Table 3.5.14 *Sensitivity and specificity for SCID-I with LEAD as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Miller et al 2001 [6] USA	*Design* Cross sectional and longitudinal *Setting* University-affiliated public psychiatric hospital in Los Angeles, USA *Population* n=75 inpatients eligible, 75% (n=56) participated, mostly chronically ill with multiple diagnoses, one patient declined to participate, other 18 excluded according to criteria below *Inclusion criteria* >18 years, sufficient English fluency and cognitive function to participate in interviews. First patient each week was screened *Exclusion criteria* Too acutely disordered *Prevalence* BP: 10.7% (6/56) MDD: 21.4% (12/56) Schizoaffective: 37.5% (21/56)	SCID-I-CV *Reference test* LEAD Consensus by raters using discharge clinical diagnoses, SCID-I, structured interview, all data *Number of patients* n=56 (45% female) Mean age: 36.9 years (SD 12.4; range 19–59) *Drop-out rate* None	*Rater of index test* Two psychiatrists *Rater training* Extensive training *Rater of reference test* Three to four psychiatrists or PhD psychologist *Training of reference test* NA *Interobserver reliability* Not reported *Outcome measure* Sensitivity and specificity values are constructed from data in the article	*Bipolar disorder (n=6)* Sensitivity: 100% Specificity: 96% *Depression (n=12)* Sensitivity: 92% Specificity: 98% *Schizoaffective disorder (n=21)* Sensitivity: 76% Specificity: 97%	Moderate *Comments* *Sampling method* Adequate, among severely ill patients but small sample *Blinding* Yes, NA for LEAD *Handling of missing data* NA *Other comments* Withdrawals explained, SCID-I interobserver reliability not reported, SCID-I was performed close to intake

The table continues on the next page

Table 3.5.14 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Ramirez Basco et al 2000 [7] USA	*Design* Cross sectional and longitudinal *Setting* Community mental health centre in Dallas, USA *Population* 210 outpatients recruited, 200 completed both interviews *Inclusion criteria* Not reported *Exclusion criteria* Not reported *Prevalence* BP: 25% (50/200) MDD: 19% (38/200) Schizoaffective: 18% (36/200)	*Index test* I1: SCID-I diagnoses. Including family history of psychiatric illness, general medical history. A life chart was constructed I2: SCID-I-plus-chart diagnoses. Above + medical records *Reference test* LEAD Interview after a week with one of prime investigators (psychiatrist or psýcholo-gist) after reviewing all data *Number of patients* n=200 Mean age: 37.6 years (SD 10) 116 (58%) females 134 (67%) Caucasian *Drop-out rate* 5%	*Rater of index test* Three psychiatric nurses *Training of index test* Yes *Rater of reference test* One of three psychiatrists or an psychologist *Training of reference test* NA *Interobserver reliability* Not reported	*Bipolar disorder (n=50)* *SCID-I* Sensitivity: 76% Specificity: 93% *SCID-I + chart* Sensitivity: 88% Specificity: 93% *Major depressive disorder (n=38)* *SCID-I* Sensitivity: 84% Specificity: 91% *SCID-I + chart* Sensitivity: 82% Specificity: 96% *Schizoaffective disorder (n=36)* *SCID-I* Sensitivity: 53% Specificity: 97% *SCID-I + chart* Sensitivity: 89% Specificity: 95%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes, and NA *Handling of missing data* Adequate *Other comments* Low drop-out rate. Raters were trained but no interrater reliability reported. Analysis of time requirements for the diagnostic steps, and effects on treatment of diag-nostic feedback, adds further to clinical relevance

BP = Bipolar disorder; LEAD = Longitudinal, Experts, All Data procedure; MDD = Major depressive disorder; NA = Not applicable; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SCID-I-CV = Structured clinical interview for DSM-IV, clinical version

Table 3.5.15 *Sensitivity and specificity for SADS with LEAD as reference.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Kosten et al 1992 [10] USA	*Design* Cross sectional as well as retrospective data are included *Setting* University-based clinic *Population* Opiate addicted probands and their spouses and/or first-degree relatives n=475 *Inclusion criteria* Over 18 years of age *Exclusion criteria* Not reported *Prevalence* MDD: 57.9% (275/475)	*Index test* SADS *Reference test* LEAD *Number of patients* n=475 *Drop-out rate* Not reported	*Rater of index test* Clinical assessors *Training of index rater* Trained in using SADS *Rater of reference test* 2 psychologists *Training of reference test* The 2 psychologists discussed the diagnoses on which they disagreed and obtained a consensus on each *Interobserver reliability* SADS: No information LEAD: 0.89 (range 0.62–1.00)	*Depressive disorder* Sensitivity: 79% Specificity: 98%	Low *Comments* *Sampling method* Probably convenience sample *Blinding* Not applicable *Handling of missing data* Information about drop-outs was not reported *Other comments* No outcome measures about sensitivity and specificity

LEAD = Longitudinal, Experts, All Data procedure; MDD = Major depressive disorder;
SADS = Schedule for affective disorders and schizophrenia

Table 3.5.16 *Sensitivity and specificity for Composite International Diagnostic Interview (CIDI) for depression with LEAD as reference.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Booth et al 1998 [16] USA	*Design* Cross sectional *Setting* University-affiliated Veterans Affairs Medical Center, February 1993 to January 1994 *Population* Convenience sample of medically ill inpatients with malignancies (29%), cardiovascular conditions (25%), general surgical conditions (13%), and other medical conditions *Inclusion criteria* Hospitalized for a medical diagnosis *Exclusion criteria* Hospitalized for psychiatric or neurological reasons *Prevalence* Current MDD: 17% (9/54) Lifetime MDD: 35% (19/54)	*Index test* UM-CIDI *Reference test* LEAD Depression, alcohol abuse and anxiety sections of SCID-I-III-R + all available, longitudinal data presented for expert panel reaching consensus diagnoses *Number of patients* n=54 Mean age: 56 years (range 30–76) 100% males 8% African-American *Drop-out rate* Not reported	*Rater of index test* Non-clinician research staff *Training of index rater* 40 hours *Rater of reference test* Two trained and clinically experienced mental health clinicians administered the SCID-I Expert panel: One SCID-I-interviewer, one research consultant-liaison psychiatrist with expertise in structured and semistructured interviews, one senior psychiatrist *Training of reference test* Experience in the SCID-I *Interobserver reliability* NA	*Current MDD* Sensitivity: 67% Specificity: 84% *Lifetime MDD* Sensitivity: 74% Specificity: 91%	Moderate *Comments* *Sampling method* Adequate, but relatively small convenience sample of male medically ill patients. Therefore results may not be generalizable to other samples *Blinding* Yes *Handling of missing data* NA *Other comments* Order of UM-CIDI and SCID-I randomly assigned to avoid bias. Study used two different CIDI scoring algoritms for MDD. Here diagnoses without exclusion rules are presented

CIDI = Composite international diagnostic interview; LEAD = Longitudinal, Experts, All Data procedure; MDD = Major depressive disorder; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SCID-I-III-R = Structured clinical interview for DSM-I-III-R; UM-CIDI = Composite international diagnostic interview, University of Michigan version

Table 3.5.17 *Sensitivity and specificity of DIS for depression with semistructured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Hasin et al 1987 [20] USA	*Design* Cross sectional *Setting* A 28-day inpatient alcohol rehabilitation unit *Population* Patients with alcohol and drug problems *Inclusion criteria* Not reported *Exclusion criteria* Not reported *Prevalence* MDD: 66.7% Dysthymic disorders/ intermittent depressive disorder: 33.3%	*Index test* DIS *Reference test* SADS *Number of patients* n=120 (randomly selected) (31% females) Mean age: 37.7 years (SD 12.2; range 17–72) 4 had drug problems *Drop-out rate* 7%	*Rater of index test* 13 interviewers (4 had no experience of psychiatric patients, 4 had done volunteer work, and 5 had extensive psychiatric nursing experience) *Training of index rater* Standard training following methods developed by Washington University *Rater of reference test* The authors of the study *Training of reference test* Adequate *Interobserver reliability* DIS: Not reported SADS: κ value was 1.0	*Major depressive disorder* Sensitivity: 25% Specificity: 90% *Dysthymic disorders/ intermittent depressive disorder* Sensitivity: 22% Specificity: 92%	Moderate *Comments* *Sampling method* Adequate, but there was no information about number of patients that did not consent *Blinding* Yes and the order of the interviews were in addition randomly assigned *Handling of missing data* Information about exclusion and missing data were reported

DIS = Diagnostic interview schedule; MDD = Major depressive disorder; SADS = Schedule for affective disorders and schizophrenia; SD = Standard deriation

Table 3.5.18 *Sensitivity and specificity of MINI for affective syndromes with structured or semistructured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Lecrubier et al 1997 [25] France (Paris) and USA (Florida)	*Design* Cross sectional *Setting* Psychiatric wards in France and USA *Population* Consecutive patients (310 from France and 40 from USA) *Inclusion criteria* Age: >18 years *Exclusion criteria* Dementia, mental retardation, language problems or serious medical illnesses *Prevalence* MDD: 50.4% Current mania: 6.1% Lifetime mania: 16.3%	*Index test* MINI *Reference test* CIDI *Number of patients* n=343 (51% females) Mean age: 42.2 years (SD 15.1) *Drop-out rate* 2%	*Rater of index test* In France: 3 psychiatrists and 2 psychologists. In USA: 10 psychiatrists, 4 research assistants and 2 medical students *Training of index test rater* Training packet for the MINI [24] *Rater of reference test* Same as index test rater *Training of reference test* The raters in France were trained and experienced by the French and Geneva WHO centres *Interobserver reliability* MINI: κ values range from 0.88 to 1.0 CIDI: Not reported	*Major depressive disorder* Sensitivity: 94% Specificity: 79% *Current mania* Sensitivity: 86% Specificity: 96% *Lifetime mania* Sensitivity: 77% Specificity: 92%	Moderate *Comments* *Sampling method* Adequate, but, there was no information about proportion that did not consent *Blinding* Acceptable *Handling of missing data* Adequate *Other comments* The interrater reliability was only reported for the index test

The table continues on the next page

Table 3.5.18 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Sheehan et al 1997 [24] USA (Florida) and France (Paris)	*Design* Cross sectional *Setting* Psychiatric outpatients in France and the USA *Population* Consecutive patients (280 from USA and 40 from France) *Inclusion criteria* Age: >18 years *Exclusion criteria* Dementia, mental retardation or serious medical illnesses *Prevalence* MDD: 45.9% Dysthymia: 1.6% Current mania: 10.3% Lifetime mania: 20.0%	*Index test* MINI *Reference test* SCID-P *Number of patients* n=320 (48% female) 56% met criteria for more than one current SCID-P diagnosis Mean age: 44.8 years (SD 15.1) *Drop-out rate* Not reported	*Rater of index test* In USA: 10 psychiatrists, 4 research assistants and 2 medical students. In France: 3 psychiatrists and 2 psychologists *Training of index rater* Training packet for the MINI *Rater of reference test* Same as index test rater *Training of reference test* One day training session for the SCID-P *Interobserver reliability* MINI-CR: All of κ values were above 0.75, and 16 of 23 were 0.90 or higher SCID-P: Not reported	*Major depressive disorder* Sensitivity: 96% Specificity: 88% *Dysthymia* Sensitivity: 67% Specificity: 99% *Current mania* Sensitivity: 82% Specificity: 95% *Lifetime mania* Sensitivity: 81% Specificity: 94%	Moderate *Comments* *Sampling method* Adequate, but there was no information about the number that did not consent *Blinding* Yes *Handling of missing data* Not reported *Other comments* The time between the MINI interviews and the SCID-P interviews were not explicitly reported but it was reasonable to assume that the period of time was short enough in the current study

CIDI = Composite international diagnostic interview; MDD = Major depressive disorder; MINI = Mini international neuropsychiatric interview; SCID-P = Structured clinical interview for DSM-IV axis I disorders, patient edition; SD = Standard deviation

Table 3.5.19 *Sensitivity and specificity of PRIME-MD for affective disorders with structured or semistructured interview as reference.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Leopold et al 1998 [27] USA	*Design* Cross sectional *Setting* Medical patients from the practices of 2 radiation oncologists *Population* Selection criteria: Availability of the SCID-I administrators n=135, of these 122 (90%) fulfilled the inclusion criteria *Inclusion criteria* Over 18 years old, English-speaking, cognitively unimpaired, and a life expectancy of at least 1 month *Prevalence* Any mood disorder: 49.1% (26/53) MDD: 18.9% (10/53)	*Index test* PRIME-MD *Reference test* SCID-I (DSM-III-R, APA 1987) *Number of patients* n=53 were administered both interviews *Drop-out rate* Not reported	*Rater of index test* 2 radiation oncologists *Training of index rater* Not reported *Rater of reference test* 1 clinical psychologist (administered 51 of the evaluations) and 1 psychiatrist (administered 2 of the evaluations) *Training of reference test* Adequate *Interobserver reliability* PRIME-MD: Not reported SCID-I: Not reported	*Any mood disorder* Sensitivity: 77% (95% CI, 56; 91) Specificity: 93% (95% CI, 76; 99) *Major depressive disorder* Sensitivity: 30% (95% CI, 7; 65) Specificity: 93% (95% CI, 76; 99)	Moderate *Comments* *Sampling method* Acceptable *Blinding* Yes *Handling of missing data* –

The table continues on the next page

Table 3.5.19 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Loerch et al 2000 [28] Germany	*Design* Cross sectional *Setting* Primary care and psychiatric clinics in Germany *Population* Medical and psychiatric out- and inpatients. Patients were contacted by their physicians and invited to participate n=924 *Inclusion criteria* Not reported *Exclusion criteria* Not reported *Prevalence* Any mood disorder: 19.9% (140/704) MDD: 12.6% (89/704)	*Index test* PRIME-MD *Reference test* Munich CIDI (M-CIDI) *Number of patients* n=704 (62% females) were administered both interviews Mean age: 46.1 years (SD 16.9) *Drop-out rate* Not reported	*Rater of index test* Physicians *Training of index rater* No training *Rater of reference test* 10 full-time research psychologists and 4 medical students with at least 4 years of medical school *Training of reference test* All were trained and supervised by 2 trained CIDI-raters *Interobserver reliability* Not reported	*Any mood disorder* Sensitivity: 69% Specificity: 85% *Major depressive disorder* Sensitivity: 68% Specificity: 84%	Moderate *Comments* *Sampling method* Mainly convenience sample. Number of patients contacted by physicians were not reported *Blinding* Yes *Handling of missing data* Not reported

APA = American journal of psychiatry; CI = Confidence interval; CIDI = Composite international diagnostic interview; DSM = Diagnostic and statistical Manual of mental disorder; MDD = Major depressive disorder; PRIME-MD = Primary health care evaluation of mental disorders; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SD = Standard deviation

Table 3.5.20 *Sensitivity and specificity of BDI-II with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Warmenhoven et al 2012 [40] The Netherlands	*Design* Cross sectional *Setting* Outpatient palliative care department *Population* Convenience sample n=61 of patients with advanced, non-curable stages of metastatic cancer *Inclusion criteria* Not reported *Exclusion criteria* Unable to read and understand the Dutch language *Prevalence* MDD: 21.7%	*Index test* BDI-II *Reference test* PRIME-MD *Number of patients* n=46 *Drop-out rate* 25%	*Rater of reference test* A physician *Training of reference test* Not reported *Interobserver reliability* Not reported	*Major depressive disorder* Sensitivity: 90% Specificity: 64%	Moderate *Comments* *Sampling method* Acceptable *Blinding* Yes *Handling of missing data* Not reported

