PMID: 21154375.
94. Ipser JC, Stein DJ, Hawkridge S, et al. Pharmacotherapy for anxiety disorders in children and adolescents. Cochrane Database Syst Rev. 2009(3)doi: 10.1002/14651858.CD005170.pub2. PMID: 19588367.
95. Samuel M, Zimovetz EA, Gabriel Z, et al. Efficacy and safety of treatments for refractory generalized anxiety disorder: a systematic review. Int Clin Psychopharmacol. 2011 Mar;26(2):63-8. doi: 10.1097/YIC.0b013e328341bb4a. PMID: 21088608.
96. Martin JL, Sainz-Pardo M, Furukawa TA, et al. Benzodiazepines in generalized anxiety disorder: heterogeneity of outcomes based on a systematic review and meta-analysis of clinical trials. J Psychopharmacol. 2007 Sep;21(7):774-82. doi: 10.1177/0269881107077355. PMID: 17881433.
97. Miyasaka SL, Atallah AN, Soares B. Valerian for anxiety disorders. Cochrane Database Syst Rev. 2006 Oct 18(4)doi: 10.1002/14651858.CD004515.pub2. PMID: 17054208
98. Wang Z, Whiteside S, Sim L, et al. Anxiety in Children Agency for Healthcare Research and Quality,. 17-EHC023-EF. Rockville, MD: 2017.
99. Slee A, Nazareth I, Bondaronek P, et al. Pharmacological treatments for generalised anxiety disorder: a systematic review and network meta-analysis. Lancet. 2019 Feb 23;393(10173):768-77. doi: 10.1016/S0140-6736(18)31793-8. PMID: 30712879
100. Anxiety and Depression Association of America. Clinical Practice Review for GAD. 2015. https://adaa.org/resources-professionals/practice-guidelines-gad. Accessed July 27, 2018.
101. Gartlehner G, Hansen RA, Morgan LC, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. Ann Intern Med. 2011;155(11):772-85.

APPENDIX 1

Search Strategies

Database: Ovid MEDLINE(R)

---

1 exp ANXIETY/di, dg, ep [Diagnosis, Diagnostic Imaging, Epidemiology] (12932)
2 exp Anxiety Disorders/di, dg, ep [Diagnosis, Diagnostic Imaging, Epidemiology] (26084)
3 1 or 2 (37783)
4 exp Mass Screening/ (114966)
5 3 and 4 (787)
6 (generaliz* adj3 anxi*).mp. (5482)
7 4 and 6 (83)
8 (generaliz* adj3 anxi* adj7 screen*).mp. (61)
9 5 or 7 or 8 (851)
10 limit 9 to female (704)
11 exp Women's Health/ (26041)
12 9 and 11 (6)
13 10 or 12 (704)
14 (screen* adj7 anxi*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (1943)
15 (screen* adj7 (women or woman or female*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (29708)
16 3 and 15 (161)
17 11 and 14 (18)
18 limit 14 to female (1540)
19 17 or 18 (1541)
20 13 or 19 (2001)
21 limit 20 to english language (1895)
22 limit 21 to (comparative study or controlled clinical trial or guideline or meta analysis or randomized controlled trial or systematic reviews) (358)
23 exp Epidemiologic Studies/ (2123303)
24 21 and 23 (742)
25 24 not 22 (615)
26 21 not (22 or 24) (922)

Database: EBM Reviews - Cochrane Database of Systematic Reviews
---

1 (generaliz* adj3 anxi*).mp. [mp=title, short title, abstract, full text, keywords, caption text] (30)

Database: EBM Reviews - Cochrane Central Register of Controlled Trials
---

1 (generaliz* adj3 anxi* adj10 (screen* or diagnos* or detect* or identif*) adj15 (woman or women or female*)).mp. [mp=title, original title, abstract, mesh headings, heading words, keyword] (63)
2 (generaliz* adj3 anxi* adj10 (tool* or survey* or instrument* or questionnair*) adj15 (woman or women or female)).mp. [mp=title, original title, abstract, mesh headings, heading words, keyword] (29)
3 1 or 2 (87)
4 (generaliz* adj3 anxi* adj10 (tool* or survey* or instrument* or questionnair*)).mp. (105)
5 (generaliz* adj3 anxi* adj10 (screen* or diagnos* or detect* or identif*)).mp. (264)
6 4 or 5 (342)
7 6 not 3 (255)

