APPENDIX 11: CASE ID – STUDY CHARACTERISTICS AND RISK OF BIAS

Study	Instrument	No. of items	Range (cut-off)	Recruitment	N	Female, n (%)	Age	Country	Prevalence	Sensitivity	Specificity
DODD2009	MDQ	13	Yes/no (7)	Community	1066	1066 (100%)	51	Australia	2.3%	0.25	0.99
HENRY2008	CMRS-P	10	4 point Likert scale. 4-40 (10)	Community and psychiatric settings	100	45 (45%)	10	USA	50%	0.92	0.82
HIRSCHFELD2003	MDQ	13	Yes/no (7)	Community (General population)	695	NR	46	USA	11.2%	0.28	0.97
TILLMAN2005	Conners' Abbreviated Parent Questionnaire	10	4 possible answers per question. 4-40 (9 for 7-8y, 8 for 9-10y, 6 for 11-16y)	Community and psychiatric settings	264	89 (34%)	11	USA	34.9%	0.73	0.86

Note. MDQ = Mood Disorder Questionnaire; CMRS-P = Child Mania Rating Scale – Parent version;

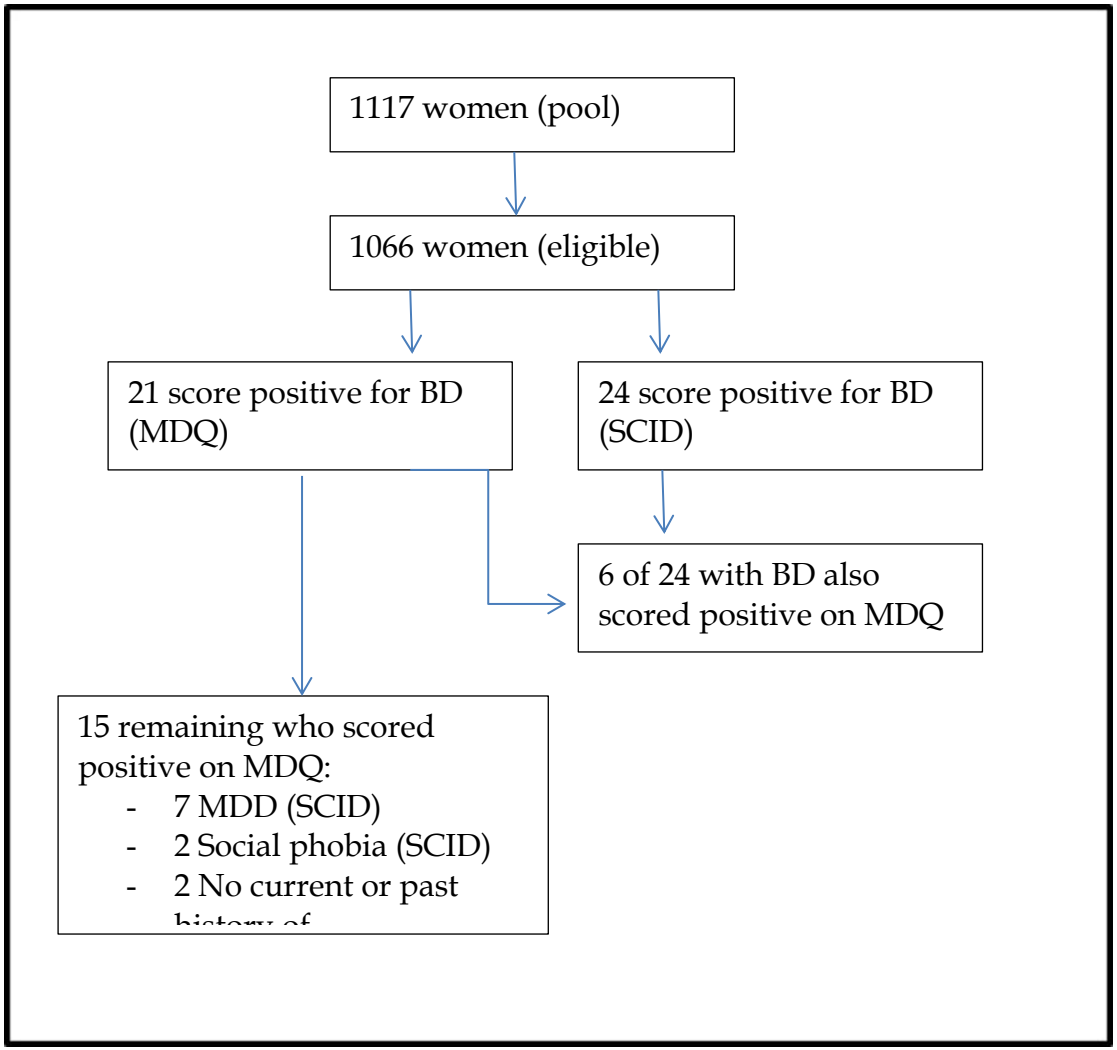
Methodology checklist: the QUADAS-2 tool for studies of diagnostic test accuracy