The table continues on the next page

Table 3.5.20 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
De Souza et al 2010 [38] United Kingdom	*Design* Cross sectional *Setting* Neuropsychiatric clinic in Birmingham *Population* Outpatients with Huntington's disease *Inclusion criteria* >18 years *Exclusion criteria* Not able to give informed consent, not fluent in English *Prevalence* MDD: 24% (12/50)	*Index test* BDI-II *Reference test* SCAN *Number of patients* n=50 Mean age: 51.2 years (SD 10.35) *Drop-out rate* Not reported	*Rater of index test* NA *Training of index rater* NA *Rater of reference test* One researcher and one psychologist *Training of reference test* Yes, but not specified *Interobserver reliability* Not reported	*Major depressive disorder* Sensitivity: 83% Specificity: 71%	Moderate *Comments* *Sampling method* Adequate *Blinding* Unclear *Handling of missing data* Not reported
Dutton et al 2004 [39] USA	*Design* Cross sectional *Setting* Waiting rooms of two primary care clinics in southeastern USA *Population* n=223 African-American patients fulfilled criteria and consented *Inclusion criteria* >18 years *Exclusion criteria* Failing to demonstrate at least a 5th grade oral comprehension level *Prevalence* MDD: 29.5% (65/220)	*Index test* BDI-II *Reference test* PRIME-MD *Number of patients* n=220 Mean age: 49.3 years (SD 10.91) *Drop-out rate* 1%	*Rater of index test* NA *Training of index rater* NA *Rater of reference test* Six research assistants who all were doctoral students in clinical psychology *Training of reference test* Yes but not specified *Interobserver reliability* Not reported	*Major depressive disorder* Sensitivity: 88% (95% CI, 77; 94) Specificity: 84% (95% CI, 77; 89)	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Not reported

The table continues on the next page

Table 3.5.20 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Poole et al 2009 [41] United Kingdom	*Design* Cross sectional *Setting* Assessment clinic of a specialist pain center in United Kingdom *Population* Convenience sample of patients with chronic pain *Inclusion criteria* Pain episode lasting >12 weeks *Exclusion criteria* Not reported *Prevalence* MDD: 72% (26/36)	*Index test* BDI-II *Reference test* SCID-I *Number of patients* n=36 (64% female) Mean age: 47.8 years (SD 12.85) *Drop-out rate* Not reported	*Rater of reference test* Two trainee clinical psychologists with experience *Training of reference test* Adequate *Interobserver reliability* κ 1.0	*Major depressive disorder* Sensitivity: 100% Specificity: 50%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Not reported

BDI-II = Beck depression inventory II; CI = Confidence interval; MDD = Major depressive disorder; NA = Not applicable; PRIME-MD = Primary health care evaluation of mental disorders; SCAN = Schedules for clinical assessment in neuropsychiatry; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SD = Standard deviation

Table 3.5.21 *Sensitivity and specificity of CES-D with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Caracciolo et al 2002 [49] Italy	*Design* Cross sectional *Setting* Rehabilitation centre *Population* Consecutive patients with orthopedic (OP) and neurological (NP) disorders *Inclusion criteria* Ability to comply with CES-D and ≥17 on the Mini Mental State Examination (MMSE) *Exclusion criteria* Dementia, aphasia, severely ill *Prevalence of MDD* OP: 12% NP: 22%	*Index test* CES-D *Reference test* SCID-I *Number of patients* OP n=101 (68% females) Mean age: 70 years (61–77) NP n=50 (52% females) Mean age: 67 years (50–73) *Drop-out rate* Not reported	*Rater* 2 psychologists *Rater training* Certified psychologist *Interobserver reliability* Not reported	*OP* Sensitivity: 100% Specificity: 57% (95% CI, 48; 67) *NP* Sensitivity: 100% Specificity: 36% (95% CI, 23; 49)	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Not reported

The table continues on the next page

Table 3.5.21 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Breslau et al 1985 [50] USA	*Design* Cross sectional *Setting* 4 pediatric specialties clinics in Cleveland, Ohio *Population* n=332 mothers to children and adolescents with disabilities (CF, CP, myelodysplasia or multiple handicaps) *Prevalence MDD* 5.2%	*Index test* CES-D, self-administered during interview *Reference test* DIS *Number of participants* n=319 came to interview complete data from 308 Mean age: 42 years *Drop-out rate* 3.4%	*Rater* Lay female interviewers *Rater training* "Trained" *Inter observer reliability* κ 0.75 (6 months retest, personal vs telephone)	*Major depressive disorder* Sensitivity: 87.5% Specificity: 73%	Moderate *Comments* *Sampling method* Adequate *Blinding* Not reported *Handling of missing data* Not reported

CES-D = Center for epidemiologic studies depression scale; CI = Confidence inerval; DIS = Diagnostic interview schedule; MDD = Major depressive disorder; NP = Neurological disorder; OP = Orthopedic disorder; SCID-I = Structured clinical interview for DSM-IV axis I disorders

Table 3.5.22 *Sensitivity and specificity of EPDS with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Eberhard-Gran et al 2001 [58] Norway	*Design* Cross sectional, stratified samples for EPDS <10 and ≥10 *Setting* Two community-based child health clinics *Population* n=362 women 6 weeks after delivery *Exclusion criteria* Not speaking or reading Norwegian language *Prevalence MDD* 16%	*Index test* EPDS *Reference test* PRIME-MD *Number of patients* n=57 EPDS score ≥10: n=25 EPDS <10: n=31 *Drop-out rate* 2%	*Rater* Three experienced general practitioners and one psychiatrist *Rater training* Unclear *Interobserver reliability* Test-retest correlation: 0.74 (n=56) Interrater reliability between the clinical interview and the audio-tape observation was 1.0 (Cohens´s κ, n=10)	*For MDD cut off ≥12* Sensitivity: 56% Specificity: 81%	High *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Yes

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Phillips et al 2009 [60] Australia	*Design* Cross sectional *Setting* Inpatients *Population* n=413 women The first 170 were asked to participate in an interview *Inclusion criteria* Mothers with complex early parenting difficulties *Exclusion criteria* Language issues *Prevalence* MDD: 25% (42/166)	*Index test* EPDS *Reference test* SCID-I *Number of patients* n=166 Mean maternal age: 31.9 years (17–44) Mean infants age: 5.4 months (1 week–12 months) *Drop-out rate* 2%	*Rater* Psychologist *Rater training* Extensive training in diagnostic interviewing *Interobserver reliability* Only one rater	*MDD cut off 12* Sensitivity: 79% Specificity: 75%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Beck et al 2001 [61] USA	*Design* Cross sectional *Setting* Research *Population* Women from childbirth classes (n=122) and newspaper advertisement recruitment (n=28) who had delivered a live healthy infant *Inclusion criteria* ≥18 years, 2 and 12 weeks post partum *Exclusion criteria* Not fluent in English language *Prevalence postnatal MDD* 18/150 (12%)	*Index test* EPDS *Reference test* SCID-I mood disorder section *Number of patients* n=150 women Mean age: 31 years (SD 4.82), range 18 to 46 years *Drop-out rate* Unclear	*Rater* Psychotherapist nurse *Rater training* Unclear *Interobserver reliability* Only one rater	*MDD cut off 12* Sensitivity: 78% Specificity: 99%	Moderate *Comments* *Sampling method* Unclear *Blinding* Yes *Handling of missing data* Unclear
Bunevicius et al 2009 [62] Lithuania	*Design* Cross sectional *Setting* An obstetric clinic *Population* All 307 pregnant women visiting the clinic were asked about participation *Prevalence MDD* 3% (second and third trimester) 5% (first trimester)	*Index test* EDS *Reference test* SCID-I-NP *Number of patients* n=230 Mean age: 29±5 years *Drop-out rate* Not reported	*Rater* Psychiatrist *Rater training* Trained but not specified *Interobserver reliability* Only one rater	*MDD cut off ≥12* *First trimester* Sensitivity: 92% Specificity: 95% *Second trimester* Sensitivity: 67% Specificity: 92% *Third trimester* Sensitivity: 63% Specificity: 92%	Moderate *Comments* *Sampling method* All pregnant women consecutively invited, 307 agreed to participate and 230 completed all assessments *Blinding* Yes *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Aydin et al 2004 [63] Turkey	*Design* Cross sectional *Setting* Primary health care clinics *Population* 1 750 women within their first post partum year, of whom 352 attended a primary health care clinic and were invited to participate, of whom 241 participated *Exclusion criteria* Psychiatric treatment history *Prevalence postnatal MDD* 14%	*Index test* EPDS *Reference test* SCID-I *Number of patients* n=341 women Mean age: 26.6±4.8 years *Drop-out rate* Not reported	*Rater* A mental health professional *Rater training* Not reported *Interobserver reliability* Not reported	*MDD, cut off 12.5* Sensitivity: 75.5% (95% CI, 71; 79) Specificity: 71.5% (95% CI, 67; 76)	Moderate *Comments* *Sampling method* Adequate, consecutively included *Blinding* Yes *Handling of missing data* Not reported

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Garcia-Esteve et al 2003 [64] Spain	*Design* Cross sectional with stratification (probable cases and control) *Setting* A public Maternity Hospital of Barcelona *Population* Women attending in a routine postnatal check-up at 6 week after delivery n=1 201 *Exclusion criteria* Language issues *Prevalence* MDD + minor depression: 30% (100/334)	*Index test* EPDS *Reference test* SCID-I-NP *Number of patients* n=334 *Drop-out rate* 16%	*Rater* First author *Rater training* Adequate *Interobserver reliability* Not reported	*MDD cut off 12/13* Sensitivity: 86% Specificity: 95%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Berle et al 2003 [65] Norway	*Design* Cross sectional *Setting* Routine post natal visits *Population* Women 6–12 weeks post partum n=411 screened for EPDS *Inclusion criteria* All women with EPDS ≥8 or higher and every tenth women who scored below *Prevalence* MDD + minor depression: 10% with (41/411)	*Index test* EPDS *Reference test* MINI *Number of patients* n=100 *Drop-out rate* Not reported	*Rater* Psychiatrist *Rater training* Not described *Interobserver reliability* κ 0.82, 0.84 and 0.78 between rater pairs	*MDD cut off 12* Sensitivity: 85% Specificity: 86%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Adouard et al 2005 [66] France	*Design* Cross sectional *Setting* Maternity hospital ultrasound visit *Population* 66 pregnant women from a consecutive sample at a Maternity Hospital *Inclusion criteria* 28–34 weeks´ gestation and high-risk pregnancy Maternal age: ≥18 years *Exclusion criteria* Current psychotic pathology, inability to read or speak French and current obstetrical hospitalization *Prevalence* MDD: 25% (15/60)	*Index test* EPDS *Reference test* MINI *Number of patients* n=60 women Mean age: 31.5±4.8 years *Drop-out rate* Not reported	*Rater* Psychiatrist *Rater training* Unclear *Interobserver reliability* Not reported	*MDD cut off 12.5* Sensitivity: 73% Specificity: 82%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Adequate

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Areias et al 1996 [76] Portugal	*Design* Cross sectional and longitudinal *Setting* Two antenatal clinics *Population* 80 first time consecutive mothers, beyond 24 weeks gestation at entry *Exclusion criteria* Unable to complete the questionnaires *Prevalence MDD* 16.7% during pregnancy	*Index test* EPDS *Reference test* SADS or SADS-L *Number of patients* n=54 *Drop-out rate* None	*Rater* Psychologist *Rater training* Yes, but not specified *Interobserver reliability* Only one rater	*EPDS cut off score 12 for women* Sensitivity: 35% Specificity: 96% *EPDS cut off score 12 for men* Sensitivity: 20% Specificity: 92%	Moderate *Comments* *Sampling strategy* Adequate *Blinding* Unclear *Handling of missing data* Unclear
Bunevicius et al 2009 [67] Lithuania	*Design* Cross sectional *Setting* Research *Population* Randomly selected pregnant women, from a larger study (n=307) interviewed two weeks post partum (n=94) *Inclusion criteria* Age ≥18 years *Prevalence MDD* 14%	*Index test* EPDS *Reference test* CIDI-SF *Number of patients* n=94 *Drop-out rate* None	*Rater* Psychiatrist *Rater training* Trained (no specific training) *Interobserver reliability* Only one psychiatrist made the interview	*MDD cut off ≥12* Sensitivity: 46%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Chaudron et al 2010 [68] USA	*Design* Cross sectional *Setting* Urban well-child care clinic *Population* n=647 consecutive mothers. Mothers were grouped by infants age (early, middle and late) *Inclusion criteria* Mothers age: ≥18 years Infant age: ≤14 months CES-D scores: ≥16 *Exclusion criteria* Language barriers *Prevalence MDD* 37% (73/198)	*Index test* EPDS *Reference test* SCID-I *Number of patients* n=198 Mean age: 24.3±5.0 years *Drop-out rate* Not reported	*Rater* Trained raters and the results were reviewed by a consensus team *Rater training* Unclear *Interobserver reliability* Not reported	*MDD cut off 13* Sensitivity: 55% Specificity: 91%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Cox et al 1987 [51] Scotland	*Design* Cross sectional *Setting* Community-based postnatal health service *Population* 72 mothers consecutively invited and 12 women invited from a local health clinic Mean age: 26 years *Inclusion criteria* Mothers identified by the health visitors at about 6 weeks following delivery as being potentially depressed 12 women without depressive symptoms *Prevalence MDD* 25% (21/84)	*Index test* EPDS *Reference test* SPI *Number of patients* n=84 women *Drop-out rate* Not reported	*Rater* One psychiatrist and one psychologist *Rater training* Trained in SPI *Interobserver reliability* Not reported	*MDD cut off 12/13* Sensitivity: 86% Specificity: 78%	Moderate *Comments* *Sampling method* Unclear *Blinding* Yes *Handling of missing data* Unclear
Matthey et al 2001 [69] Australia	*Design* Cross sectional *Setting* At home 6–7 weeks post partum *Population* 251 couples recruited from preparation for parenthood classes *Prevalence MDD* 10.4%	*Index test* EPDS *Reference test* DIS *Number of patients* n=238 mothers *Drop-out rate* 5%	*Rater* Research assistant *Rater training* Adequate *Interobserver reliability* Done but not reported	*MDD cut off 12.5* Sensitivity: 38% Specificity: 94% *MDD cut off 12.5 for men* Sensitivity: 43% Specificity: 98%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Adequate

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Rowe et al 2008 [70] Australia	*Design* Cross sectional *Setting* Residential early parenting centers for mothers with caretaking difficulties *Population* 185 consecutive women *Inclusion criteria* English speaking women with infants up to one year old *Prevalence MDD* 26.8%	*Index test* EPDS *Reference test* CIDI-Auto *Number of patients* n=145 Mean age: 33.3±4.3 years *Drop-out rate* 5%	*Rater* Trained interviewers *Rater training* Yes, but not specified *Interobserver reliability* Not reported	*MDD cut off score 12* Sensitivity: 76% Specificity: 39%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* 5% with missing data was excluded *Other comments* Time between reference and index test was probably in 5 days but not clearly stated
Leonardoua et al 2009 [72] Greece	*Design* Cross sectional *Setting* Two maternity hospitals *Population* 109 women of whom 95 participants were recruited on the second day post partum *Exclusion criteria* Language issues *Prevalence MDD* 5%	*Index test* EPDS *Reference test* SCID-I-NP *Number of patients* n=81 Mean age: 31.5±4.1 years *Drop-out rate* 15%	*Rater* Principal investigator *Rater training* Trained but procedure not described *Interobserver reliability* NA	*MDD cut off 12/13* Sensitivity: 80% Specificity: 99%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Unclear *Other comments* Time between the reference and the index test is unclear

