APPENDIX 10: COMPLETED METHODOLOGY CHECKLISTS FOR CASE IDENTIFICATION AND ASSESSMENT

CONNOR2001 2
DALRYMPLE2008 4
KROENKE2007 6
MCQUAID2000 8
MEANS-CHRISTENSEN2006 10
OSÓRIO2007 11
SEELEY-WAIT2009 13
WEEKS2007 15

All references for the study IDs can be found in the full guideline.

Abbreviations

AACT	Adult Anxiety Clinic of Temple
ADD	Anxiety and Depression Detector
ADIS (-IV-L)	Anxiety Disorders Interview Schedule (for DSM-IV Lifetime Version)
CIDI (-Auto)	Composite International Diagnostic Interview (self-administered)
DSM-IV	*Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition
GAD	generalised anxiety disorder
GAD (-2, -7)	Generalised Anxiety Disorder scale (two-item, seven-item)
GSAD	generalised social anxiety disorder
MS	Mini-SPIN
(Mini-) SPIN	(Mini) Social Phobia Inventory
SAD	social anxiety disorder
SCID (-IV)	Structured Clinical Interview for DSM Disorders (Fourth Version)
SPQ	Social Phobia Questionnaire

Study ID: **CONNOR2001**	
DOMAIN 1: PATIENT SELECTION	
A. Risk of bias	
Describe methods of patient selection:	'A random sample of adult participants in a managed health care organization (n = 7,165) were contacted… For comparison, a random sample of individuals who were Mini-SPIN negative and depression positive according to the three-item depression screener (n = 397) or Mini-SPIN and depression negative (n = 276) were selected as control groups and interviewed in the same manner.'
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	CONCERN: LOW
DOMAIN 2: INDEX TEST(S)	
A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	'Those who screened positive for GSAD on the Mini-SPIN (score of at least 6) and agreed to further participation were interviewed by telephone after providing verbal informed consent'.
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	'To confirm the diagnosis of GSAD (n = 344), the interview consisted of a battery of assessments including the social phobia module of the Structured Clinical Interview for DSM-IV (SCID) [First *et al*., 1995]'.
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: UNCLEAR
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW
DOMAIN 4: FLOW AND TIMING	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	'Because approximately every fourth person who was negative on the Mini-SPIN was chosen to be interviewed, each of these subjects was weighted in the analyses to represent 3.97 subjects, to approximate the total sample of this "negative screen" group.'
Describe the time interval and any interventions between index test(s) and reference standard:	Unclear
Was there an appropriate interval between index test(s) and reference standard?	Unclear
Did all patients receive a reference standard?	No
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	RISK: LOW

Study ID: **DALRYMPLE2008**	
DOMAIN 1: PATIENT SELECTION	
A. Risk of bias	
Describe methods of patient selection:	'...psychiatric outpatients presenting for treatment at the outpatient practice of the Rhode Island Hospital Department of Psychiatry… Individuals presenting for an intake appointment were asked to participate in a diagnostic evaluation before meeting with their treating clinician.'
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	CONCERN: LOW
DOMAIN 2: INDEX TEST(S)	
A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	'The SAD module of the SCID begins with a standard screening question: "Was there ever anything that you have been afraid to do or felt uncomfortable doing in front of other people, like speaking, eating, or writing?" If the screening question is answered in the affirmative, then the remaining SAD criteria are evaluated.'
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	'The SAD module was modified for the current study by adding a list of social fears to cue the participants. All patients were administered the social fears list after the screening question, regardless of whether they provided a negative or affirmative response to the screening question.'
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	No
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW
DOMAIN 4: FLOW AND TIMING	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	'All patients were evaluated with the full SCID.' 'Three participants had missing data and were therefore excluded from the analyses, yielding a sample size of 1797.'
Describe the time interval and any interventions between index test(s) and reference standard:	None
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive a reference standard?	Yes
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	RISK: LOW

Study ID: **KROENKE2007**	
DOMAIN 1: PATIENT SELECTION	
A. Risk of bias	
Describe methods of patient selection:	'Of the 2740 participants, the first 2149 were used for development and validation of the GAD-7 scale, whereas the last 591 were used to determine the test–retest reliability of the scale.'
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	CONCERN: LOW
DOMAIN 2: INDEX TEST(S)	
A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	'Before seeing their physicians, patients completed a 4-page questionnaire that included the GAD-7… Scores on the GAD-7 range from 0 to 21; scores of 5, 10, and 15 represent mild, moderate, and severe anxiety symptoms, respectively. The first 2 items of the GAD-7 represent core anxiety symptoms, and scores on this GAD-2 subscale range from 0 to 6.'
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	No
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	'They based a diagnosis of social anxiety disorder on whether the patient met SCID diagnostic criteria and had a Mini-SPIN score of 8 or greater, because this improves the accuracy of social anxiety disorder diagnoses.' 'The 2 mental health professionals, while blinded to the results of the self-report research questionnaire, conducted structured psychiatric interviews by telephone…'
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW
DOMAIN 4: FLOW AND TIMING	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	'1654 agreed to a telephone interview, of whom 965 were randomly selected to undergo this interview'. Describe the time interval and any interventions between index test(s) and reference standard: 'within 1 week of their clinic visit'.
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive a reference standard?	No
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	RISK: LOW