Study ID: DODD2009

Phase 1: State the review question

Patients (setting, intended use of index test, presentation, prior testing): Children and young people (aged 18 years and younger) and adults with suspected bipolar disorder
Index test(s): Brief screening questionnaires (<15 items) identified by the GDG
Reference standard and target condition: DSM or ICD diagnosis of bipolar disorder.

Phase 2: Draw a flow diagram for the primary study



Phase 3: Risk of bias and applicability judgements

QUADAS-2 is structured so that four key domains are each rated in terms of the risk of bias and the concern regarding applicability to the review question (as stated in Phase 1). Each key domain has a set of signalling questions to help reach the judgements regarding bias and applicability.

Domain 1: Patient selection

A. Risk of bias	
Describe methods of patient selection: "Included data collected from women participating in the Geelong Osteoporosis Study (GOS), a large epidemiological study involving age-stratified community-based samples of women randomly recruited from the electoral roll for the region (Barwon Statistical Division, South-Eastern Australia). The initial sample was recruited between 1994 and 1997, with 1494 women (median age=54 years, interquartile range (IQR)37-72) agreeing to participate [18]. A further cohort of 200 women aged 20-29 years was also recruited at the time of the 10 year follow up."	
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
Describe included patients (prior testing, presentation, intended use of index test and setting): “From a potential pool of 1117 women, enrolled in the GOS at the time of the study, 23 women who did not participate in the clinical interview and a further 28 who had not completed the MDQ were excluded from the analyses, resulting in a sample of 1066 women (95%) aged 21-94 years eligible for inclusion.” . .
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 MDQ is a self-report screening instrument for detecting previous mania and hypomania, characteristic of bipolar I and II disorders, respectively[6]. The MDQ includes 13 questions relating to manic or hypomanic symptoms derived from DSM-IV criteria plus a further two items assessing the temporal clustering of symptoms and functional impairment. This questionnaire has been used and validated in both primary care and community settings [1,7]."	
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: Unclear**	
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question? **Concern: High (women only)**	

Domain 3: Reference standard

A. Risk of bias

Describe the reference standard and how it was conducted and interpreted: "Psychiatric status was assessed using the Structured Clinical Interview for DSM-IV-TR Research Version, Non-patient edition (SCID). The SCID-I/NP was used to identify those who had ever experienced a depressive disorder, including bipolar disorder (I, II and not otherwise specified or NOS), major depressive disorder, minor depression, dysthymia, mood disorder due to a general medical condition and substance-induced mood disorder and/or any anxiety disorder. All psychiatric interviews were conducted by trained personnel who were blind to the results from the MDQ screen." .	
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): . . Describe the time interval and any interventions between index test(s) and reference standard: Not described. . . .	
Was there an appropriate interval between index test(s) and reference standard?	**Unclear**
Did all patients receive a reference standard?	**Yes**