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Boyce et al 1993 [71] Australian and New Zealand	*Design* Cross sectional *Setting* Research *Population* n=202 eligible women from the Mothers' Advisory Clinics and from hospital outpatients psychiatric department *Inclusion criteria* Within 6 months post partum *Exclusion criteria* Treatment for puerperal psychosis *Prevalence MDD* 9%	*Index test* EPDS *Reference test* DIS *Number of patients* n=107 women Mean age: 28.4 years (SD 4.18) *Drop-out rate* 4%	*Rater* Psychologist *Rater training* Not reported *Interobserver reliability* Not reported	*MDD cut off 12.5* Sensitivity: 100% Specificity: 96%	Moderate *Comments* *Sampling method* Adequate *Blinding* Unclear *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Leverton et al 2000 [73] United Kingdom	*Design* Cross sectional *Setting* Antenatal clinic attendees' recruitment *Population* 999 consecutive women of whom 454 were eligible *Inclusion criteria* Age between 18 and 40 years. Assessment at 3 months post partum *Exclusion criteria* Language difficulties Single parent >18 weeks pregnant >1 child *Prevalence MDD* 3% (5/199)	*Index test* EPDS *Reference test* PSE *Number of patients* n=199 women *Drop-out rate* Not reported	*Rater* Psychiatrists (unknown number) *Rater training, reference test* Adequate *Interobserver reliability* Not reported	*MDD cut off 12/13* Sensitivity: 70% Specificity: 93%	Moderate *Comments* *Sampling method* Yes but high external drop-out *Blinding* Yes *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Ji et al 2011 [74] USA	*Design* Cross sectional *Setting* Research setting: Women's Mental Health Program *Population* 708 women with a lifetime history of mental illness, participating in one of two prospective longitudinal perinatal investigations *Inclusion criteria* <28 weeks of gestation evaluated at 4–6 week intervals across pregnancy and through 26 weeks post partum *Prevalence MDD* First trimester: 15% (24/156) Late post partum: 10% (50/497)	*Index test* EPDS *Reference test* SCID-I, mood module *Number of patients* n=534, varies between different trimesters *Drop-out rate* 24%	*Rater* Research interviewers; no information of level of competence or number of raters *Rater training* Not described *Interobserver reliability* Quarterly interrater reliability assessments conducted throughout the course, κ >0.8 on all instruments	*MDD cut off 12/13* *First trimester* Sensitivity: 83% Specificity: 81% *Late post partum* Sensitivity: 76% Specificity: 85%	Moderate *Comments* *Sampling method* Probably *Blinding* Not reported *Handling of missing data* Exclusion of patients with missing data *Other comments* Drop-out analysis reported

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Bergink et al 2011 [75] The Netherlands	*Design* Cross sectional *Setting* Research setting *Population* 1 507 pregnant women consecutively recruited to a thyroid screening study. Assessment at 12, 24 and 36 weeks' gestation. 1 085 women agreed to participate *Inclusion criteria* Caucasian Appropriate knowledge of the Dutch language *Prevalence MDD* 5.6% 12 weeks' gestation 5.4% 24 weeks' gestation 3.4% 36 weeks' gestation	*Index test* EPDS *Reference test* CIDI *Number of patients* n=845 *Drop-out rate* 22%	*Rater* One midwife and five experienced psychology students *Rater training* Extensive CIDI training *Interobserver reliability* Not reported	*MDD cut off 12* *12 weeks* Sensitivity: 55% Specificity: 98% *24 weeks* Sensitivity: 48% Specificity: 98% *36 weeks* Sensitivity: 45% Specificity: 97%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Questionnaires with missing data were excluded (n=127). There were 113 women lost at follow-up and their data were excluded

The table continues on the next page

Table 3.5.22 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Navarro et al 2007 [59] Spain	*Design* Cross sectional *Setting* Clinical *Population* 1 591 women consecutively visiting for routine postnatal follow up at 6 weeks after delivery. A stratified subsample of 428 women was drawn *Exclusion criteria* Illiterate women Dead newborn *Prevalence MDD* 28% (115/405)	*Index test* EPDS *Reference test* SCID-I-NP *Number of patients* n=405 *Drop-out rate* 5% (23/428)	*Rater* Two senior clinicians (one psychiatrist and one psychologist) *Rater training* Trained, and before initiation of the study they both assessed 40 post partum women *Interobserver reliability* κ 0.80 for mood disorders	*MDD cut off 12/13* Sensitivity: 66% Specificity: 96%	High *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Unclear

CES-D = Center for epidemiologic studies depression scale for children; CI = Confidence interval; CIDI = Composite international diagnostic interview; CIDI-Auto = Composite international diagnostic interview, computer assisted version; CIDI-SF = Composite international diagnostic interview, short form; DIS = Diagnostic interview schedule; EDS = Edinburgh depression scale; EPDS = Edinburgh post natal depression scale; MDD = Major depressive disorder; MINI = Mini international neuropsychiatric interview; n = Number of patients; NA = Not applicable; PRIME-MD = The primary care evaluation of mental disorders; PSE = Present state examination; SADS = Schedule for affective disorders and schizophrenia; SADS-L = Schedule for affective disorders and schizophrenia, lifetime version; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SCID-I-NP = Structured clinical interview for DSM-IV, non patient version; SD = Standard deviation; SPI = Standardized psychiatric interview

Table 3.5.23 *Sensitivity and specificity of HADS-D with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Sultan et al 2010 [79] France	*Design* Cross sectional *Setting* Diabetes clinic *Population* n=370 consecutive patients with type 2 diabetes *Inclusion criteria* Age: 20–75 years Diabetes diagnosis ≥1 year before test *Exclusion criteria* No major co-morbidity *Prevalence MDD* 10.5%	*Index test* HADS *Reference test* MINI *Number of patients* n=302 Mean age: 59.4±10.7 years *Drop-out rate* 1.3%	*Rater* Psychology interns *Rater training* Unclear *Interobserver reliability* Not stated	*MDD* Sensitivity: 53% (95% CI, 35; 71) Specificity: 86% (95% CI, 81; 90)	Moderate *Comments* *Sampling method* Adequate *Blinding* Adequate *Handling of missing data* Adequate

The table continues on the next page

Table 3.5.23 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Terluin et al 2009 [80] The Netherlands	*Design* Cross sectional *Setting* General practices with 70 physicians *Population* n=370 Patients on sick leave because of psychological problems *Inclusion criteria* Age: 18–60 years Sick leave <3 months *Exclusion criteria* Recognized depressive or anxiety disorder *Prevalence MDD* 49%	*Index test* HADS self rating questionnaire *Reference test* CIDI *Number of patients* n=307 (253 female) Mean age: 39.5±9.2 years *Drop-out rate* 3.9%	*Rater* 5 lay raters *Rater training* Dutch WHO CIDI training and reference centre *Interobserver reliability* Not stated	*MDD* Sensitivity: 93% Specificity: 39%	High *Comments* *Blinding* Rater blinded

The table continues on the next page

Table 3.5.23 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Whelan-Goodinson et al 2009 [81] Australia	*Design* Cross sectional *Setting* Community-based participants *Population* n=157 Patients with traumatic brain injury (TBI) *Inclusion criteria* Age 19–74 years TBI 0.5–5.5 years previously Glasgow Coma Scale 3–14 English speaking *Exclusion criteria* No previous TBI or neurological disorder *Prevalence MDD* 34%	*Index test* HADS self rating scale *Reference test* SCID-I *Number of patients* n=100 (29% female) Mean age: 37.18±14.19 years *Drop-out rate* None	*Rater* 2 doctoral students trained in psychopathology *Rater training* Trained in SCID-I administration supervised by clinical psychologist *Interobserver reliability* 12 subjects tested of other rater, r=0.92	*MDD* Sensitivity: 62% Specificity: 92%	Moderate *Comments* *Sampling method* Adequate *Blinding* Unclear

The table continues on the next page

Table 3.5.23 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Silverstone et al 1994 [85] Canada	*Design* Cross sectional *Setting* Accident and emergency departments of medical wards *Population* n=189 acutely ill patients consecutively included and still in hospital on the 7th day after injury *Inclusion criteria* Age: ≥18 years *Exclusion criteria* MMSE <23 *Prevalence MDD* 5.9%	*Index test* HADS administered during interview *Reference test* SCAN *Number of patients* n=153 *Drop-out rate* 2.5%	*Rater* Not reported *Rater training* Not reported *Inter observer reliability* Not reported	*MDD* Sensitivity: 100% Specificity: 73%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes
Hall et al 1999 [84] United Kingdom	*Design* Prospective study with a diagnostic test after 3 months *Setting* Cancer Research Campaign Interview in patients' homes *Population* Women <75 years with early breast cancer (stage I–II) *Prevalence MDD* 37.2% (99/266)	*Index test* HADS *Reference test* PSE *Number of patients* n=269 *Drop-out rate* 1%	*Rater* 3 interviewers *Rater training* "Trained" *Interobserver reliability* Tested by videotapes in 30% of interviews by independent rater κ performed but not reported	*MDD* Sensitivity: 33% Specificity: 93%	Moderate *Comments* *Blinding* Yes *Handling of missing data* Not reported

The table continues on the next page

Table 3.5.23 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Orive et al 2010 [86] Spain	*Design* Cross sectional *Setting* One hospital *Population* n=167 consecutive patients in waiting rooms to clinics for somatic disease or psychiatry units (50% on psychotropic drugs) *Inclusion criteria* Patients >18 years, attending services for a medical disorder *Exclusion criteria* Severe physical disease, cognitive deterioration, neurological or psychotic disorder compromising the ability to complete the questionnaires and patients with more than 50% incomplete questionnaires *Prevalence* 52% (35/67)	*Index test* HADS *Reference test* Clinical interview, PRIME-MD *Number of patients* n=67 completed HADS (87% women) *Drop-out rate* 1%	*Rater* Two of the authors Interviews by 10 mental health professionals blinded to the result of index test *Rater training* Evaluation of diagnosis of depression of 10 cases from a psychiatrist *Interobserver reliability* κ 0.67–1.0	*MDD* Sensitivity: 0.86% (95% CI, 0.70; 0.95) Specificity: 0.75% (95% CI, 0.57; 0.88)	High *Comments* *Sampling method* Adequate *Blinding* Yes

The table continues on the next page

Table 3.5.23 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Lowe et al 2004 [82] Germany	*Design* Cross sectional *Setting* Outpatient hospital clinic and 12 family practices *Population* Random sample of n=2 050 outpatients at visit ≥18 years with somatic disease *Prevalence MDD* 13.2%, 66/501	*Index test* HADS *Reference test* SCID-I *Number of patients* n=1 619 consented and filled in questionnaires n=528 completed SCID-I *Drop-out rate* 4.9%	*Rater* 4 trained raters, blind to the result of the questionnaires *Rater training* Reliability tested by video-tapes *Inter observer reliability* κ (SCID-I) 0.88 (0.47–1.0)	*MDD* Sensitivity: 88% (95% CI, 78; 95) Specificity: 69% (95% CI, 64; 73)	Moderate *Comments* *Sampling method* Adequate
Zoger et al 2004 [83] Sweden	*Design* Cross sectional *Setting* Audiological physician at a university hospital *Population* n=98 consecutive outpatients with tinnitus (interquartile range 30–71 months). Investigated 24 months after initial visit *Prevalence MDD* 18%	*Index test* HADS *Reference test* SCID-I *Number of patients* n=82 in study group consented (36% women) Mean age: 50 years *Drop-out rate* Not reported	*Rater* One experienced phychiatrist *Rater training* Trained in SCID-I procedure *Inter observer reliability* Not reported	*MDD* Sensitivity: 80% Specificity: 94%	Moderate *Comments* *Sampling method* Adequate

The table continues on the next page

Table 3.5.23 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Stafford et al 2007 [87] Australia	*Design* Cross sectional *Setting* One hospital *Population* n=528 consecutive patients admitted for PTCA, AMI or coronary artery bypass graft surgery. Recruitment by postal invitation and follow-up phone call 6 weeks after discharge. Assessment took place 3 months post discharge *Prevalence MDD* 4.7%	*Index test* HADS at home sent by postal mail *Reference test* MINI *Number of patients* n=193 fulfilled both tests *Drop-out rate* 15.7%	*Rater* First author did all the interviews *Rater training* Not reported *Interobserver reliability* Not reported	*MDD* Sensitivity: 45.7% Specificity: 91.8%	Moderate *Comments* *Sampling method* Adequate

AMI = Acute myocardinal infarction; CI = Confidence interval; CIDI = Composite international diagnostic interview; HADS = Hospital anxiety and depression scale; MDD = Major depressive disorder; MINI = Mini international neuropsychiatric interview; MMSE = Mini-mental state examination; PTCA = Percutaneous transluminal coronary angioplasty; SCAN = Schedules for clinical assessment in neuropsychiatry; SCID-I = Structured clinical interview for DSM-IV axis I disorders; PSE = Present state examination; PRIME-MD = Primary care evaluation of mental disorders

Table 3.5.24 *Sensitivity and specificity of, PHQ-9 with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Wittkampf et al 2009 [89] The Netherlands	*Design* Cross sectional *Setting* 23 family practitioners (FP) *Population* Patients from high risk groups (unexplained somatic complaints, frequent attendees and new mental health problems) were selected by FP from consultation lists or patient databases *Inclusion criteria* Symptoms within 3 months *Exclusion criteria* Other psychiatric diseases, somatic problems, language problems *Prevalence (inverse probability weighting, IPW)* 12.3% (9.9–15.2), 82/664	*Index test* PHQ-9 *Reference test* SCID-I Telephone interviews *Number of patients* n=664 (443 females) with PHQ-9 score >5 280 with PHQ-9 score <5 *Drop-out rate* 27% did not complete both tests 5 tests were excluded because of insufficient data	*Rater* Researchers *Rater training* Skilled professional. Ongoing supervision during study Monthly consensus *Interobserver reliability* κ 0.73, 90% agreement	*Depression (algorithm)* (Corrected with IPW) Sensitivity: 68% (95% CI, 57; 78) Specificity: 95% (95% CI, 93; 97) *Depression (cut off 10)* Sensitivity: 100% (95% CI, 92; 100) Specificity: 45% (95% CI, 30; 60)	High *Comments* *Sampling method* Adequate *Blinding* Yes *Other comments* Time between index and reference test stated as "just before" Not stated that the diagnosis was major depression

The table continues on the next page

Table 3.5.24 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Löwe et al 2004 [82] Germany	*Design* Cross sectional *Setting* Outpatient clinic in a medical hospital and 12 family practices *Population* Random sample: n=2 050 outpatients at visit ≥18 years with somatic disease *Prevalence MDD* 13.2%	*Index test* PHQ-9 *Reference test* SCID-I *Number of patients* n=1 619 consented and filled in questionnaires n=528 completed SCID-I *Drop-out rate* Missing data on 27 (5%)	*Rater* 4 trained raters psychologists, blind to the result of the questionnaires *Rater training* Reliability tested by video-tapes *Interobserver reliability* κ (SCID-I) 0.88 (0.47–1.0)	*MDD (algorithm)* Sensitivity: 83% (95% CI, 72; 91) Specificity: 90% (95% CI, 87; 93) *MDD (cut off 11)* Sensitivity: 90% (95% CI, 80; 96) Specificity: 77% (95% CI, 73; 81)	Moderate *Comments* *Sampling method* Adequate
Carballeira et al 2007 [90] Switzerland	*Design* Cross sectional *Setting* Internal medicine units at a university hospital *Population* n=1 053 consecutive patients *Inclusion criteria* Age 18–64 years, fluent in French *Exclusion criteria* Mental and physical disorders *Prevalence MDD* 11.3%	*Index test* PHQ-9 *Reference test* Psychiatric assessment using DSM-IV criteria *Number of patients* n=318 (38% female) *Drop-out rate* 8.2%	*Rater* One licensed psychologist (PHQ-9) 2 psychiatrists and 1 pre-registration psychiatrist (reference test) *Rater training* Trained before trial and ongoing evaluation during study every 20th patient *Interobserver reliability* κ 0.75 DSM-IV diagnosis	*MDD (algorithm)* Sensitivity: 50% (95% CI, 30; 70) Specificity: 86% (95% CI, 81; 91)	Moderate *Comments* *Samling methods* Adequate *Blinding* Yes *Indeterminate results* Not reported *Other comments* Sponsored by Pfizer