Database: Health and Psychosocial Instruments
---

1 (generaliz* adj3 anxi* adj10 (tool* or survey* or instrument* or questionnair*)).mp. [mp=title, acronym, descriptors, measure descriptors, sample descriptors, abstract, source] (17)
2 (generaliz* adj3 anxi* adj10 (screen* or diagnos*)).mp. [mp=title, acronym, descriptors, measure descriptors, sample descriptors, abstract, source] (30)
3 1 or 2 (38)

APPENDIX 2

Inclusion/Exclusion Criteria

Category	Included	Excluded
Populations	Adolescent and adult women age 13 and older without current diagnosis of anxiety disorders	<13 years old; studies enrolling <50% women
Conditions	Generalized anxiety disorder or anxiety not yet defined	Specific to PTSD, OCD, panic disorder, anxiety associated with a disease or illness, other mental health condition
Interventions	Screening instrument or method used in primary care applicable settings to identify patients with anxiety; treatments for anxiety disorders	• Other types of interventions • Unclear intervention description
Reference standard	• DSM criteria • Other diagnostic criteria, clinical diagnosis, or adaptations of established criteria	• Screening instrument or method of interest used as reference standard • No reference standard • Inadequate description of reference standard
Outcomes	KQ 1: Improvement in symptoms, quality of life, and function KQ 2: Diagnostic accuracy (sensitivity, specificity, PPV, NPV, AUC) KQ 3: False positive results, patient distress, any potential harms reported by the study	Prevalence, risk factors, cost, cost-effectiveness
Study Design	• Diagnostic accuracy studies • Randomized controlled trials • Prospective cohort studies • Case-control studies • Systematic reviews	• Case reports • Cost effectiveness studies • Modeling studies

AUC=area under the receiver operating characteristic curve; DSM=Diagnostic and Statistical Manual of Mental Disorders; KQ=key question; NPV=negative predictive value; OCD=obsessive compulsive disorder; PPV=positive predictive value; PTSD=post-traumatic stress disorder.

APPENDIX 3

Strength of Evidence

The strength of evidence for each key question is assessed by using the approach described in the AHRQ Methods Guide.[21] Grades are based on:

- Study limitations (low, medium, or high level of study limitations)
- Consistency (consistent, inconsistent, or unknown/not applicable)
- Directness (direct or indirect)
- Precision (precise or imprecise)
- Reporting bias (suspected or undetected)

An overall grade of high, moderate, low, or insufficient is assigned according to a four-level scale by evaluating and weighing the combined results of the above domains:

- High: Very confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has few or no deficiencies. The findings are stable, i.e., another study would not change the conclusions.
- Moderate: Moderately confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has some deficiencies. The findings are likely to be stable, but some doubt remains.
- Low: Limited confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has major or numerous deficiencies (or both). Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.
- Insufficient: No evidence, are unable to estimate an effect, or have no confidence in the estimate of effect for this outcome. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion.

Applicability

Applicability (external validity) is estimated by examining the characteristics of the patient populations; the sample size of the studies; clinical settings (e.g., primary care, community setting); and clinical relevance of the screening approach. Variability in the studies may limit the ability to generalize the results to other populations and settings. Applicability is rated high, moderate, or low.