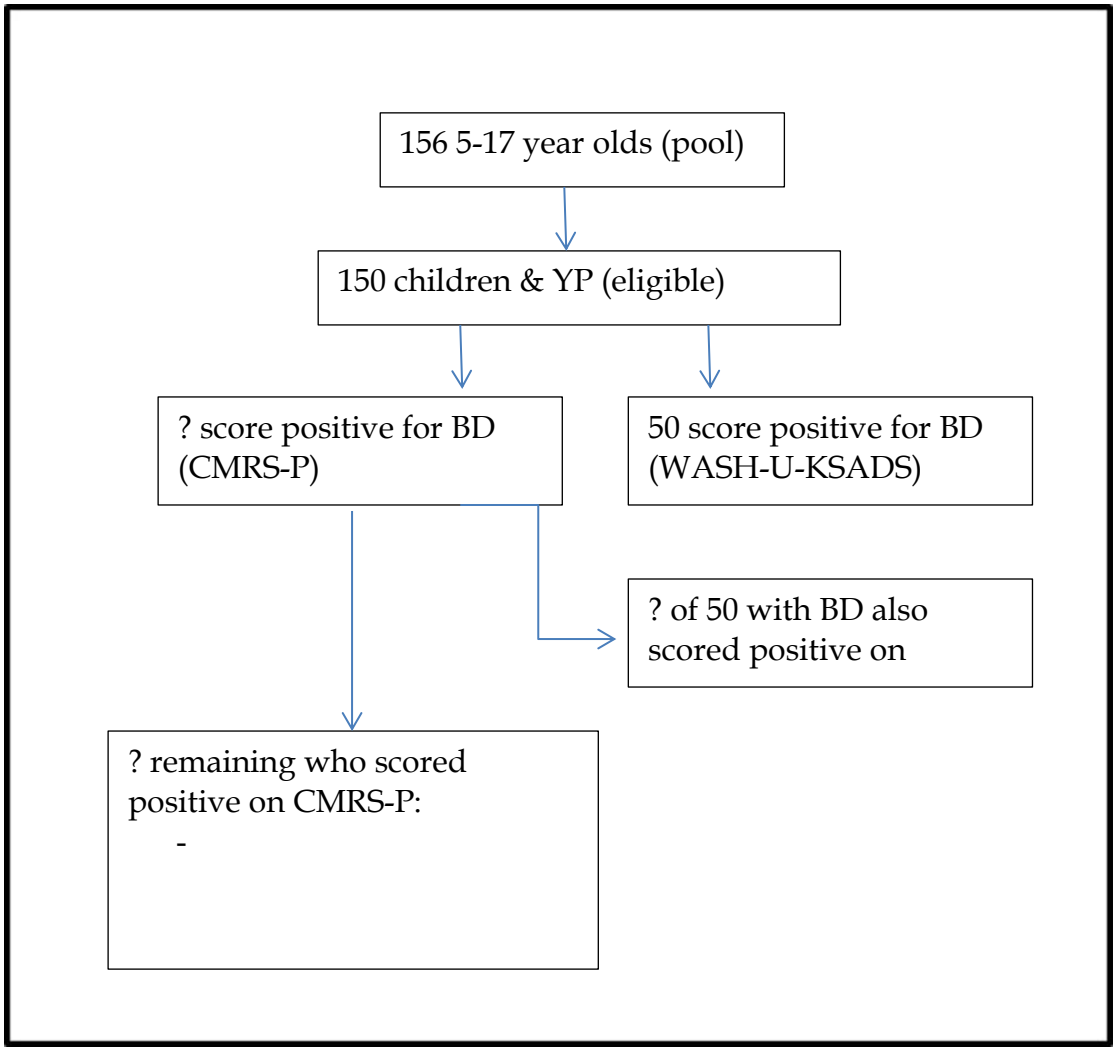
Did patients receive the same reference standard?	**Yes**
Were all patients included in the analysis?	**Yes**
Could the patient flow have introduced bias? **Risk: Low**	

Study ID: HENRY2008

Phase 1: State the review question

Patients (setting, intended use of index test, presentation, prior testing): Children and young people (aged 18 years and younger) and adults with suspected bipolar disorder
Index test(s): Brief screening questionnaires (<15 items) identified by the GDG
Reference standard and target condition: DSM or ICD diagnosis of bipolar disorder.

Phase 2: Draw a flow diagram for the primary study



Phase 3: Risk of bias and applicability judgements

QUADAS-2 is structured so that four key domains are each rated in terms of the risk of bias and the concern regarding applicability to the review question (as stated in Phase 1). Each key domain has a set of signalling questions to help reach the judgements regarding bias and applicability.