The table continues on the next page

Table 3.5.24 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Picardi et al 2005 [91] Italy	*Design* Cross sectional *Setting* Inpatient ward in a dermatological hospital *Population* Patients consecutively admitted to 5 of 8 inpatients' wards at the hospital *Inclusion criteria* Patients ≥18 years *Exclusion criteria* Patients with cognitive impairment or dementia *Prevalence MDD* 8.4%	*Index test* PHQ-9 *Reference test* SCID-I *Number of patients* 141 (79% of eligible) 56% women *Drop-out rate* 9.9%	*Rater* Trained mental health personnel *Rater training* Reliability not reported *Interobserver reliability* Not reported	*MDD (algorithm)* Sensitivity: 55% (95% CI, 46; 63) Specificity: 91% (95% CI, 86; 96)	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Other comments* Sponsored by Pfizer
Henkel et al 2004 [92] Germany	*Design* Cross sectional *Setting* 18 primary care units *Population* Consecutive patients who routinely presented in waiting room *Prevalence* 10.2%	*Index test* PHQ-9 *Reference test* CIDI *Number of patients* n=470 Completed CIDI *Drop-out rate* 4.7%	*Rater* 6 psychologists and 1 psychiatrist *Rater training* Yes by a designated training centre *Interobserver reliability* "High standard of inter-rater reliability"	*MDD (algorithm)* Sensitivity: 78% (95% CI, 0.66; 0.87) Specificity: 85% (95% CI, 0.81; 0.89)	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes

The table continues on the next page

Table 3.5.24 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Thekkumpurath et al 2011 [93] United Kingdom	*Design* Cross sectional *Setting* Clinics of Regional cancer centers *Population* n=4 264 consecutive patients with cancer *Inclusion criteria* ≥15 on HADS *Exclusion criteria* Patients too ill or with significant communication and cognitive difficulties *Prevalence MDD* 34.5% (270/782)	*Index test* PHQ-9 *Reference test* SCID-I over telephone *Number of patients* n=782 *Drop-out rate* Not reported	*Rater* Specially trained psychology graduates and nurses *Rater training* Weekly supervision and review of audiotapes by psychiatrist *Interobserver reliability* Not reported	*MDD (algorithm)* Sensitivity: 56% (95% CI, 0.55; 0.57) Specificity: 96% (95% CI, 0.95; 0.97) *MDD cut off 10* Sensitivity: 73% (95% CI, 68; 78) Specificity: 88% (95% CI, 87; 89)	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Not reported *Other comments* Time between tests unclear “several days”

The table continues on the next page

Table 3.5.24 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Orive et al 2010 [86] Spain	*Design* Cross sectional *Setting* One hospital *Population* n=167 Patients in waiting rooms to clinics for somatic or psychiatric diseases *Inclusion criteria* Patients >18 years, attending with a medical disorder *Exclusion criteria* Severe physical disease, cognitive deterioration, neurological or psychotic disorder compromising the ability to complete the questionnaires and patients with more than 50% incomplete questionnaires *Prevalence* 47%	*Index test* PHQ-9 *Reference test* Clinical interview, PRIME-MD *Number of patients* n=53 *Drop-out rate* None	*Rater training* Evaluation of diagnosis of depression of 10 cases from a psychiatrist *Inter observer reliability* κ 0.67–1.0	*MDD cut off 10* Sensitivity: 68% (95% CI, 0.47; 0.85) Specificity: 89% (95% CI, 0.72; 0.98)	High *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* >50% exclusion

The table continues on the next page

Table 3.5.24 *continued*

Author Year Reference Country	Study design Setting Population Inclusion & exclusion criteria Prevalence	Index test Reference test Number of patients Drop-out rate	Rater Rater training Interobserver reliability	Sensitivity Specificity	Study quality Comments
Stafford et al 2007 [87] Australia	*Design* Cross sectional *Setting* One hospital *Population* n=528 patients admitted for PTCA, AMI or coronary artery bypass graft surgery. Recruitment by postal invitation and follow-up phone call 6 weeks after discharge. Assessment 3 months post discharge *Prevalence MDD* 18.1%	*Index test* PHQ-9 at home sent by postal mail *Reference test* MINI by telephone within 2–3 days *Number of patients* n=193 fulfilled criteria, consented and returned the PHQ-9 *Drop-out rate* 15.7%	*Rater* First author did all the interviews *Rater training* Not stated *Interobserver reliability* –	*MDD (algorithm)* Sensitivity: 34% Specificity: 97% *MDD cut off 10* Sensitivity: 54% Specificity: 91%	Moderate
Spitzer et al 1999 [96] USA Kroenke et al 2001 [191] USA	*Design* Cross sectional *Setting* Waiting room for 8 primary care clinics, 62 primary care and 21 general internal medicine physicians *Population* n=3 890 patients with hypertension 25%, arthritis 11%, diabetes 8%, pulmonary disease 7% *Prevalence MDD* 7.1%	*Index test* PHQ-9 *Reference test* SCID-I by telephone *Number of patients* n=3 000 (66% females) Mean age: 46±17.2 years *Drop-out rate* 23.3%	*Rater* One PhD clinical psychologist and three senior psychiatric social workers *Rater training* Not reported *Interobserver reliability* Not reported	*MDD (algorithm)* Sensitivity: 73% (95% CI, 0.59; 0.87) Specificity: 98% (95% CI, 0.96; 1.00) *MDD cut off 10* Sensitivity: 88% (95% CI, 74; 96) Specificity: 88% (95% CI, 85; 91)	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Incomplete questionnaires excluded

The table continues on the next page

Table 3.5.24 *continued*

Author Year Reference Country	Study design Setting Population Inclusion & exclusion criteria Prevalence	Index test Reference test Number of patients Drop-out rate	Rater Rater training Interobserver reliability	Sensitivity Specificity	Study quality Comments
Persoons et al 2003 [94] Belgium	*Design* Cross sectional *Setting* The otolaryngology department of a tertiary care hospital *Population* Consecutive outpatients with dizziness *Inclusion criteria* Patients 15–75 years *Exclusion criteria* Not reading or writing Dutch, psycho-organic or psychotic disorders. Consulting for medico-legal reasons *Prevalence MDD* 16.5% (16/97)	*Index test* PHQ-9 *Reference test* MINI *Number of patients* 310 met inclusion criteria n=268 agreed and completed PHQ-9 n=97 (random sample) were interviewed *Drop-out rate* Unclear	*Rater* A trained psychologist (one of the authors) *Rater training* Not reported *Interobserver reliability* Not reported	*MDD (algorithm)* Sensitivity: 68.8% Specificity: 94.4%	Moderate

The table continues on the next page

Table 3.5.24 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Diez-Quevedo et al 2001 [95] Spain	*Design* Cross sectional *Setting* University hospitals in Spain *Population* n=15 235 medical and surgical inpatients n=9 590 randomly selected *Inclusion criteria* Patients 18–74 years *Exclusion criteria* Intensive care patients. Those admitted to psychiatry and obstetric department or department of substance abuse. Physical impairment or illiteracy or non-Spanish speaking *Prevalence MDD* 8.3%	*Index test* PHQ-9 *Reference test* SCID-I *Number of patients* n=1 003 fulfilled criteria and consented *Drop-out rate* No drop-out	*Rater* One of the authors (a clinical psychologist) interviewed all patients face-to-face 48 hours after completion of PHQ-9 *Rater training* Not reported “experienced clinical psychologist” *Interobserver reliability* NA	*MDD (algorithm)* Sensitivity: 84% (95% CI, 0.75; 0.94) Specificity: 92% (95% CI, 0.91; 0.92)	Moderate *Comments* *Blinding* Yes

AMI = Acute myocardinal infarction; CI = Confidence interval; CIDI = Composite international diagnostic interview; DSM-IV = Diagnostic and statistical manual of mental disorder, fourth edition; HADS = Hospital anxiety and depression scale; MDD = Major depressive disorder; MINI = Mini international neuropsychiatric interview; PHQ-9 = Patient health questionnaire; PRIME-MD = Primary care evaluation of mental disorders; PTCA = Percutaneous transluminal coronary angioplasty; SCID-I = Structured clinical interview for DSM-IV axis I disorders

Table 3.5.25 *Sensitivity and specificity of the Seasonal Health Questionnaire (SHQ) with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Thompson et al 2004 [101] United Kingdom	*Design* Cross sectional *Setting* Waiting room. Morning and evening clinics of two general medical practices *Population* n=1 041 consecutive patients n=803 filled the SHQ *Inclusion criteria* Age over 26 years. Patients were diagnosed (SAD, reccurent MDD, MDD, "not depressed") based on the SHQ 25 patients from SAD and MDD groups were selected for interview *Prevalence, based on SCID-I interview with selected patients* SAD: 39% (22/56)	*Index test* SHQ *Reference test* SCID-I (DSM-III-R) by telephone *Number of patients* n=75 *Drop-out rate* 25%	*Rater of reference test* One psychiatrist *Training of reference test* Trained but not specified *Interobserver reliability* NA	*SAD* Sensitivity: 59% (95% CI, 39; 77) Specificity: 97% (95% CI, 85; 99)	Low *Comments* *Sampling method* No patients were selected according to results of the index test and algorithm is not available *Blinding* Yes *Handling of missing data* Not reported

CI = Confidence interval; DSM-III-RR = Diagnostic and statistical manual of mental disorder, third edition; MDD = Major depressive disorder; NA = Not applicable; SAD = Seasonal affective disorder; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SHQ = Seasonal health questionnaire

Table 3.5.26 *Sensitivity and specificity of the Seasonal Pattern Assessment Questionnaire (SPAQ) with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Thompson et al 2004 [101] United Kingdom	*Design* Cross sectional *Setting* Waiting room. Morning and evening clinics of two general medical practices *Population* n=1 041 consecutive patients n=803 filled the SHQ *Inclusion criteria* Age over 26 years. Patients were diagnosed (SAD, reccurent MDD, MDD, "not depressed") based on the SHQ 25 patients from SAD and MDD groups were selected for interview *Prevalence, based on SCID-I interview with selected patients* SAD: 39% (8/54)	*Index test* SPAQ *Reference test* SCID-I (DSM-III-R) by telephone *Number of patients* n=75 *Drop-out rate* 28%	*Rater of reference test* One psychiatrist *Training of reference test* Trained but not how *Interobserver reliability* Not reported	*SAD* Sensitivity: 38% (95% CI, 21; 59) Specificity: 79% (95% CI, 62; 89)	Low *Comments* *Sampling method* No, patients were selected according to results of the index test and algorithm is not available *Blinding* Yes *Handling of missing data* Not reported

The table continues on the next page

Table 3.5.26 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Mersch et al 2004 [108] The Netherlands	*Design* Cross sectional, case-control study *Setting* Research at a university hospital in the Netherlands *Population* n=45 patients from the SAD treatment program n=48 depressed patients (BDI >16) n=46 non-depressed outpatients (BDI <16) n=37 controls recruited by advertisement *Prevalence, based on clinical interview with selected patients* SAD: 26%	*Index test* SPAQ *Reference test* Checklist criteria for SAD BDI cutoff 16 *Number of patients* 164 *Drop-out rate* 20%	*Rater of reference test* Experienced clinical psychologist for SAD patients and depressed patients *Training of reference test* – *Interobserver reliability* Not reported	*SAD* Sensitivity: 44% Specificity: 94%	Low *Comments* *Sampling method* No, patients were selected according to checklist and BDI scores *Blinding* Yes *Handling of missing data* Not reported *Other comments* Not reliable reference test

BDI = Beck depression inventory; CI = Confidence interval; DSM-III-RR = Diagnostic and statistical manual of mental disorder, third edition; MDD = Major depressive disorder; SAD = Seasonal affective disorder; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SHQ = Seasonal health questionnaire; SPAQ = Seasonal pattern assessment questionnaire

Table 3.5.27 *Sensitivity and specificity of Hypomania checklist (HCL-32) with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Angst et al 2005 [109] Italy & Sweden	*Design* Cross sectional *Setting* Italy: Private outpatient practice. Sweden: Affective units at two psychiatric clinics *Population* Italy: n=220 consecutive patients with putative MDD Sweden: n=258 *Inclusion criteria* BPI, BPII or MDD *Exclusion criteria* Schizoaffective disorder *Prevalence* BP: 62.4% (266/426) BPI: 24% (102/426) BPII: 38.4% (164/426) MDD: 37.6% (160/426)	*Index test* HCL-32 Cut off ≥14 *Reference test* Italy: Modified SCID-I (Benazzi & Akiskal) to increase sensitivity to BPII disorders Sweden: Semistructured interview based on DSM-IV criteria, modified with duration criteria for hypo-manic episode left out *Number of patients* Italy: 186 (62% female) Mean age: 43.2 years (SD 13.1) Sweden: 240 (59% female) Mean age: 51.8 years (SD 18.3) *Drop-out rate* Italy: 0 Sweden: 1%	*Rater of reference test* *Italy* One senior psychiatrist *Sweden* Two senior psychiatrists or trained research nurses *Training of reference test* Trained, but not specified *Interobserver reliability* Not reported	*BP* Sensitivity: 80% Specificity: 51%	Moderate *Comments* *Sampling method* Yes, but different criteria and sampling in two different settings. Referral procedure unclear in Swedish sample *Blinding* Yes *Handling of missing data* Described for the Swedish but not for the Italian sample *Other comments* Time between index test and reference test is unclear

The table continues on the next page

Table 3.5.27 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Carta et al 2006 [111] Italy	*Design* Cross sectional *Setting* One psychiatric clinic in Italy *Population* n=123 consecutive patients seeking psychiatric care, or psychiatric evaluation before medical treatment, or applying for certification of mental status (gun, driving license) *Prevalence* BP or schizoaffective: 21% (26/123)	*Index test* HCL-32 Cut off ≥14 *Reference test* SCID-I *Number of patients* n=123 (83% female) Mean age: 37.9 ±12.4 years *Drop-out rate* Not reported	*Rater of reference test* Unclear number of physicians in psychiatry *Rater training* SCID-I training, but not specified *Interobserver reliability* Not reported	*BP* Sensitivity: 73% Specificity: 69%	Moderate *Comments* *Sampling method* Adequate *Blinding* Probable *Handling of missing data* Unclear
Meyer et al 2010 [112] Germany	*Design* Cross sectional *Setting* Psychiatric clinics and outpatient clinics across Germany *Population* n=488 consecutive patients Mean age: 39.22 years (SD 13.37, range 17–76) (62.7% females) *Prevalence* BP: 32% (116/361)	*Index test* HCL-32 Cut off ≥14 *Reference test* SCID-I *Number of patients* n=361 *Drop-out rate* 26%	*Rater of reference test* Unclear number of psychologists and psychiatrists *Rater training of reference test* Trained, but not specified *Interobserver reliability* Reported only for 1/7 centres; BPI κ 0.95 BPII κ 0.89 Anxiety disorders κ 0.95	*BP* Sensitivity: 88% Specificity: 36%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Patients with missing data were excluded from analysis

BP = Bipolar disorder; BPI/II = Bipolar type I/II; DSM-IV = Diagnostic and statistical manual of mental disorder, fourth edition; HCL-32 = Hypomania/mania symtom checklist; MDD = Major depressive disorder; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SD = Standard deviation

Table 3.5.28 *Sensitivity and specificity for Mood Disorder Questionnaire (MDQ) with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Gervasoni et al 2009 [117] Switzerland	*Design* Cross sectional and longitudinal *Setting* General outpatient psychiatric clinic *Population* n=183 *Inclusion criteria* Symptoms of mood disorder, newly referred for treatment, not previously treated in specialized unit for mood disorders *Prevalence* BP: 30%	*Index test* MDQ Standard criteria *Reference test* SCID-I *Number of patients* 146 (53% females) Median age: 40 years (range 19–64) *Drop-out rate* 20%	*Rater of reference test* Psychiatrist or psychologist *Rater training* Yes but not specified how *Interobserver reliability* Not reported	*BP* Sensitivity: 63.6% Specificity: 83.3%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Unclear *Other comments* Unclear training and expertise of raters performing the SCID-I