APPENDIX 4

Evidence Table of Studies of the Accuracy of Screening Instruments

Study, Year	Design	Enrolled, *n*	Age, *y*	Character-istics	Setting	Screening instrument and threshold	Reference standard	Accuracy measures (95% CI)
Adolescents								
Birmaher, 1997[43]	Cohort	341	9-18	59% female	Mood/ anxiety disorders clinic	SCARED (score ≥15)	Clinical interview using DSM-IV diagnosis criteria or K-SADS-P diagnosis	Sensitivity 72%; specificity 64%
Birmaher, 1999[44]	Cohort	190	9-19	52% female	Mood/ anxiety disorders clinic	SCARED-41 (score ≥25); 5-item SCARED (score ≥3)	Comprehensive symptom checklist for DSM-IV diagnostic criteria	SCARED-41: sensitivity 71%; specificity 67%; 5-item SCARED: sensitivity 74%; specificity 73%
Bodden, 2009[45]	Case-control	176	8-18	Clinically anxious; 60% female	General population not further described	SCARED-71 (GAD sub score ≥8)	ADIS-C and ADIS-P	Sensitivity 62%; specificity 69%
Dierker, 2001[46]	Cohort	632	9th graders	55% female	5 high schools in the U.S.	RCMAS; MASC; (thresholds not reported)	Diagnostic interview modules selected from the DSM-IV diagnosis criteria	RCMAS: AUROC for girls 0.62; MASC: AUROC for girls 0.82
Piqueras, 2017[47]	Systematic review of 146 studies	88,648	6-18	Multiple studies	Mixed	RCADS (thresholds vary by study)	Vary by study	Reliability (43 studies) 0.91 (0.90-0.92)*
Adults								
Behar, 2003[28]	Case-control	2449	Young adults	71% female	General population not further described	PSWQ (score ≥62)	GAD-Q-IV	Sensitivity: 75%; specificity: 86%
Cano-Vindel, 2018[71]	Cohort	1052	>18	77% female	Primary care clinics	GAD-2 (score ≥3)	SCID-I	Sensitivity 77%; specificity 80%; AUROC 0.81
Donker, 2009[29]	Cohort	502	18-80	57% female	Internet	WSQ (score ≥10)	CIDI diagnosis with live phone interviews	Sensitivity 89%; specificity 82%

Study, Year	Design	Enrolled, *n*	Age, *y*	Character-istics	Setting	Screening instrument and threshold	Reference standard	Accuracy measures (95% CI)
Donker, 2010[30]	Cohort	1607	18-65	67% female	Primary care clinics	Kessler-10 (K10); EK-10 (extended version); (score ≥20)	CIDI interview, DSM-IV diagnosis	K10-20: sensitivity 94%; specificity 67%; EK10-20: sensitivity 95%; specificity 61%
Donker, 2011[31]	Cohort	522	18-80	57% female	Web-based	GAD-2 (score ≥3); GAD-7 (score ≥10)	DSM-IV CIDI GAD	GAD-2: sensitivity 70%; specificity 76%; AUROC 0.78 (0.69-0.86); GAD-7: sensitivity 87-89%; specificity 50-82%; AUROC 0.77 (0.68-0.85)
Farvolden, 2003[32]	Cohort	32	>18	59% female	Web-based	WB-DAT (threshold not reported)	SCID-I/P interview	Sensitivity 63%; specificity 94%
García-Campayo, 2012[33]	Cross-sectional	220	>18	72% female	Primary care clinics	GAD-2 (score ≥3)	HAM-A, HADS, and WHODAS II	Sensitivity 91.5%; specificity: 85.8%; AUROC 0.937
Houston, 2011[34]	Cross-sectional	24	>18	>60% female	Primary care clinics	PDI-4 (threshold not reported)	SCID/ACDS assessment, DSM-IV	Sensitivity 83%; specificity 75%; follow up with GAD-7: sensitivity 89%; specificity 82%
Kiely, 2015[35]	Cohort	1015	32-36; 52-58	59% female	Community	GAS (score ≥7)	CIDI	Sensitivity 84%; specificity 86%; AUROC 0.8957
Kroenke, 2009[36]	Cohort	2149	18-95	66% female	Primary care clinics	GAD-2 (threshold not reported)	Structured interview using DSM-IV criteria	AUROC 0.908 (0.876-0.940)
Leyfer, 2005[37]	Cohort	193	17-76	76% female	General population	BAI (score ≥3.5)	ADIS-IV	Sensitivity 75%; specificity 73%
Moore, 2014[38]	Cohort	104	18-45	69% female	Primary care clinics	GAD-Q-IV (DSM-based algorithm; score ≥7.6)	SCID-IV	AUROC 0.85 (0.76-0.93); DSM-based algorithm: sensitivity 97%; specificity 86%; score ≥7.6: sensitivity 89%; specificity 72%
Munoz-Navarro, 2017[39]	Cohort	178	18-65	71% female	Primary care clinics	GAD-7 (score ≥10)	CIDI for DSM-IV	Sensitivity 87%; specificity 78%
Newman, 2002[40]	Cohort	143	Young adult	80% female	Undergraduates	GAD-Q-IV (DSM-based algorithm; score ≥5.7)	DSM structured interview	Sensitivity 83%; specificity 89%
Parkerson, 1997[41]	Cross-sectional	481	18-64	72% female	Primary care clinics	DUKE-AD (score ≥30)	DSM	Sensitivity 71.4%; specificity 59.2%; AUROC 0.723