Domain 1: Patient selection

A. Risk of bias
Describe methods of patient selection: “Subjects were included in this sample if they were: between 5 and 17 years of age, inclusive; had been diagnosed with ADHD, BD I or II, or BD-NOS; or were healthy controls (HC) with no psychiatric symptoms, based on the WASH-U- KSADS (Geller, Zimerman et al., 2001) interview. Potential participants were excluded if they: suffered from head injury, epilepsy, a pervasive development disorder (PDD), or mental retardation; had significant medical illness; or were taking any medications or substances that could alter their moods. Patients currently under treatment were excluded as the purpose of the study was to screen for subjects at the intake phase, regardless of the severity of their disorders.” . “Subjects with BD were recruited from among the community, pediatricians, child psychiatrists, the Child and Adolescent Bipolar Foundation (CABF), and our mood disorders clinic.” .

Was a consecutive or random sample of patients enrolled?	**Unclear**
Was a case-control design avoided?	**No**
Did the study avoid inappropriate exclusions?	**Unclear**
Could the selection of patients have introduced bias? **Risk: Unclear**	
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting): "The final sample consisted of 150 subjects (BD=50; ADHD=50; HC=50)."	
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 research staff (n=6) who administered the demographic and parent/self- report measures were different from those who conducted the diagnostic interviews (n=4). The CMRS-P was administered prior to conducting the diagnostic interview to minimize bias and fatigue effects." . "To create a brief CMRS-P, we selected items that allowed us to cover the entire range of severe mania symptoms and were separated by relatively equal intervals."	
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: Unclear**	

Domain 3: Reference standard

A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted: "The mania section of the WASH-U-KSADS is a semistructured diagnostic interview used to diagnose mania. Symptoms are rated on a continuous severity scale by interviewing parents and children (Geller, Zimerman et al., 2001). The WASH-U-KSADS includes specific questions about onset and offset of symptoms, ADHD assessment, and criteria for diagnosis of manic behavior or thinking. The WASH-U KSADS has demonstrated 100% interrater reliability (Geller, Zimerman et al., 2001)."	
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): “Six subjects were excluded due to primary diagnosis of pervasive developmental disorder (n 5 3), medication for posttraumatic stress disorder (n 5 1), aging-out at 18th birthday (n51), and father’s withdrawal of consent during a custody battle (n 5 1).” . Describe the time interval and any interventions between index test(s) and reference standard: Not described. . . .	
Was there an appropriate interval between index test(s) and reference standard?	**Unclear**

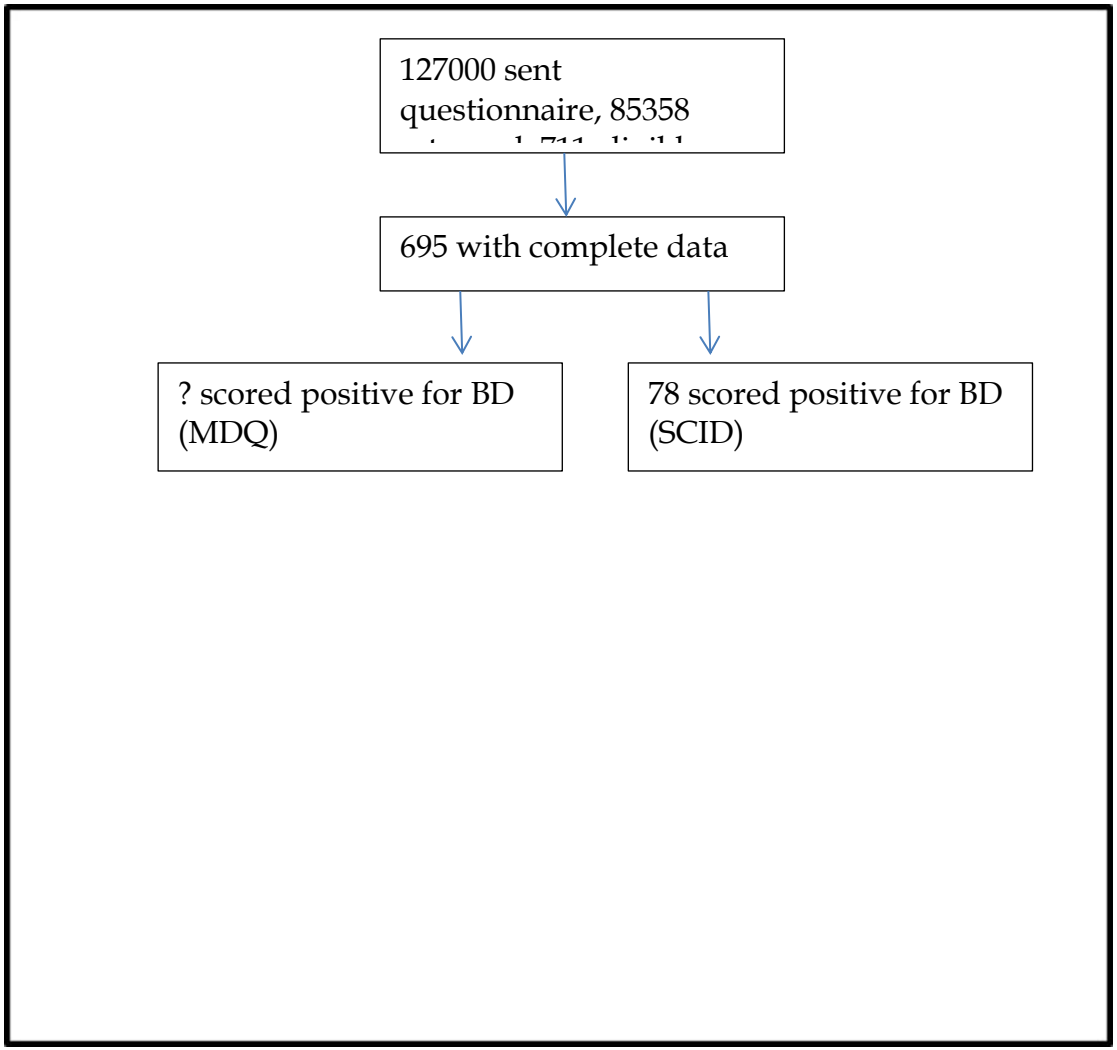
Did all patients receive a reference standard?	**Yes**
Did patients receive the same reference standard?	**Yes**
Were all patients included in the analysis?	**Yes**
Could the patient flow have introduced bias? **Risk: Low**	