The table continues on the next page

Table 3.5.28 *continued*

Author Year Reference Country	Study design Setting Population Inclusion & exclusion criteria Prevalence	Index test Reference test Number of patients Drop-out rate	Rater Rater training Interobserver reliability	Sensitivity Specificity	Study quality Comments
Graves et al 2007 [119] USA	*Design* Cross sectional *Setting* Waiting room for 3 primary care clinics *Population* n=1 441 patients were invited to participate in screening with MDQ and CAPS n=579 accepted n=356 had experienced at least one significant trauma *Inclusion criteria* Informed consent and at least one significant trauma *Prevalence* BP: 9% (21/228)	*Index test* MDQ Standard criteria *Reference test* SCID-I interview *Number of patients* n=228 (66% females) Mean age: 41.8 years (SD 13.5, range 18–78) *Drop-out rate* Not reported	*Rater of reference test* Experienced research staff *Rater training* Exensively trained until minimum κ of 0.75 *Interobserver reliability* Final SCID-I diagnoses determined in consensus meeting with one certified psychiatrist	*BP* Sensitivity: 61.9% Specificity: 69%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handing of missing data* Unclear *Other comments* Number of assessors not given, but highly trained
Hardoy et al 2005 [118] Italy	*Design* Cross sectional *Setting* Outpatient psychiatric clinic in Italy *Population* n=154 consecutive patients seeking psychiatric care, coming for psychiatric evaluation from general hospital, or applying for certification of mental capacities (gun, driving license) *Prevalence (SCID-I)* BP: 33.1% (51/154)	*Index test* MDQ Cut off ≥7 symptoms + unclear if other standard criteria *Reference test* SCID-I interview *Number of patients* n=154 (60% females) Mean age: 37.2±12.4 years *Drop-out rate* Not reported	*Rater of reference test* Physicians working at least three years in psychiatry *Rater training* Specific training, but not how *Interobserver reliability* Not reported	*BP* Sensitivity: 67% Specificity: 86%	Moderate *Comments* *Sampling* Adequate *Blinding* Unclear, but probable *Handling of missing data* Unclear *Other comments* Not clear if standard criteria for diagnosis of BP with MDQ

The table continues on the next page

Table 3.5.28 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Kemp et al 2008 [120] USA	*Design* Cross sectional *Setting* County jail in *Population* n=597 detainees, n=526 completed MDQ at intake, n=146 accepted interview *Prevalence* BP: 33.5% (55/164)	*Index test* MDQ; standard criteria *Reference test* MINI *Number of patients* n=164 (13% females) Mean age: 33.4 years (SD 10.9) *Drop-out rate* Not reported	*Rater of reference test* Research assistant *Rater training* Yes, certified by 100% agreement with expert in 10 interviews and maintained interrater agreement for at least 80% of individual items *Interobserver reliability* Not reported, see above	*BP* Sensitivity: 47% Specificity: 94%	Moderate *Comments* *Sampling method* Yes, it is a special population but clinically relevant during intake at jail *Blinding* Yes *Handling of missing data* Unclear *Other comments* Structured interview as reference test. Adequate training but numbers of raters not reported. Reliable analysis of non-participants
Meyer et al 2010 [112] Germany	*Design* Cross sectional *Setting* Seven psychiatric clinics and outpatient clinics across Germany *Population* n=488 consecutive patients Mean age: 39.22 years (SD 13.37, range 17–76) *Prevalence* BP: 29% (126/440)	*Index test* MDQ; standard criteria *Reference test* SCID-I *Number of patients* n=440 (62.7% females) *Drop-out rate* 10%	*Rater of reference test* Unclear number of psychologists and psychiatrists *Rater training of reference test* Trained, but not described how *Interobserver reliability* Reported only for 1/7 centers	*BP* Sensitivity: 80% Specificity: 64%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Patients with missing data were excluded from analysis

The table continues on the next page

Table 3.5.28 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Sharma et al 2011 [122] Canada	*Design* Cross sectional *Setting* Prospective study on the course of MDD and BP during pregnancy and post partum period at a perinatal clinic in a psychiatric hospital *Population* n=224 women were referred 125 completed MDQ 2–4 weeks after delivery *Inclusion criteria* At least 18 years old, history of DSM-IV diagnosis prior to index pregnancy of MDD, BPI or BPII *Exclusion criteria* Not speaking English, comorbid for a current Axis I diagnosis, major physical illness or Axis II diagnosis, having stillbirth or children with major congenital anomalies *Prevalence* BP: 46% (57/125) MDD: 54.4% (68/125)	*Index test* MDQ; standard criteria *Reference test* SCID-I *Number of patients* n=125 women Mean age: 28 years (SD 5.17) *Drop-out rate* None	*Rater of reference test* An experienced research coordinator *Rater training of reference test* Extensive training during 8 years *Interobserver reliability* Unclear	*BP* Sensitivity: 75.44% (95% CI, 62.24; 85.87) Specificity: 86.76% (95% CI, 76.36; 93.77)	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Excluded from analysis *Other comments* Information about time period between index and reference test, and the rater's competence was achieved from corresponding author

The table continues on the next page

Table 3.5.28 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Weber Rouget et al 2005 [128] Switzerland	*Design* Cross sectional and longitudinal *Setting* Outpatient psychiatric clinic in Switzerland *Population* n=96 Mean age: 45 years (SD 9.8, range 18–63) *Inclusion criteria* Patients with mood disorders *Prevalence* BP: 56.3% (54/96)	*Index test* MDQ Standard criteria *Reference test* SCID-I *Number of patients* 96 (60% females) *Drop-out rate* Not reported	*Rater of reference test* one psychiatrist and one psychologist *Rater training* Trained, but not described how *Interobserver reliability* Not reported	*BP* Sensitivity: 74.1% Specificity: 90.5%	Moderate *Comments* *Sampling method* Yes, but from specialized mood disorder clinic and selection criteria not clearly described *Blinding* Yes *Handling of missing data* Unclear
Zimmerman et al 2011 [116] USA	*Design* Cross sectional *Setting* Community-based outpatient practice affiliated with academic center; The Rhode Island MIDAS project *Population* n=773 consecutive outpatients *Prevalence* Lifetime BP: 10.6% (80/752)	*Index test* MDQ Standard criteria *Reference test* SCID-I *Number of patients* 752 (59.3% females) Mean age: 39.4 years (SD 13.9) *Drop-out rate* 3%, 21 with incomplete MDQ	*Rater of reference test* PhD level psychologists and research assistants *Rater training* Highly trained, monitored throughout project *Interobserver reliability* Reliability diagnosing BP κ 0.75	*BP* Sensitivity: 67.5% Specificity: 84.5%	High *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Patients with incomplete data not included in analysis, number reported and drop-out analysis

BP = Bipolar disorder; BPI/II = Bipolar type I/II; CAPS = Clinical administered post-traumatic stress disorder scale; CI = Confidence interval; MDD = Major depressive disorder; MDQ = Mood disorder questionnaire; MINI = Mini international neuropsychiatric interview; SCID-I = Structured clinical interview for DSM-IV axis I disorders; UP = Unipolar; SD = Standard deviation

Table 3.5.29 *Sensitivity and specificity of Mood Disorder Questionnaire (MDQ) with modified criteria with structured or semistructured interviews as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Gervasoni et al 2009 [117] Switzerland	*Design* Cross sectional *Setting* General outpatient psychiatric clinic in Geneva, Switzerland *Population* n=183 patients with symptoms of mood disorder, newly referred for treatment *Exclusion criteria* Previously treated in specialized unit for mood disorders *Prevalence* BP: 30% (44/146)	*Index test* MDQ with standard criteria or modified Geneva algorithm *Reference test* SCID-I *Number of patients* n=146 (53% females) Median age: 40 years (range 19–64) *Drop-out rate* 20%	*Rater of reference test* Psychiatrist or psychologist *Rater training* Well-trained, but not described how *Interobserver reliability* Not reported	*BP vs UP* *Standard criteria* Sensitivity: 63.6% Specificity: 83.3% *Geneva algorithm* Sensitivity: 72.7% Specificity: 78.4% *BP I vs UP* *Standard criteria* Sensitivity: 85.0% *Geneva algorithm* Sensitivity: 90.0% *BP II vs UP* *Standard criteria* Sensitivity: 45.8% *Geneva algorithm* Sensitivity: 58.3%	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.29 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Weber Rouget et al 2005 [128] Switzerland	*Design* Cross sectional *Setting* Outpatient psychiatric clinic in Switzerland *Population* n=96 patients with mood disorders (60% females) Mean age: 45 years (SD 9.8, range 18–63) *Prevalence* BP: 56.3% (54/96)	*Index test* MDQ, standard criteria or modified Geneva criteria *Reference test* SCID-I *Number of patients* n=96 *Drop-out rate* Not reported	*Rater of reference test* One psychiatrist and one psychologist *Rater training* Trained, but not described how *Interobserver reliability* Not reported	*BP vs UP* *Standard criteria* Sensitivity: 74.1% Specificity: 90.5% *Geneva algorithm* Sensitivity: 87.0% Specifivity: 85.7% *BPI vs UP* *Standard criteria* Sensitivity: 90.3% Specificity: 90.5% *Geneva algorithm* Sensitivity: 96.8% Specificity: 85.7% *BPII vs UP* *Standard criteria* Sensitivity: 52.4% Specificity: 90.5% *Geneva algorithm* Sensitivity: 76.2% Specificity: 87.5%	Moderate *Comments* *Sampling method* Yes, but from specialized mood disorder clinic and selection criteria not clearly described *Blinding* Yes *Handling of missing data* Unclear

The table continues on the next page

Table 3.5.29 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Zimmerman et al 2009 [115] USA	*Design* Cross sectional *Setting* Community-based outpatient practice affiliated with academic center; The Rhode Island MIDAS project *Population* n=534 outpatients *Prevalence* Lifetime BP: 10.8% (52/480) BPI: n=18 BPII: n=21 BPNOS: n=8 Cyclothymia: n=5	*Index test:* MDQ, standard criteria and modified Geneva algorithm *Reference test* SCID-I interview *Number of patients* n=480 (61% females Mean age: 39.5 years (SD 13.7) *Drop-out rate* 10% (54 /534)	*Rater of reference test* PhD level psychologists and research assistants *Rater training* Highly trained, monitored throughout project *Interobserver reliability* Reliability diagnosing BP κ 0.85	*BP vs non-BP* *Standard criteria* Sensitivity: 63.5% Specificity: 84.8% *Modified criteria* Sensitivity: 73.1% Specificity: 81.5% *BPI vs non-BP* *Standard criteria* Sensitivity: 61.1% *Modified criteria* Sensitivity: 72.2% *BPII vs non-BP* *Standard criteria* Sensitivity: 57% *Modified criteria* Sensitivity: 66.7%	High *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Excluded from analysis, drop-out analysis reported

BP = Bipolar disorder; BPI/II = Bipolar type I/II; BPNOS = Bipolar not otherwise specified; MDQ = Mood disorder questionnaire; SCID-I = Structured clinical interview for DSM-IV axis I disorders; UP = Unipolar

Table 3.5.30 *Which diagnoses are associated to false positive screen by Mood Disorder Questionnaire (MDQ)?*

Author **Year** **Reference nb** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Zimmerman et al 2010 [129] USA	*Design* Cross sectional *Setting* Community-based outpatient practice affiliated with academic center; The Rhode Island MIDAS project *Population* Participants in the MIDAS Project n=534 patients filled MDQ, 90% (480) had complete data Mean age: 39.5 years (SD 13.7) (61% females) *Inclusion criteria* All patients referred *Exclusion criteria* Bipolar patients *Prevalence* Lifetime BP: 10.8% (52/480)	*Index test* MDQ Standard criteria: Cut off 7 + at least moderate impairment *Reference test* SCID-I SIDP-IV interview *Number of patients* 480–52 (bipolar) = 428 *Drop-out rate* 10% (54/534)	*Rater of index test* PhD Level psychologist and research assistents *Rater training* Highly trained, monitored throughout project *Interobserver reliability* Reliability diagnosing BP κ 0.85 Reliability diagnosing dimensional BPD scores ICC=0.96	*OR BP for MDQ+ vs MDQ* 6.4 (95% CI, 2.9; 13.9)	High *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Reported with drop-out analysis *Other comments* Patients with bipolar disorder excluded in this analysis

BP = Bipolar disorder; CI = Confidence interval; ICC = Interclass correlation; MDQ = Mood disorder questionnaire; OR = Odds ratio; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SD = Standard deviation; SIDP-IV = Structured interview of DSM-IV personality

Table 3.5.31 *BDI for assessment of severity of depression with DSM-IV criteria as reference.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Correlation**	**Study quality** **Comments**
Sprinkle et al 2002 [134] USA	*Design* Cross sectional *Setting* Public university, USA *Population* n=137 students visiting the university counselling centre sample *Prevalence* Mood disorder: 48%	*Index test* BDI-II *Reference test* SCID-I *Number of patients* n=137 Mean age: 22.0 years (SD 3.44) *Drop-out rate* No information	*Rater of reference test* 9 counsellors *Training of reference test* Yes, but not described how *Interobserver reliability* Not reported *Outcome measure* Correlation between BDI-II scores and number of MDD symptoms (0–9) and degree of impairment (GAF; mild, moderate and severe)	*Correlation coefficient* r=0.83; r^2=0.69, ie, 69% explained variance	Moderate *Comments* *Sampling method* Acceptable *Blinding* Yes *Handling of missing data* Not reported

BDI-II = Beck depression inventory II; GAF = Global assessment of functioning; MDD = Major depressive disorder; ; SCID-I = Structured clinical interview for DSM-IV axis I disorders; SD = Standard deviation

Table 3.5.32 *Calgary Depression Scale for Schizophrenia (CDSS) for assessment of severity of depression with DSM-IV criteria or CGI-S as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Correlation**	**Study quality** **Comments**
Müller et al 2006 [135] Germany	*Design* Cross sectional *Setting* Psychiatric clinic in Germany *Population* n=119 inpatients (38% females) with schizophrenia Mean age: 31.9 years (SD 10.7) *Inclusion criteria* Stabilized acute schizophrenia *Prevalence (reference test)* 31% no depression 19% mild depression 31% moderate depression 19% severe depression	*Index test* CDSS *Reference test* Global 4-point depression severity rating (DEP-SEV) based on CGI-S ratings + DSM-IV criteria for MDD with modifications that severity was rated independently of psychotic features and a time frame of 1 week *Number of patients* n=119 *Drop-out rate* Not reported	*Rater of index test* 5 experienced psychiatrists *Training of index rater* Continuous rater training *Rater of reference test* Same 5 experienced psychiatrists *Training of reference test* Continuous rater training *Interobserver reliability* ICC ≥0.87 for CDSS and DEP-SEV	*CDSS vs DEP-SEV* r=0.80 ($p<0.0001$)	Low *Comments* *Sampling method* Yes, enough variation of severity of depression. However, selection criteria were not clearly described and there might be a sampling bias *Blinding* No, same rater for all assessments *Handling of missing data* NA *Other comments* The reference standard was likely to classify severity of depression correct

CDSS = Calgary depression scale for schizophrenia; CGI-S = Clinical global impressions scale; DSM-IV = Diagnostic and statistical manual of mental disorder, fourth edition; ICC = Intraclass correlation; MDD = Major depressive disorder; NA = Not applicable; SD = Standard deviation

Table 3.5.33 *IDS/QIDS for assessment of severity of depression with DSM-IV criteria or CGI-S as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Correlation**	**Study quality** **Comments**
Dunlop et al 2010 [144] USA	*Design* RCT, double-blind multicenter trial of venlafaxin ER treatment *Setting* 29 sites in the US *Population* Outpatients with recurrent depression, with MDD one month before trial and with 3 or more episodes of MDD, 2 of which within the past 5 years *Inclusion criteria* Patients >18 years HDRS-17 ≥20 at screening and ≥18 at randomisation one week prior to trial start *Test situation, period between performance* Acute-treatment phase of 10 weeks and continuation phase of 6 months Tests were administrated at 1, 2, 3, 4, 5, 6, 8 and 10 weeks during acute phase and once monthly at continuation phase	*Index test* IDS-sr and QIDS-sr *Reference tests* Clinician rating by CGI-C *Number of patients* n=1 047 acute phase n=715 continuation phase Mean age: 40 years (65% females) *Drop-out rate* –	*Rater* Treating physicians *Rater training* Unclear *Interobserver reliability* Not reported *Outcome measure* Correlations between patients and clinicians rating of MDD severity	*IDS-sr vs CGI* *Pearsons correlation* Baseline: 0.27 Week 10: 0.67 Month 6: 0.65 *QIDS-sr vs CGI* *Pearsons correlation* Baseline: 0.24 Week 10: 0.63 Month 6: 0.64	Low *Comments* *Sampling method* *Blinding* Unclear *Other* Time between tests unclear Secondary analysis of studies