Study, Year	Design	Enrolled, *n*	Age, *y*	Character-istics	Setting	Screening instrument and threshold	Reference standard	Accuracy measures (95% CI)
Puddifoot, 2007[42]	Cohort	982	16-93	72% female	Primary care clinics	2 screening questions (threshold not reported)	HADS anxiety score >11	Two screening questions: sensitivity 58%; specificity 87%; worry question alone: sensitivity 76%; specificity 82%
Schroder, 2019[72]	Cohort	1191	>18	73% female	Undergrad uates and community	PSWQ (score >62); PSWQ-3 (score >11); PSWQ – Item 15	GAD-7	PSWQ: sensitivity 68%; specificity 89%; PSWQ-3: sensitivity 68%; specificity 92%; PSWQ-item 15: sensitivity 64%; specificity 92%
Spitzer, 2006[20]	Cohort	2740	18-95	65% female	Primary care clinics	GAD-7 (score ≥10)	Structured interviews for 965	Sensitivity 89%; specificity 82%
Older adults								
Krasucki, 1999[55]	Cohort	88	>65	64% female	Primary care clinics	ADS (score = 2-3); FEAR (threshold not reported)	Clinical Interview, ICD-10 diagnosis	ADS: sensitivity 85%; specificity 71%; FEAR: sensitivity 74%; specificity 85%
Vasiliadis, 2015[56]	Cross-sectional	1775	≥65	57% female	Primary care clinics	GAD-7 (score ≥5)	DSM-IV diagnosis, in person interview	Sensitivity 71%; specificity 57%; AUROC 0.695
Wetherell, 2007[57]	Cohort	68	>60	67% female	Primary care clinics	HADS-A (score ≥8); BSI-18 (score >8)	ADIS-IV interview, DSM diagnosis	HADS: sensitivity 97%; specificity 67%; AUROC 0.80; BSI-18: AUROC 0.573 (SE 0.092)

Study, Year	Design	Enrolled, *n*	Age, *y*	Character-istics	Setting	Screening instrument and threshold	Reference standard	Accuracy measures (95% CI)
Pregnant and postpartum women								
Fairbrother, 2019[58]	Cohort	115	Mean 33.2	Postpartum	Prenatal clinics, physician offices, midwifery clinics	EPDS (score >6); EDS-3a (score >4); GAD-7 (score >6); GAD-2 (score >3); AD-13 (score >11)	SCID-IV diagnosis	EPDS: sensitivity 73.7%; specificity 63.7%; PPV 0.224; NPV 0.944; AUROC 0.750 (0.663-0.824); EDS-3a: sensitivity 65.8%; specificity 76.0%; PPV 0.281; NPV 0.940; AUROC 0.757 (0.678-0.836); GAD-7: sensitivity 55.3%; specificity 83.2%; PPV 0.318; NPV 0.929; AUROC 0.719 (0.619-0.818): GAD-2: sensitivity 81.6%; specificity 50.9%; PPV 0.193; NPV 0.951; AUROC 0.718 (0.675-0.829); AD-13: sensitivity 86.5%; specificity 68.2%; PPV 0.274; NPV 0.973; AUROC 0.834 (0.776-0.893)
Matthey, 2013[48]	Cohort	391	Mean 28.8	Pregnant	First prenatal visit	EDS-3a, score of ≥5; HADS-A, score of ≥9; PRT, score of ≥22; PRAQ-R, score of ≥26; MGMQ, threshold score not applicable	MINI diagnosis	EDS-3a: sensitivity 54%; HADS-A: sensitivity 35%; PRT: sensitivity 20%; PRAQ-R: sensitivity 33%; MGMQ: sensitivity 80%
Matthey, 2019[68]	Cohort	252	Mean 28.4	Pregnant	First prenatal visit	MGMQ, threshold score not applicable	DSM-IV criteria	Bother impact of a little or more (30% of participants): sensitivity 72%-75%; specificity 80%; PPV: 39%-43%; bother impact of moderately or more (12.6% of participants): sensitivity 51%-56%; specificity 94%-96%; PPV: 67%-73%