Study ID: HIRSCHFELD2003

Phase 1: State the review question

Patients (setting, intended use of index test, presentation, prior testing): Children and young people (aged 18 years and younger) and adults with suspected bipolar disorder
Index test(s): Brief screening questionnaires (<15 items) identified by the GDG
Reference standard and target condition: DSM or ICD diagnosis of bipolar disorder.

Phase 2: Draw a flow diagram for the primary study



Phase 3: Risk of bias and applicability judgements

QUADAS-2 is structured so that four key domains are each rated in terms of the risk of bias and the concern regarding applicability to the review question (as stated in Phase 1). Each key domain has a set of signalling questions to help reach the judgements regarding bias and applicability.

Domain 1: Patient selection

A. Risk of bias	
Describe methods of patient selection: “Subjects for this study were a subset of respondents in a large general population epidemiological study of bipolar I and II disorders (the “prevalence study”). The target sample for the current study was 700 randomly selected subjects stratified by Mood Disorder Questionnaire score. Approximately 40 subjects were selected for each Mood Disorder Questionnaire.”	
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
Describe included patients (prior testing, presentation, intended use of index test and setting): "The sample's mean age was 46.1 years (weighted). A total of 95% (weighted) reported high school completion or the equivalent. A total of 89% (weighted) were white non-His- panic, 5% were black non-Hispanic, 2.3% were Hispanic, and the remainder of subjects were of other ethnic back- grounds."
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 Mood Disorder Questionnaire is a self-report inventory that screens for bipolar I and II disorders with 13 yes/no items derived from both DSM-IV criteria and clinical experience (1). A positive screen requires that seven or more items be endorsed, that at least several of the items co-occurred, and that the symptoms caused at least moderate psychosocial impairment."

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: "A team of 10 doctoral and two master's-level clinical and psy- chiatric research interviewers were recruited and trained to administer an abbreviated lifetime version of the SCID for Axis I Disorders (4). Each subject was contacted by survey staff and scheduled for the SCID as a computer-aided telephone interview. The interviewers were blind to the results of the initial Mood

Dis- order Questionnaire.”	
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): "A total of 711 subjects were identified by National Family Opinion as meeting the study entry criteria. Of these, 12 refused the reinterview. An additional two had incomplete data and were not included in the analyses. A total of 695 subjects were left who completed the telephone research interview and whose data were complete for analyses." Describe the time interval and any interventions between index test(s) and reference standard: Not described. . . .	
Was there an appropriate interval between index test(s) and reference standard?	**Unclear**
Did all patients receive a reference standard?	**Yes**
Did patients receive the same reference standard?	**Yes**
Were all patients included in the analysis?	**No**

Could the patient flow have introduced bias?