CGI = Clinical global impression; CGI-C = Clinical global impression-severity; HDRS = Hamilton rating scale for depression; IDS-sr = Inventory of depressive symptomatology; MDD = Major depressive disorder; QIDS-sr = Quick inventory of depressive symptomatology; RCT = Randomised controlled trial

Table 3.5.34 *MADRS for assessment of severity of depression with DSM-IV criteria or CGI-S as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability**	**Correlation**	**Study quality** **Comments**
Bandelow et al 2006 [147] Several countries	*Design* Post hoc analysis of pooled data from 5 RCT *Setting* All studies in which also CGI was used *Population* Outpatients subjected to acute treatment with escitalopram for MDD *Test situation, period between performance* 8 weeks	*Index test* MADRS *Reference tests* CGI-S CGI-I *Number of patients* n=1 992 (pooled data) MDD *Drop-out rate* Not reported	*Rater* Clinicians/investigators *Rater training* Not reported *Interobserver reliability* Not reported *Outcome measure* Correlation between tests (Thresholds for response and remission)	*MADRS vs CGI-S* *Spearman rank correlation* 0.82 across all groups *MADRS vs CGI-I* *Spearman rank correlation* 0.86 across all groups	Low *Comments* Study sponsored by Lundbeck A/S, Denmark or Forest Laboratories Inc New York
Müller et al 2003 [141] Germany	*Design* Longitudinal study of MDD, continuous clinical assessment *Setting* Hospitalized patients treated according to clinical routine *Population* Inpatients with MDD *Test situation, period between performance* Tests administered at the same time of day within one week after admission or before discharge	*Index test* MADRS *Reference test* CGI *Number of patients* n=85 (69% female) Mean age: 51.4±14.6 years *Drop-out rate* Not reported	*Rater* One of five treating physicians *Rater training* "Trained" *Interobserver reliability* ICC >0.85 *Outcome measure* To assess depression severity using MADRS	*MADRS vs CGI* *Pearsons correlation coefficient* 0.87 p<0.0001	Low *Comments* *Sampling method* *Blinding* No *Other comments* Time between tests unclear

CGI = Clinical global impressions; CGI-I = Clinical global impressions scale improvement; CGI-S = Clinical global impressions scale; ICC = Intraclass correlation; MADRS = Montgomery-Åsberg depression rating scale; MDD = Major depressive disorder; RCT = Randomised controlled trial

Table 3.5.35 *AS-18 for assessment of severity of bipolar disorder with CGI-BP as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Correlation**	**Study quality** **Comments**
Adler et al 2008 [150] Sweden	*Design* Cross sectional *Setting* Affective disorder outpatient clinic and inpatient wards at a University Hospital in Sweden *Population* n=61, opportunistically recruited patients (61% females) Mean age: 44 years, (range 17–76) *Inclusion criteria* Not reported *Exclusion criteria* Not reported *Prevalence* BPI: 61% (n=37) BPII: 13% (n=8) BPNOS: 13% (n=8) MDD: 13% (n=8)	*Index test* AS-18 *Reference test* CGI-BP *Number of patients* n=61 *Drop-out rate* Not reported	*Rater of reference test* Clinicians *Training of reference test* Not reported *Interobserver reliability* Not reported *Outcome measure* Spearman correlations between depression and mania subscores of the AS-18 with CGI-BP depression and mania scores	*AS-18-Dep vs CGI-BP-D* r=0.68 *AS-18-Man vs CGI-BP-M* r=0.80 *AS-18-Dep vs CGI-BP-M* r=–0.01 *AS-18-Man vs CGI-BP-D* r=0.10	Moderate *Comments* *Sampling method* Yes, but small convenience sample and no structured diagnostic interviews were used for making diagnoses *Blinding* Yes *Handling of missing data* Not reported

AS-18 = Affective self rating scale 18; BP = Bipolar disorder; BPNOS = Bipolar not otherwise specified ; CGI = Clinical global impression; CGI-BP = Clinical global impression for Bipolar disorder; CGI-BP-M/CGI-BP-D = Clinical global impression for Bipolar disorder, modified version; MDD = Major depressive disorder

Table 3.5.36 *MAS for assessment of severity of bipolar disorder with CGI-BP as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Populatio** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Correlation**	**Study quality** **Comments**
Vieta et al 2008 [158] Spain	*Design* Cross sectional *Setting* 18 psychiatric centers across Spain *Population* n=215 consecutive outpatients with BPI or BPII on treatment as usual *Inclusion criteria* BPI or BPII, stable patients without restriction in scores, unstable patients with YMRS ≥18 and MADRS ≥22 *Exclusion criteria* – *Prevalence* Not applicable	*Index test* MAS *Reference test* CGI-BP *Number of patients* n=113 presenting with manic episode Mean age: 43.1 years *Drop-out rate* Not reported	*Rater of index test* Clinicians *Training of index rater* Not reported *Rater of reference test* Same clinicians *Training of reference test* Not reported *Interobserver reliability* ICC (95% CI): 0.89 (0.80; 0.97) (n=46) *Outcome measure* One-way ANOVAs of MAS baseline scores with CGI categories as dependent variables	*MAS vs CGI-BP* Adequate	Low *Comments* *Sampling method* Adequate *Blinding* No *Handling of missing data* Not reported *Other comments* Neither correlation coefficients or Intraclass coefficients reported. No post hoc (Eta2) analysis of ANOVA statistics

BP = Bipolar disorder; CGI-BP = Clinical global impression for Bipolar disorder; ICC = Intraclass correlation; MADRS = Montgomery-Åsberg depression rating scale; MAS = Bech-Rafaelsen mania scale; SD = Standard deviation; YMRS = Young mania rating scale

Table 3.5.37 *Young Mania Rating Scale (YMRS) for assessment of severity of bipolar disorder with CGI-BP as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Correlation**	**Study quality** **Comments**
Vieta et al 2008 [158] Spain	*Design* Cross sectional *Setting* 18 psychiatric centers across Spain *Population* n=215 consecutive outpatients with BPI and BPII on treatment as usual *Inclusion criteria* BPI or BPII, stable patients without restriction in scores, unstable patients with YMRS ≥18 and MADRS ≥22 *Exclusion criteria* – *Prevalence* Not applicable	*Index test* YMRS *Reference test* CGI-BP *Number of patients* n=113 presenting with manic episode Mean age: 43.1 years (SD 13.2), 60.7% women *Drop-out rate* Not reported	*Rater of index test* Clinicians *Training of index rater* Not reported *Rater of reference test* Same clinicians *Training of reference test* Not reported *Interobserver reliability* Not reported *Outcome measure* One-way ANOVAs of baseline YMRS scores with CGI categories as dependent variables	*YMRS vs CGI-BP* Adequate	Low *Comments* *Sampling method* Adequate *Blinding* No *Handling of missing data* Not reported *Other comments* Neither correlation coefficients or Intraclass coeffiecients reported. No post hoc (Eta2) analysis of ANOVA statistics

BP = Bipolar disorder; CGI-BP = Clinical global impression for Bipolar disorder; MADRS = Montgomery-Åsberg depression rating scale; SD = Standard deviation; YMRS = Young mania rating scale

Table 3.5.38 *LCM-p for assessment of severity of bipolar disorder with CGI-BP as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Correlation**	**Study quality** **Comments**
Meaden et al 2000 [164] USA	*Design* Cross sectional and longitudinal *Setting* Participants were recruited from community through advertisements in newsletters and websites in Illinois *Population* n=40 (70% females) Mean age: 43.5 years (SD 12.2), range 27–72 *Inclusion criteria* Current diagnosis of bipolar disorder I or II. At least 18 years old *Exclusion criteria* None	*Index test* LCM-p *Reference test* CGI-BP ratings immediately after telephone ratings of HDRS, YMRS and GAF at three occasions a month apart *Number of patients* Time 1: 35 Time 2: 36 Time 3: 32 29 patients had complete data on all occation *Drop-out rate* See above	*Rater of reference test* One researcher *Training of reference test* Not reported *Interobserver reliability* NA *Outcome measure* Spearman correlations between severity of mania with LCM-HIGH and CGI-mania, severity of depression with LCM-LOW and CGI-BP-depression, and the opposite subscales	*LCM-HIGH vs CGI-mania* Time 1: 0.318 Time 2: 0.383 Time 3: 0.556 *LCM-LOW vs CGI-depression* Time 1: 0.430 Time 2: 0.593 Time 3: 0.516 *LCM-LOW vs CGI-mania* Time 1: –0.193 Time 2: 0.054 Time 3: 0.144 *LCM-HIGH vs CGI-depression* Time 1: –0.072 Time 2: 0.019 Time 3: 0.055	Moderate *Comments* *Sampling method* Adequate *Blinding* Yes *Handling of missing data* Excluded from analysis *Other comments* Telephone interview makes rating of reference test less reliable. Rater of reference test used the structured interview guide for the HDRS (SIGH-D), however the YMRS observable item had to be excluded. It is probable that these ratings influenced ratings of CGI-BP

BP = Bipolar disorder; CGI = Clinical global impressions scale; CGI-BP = Clinical global impression for Bipolar disorder; GAF = Global assessment of functioning; HDRS = Hamilton rating scale for depression; LCM = Life-chart methodology; LCM-p = Life-chart methodology prospective; n = Number of patients; NA = Not applicable; SD = Standard deviation; YMRS = Young mania rating scale

Table 3.5.39 *Correlation between clinician and patient rating with IDS.*

Author **Year** **Ref number** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Main findings** **(Correlations, p-value)**	**Study quality** **Comments**
Biggs et al 2000 [172] USA	*Design* Cross sectional *Setting* Texas Medication Algorithm Project (TMAP), 8 public sectors *Population* Outpatients and inpatients with major depressive disorder *Test situation, period between performance* Tests administered at the same visit in variable order 2.8 and 3.8 assessments for inpatients and outpatients respectively	*Index test* IDS-SR *Reference test* IDS-C *Number of patients* n=62 (59.7% female) 28 inpatients 34 outpatients Age: 20–63 years *Drop-out rate* Not reported 2–3 ratings not accounted for	*Rater* Trained physicians in unclear whether it was the treating physician *Rater training* Trained *Interobserver reliability* ICC not reported *Outcome measure* Kendall correlation ICC between self vs clinician ratings	*Inpatients* ICC: 0.90 Kendall: 0.75 p<0.0001 *Outpatients* ICC: 0.77 Kendall: 0.59 p<0.0001 *All patients* ICC: 0.85 Kendall: 0.68 p<0.0001	Moderate Correlation only tested at baseline
Domken et al 1994 [173] United Kingdom	*Design* Cross sectional *Setting* 5 acute psychiatric wards in Newcastle *Population* n=60 Consecutively admitted patients *Inclusion criteria* MDD *Exclusion criteria* Psychosis, drug or alcohol abuse, cognitive impairment, not willing to participate *Test situation, period between performance* Self rating test first rater blinded	*Index test* IDS-SR *Reference test* IDS-C *Number of patients* 48 included *Drop-out rate* None	*Rater* Not described, but blinded *Rater training* Not described *Interobserver reliability* Not described *Outcome measure* Pearson correlation (95% CI) between self vs clinician ratings	*IDS-C vs IDS-SR* Pearson r 0.65 p<0.001	Moderate

The table continues on the next page

Table 3.5.39 *continued*

Author Year Ref number Country	Study design Setting Population Inclusion & exclusion criteria Prevalence	Index test Reference test Number of patients Drop-out rate	Rater Rater training Interobserver reliability Outcome measure	Main findings (Correlations, p-value)	Study quality Comments
Corruble et al 1999 [174] France	*Design* Cross sectional *Setting* Department of psychiatry at a hospital in Paris *Population* Consecutively admitted inpatients before treatment with antidepressants *Exclusion criteria* Bipolar and schizophrenic disorders, alcohol and drug abuse, neurological disorders, unstable medical disorders *Test situation, period between performance* 4 weeks follow-up; day 10 and 28 Clinician test first	*Index test* IDS-SR *Reference test* IDS-C *Number of patients* n=68 (68.4% females) Mean age: 41.7±12.6 years *Drop-out rate* Day 10: n=4 Day 28: n=5	*Rater* One psychiatrist *Rater training* Not stated *Outcome measure* Pearson correlations, ICC (95% CI) between self vs clinician ratings	*Day 0* Pearson r: 0.79 ICC: 0.76 (95% CI, 0.62; 0.83) *Day 10* Pearson r: 0.93 ICC: 0.91 (95% CI, 0.85; 0.94) *Day 28* Pearson r: 0.89 ICC: 0.85 (95% CI, 0.76; 0.90)	Moderate

The table continues on the next page

Table 3.5.39 *continued*

Author Year Ref number Country	Study design Setting Population Inclusion & exclusion criteria Prevalence	Index test Reference test Number of patients Drop-out rate	Rater Rater training Interobserver reliability Outcome measure	Main findings (Correlations, p-value)	Study quality Comments
Rush et al 2006 [170] USA	*Design* Cross sectional *Setting* 14 public mental health clinics Texas medication algorithm project *Population* Outpatients with psychotic and non-psychotic depression *Inclusion criteria* ≥18 years DSM-IV criteria for MDD *Exclusion criteria* Patients receiving mental retardation services or assertive community treatment *Test situation, period between performance* Tests at same time (personal communication)	*Index test* IDS-SR QIDS-16-SR *Reference test* IDS-C QIDS-16-C QIDS-16 results were extracted from the IDS tests *Number of patients* n=544 (79% female) Mean age: 42.2±11.1 years *Drop-out rate* Not reported	*Rater* Research outcome assessors independent from clinicians without knowledge of the self-rate results *Rater training* Not stated *Outcome measure* Pearson correlations, ICC between self vs clinician ratings	*IDS-30* ICC: 0.88 Pearson r: 0.89 p<0.001 *IDS-30* ICC: 0.87 Pearson r: 0.87 p<0.001	Moderate

The table continues on the next page

Table 3.5.39 *continued*

Author Year Ref number Country	Study design Setting Population Inclusion & exclusion criteria Prevalence	Index test Reference test Number of patients Drop-out rate	Rater Rater training Interobserver reliability Outcome measure	Main findings (Correlations, p-value)	Study quality Comments
Rush et al 2006 [171] USA	*Design* Cross sectional *Setting* 18 primary and 23 speciality care settings participating in the STAR*D clinical trial *Population* n=1 500 consecutive out-patients with non-psychotic major depressive disorder *Inclusion criteria* Age 18–75 years *Exclusion criteria* Patient with psychotic and compulsive disorders, bulimia nervosa, intolerance to protocol treatment and medical conditions such as seizures *Test situation, period between performance* Tests at baseline or after first treatment step Tests at same time, clinical rating by blinded assessors by telephone before collection of clinician and self rating tests	*Index test* QIDS-16-SR *Reference test* QIDS-16-C *Number of patients* Exit: n=582 (63.1% females) Mean age: 42.6±13.1 years *Drop-out rate* Not reported	*Rater* Clinical research coordi-nators independent from clinicians without knowledge of treatment *Rater training* Research outcome assessors trained and certified *Outcome measure* Pearson correlations between *self vs clinician ratings*	Pearson r: 0.89	Moderate Structured interview before instrument QIDS-sr at clinic visit, all three QIDS measures within 2 days or less