Study, Year	Design	Enrolled, *n*	Age, *y*	Character-istics	Setting	Screening instrument and threshold	Reference standard	Accuracy measures (95% CI)
McDonald, 2012[49]	Cohort	567	>18	Pregnant <24 weeks	Community	EPDS, score of ≥10; McDonald Prenatal Screening Tool, score of ≥2	STAI-state anxiety scale	EPDS: sensitivity 41% (27-61); specificity 88% (82-91); PPV 0.34 (0.20-0.49); NPV 0.91 (0.87-0.95); AUROC 0.73 (0.62-0.83); McDonald: sensitivity 44% (29-60); specificity 88% (82-91); PPV 0.34 (0.20-0.49); NPV 0.91 (0.87-0.95); AUROC 0.71 (0.61-0.82)
Meades, 2011[50]	Systematic review	441 HADS-A; 100 STAI; 2525 GHQ	>18	Pregnant	Mixed	HADS-A (score ≥8); STAI (score >40); GHQ-12 (score 3 or 4-5); GHQ-28 (score of 3-4 or 7-8); GHQ-30 (score of 5-6, 6-7, or 7-8)	HADS-A: MINI plus, semi-structured interview, or SCID diagnosis; STAI: MINI plus, semi-structured interview, or SCID diagnosis; GCQ: clinical interview schedule, SCID, or ICD-o diagnosis; SADS, PAS, or ICD-9	HADS-A (one study): sensitivity 92.9%; specificity 90%; STAI (one study): sensitivity 80.95%; specificity 79.75%; PPV 0.52; NPV 0.94; GHQ-30 (3 studies): sensitivity 77-83%; specificity 71-89%; PPV 0.37-0.53; NPV 0.90-0.97; GHQ-28 (2 studies): sensitivity 75%, 82%; specificity 83%, 85%; PPV 0.46,0.53; NPV 0.95, 0.96; GHQ-12 (2 studies): sensitivity 83%, 81%; specificity 80%, 81%
O'Hara, 2012[51]	Cohort	353	Mean 27	Mean 21 weeks postpartum	Community and maternal and child health centers	BAI-Subj (score >4 or >6)	SCID diagnosis	Score >4: sensitivity 76%; specificity 71%; PPV 0.31; AUROC 0.78; score >6: sensitivity 56%; specificity 82%; PPV 0.35
Simpson, 2014[52]	Cohort	155 pregnant and 85 postpartum	Mean 30.5	Pregnant and postpartum	Psychiatric referral	EPDS (score =10-13); EDS-3a (score >4); GAD-7 (score >10 or >13)	DSM-IV diagnosis	EPDS: sensitivity 77.3-89.3%; specificity 26.7-40.3%; PPV 0.36-0.38; NPV 0.79-0.84; EDS-3a: sensitivity 68.0%; specificity 63.5%; PPV 0.46; NPV 0.81; GAD-7 (score >10): sensitivity 76.0%; specificity 51.5%; PPV 0.42; NPV 0.83; GAD-7 (score >13): sensitivity 61.3%; specificity 72.7%; PPV 0.51; NPV 0.81

Study, Year	Design	Enrolled, *n*	Age, *y*	Character-istics	Setting	Screening instrument and threshold	Reference standard	Accuracy measures (95% CI)
Somerville, 2014[53]	Cohort	53	≥18	Pregnant and postpartum	Prenatal clinic	PASS (score >26)	ICD-10 diagnosis	Sensitivity 70%; specificity 30%; AUROC 0.7 (SE 0.04)
Tendais, 2014[54]	Cohort	35	Mean 28	Pregnant	Obstetrics outpatient unit	EPDS (score >9 pregnancy or >7 postpartum); STAI (score >40 pregnancy or >34 postpartum)	SCID diagnosis	EPDS pregnancy: sensitivity 73.7% (56.9%-86.6%); specificity 70.0% (60.5%-78.4%); PPV 0.46 (0.33-0.59); NPV 0.89 (0.80-0.94); EPDS postpartum: sensitivity 78.3% (56.3%-92.5%); specificity 81.6% (71.0%-89.5%); PPV 0.56 (0.38-0.74); NPV 0.93 (0.83-0.98); STAI pregnancy: sensitivity 65.7% (47.8%-80.9%); specificity 67.3% (57.8%-75.8%); PPV 0.38 (0.26-0.52); NPV 0.86 (0.77-0.93); STAI postpartum: sensitivity 71.4% (66.1%-99.8%); specificity 67.1% (56.0%-76.9%); PPV 0.26 (0.13-0.43); NPV 0.93 (0.84-0.98)