Study ID: **MCQUAID2000**	
DOMAIN 1: PATIENT SELECTION	
A. Risk of bias	
Describe methods of patient selection:	'213 primary care patients attending scheduled appointments at UCSD [University of California, San Diego] Healthcare Mira Mesa Family Practice Clinic.'
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	CONCERN: LOW
DOMAIN 2: INDEX TEST(S)	
A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	'Social Phobia Questionnaire (SPQ) [Stein et al., 1999b]'. 'endorsing high anxiety or avoidance to one of the 10 SPQ items (for the first item, speaking in public, both anxiety and avoidance needed to be endorsed as nearly always occurring).'
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	'Interviewers were research assistants with a bachelor degree in psychology. Both were trained by the first author, who in turn had been trained on the CIDI by Sergio Aguilar-Gaxiola, M.D., Ph.D., at California State University, Fresno, one of the World Health Organization (WHO) certified CIDI training sites. Interviewers reviewed all cases in which there was a question regarding diagnosis with the first author.'
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW
DOMAIN 4: FLOW AND TIMING	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	'We approached 2,218 (94%) of the 2,358 clinic visitors present when a research assistant was available. Of those initially contacted, 1,024 met criteria for participation (i.e., were patients of the clinic, 18 years of age or older, spoke English, and had not participated in the study previously), 566 (55%) consented to participate, and 511 (49.9%) completed the screening measures before being called for their appointment.' 'Patients were contacted if they met a cutoff score on one of the screening measures.' 'In addition, we contacted 29 participants who did not meet screening criteria on any of the measures and who served as a control sample, for a total of 213 participants.' Describe the time interval and any interventions between index test(s) and reference standard: 'in the 2 weeks following their screening'.
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive a reference standard?	No
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No

Could the patient flow have introduced bias?	RISK: HIGH

Study ID: **MEANS-CHRISTENSEN2006**	
DOMAIN 1: PATIENT SELECTION	
A. Risk of bias	
Describe methods of patient selection:	'12,724 patients were approached and asked to participate'. '4409 patients (34.7%) did not complete the screening questionnaire, either because they refused or were ineligible.'
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	CONCERN: LOW
DOMAIN 2: INDEX TEST(S)	
A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	'Patients completed a pencil-and-paper screening questionnaire. There were two versions of the screening questionnaire, and each patient completed one of the two versions. Patients recruited in the first 20 months of the study completed Version A, and patients in the final 5 months of the study completed Version B.'
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	RISK: HIGH
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	CONCERN: LOW
DOMAIN 3: REFERENCE STANDARD	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	'The reference standard for establishing diagnostic accuracy of the ADD was the interviewer-administered computerized version of the CIDI-Auto, version 2.1. The CIDI-Auto was administered during a telephone interview by one of 19 trained CIDI interviewers, and diagnoses were obtained for major depression, GAD, panic disorder, social

	phobia and PTSD [post-traumatic stress disorder].' 'Interviewers were not blind to the results of the screening questionnaire as some of the content in the remainder of the telephone interview (not pertinent to this study) was determined by their screening status.'
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	No
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW
DOMAIN 4: FLOW AND TIMING	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	'Selected patients who screened negatively on all items were less likely to agree to participate in the interview [41% vs. 58%; v2(1)=75.27, Pb.01] and were less likely to actually complete the interview after having agreed to participate [51% vs. 66%; v2(1)=25.48, Pb.01].'
Describe the time interval and any interventions between index test(s) and reference standard:	'median time between screening and interview = 14 days'.
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive a reference standard?	No
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	RISK: LOW

Study ID: **OSÓRIO2007**	
DOMAIN 1: PATIENT SELECTION	
A. Risk of bias	
Describe methods of patient selection:	'Students (N=656) from a private and a public university in a city in the interior of the State of Sao Paulo-Brazil were randomly selected for the study.' 'Based on the proposal by Connor et al. [4] (a response equal to or above six to items 6, 9 and 15 of the SPIN), the researchers identified all subjects who met the criteria for a positive MS (N=473; 20.4% of the studied sample). As a comparison group, 183 subjects with negative MS were selected