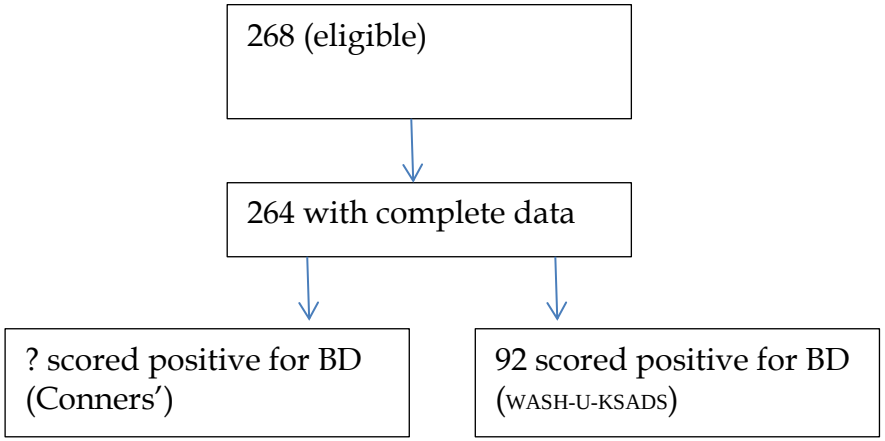
Risk: Low

Study ID: TILLMAN2005

Phase 1: State the review question

Patients (setting, intended use of index test, presentation, prior testing): Children and young people (aged 18 years and younger) and adults with suspected bipolar disorder
Index test(s): Brief screening questionnaires (<15 items) identified by the GDG
Reference standard and target condition: DSM or ICD diagnosis of bipolar disorder.

Phase 2: Draw a flow diagram for the primary study



Phase 3: Risk of bias and applicability judgements

QUADAS-2 is structured so that four key domains are each rated in terms of the risk of bias and the concern regarding applicability to the review question (as stated in Phase 1). Each key domain has a set of signalling questions to help reach the judgements regarding bias and applicability.

Domain 1: Patient selection

A. Risk of bias	
Describe methods of patient selection: "To optimize generalization, subjects with a prepubertal and early adolescent bipolar disorder phenotype and subjects with ADHD were identified through consecutive new case as- certainment from outpatient child psychiatric and pediatric sites. The healthy comparison group was identified through a random survey that matched the comparison subjects to subjects with a prepubertal and early adolescent bipolar disorder phenotype by age, gender, socioeconomic status, ethnicity, and zip code." .	
Was a consecutive or random sample of patients enrolled?	**Yes**
Was a case-control design avoided?	**No**
Did the study avoid inappropriate exclusions?	**Yes**

Could the selection of patients have introduced bias? **Risk: Low**
B. Concerns regarding applicability
Describe included patients (prior testing, presentation, intended use of index test and setting): Participants had bipolar, ADHD or were healthy controls, 7-16 years old. . .
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: “Parents completed the Conners’ Abbreviated Parent Questionnaire about their children before the WASH-U-KSADS interviews. The Conners’ Abbreviated Parent Questionnaire, an instrument for assessing ADHD in children and adolescents, has 10 items,

each with four possible answers. Each item has a minimum score of 1 and a maximum score of 4." . .	
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: "The Washington University in St. Louis Kiddie Schedule for Affective Disorders and Schizophrenia (WASH-U-KSADS) (6), a semistructured interview, was administered by experienced re- search nurses separately to parents about their children and to children about themselves. The nurses were blind to the diagnostic group of the subjects." .	
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): "Subjects enrolled in the study included 93 with a prepubertal and early adolescent bipolar disorder phenotype, 81 with ADHD, and 94 healthy comparison subjects. Four subjects with missing values for the Conners' Abbreviated Parent Questionnaire were excluded from the analysis, leaving 92 prepubertal and early adolescent bipolar disorder phenotype, 80 ADHD, and 92 healthy comparison subjects." . Describe the time interval and any interventions between index test(s) and reference standard: Not described.	
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?	**No**
Could the patient flow have introduced bias? **Risk: Unclear**	