AD-13=Anxiety Disorders-13; ADIS-IV=Anxiety Disorders Interview Schedule; ADIS-C=Anxiety Disorder Interview Schedule-Child scale; ADIS-P=Anxiety Disorder Interview Schedule-Parent scale; ADS=Anxiety Disorder Scale; AUROC=area under the receiver-operating characteristic curve; BAI=Beck Anxiety Inventory; BSI=Brief Symptom Inventory; CI=confidence interval; CIDI=Composite International Diagnostic Interview; DSM=Diagnostic and Statistical Manual of Mental Disorders; DUKE-AD=Duke Anxiety-Depression Scale; EDS-3a=Edinburgh Depression Scale-anxiety subscale; EPDS=Edinburgh Postnatal Depression Scale; FEAR=Frequency of anxiety; Enduring nature of anxiety; Alcohol or sedative use; Restlessness or fidgeting; GAD=generalized anxiety disorder; GAD-7=Generalized Anxiety Disorder 7-item scale; GAS=Goldberg Anxiety Scale; GHQ=General Health Questionnaire; HADS=Hospital Anxiety and Depression Scale; HADS-A=Hospital Anxiety and Depression Scale-anxiety subscale; HAM-A=Hamilton Anxiety Scale; ICD=International Statistical Classification of Diseases; K-SADS-P=Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present Episode; MASC=Multidimensional Anxiety Scale for Children; MINI=Mini-International Neuropsychiatric Interview; MGMQ=Matthey Generic Mood Question; NPV=negative predictive value; OCD=obsessive compulsive disorder; OCI-R=Obsessive Compulsive Inventory-Revised; PASS=Perinatal Anxiety Screening Scale; PCL=PTSD Checklist; PDI-4=Provisional Diagnostic Instrument-4; PHQ=Patient Health Questionnaire; PSWQ=Penn State Worry Questionnaire; PPV=positive predictive value; PRAQ-R=Pregnancy Related Anxiety Questionnaire-Revised; PRT=Pregnancy-Related Thoughts; PTSD=posttraumatic stress disorder; RCADS=Revised Child Anxiety and Depression Scale; RCMAS=Revised Children's Manifest Anxiety Scale; SCARED=Screen for Child Anxiety Related Emotional Disorders; SCID=Structured Clinical Interview for DSM-IV; SCID-IV=Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, fourth edition axis I disorders; SCID/ACDS =Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, fourth edition axis I disorders and the Adult ADHD Clinician Diagnostic Scale version 1.2; SE=standard error; SPIN=Social Phobia Inventory; STAI=State Trait Anxiety Inventory; WB-DAT=Web-Based Depression and Anxiety Test; WHODAS II=World Health's Organization Disability Assessment Scale; WSQ=Web Screening Questionnaire; y=years.
*Determined by Cronbach's alpha measure to estimate internal consistency.

APPENDIX 5

Quality and Applicability Ratings of Diagnostic Accuracy Studies

	Patient selection*						Reference standard*				Ratings	
Author, year	1. Spectrum	1. Sample size >100	2. Sample selection	3. Eligibility criteria	4. Minimal attrition	5. Test & threshold described	6. Credible & replicable	7. Blinding	8. Applied to all	Sens; spec; AUROC	Quality†	Applicability‡
Adolescents												
Birmaher, 1997[43]	No	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	No	Yes	Fair	Low
Birmaher, 1999[44]	No	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Fair	Low
Bodden, 2009[45]	No	Yes	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Fair	Low
Dierker, 2001[46]	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Unclear	No	Yes	Poor	High
Adults												
Behar, 2003[28]	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Fair	High
Cano-Vindel, 2018[71]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Fair	High
Donker, 2009[29]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Fair	High
Donker, 2010[30]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Fair	High
Donker, 2011[31]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Fair	High
Farvolden, 2003[32]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Poor	Low
García-Campayo, 2012[33]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Fair	High
Houston, 2011[34]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Poor	Low
Kiely, 2015[35]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good	High
Kroenke, 2009[36]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Good	High
Leyfer, 2005[37]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Fair	High