CI = Confidence interval; ICC = Intraclass correlation; IDS-C = Inventory of depressive symptomatology, clinician rated; IDS-SR = Inventory of depressive symptomatology, self report; MDD = Major depressive disorder; QIDS = Quick inventory of depressive symptomatology

Table 3.5.40 *Assessment of severity of major depression using Hamilton Rating Scale for Depression (HDRS-17).*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Main findings** **(CI=95%)**	**Study quality** **Comments**
Kørner et al 2007 [142] Denmark	*Design* Cross sectional *Setting* 3 psycho-geriatric outpatient clinics *Population* Patients with dementia Elderly controls without dementia *Inclusion criteria* Controls were relatives to patients or members of local organisations for elderly *Exclusion criteria* Major psychiatric illness or aphasic disorder *Prevalence* Depression: 67% Depression + dementia: 33%	*Index test* HDRS-17 *Reference test* CGI *Number of patients* n=145 (102 females, 78.6±6.8 years, 43 male 72.4±5.6 years) Dementia: 47 Non-dementia: 98 *Drop-out rate* Not reported	*Rater* 7 psycho-geriatricians *Rater training* Co-rating sessions prior to and during the study *Interobserver reliability* ICC: 0.98 *Outcome measure* The correlation of HDRS-17 ratings of depression	*ICC* *All* 0.85 *Non-dementia* 0.87 *Dementia* 0.78	Moderate *Comments* *Blinding* Raters blinded *Handling of missing data* Not reported

CGI = Clinical global impressions scale; CI = Confidence interval; HDRS = Hamilton depression rating scale; ICC = Intraclass correlation

Table 3.5.41 *Correlation between clinician and patient rating with Montgomery-Åsberg Depression rating scale (MADRS).*

Author **Year** **Ref number** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Main findings** **(CI=95%)**	**Study quality** **Comments**
Bondolfi et al 2010 [176] Switzerland	*Design* Cross sectional *Setting* 2 mental health care centres *Population* Outpatients with ongoing or starting treatment for mood disorders *Test situation, period between performance* Assessment at same instance, self-rate first at inclusion and after 4 weeks	*Index test* MADRS-S *Reference test* MADRS Face-to-face interview by psychiatrist before MADRS *Number of patients* n=63 (50% females) Mean age: 43.9±11.2 years *Drop-out rate* Not reported	*Rater* Psychiatrist and one senior psychiatrist *Rater training* Interrater reliability assessed through 23 videotapes ICC: 0.87 *Outcome measure* Concurrent validity correlations between MADRS and MADRS-S (Pearson's correlation and Cohen's κ)	*Concurrent validity sum of scores* *At inclusion (total score)* 0.81, p<0.001 *After 4 weeks* 0.91, p<0.001 *Agreement MADRS-S and MADRS* 0.69 (κ-statistics) *Correlation between changes of scores on both instrument* 0.71, p< 0.001	Moderate
Cunningham et al 2011 [177] Sweden	*Design* Longitudinal drug trial *Setting* General practice *Population* Patients with MDD *Inclusion criteria* Age: ≥18–70 years DSM-III-R criteria for MDD *Test situation, period between performance* Follow-up 6 and 12 months Both tests at same visit MADRS-S in waiting-room before visit	*Index test* MADRS-S *Reference test* MADRS *Number of patients* n=400 (71,8% female) Mean age: 47.2±12.3 years *Drop-out rate* At baseline n=10 At 52 weeks n=82 Complete data from MADRS-S and MADRS from 249 patients	*Rater* General practitioners rated their own patients *Rater training* Before trial co-rating meetings including videotaped interviews Interrater reliability ICC: 0.89 *Outcome measure* Agreement of patient (MADRS-S) and general practitioner (MADRS) assessment of response and remission	*ICC* *Baseline* 47% (95% CI, 0.39; 0.54) *2 weeks* 60% (95% CI, 0.53; 0.66) *4 weeks* 71% (95% CI, 0.65; 0.76) *8 weeks* 75% (95% CI, 0.70; 0.79) *12 weeks* 61% (95% CI, 0.54; 0.67) *16 weeks* 69% (95% CI, 0.63; 0.74) *20 weeks* 67% (95% CI, 0.60; 0.73) *24 weeks* 0.60% (95% CI, 0.53; 0.66) *52 weeks* 0.55% (95% CI, 0.47; 0.62)	Moderate Withdrawals from study not explained clearly *Blinding* Yes

CI = Confidence interval; DSM-III-RR = Diagnostic and statistical manual of mental disorder, third edition; ICC = Intraclass correlation; MADRS = Montgomery-Åsberg depression rating scale; MADRS-S = Montgomery-Åsberg depression rating scale, self assessment; MDD = Major depressive disorder

Referenser

1. First M, Spitzer R, Gibbon M, Williams JBW. Structured Clinical Interview for DSM-IV Axis I disorders (SCID I). New York: Biometric Research Department; 1997.

2. APA. American psychiatric Association. Diagnostic and statistical manual of mental disorders, fourth edition. Washington; 1994.

3. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. Arch Gen Psychiatry 1992;49:624-9.

4. First MB, Gibbon M, Spitzer RL, Williams JBW, Benjamin LS. Handbok SCID-I och SCID-II för DSM-IV. Svensk bearbetning av Jörgen Herlofson. Pilgrim Press; 1999.

5. Torrens M, Serrano D, Astals M, Perez-Dominguez G, Martin-Santos R. Diagnosing comorbid psychiatric disorders in substance abusers: validity of the Spanish versions of the Psychiatric Research Interview for Substance and Mental Disorders and the Structured Clinical Interview for DSM-IV. Am J Psychiatry 2004;161:1231-7.

6. Miller PR, Dasher R, Collins R, Griffiths P, Brown F. Inpatient diagnostic assessments: 1. Accuracy of structured vs. unstructured interviews. Psychiatry Res 2001;105:255-64.

7. Ramirez Basco M, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W, et al. Methods to improve diagnostic accuracy in a community mental health setting. Am J Psychiatry 2000;157:1599-605.

8. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960:37-40.

9. Endicott J, Spitzer RL. A diagnostic interview: The schedule for affective disorders and schizophrenia. Arch Gen Psychiatry 1978;35:837-44.

10. Kosten TA, Rounsaville BJ. Sensitivity of psychiatric diagnosis based on the best estimate procedure. Am J Psychiatry 1992; 149:1225-7.

11. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 1990;47:589-93.

12. Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). Int J Methods Psychiatr Res 2004;13:93-121.

13. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al. The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 1988;45:1069-77.

14. Wing JK, Cooper JE, Sartorius N. The Description and classification of psychiatric symptoms: An instruction manual for the PSE and CATEGO system. London: Cambridge University Press; 1974.

15. Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD, Regier D. Cross-cultural feasibility, reliability and

sources of variance of the composite inter-national diagnostic interview (CIDI). Br J Psychiatry 1991;159:645-53.

16. Booth BM, Kirchner JE, Hamilton G, Harrell R, Smith GR. Diagnosing depression in the medically ill: validity of a lay-administered structured diagnostic interview. J Psychiatr Res 1998;32:353-60.

17. Kessler RC, Wittchen HU, Abelson JM, Kendler KS, Knauper B, Zhao S. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US national comorbidity survey (NCS). Int J Methods Psychiatr Res 1998;7:33-55.

18. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Arch Gen Psychiatry 1981;38:381-9.

19. Hesselbrock V, Stabenau J, Hesselbrock M, Mirkin P, Meyer R. A comparison of two interview schedules. The Schedule for Affective Disorders and Schizophrenia-Lifetime and the National Institute for Mental Health Diagnostic Interview Schedule. Arch Gen Psychiatry 1982; 39:674-7.

20. Hasin DS, Grant BF. Diagnosing depressive disorders in patients with alcohol and drug problems: a comparison of the SADS-L and the DIS. J Psychiatr Res 1987;21:301-11.

21. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34-57.

22. Amorim P, Lecrubier Y, Weiller E, Hergueta T, Sheehan D. DSM-IH-R Psychotic Disorders: procedural validity of the Mini International Neuropsychiatric Interview (MINI). Concordance and causes for discordance with the CIDI. Eur Psychiatry 1998;13:26-34.

23. Jones JE, Hermann BP, Barry JJ, Gilliam F, Kanner AM, Meador KJ. Clinical assessment of Axis I psychiatric morbidity in chronic epilepsy: a multicenter investigation. J Neuropsychiatry Clin Neurosci 2005;17:172-9.

24. Sheehan D, Lecrubier Y, Harnett Sheehan K, Janavs J, Weiller E, Keskiner A, et al. The validity of the Mini International Neuropsychitric Interview (MINI) according to the SCID-P and its reliability. Eur Psychiatry 1997;12:232-41.

25. Lecrubier Y, Sheehan DV, Weiller E, Amorim P, Bonora I, Harnett Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur Psychiatry 1997;12:224-31.

26. Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV, 3rd, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. JAMA 1994; 272:1749-56.

27. Leopold KA, Ahles TA, Walch S, Amdur RJ, Mott LA, Wiegand-Packard L, et al. Prevalence of mood disorders and utility of the PRIME-MD in patients undergoing radiation therapy. Int J Radiat Oncol Biol Phys 1998;42:1105-12.

28. Loerch B, Szegedi A, Kohnen R, Benkert O. The primary care evalua-

tion of mental disorders (PRIME-MD), German version: a comparison with the CIDI. J Psychiatr Res 2000;34:211-20.

29. Dahl AA, Kruger MB, Dahl NH, Karlson H, Knorring LV, Stordal E. SPIFA-A presentation of the Structured Psychiatric Interview for General Practice. Nord J Psychiatry 2009:1-11.

30. Hedlund M, Stalenheim G, Ekselius L, Carlsson M. Diagnostic agreement between a doctor and a nurse for psychiatric disorders: a pilot study. Nord J Psychiatry 2005;59:339-42.

31. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561-71.

32. Beck AT, Steer RA, Brown GK. Manual for the BDI-II. San Antonio, TX: The psychological Corporation; 1996.

33. Homaifar BY, Brenner LA, Gutierrez PM, Harwood JF, Thompson C, Filley CM, et al. Sensitivity and specificity of the Beck Depression Inventory-II in persons with traumatic brain injury. Arch Phys Med Rehabil 2009;90:652-6.

34. Di Benedetto M, Lindner H, Hare DL, Kent S. Depression following acute coronary syndromes: a comparison between the Cardiac Depression Scale and the Beck Depression Inventory II. J Psychosom Res 2006;60:13-20.

35. Ailey SH. The sensitivity and specificity of depression screening tools among adults with intellectual disabilities. J Ment Health Res Intellect Disabil 2009;2:45-64.

36. Huffman JC, Doughty CT, Januzzi JL, Pirl WF, Smith FA, Fricchione GL. Screening for major depression in post-myocardial infarction patients: operating characteristics of the Beck Depression Inventory-II. Int J Psychiatry Med 2010;40:187-97.

37. Hopko DR, Bell JL, Armento ME, Robertson SM, Hunt MK, Wolf NJ, et al. The phenomenology and screening of clinical depression in cancer patients. J Psychosoc Oncol 2008;26:31-51.

38. De Souza J, Jones LA, Rickards H. Validation of self-report depression rating scales in Huntington's disease. Mov Disord 2010;25:91-96.

39. Dutton GR, Grothe KB, Jones GN, Whitehead D, Kendra K, Brantley PJ. Use of the Beck Depression Inventory-II with African American primary care patients. Gen Hosp Psychiatry 2004; 26:437-42.

40. Warmenhoven F, van Rijswijk E, Engels Y, Kan C, Prins J, van Weel C, et al. The Beck Depression Inventory (BDI-II) and a single screening question as screening tools for depressive disorder in Dutch advanced cancer patients. Support Care Cancer 2012;20:319-24. Epub Jan 18.

41. Poole H, White S, Blake C, Murphy P, Bramwell R. Depression in chronic pain patients: prevalence and measurement. Pain Pract 2009;9:173-80.

42. Addington D, Addington J, Maticka-Tyndale E, Joyce J. Reliability and validity of a depression rating scale for schizophrenics. Schizophr Res 1992; 6:201-8.

43. Addington D, Addington J, Maticka-Tyndale E. Assessing depression in schizophrenia: the Calgary Depression Scale. Br J Psychiatry Suppl 1993:39-44.

44. Sarro S, Duenas RM, Ramirez N, Arranz B, Martinez R, Sanchez JM, et al. Cross-cultural adaptation and validation of the Spanish version of the Calgary Depression Scale for Schizophrenia. Schizophr Res 2004;68:349-56.

45. Kim SW, Kim SJ, Yoon BH, Kim JM, Shin IS, Hwang MY, et al. Diagnostic validity of assessment scales for depression in patients with schizophrenia. Psychiatry Res 2006;144:57-63.

46. Bressan RA, Chaves AC, Shirakawa I, de Mari J. Validity study of the Brazilian version of the Calgary Depression Scale for Schizophrenia. Schizophr Res 1998; 32:41-9.

47. Chiu S, Webber MP, Zeig-Owens R, Gustave J, Lee R, Kelly KJ, et al. Validation of the Center for Epidemiologic Studies Depression Scale in screening for major depressive disorder among retired firefighters exposed to the World Trade Center disaster. J Affect Disord 2010;121:212-9.

48. McHale M, Hendrikz J, Dann F, Kenardy J. Screening for depression in patients with diabetes mellitus. Psychosom Med 2008;70:869-74.

49. Caracciolo B, Giaquinto S. Criterion validity of the center for epidemiological studies depression (CES-D) scale in a sample of rehabilitation inpatients. J Rehabil Med 2002;34:221-5.

50. Breslau N. Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES-D and results from diagnostic interviews. Psychiatry Res 1985;15:219-29.

51. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987; 150:782-6.

52. Wickberg B, Hwang CP. The Edinburgh Postnatal Depression Scale: validation on a Swedish community sample. Acta Psychiatr Scand 1996;94:181-4.

53. Hanusa BH, Scholle SH, Haskett RF, Spadaro K, Wisner KL. Screening for depression in the postpartum period: a comparison of three instruments. J Womens Health (Larchmt) 2008;17:585-96.

54. Benvenuti P, Ferrara M, Niccolai C, Valoriani V, Cox JL. The Edinburgh Postnatal Depression Scale: validation for an Italian sample. J Affect Disord 1999;53:137-41.

55. Carpiniello B, Pariante CM, Serri F, Costa G, Carta MG. Validation of the Edinburgh Postnatal Depression Scale in Italy. J Psychosom Obstet Gynaecol 1997;18:280-5.

56. Murray L, Carothers AD. The validation of the Edinburgh Post-natal Depression Scale on a community sample. Br J Psychiatry 1990;157: 288-90.

57. Murray D, Cox JL. Screening for depression during pregnancy with the Edinburgh Depression Scale (EPDS). J Reprod Infant Psychol 1990;8:99-107.

58. Eberhard-Gran M, Eskild A, Tambs K, Schei B, Opjordsmoen S. The Edinburgh

Postnatal Depression Scale: validation in a Norwegian community sample. Nord J Psychiatry 2001;55:113-7.

59. Navarro P, Ascaso C, Garcia-Esteve L, Aguado J, Torres A, Martin-Santos R. Postnatal psychiatric morbidity: a validation study of the GHQ-12 and the EPDS as screening tools. General Hospital Psychiatry 2007;29:1-7.

60. Phillips J, Charles M, Sharpe L, Matthey S. Validation of the subscales of the Edinburgh Postnatal Depression Scale in a sample of women with unsettled infants. J Affect Disord 2009;118:101-12.

61. Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. Nurs Res 2001;50: 242-50.

62. Bunevicius A, Kusminskas L, Pop VJ, Pedersen CA, Bunevicius R. Screening for antenatal depression with the Edinburgh Depression Scale. J Psychosom Obstet Gynaecol 2009;30:238-43.

63. Aydin N, Inandi T, Yigit A, Hodoglugil NN. Validation of the Turkish version of the Edinburgh Postnatal Depression Scale among women within their first postpartum year. Soc Psychiatry Psychiatr Epidemiol 2004;39:483-6.