	(7.9% of the studied group)… The comparison group (negative MS) was matched to the positive MS group for gender, age, institution and study field. A total of 656 subjects (28.1% of the total studied group) were selected to participate in the second phase of data collection.'
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	CONCERN: LOW
DOMAIN 2: INDEX TEST(S)	
A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	'In the first phase, as part of the prevalence study, the SPIN was administered to 2320 students.'
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	CONCERN: LOW
DOMAIN 3: REFERENCE STANDARD	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	'In the second phase, the subjects were contacted by telephone in order to answer the SCID-IV SAD (module F), which was used as a gold standard for SAD diagnosis. The raters (N=4) were mental health professionals with previous experience in using rating scales and the SCID-IV in clinical and research contexts. The raters, who did not know the score of the subjects in the previously administered tools… set up a time when the participants could talk in private, with nobody else being present during the assessment. The duration of the interviews ranged from three to five minutes.'
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes
Could the reference standard, its conduct, or its	RISK: LOW

interpretation have introduced bias?	
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW
DOMAIN 4: FLOW AND TIMING	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	'Of the selected subjects, 59 could not be located and 7 said they were no longer interested in participating in the study.'
Describe the time interval and any interventions between index test(s) and reference standard:	Unclear
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive a reference standard?	Yes
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	RISK: LOW

Study ID: **SEELEY-WAIT2009**	
DOMAIN 1: PATIENT SELECTION	
A. Risk of bias	
Describe methods of patient selection:	'A total of 186 (51% female) clinical participants presented to the Macquarie University Anxiety Research Unit for assessment and treatment for social anxiety disorder. An additional 56 (63% female) nonclinical participants were recruited to serve as comparisons.'
Was a consecutive or random sample of patients enrolled?	No
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	RISK: HIGH
B. Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	CONCERN: LOW
DOMAIN 2: INDEX TEST(S)	
A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	'All suitable participants were sent a questionnaire pack by mail for completion at home that included the Mini-SPIN and other symptom measures. Participants returned completed measures when they presented at the unit for their ADIS-IV

	interview.'
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	CONCERN: LOW
DOMAIN 3: REFERENCE STANDARD	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	'All participants were interviewed by trained clinicians using the Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV)'.
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW
DOMAIN 4: FLOW AND TIMING	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	None reported.
Describe the time interval and any interventions between index test(s) and reference standard:	'...participants were sent a questionnaire pack by mail for completion at home that included the Mini-SPIN and other symptom measures. Participants returned completed measures when they presented at the unit for their ADIS-IV interview.'
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive a reference standard?	Yes
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Unclear
Could the patient flow have introduced bias?	RISK: UNCLEAR

Study ID: **WEEKS2007**	
DOMAIN 1: PATIENT SELECTION	
A. Risk of bias	
Describe methods of patient selection:	'Participants were 291 individuals seeking treatment at the AACT for social anxiety or generalized anxiety and worry. These individuals first called the AACT and completed a telephone screening interview. Callers were referred to the AACT through advertisements, its website, or by mental health professionals in the community.'
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the included patients do not match the review question?	CONCERN: LOW
DOMAIN 2: INDEX TEST(S)	
A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	'We evaluated the sensitivity, specificity, and diagnostic efficiency of the Mini-SPIN in the subsample of individuals who arrived for diagnostic interview, using cutoff scores ranging from 4 to 9.3. However, only one participant obtained a score of 5 on the Mini-SPIN; thus, this cutoff score was not evaluated in this sample.'
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	No
Could the conduct or interpretation of the index test have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question?	CONCERN: LOW

DOMAIN 3: REFERENCE STANDARD	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	'Assessed using the ADIS-IV-L'.
Is the reference standard likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	RISK: LOW
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question?	CONCERN: LOW
DOMAIN 4: FLOW AND TIMING	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):	'The majority of callers (n = 238, 81.8%) were invited to the AACT for further evaluation with the Anxiety Disorders Interview Schedule for DSM-IV: Lifetime Version [ADIS-IV-L; DiNardo et al., 1994]'.
Describe the time interval and any interventions between index test(s) and reference standard:	'Of the 200 individuals scheduled for pretreatment evaluation, 135 (67.5%) arrived at the AACT and were assessed using the ADIS-IV-L.'
Was there an appropriate interval between index test(s) and reference standard?	Yes
Did all patients receive a reference standard?	No
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	RISK: LOW