	Patient selection*						Reference standard*				Ratings	
Author, year	1. Spectrum	1. Sample size >100	2. Sample selection	3. Eligibility criteria	4. Minimal attrition	5. Test & threshold described	6. Credible & replicable	7. Blinding	8. Applied to all	Sens; spec; AUROC	Quality†	Applicability‡
Moore, 2014[38]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	No	Yes	Fair	High
Munoz-Navarro, 2017[39]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Fair	High
Newman, 2002[40]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Fair	High
Parkerson, 1997[41]	Yes	Yes	No	Unclear	Yes	Yes	Yes	Unclear	Unclear	Yes	Poor	High
Puddifoot, 2007[42]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good	High
Schroder, 2019[72]	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Fair	High
Spitzer, 2006[20]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good	High
Older adults												
Krasucki, 1999[55]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Poor	Low
Vasiliadis, 2015[56]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good	High
Wetherell, 2007[57]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Poor	Low
Pregnant and postpartum women												
Fairbrother, 2019[58]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Fair	High
Matthey, 2013[48]	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Fair	High
Matthey, 2019[68]	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Fair	High
McDonald, 2012[49]	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Poor	High
O'Hara, 2012[51]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Fair	High
Simpson, 2014[52]	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Poor	High
Somerville, 2014[53]	Yes	No	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Poor	Low
Tendais, 2014[54]	Yes	No	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Poor	Low

Abbreviations: AUROC = area under the receiver-operating characteristic curve.
*Quality criteria definitions:[26] 1: Test applied to an appropriate number and spectrum of patients (>100 participants). 2: Population tested was consecutive or random. 3: Clear eligibility criteria described. 4: Attrition reported and minimal loss to follow-up. 5: Test and threshold adequately described and reproducible. 6: Reference standard was credible and replicable. 7: Blinding of outcome assessors to the reference standard. 8: Reference standard was applied to all patients or a random subset.[26]
†Definition of ratings based on quality criteria: Good: Evaluates relevant available screening test; uses a credible reference standard; interprets reference standard independently of screening test; reliability of test assessed; has few or handles indeterminate results in a reasonable manner; includes large number (>500) broad-spectrum patients with and without the condition; study attempts to enroll a random or consecutive sample of patients who meet inclusion criteria; screening cutoffs pre-specified. Fair: Evaluates relevant available screening test; uses reasonable although not best standard; interprets reference standard independent of screening test; moderate sample size (100 to 500 subjects) and a “medium” spectrum of patients (i.e. applicable to many settings where the diagnostic test would be applied). Poor: Has important limitations such as: uses inappropriate reference standard; screening test improperly administered; biased ascertainment of reference standard; small sample size (<100) of very narrow selected spectrum of patients; or these components of study not well described.[26]
‡High: Participants were selected from the community, primary care, or non-specialty care clinics; test and reference standards are relevant to clinical practice in the U.S.; test is feasible for screening by non-specialists in clinical settings. Low: Participants were selected exclusively from referral clinics; test and reference standards are not relevant to clinical practice in the U.S.; test may not be feasible for screening by non-specialists in clinical settings; or these components of the study were not described.[21]

APPENDIX 6

Quality Ratings of Systematic Reviews

Author, year	1. Includes PICO	2. A priori methods	3. Deviation from protocol	4. Explains study design inclusion	5. Compre-hensive search	6. Duplicate selection & extraction	7. List of included & excluded studies	8. Study character-istics provided	9. Risk of bias assessed
Andrews, 2018[87]	Yes	Yes	None	Yes	Yes	Unclear; Yes	Yes; No	Yes	Yes
Hunot, 2010[88]	Yes	Yes	None	Yes	Yes	Yes; Yes	Yes; Yes	Yes	Yes
Ipser, 2010[94]	Yes	Yes	None	Yes	Yes	Yes; Yes	Yes; Yes	Yes	Yes
James, 2015[89]	Yes	Yes	None	Yes	Yes	Yes; Yes	Yes; Yes	Yes	Yes
Mayo-Wilson, 2013[90]	Yes	Yes	None	Yes	Yes	Yes; Yes	Yes; Yes	Yes	Yes
Meades, 2011[50]	Unclear	No	Unclear	Yes	Yes	Not reported	Yes; No	Yes	Yes
Piqueras, 2017[47]	Yes	Yes	Unclear	Yes	Yes	Unclear; Unclear	Yes; No	Yes	Unclear
Slee, 2019[99]	Yes	Yes	Unclear	Yes	Yes	Unclear; Yes	Yes; No	Yes	Yes
van Dis, 2019[16]	Yes	Yes	Unclear	Yes	Yes	Yes; Yes	Yes; No	Yes	Yes
Wang, 2017[98]	Yes	Yes	Unclear	Yes	Yes	Yes; Data checked	Yes; Yes	Yes	Yes

PICO=populations, interventions, comparators, and outcomes

Quality Ratings of Systematic Reviews (continued)

Author, year	10. Conflict of interest stated (review; studies)	11. Appropriate meta-analysis methods	12. Risk of bias considered in meta-analysis	13. Risk of bias used in conclusions	14. Explains hetero-geneity	15. Publication bias assessed	Quality rating†	Search dates‡
Andrews, 2018[87]	No; No	Yes	Yes	Yes	Yes	Yes	Good	Sept 2016
Hunot, 2010[88]	No; No	Yes	Yes	Yes	Yes	Yes	Good	Feb 2006
Ipser, 2010[94]	Yes; Yes	Yes	Yes	Yes	Yes	Yes	Good	Aug 2008
James, 2015[89]	Yes; No	Yes	Yes	Yes	Yes	Yes	Good	July 2012
Mayo-Wilson, 2013[90]	Yes; No	Yes	Yes	Unclear	Yes	Unclear	Good	Jan 2013
Meades, 2011[50]	Yes; No	Not applicable	Not applicable	Unclear	No	Unclear	Poor	Sept 2010
Piqueras, 2017[47]	No; No	Yes	Unclear	Unclear	Yes	Unclear	Fair	Jan 2000 to June 2016
Slee, 2019[99]	Yes; No	Yes	Unclear	Yes	Yes	Yes	Good	Jan 1, 1994 to Aug 1, 2017
van Dis, 2019[16]	Yes; No	Yes	Yes	Yes	Yes	Yes	Good	Jan 1980 to Jan 1, 2019
Wang, 2017[98]	Yes; No	Yes	Unclear	Yes	Yes	Yes	Good	Feb 21, 2017

*Quality criteria definitions: 1: Research questions and inclusion include components of PICO. 2: Explicit statement of a priori development of methods. 3: No deviations from protocol, if so, they are justified. 4: Explanation of study design inclusion. 5: Comprehensive literature search. 6: Duplicate study selection and data abstraction. 7: List of studies (included and excluded) provided. 8: Characteristics of the included studies provided. 9: Satisfactory technique used for assessing risk of bias in individual studies. 10: Conflict of interest (including funding sources) for systematic review and individual studies. 11: If meta-analysis performed, appropriate methods used for combination of results. 12: If meta-analysis performed, potential impacts of risk of bias on meta-analysis or other evidence synthesis assessed. 13: Risk of bias taken into account when interpreting/discussing results. 14: Satisfactory explanation for, and discussion of, any heterogeneity observed in the results. 15: If quantitative synthesis, there was adequate investigation of publication bias (small study bias) and discussion of its likely impact on the results.
†Definition of ratings based on quality criteria: Good: Recent, relevant review with comprehensive sources and search strategies; explicit and relevant selection criteria; standard appraisal of included studies; and valid conclusions. Fair: Recent, relevant review that is not clearly biased but lacks comprehensive sources and search strategies. Poor: Outdated, irrelevant, or biased review without systematic search for studies, explicit selection criteria, or standard appraisal of studies.
‡Most searches included the earliest dates provided by specific databases, others defined a range of date