64. Garcia-Esteve L, Ascaso C, Ojuel J, Navarro P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. J Affect Disord 2003; 75:71-6.

65. Berle JO, Aarre TF, Mykletun A, Dahl AA, Holsten F. Screening for postnatal depression. Validation of the Norwegian version of the Edinburgh Postnatal Depression Scale, and assessment of risk factors for postnatal depression. J Affect Disord 2003;76:151-6.

66. Adouard F, Glangeaud-Freudenthal NM, Golse B. Validation of the Edinburgh postnatal depression scale (EPDS) in a sample of women with high-risk pregnancies in France. Arch Womens Ment Health 2005;8:89-95.

67. Bunevicius A, Kusminskas L, Bunevicius R. Validation of the Lithuanian version of the Edinburgh Postnatal Depression Scale. Medicina (Kaunas) 2009;45: 544-8.

68. Chaudron LH, Szilagyi PG, Tang W, Anson E, Talbot NL, Wadkins HI, et al. Accuracy of depression screening tools for identifying postpartum depression among urban mothers. Pediatrics 2010;125:e609-17. Epub 2010 Feb 15.

69. Matthey S, Barnett B, Kavanagh DJ, Howie P. Validation of the Edinburgh Postnatal Depression Scale for men, and comparison of item endorsement with their partners. J Affect Disord 2001;64:175-84.

70. 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; 11:103-8.

71. Boyce PM, Stubbs J, Todd AL. The Edinburgh Postnatal Depression Scale: Validation for an Australian sample. Aust N Z J Psychiatry 1993;27:472-476.

72. Leonardoua AA, Zervas YM, Papageorgiou CC, Marks MN, Tsartsara EC, Antsaklis A, et al. Validation of the Edinburgh Postnatal Depression Scale

and prevalence of postnatal depression at two months postpartum in a sample of Greek mothers. J Reprod Infant Psychol 2009;27:28-39.

73. Leverton TJ, Elliott SA. Is the EPDS a magic wand?: 1. A comparison of the Edinburgh Postnatal Depression Scale and health visitor report as predictors of diagnosis on the Present State Examination. J Reprod Infant Psychol 2000;18:279-96.

74. Ji S, Long Q, Jeffrey Newport D, Na H, Knight B, Zach EB, et al. Validity of depression rating scales during pregnancy and the postpartum period: Impact of trimester and parity. J Psychiatr Res 2011;45:213-9. Epub 2010 Jun 9.

75. Bergink V, Kooistra L, Lambregtse-van den Berg MP, Wijnen H, Bunevicius R, van Baar A, et al. Validation of the Edinburgh Depression Scale during pregnancy. J Psychosom Res 2011;70:385-9. Epub 2010 Dec 10.

76. Areias ME, Kumar R, Barros H, Figueiredo E. Comparative incidence of depression in women and men, during pregnancy and after childbirth. Validation of the Edinburgh Postnatal Depression Scale in Portuguese mothers. Br J Psychiatry 1996;169:30-5.

77. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361-70.

78. Henderson M, Tannock C. Use of depression rating scales in chronic fatigue syndrome. J Psychosom Res 2005;59: 181-4.

79. Sultan S, Luminet O, Hartemann A. Cognitive and anxiety symptoms in screening for clinical depression in diabetes A systematic examination of diagnostic performances of the HADS and BDI-SF. J Affect Disord 2010;123:332-6.

80. Terluin B, Brouwers EP, van Marwijk HW, Verhaak PF, van der Horst HE. Detecting depressive and anxiety disorders in distressed patients in primary care; comparative diagnostic accuracy of the Four-Dimensional Symptom Questionnaire (4DSQ) and the Hospital Anxiety and Depression Scale (HADS). BMC Fam Pract 2009;10:58.

81. Whelan-Goodinson R, Ponsford J, Schonberger M. Validity of the Hospital Anxiety and Depression Scale to assess depression and anxiety following traumatic brain injury as compared with the Structured Clinical Interview for DSM-IV. J Affect Disord 2009;114:94-102.

82. Lowe B, Spitzer RL, Grafe K, Kroenke K, Quenter A, Zipfel S, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. J Affect Disord 2004;78:131-40.

83. Zoger S, Svedlund J, Holgers KM. The Hospital Anxiety and Depression Scale (HAD) as a screening instrument in tinnitus evaluation. Int J Audiol 2004; 43:458-64.

84. Hall A, A'Hern R, Fallowfield L. Are we using appropriate self-report questionnaires for detecting anxiety and depression in women with early breast cancer? Eur J Cancer 1999;35:79-85.

85. Silverstone PH. Poor efficacy of the Hospital Anxiety and Depression Scale in the diagnosis of major depressive disorder

in both medical and psychiatric patients. J Psychosom Res 1994;38:441-50.

86. Orive M, Padierna JA, Quintana JM, Las-Hayas C, Vrotsou K, Aguirre U. Detecting depression in medically ill patients: Comparative accuracy of four screening questionnaires and physicians' diagnoses in Spanish population. J Psychosom Res 2010;69:399-406.

87. Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. Gen Hosp Psychiatry 2007;29:417-24.

88. Fann JR, Bombardier CH, Dikmen S, Esselman P, Warms CA, Pelzer E, et al. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. J Head Trauma Rehabil 2005;20:501-11.

89. Wittkampf K, van Ravesteijn H, Baas K, van de Hoogen H, Schene A, Bindels P, et al. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. Gen Hosp Psychiatry 2009;31: 451-9.

90. Carballeira Y, Dumont P, Borgacci S, Rentsch D, de Tonnac N, Archinard M, et al. Criterion validity of the French version of Patient Health Questionnaire (PHQ) in a hospital department of internal medicine. Psychol Psychother 2007;80:69-77.

91. Picardi A, Adler DA, Abeni D, Chang H, Pasquini P, Rogers WH, et al. Screening for depressive disorders in patients with skin diseases: a comparison of three screeners. Acta Derm Venereol 2005;85:414-9.

92. Henkel V, Mergl R, Kohnen R, Allgaier AK, Moller HJ, Hegerl U. Use of brief depression screening tools in primary care: consideration of heterogeneity in performance in different patient groups. Gen Hosp Psychiatry 2004;26:190-8.

93. Thekkumpurath P, Walker J, Butcher I, Hodges L, Kleiboer A, O'Connor M, et al. Screening for major depression in cancer outpatients: the diagnostic accuracy of the 9-item patient health questionnaire. Cancer 2011;117:218-27.

94. Persoons P, Luyckx K, Desloovere C, Vandenberghe J, Fischler B. Anxiety and mood disorders in otorhinolaryngology outpatients presenting with dizziness: validation of the self-administered PRIME-MD Patient Health Questionnaire and epidemiology. Gen Hosp Psychiatry 2003; 25:316-23.

95. Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. Psychosom Med 2001;63:679-86.

96. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999;282: 1737-44.

97. APA. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, third edition, revised. Washington; 1987.

98. WHO. The ICD-10 Classification of mental and Behavioural Disorders Diagnostic criteria for resarch. Geneva; 1993.

99. Thompson C, Cowan A. The Seasonal Health Questionnaire: A preliminary validation of a new instrument to screen for Seasonal Affective Disorder. J Affect Disord 2001;64:89-98.

100. Rosenthal NE, Sack DA, Gillin JC, Lewy AJ, Goodwin FK, Davenport Y, et al. Seasonal affective disorder. A description of the syndrome and preliminary findings with light therapy. Arch Gen Psychiatry 1984;41:72-80.

101. Thompson C, Thompson S, Smith R. Prevalence of seasonal affective disorder in primary care; a comparison of the seasonal health questionnaire and the seasonal pattern assessment questionnaire. J Affect Disord 2004;78:219-26.

102. Rosenthal N, Genhart MJ, Sack DA, Skewrer RG, Wehr TA. Seasonal affective disorder and its relevance for the understanding and treatment of bulimia. In Hudson JL, Pope HG editors. The psychobiology of bulimia. Washington DC: American Psychiatric Press; 1987. p205-28.

103. Steinhausen HC, Gundelfinger R, Winkler Metzke C. Prevalence of self-reported seasonal affective disorders and the validity of the seasonal pattern assessment questionnaire in young adults Findings from a Swiss community study. J Affect Disord 2009;115:347-54.

104. Young MA, Blodgett C, Reardon A. Measuring seasonality: Psychometric properties of the Seasonal Pattern Assessment Questionnaire and the Inventory for Seasonal Variation. Psychiatry Res 2003; 117:75-83.

105. Magnusson A. Validation of the Seasonal Pattern Assessment Questionnaire (SPAQ). J Affect Disord 1996;40:121-9.

106. Raheja SK, King EA, Thompson C. The Seasonal Pattern Assessment Questionnaire for identifying seasonal affective disorders. J Affect Disord 1996;41:193-9.

107. Murray G. The Seasonal Pattern Assessment Questionnaire as a measure of mood seasonality: A prospective validation study. Psychiatry Res 2003;120:53-59.

108. Mersch PPA, Vastenburg NC, Meesters Y, Bouhuys AL, Beersma DGM, van den Hoofdakker RH, et al. The reliability and validity of the Seasonal Pattern Assessment Questionnaire: A comparison between patient groups. J Affect Disord 2004;80:209-219.

109. Angst J, Adolfsson R, Benazzi F, Gamma A, Hantouche E, Meyer TD, et al. The HCL-32: Towards a self-assessment tool for hypomanic symptoms in outpatients. J Affect Disord 2005;88: 217-33.

110. Vieta E, Sanchez-Moreno J, Bulbena A, Chamorro L, Ramos JL, Artal J, et al. Cross validation with the mood disorder questionnaire (MDQ) of an instrument for the detection of hypomania in Spanish: the 32 item hypomania symptom check list (HCL-32). J Affect Disord 2007;101: 43-55.

111. Carta MG, Hardoy MC, Cadeddu M, Murru A, Campus A, Morosini PL, et al. The accuracy of the Italian version of the Hypomania Checklist (HCL-32) for the

screening of bipolar disorders and comparison with the Mood Disorder Questionnaire (MDQ) in a clinical sample. Clin Pract Epidemol Ment Health 2006;2:2.

112. Meyer TD, Bauer M. Testing the hypomania checklist-32 (HCL-32) and mood disorder questionnaire (MDQ): Going beyond samples of patients with affective disorders. J Affect Disord 2010; 122:S25.

113. Hirschfeld RM, Williams JB, Spitzer RL, Calabrese JR, Flynn L, Keck PE, Jr., et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. Am J Psychiatry 2000;157:1873-5.

114. Hirschfeld RM, Holzer C, Calabrese JR, Weissman M, Reed M, Davies M, et al. Validity of the mood disorder questionnaire: a general population study. Am J Psychiatry 2003;160:178-80.

115. Zimmerman M, Galione JN, Ruggero CJ, Chelminski I, McGlinchey JB, Dalrymple K, et al. Performance of the mood disorders questionnaire in a psychiatric outpatient setting. Bipolar Disord 2009;11:759-65.

116. Zimmerman M, Galione JN, Ruggero CJ, Chelminski I, Dalrymple K, Young D. Are screening scales for bipolar disorder good enough to be used in clinical practice? Compr Psychiatry 2011;52:600-6.

117. Gervasoni N, Weber Rouget B, Miguez M, Dubuis V, Bizzini V, Gex-Fabry M, et al. Performance of the Mood Disorder Questionnaire (MDQ) according to bipolar subtype and symptom severity. Eur Psychiatry 2009;24:341-4.

118. Hardoy MC, Cadeddu M, Murru A, Dell'Osso B, Carpiniello B, Morosini PL, et al. Validation of the Italian version of the "Mood Disorder Questionnaire" for the screening of bipolar disorders. Clin Pract Epidemol Ment Health 2005;1:8.

119. Graves RE, Alim TN, Aigbogun N, Chrishon K, Mellman TA, Charney DS, et al. Diagnosing bipolar disorder in trauma exposed primary care patients. Bipolar Disord 2007;9:318-23.

120. Kemp DE, Hirschfeld RM, Ganocy SJ, Elhaj O, Slembarski R, Bilali S, et al. Screening for bipolar disorder in a county jail at the time of criminal arrest. J Psychiatr Res 2008;42:778-86.

121. Struken referens

122. Sharma V, Xie B. Screening for postpartum bipolar disorder: Validation of the Mood Disorder Questionnaire. J Affect Disord 2011;131:408-11. Epub 2010 Dec 24.

123. de Dios C, Ezquiaga E, Garcia A, Montes JM, Avedillo C, Soler B. Usefulness of the Spanish version of the mood disorder questionnaire for screening bipolar disorder in routine clinical practice in outpatients with major depression. Clin Pract Epidemol Ment Health 2008;4:14.

124. Sanchez-Moreno J, Villagran JM, Gutierrez JR, Camacho M, Ocio S, Palao D, et al. Adaptation and validation of the Spanish version of the Mood Disorder Questionnaire for the detection of bipolar disorder. Bipolar Disord 2008;10:400-12.

125. Isometsa E, Suominen K, Mantere O, Valtonen H, Leppamaki S, Pippingskold

M, et al. The mood disorder questionnaire improves recognition of bipolar disorder in psychiatric care. BMC Psychiatry 2003;3:8.

126. Miller CJ, Klugman J, Berv DA, Rosenquist KJ, Ghaemi SN. Sensitivity and specificity of the Mood Disorder Questionnaire for detecting bipolar disorder. J Affect Disord 2004;81:167-71.

127. Twiss J, Jones S, Anderson I. Validation of the Mood Disorder Questionnaire for screening for bipolar disorder in a UK sample. J Affect Disord 2008;110:180-4.

128. Weber Rouget B, Gervasoni N, Dubuis V, Gex-Fabry M, Bondolfi G, Aubry JM. Screening for bipolar disorders using a French version of the Mood Disorder Questionnaire (MDQ). J Affect Disord 2005;88:103-8.

129. Zimmerman M, Galione JN, Ruggero CJ, Chelminski I, Young D, Dalrymple K, McGlinchey JB. Screening for Biopolar Disorder and Finding Borderline Personality Disorder. J Clin Psychiatr 2010;71: 1212-7.

130. Zimmerman M, Galione JN, Chelminski I, Young D, Dalrymple K. Psychiatric diagnoses in patients who screen positive on the Mood Disorder Questionnaire: Implications for using the scale as a case-finding instrument for bipolar disorder. Psychiatry Res 2011;185:444-9.

131. Hiroe T, Kojima M, Yamamoto I, Nojima S, Kinoshita Y, Hashimoto N, et al. Gradations of clinical severity and sensitivity to change assessed with the Beck Depression Inventory-II in Japanese patients with depression. Psychiatry Res 2005;135:229-35.

132. Steer RA, Brown GK, Beck AT, Sanderson WC. Mean Beck Depression Inventory-II scores by severity of major depressive episode. Psychol Rep 2001;88: 1075-6.

133. Svanborg P, Asberg M. A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Asberg Depression Rating Scale (MADRS). J Affect Disord 2001; 64:203-16.

134. Sprinkle SD, Lurie D, Insko SL, Atkinson G, Jones GL, Logan AR, et al. Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a university counseling center sample. J Couns Psychol 2002;49: 381-5.

135. Müller MJ, Müller K-M, Fellgiebel A. Detection of Depression in Acute Schizophrenia: Sensitivity and Specificity of 2 Standard Observer Rating Scales. Can J Psychiatry 2006;51:387-92.

136. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56-62.

137. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;6:278-96.

138. Furukawa TA, Akechi T, Azuma H, Okuyama T, Higuchi T. Evidence-based guidelines for interpretation of the Hamilton Rating Scale for Depression. J Clin Psychopharmacol 2007;27:531-4.

139. Ruhe HG, Dekker JJ, Peen J, Holman R, de Jonghe F. Clinical use of the Hamilton Depression Rating Scale: is increased efficiency possible? A post

hoc comparison of Hamilton Depression Rating Scale, Maier and Bech subscales, Clinical Global Impression, and Symptom Checklist-90 scores. Compr Psychiatry 2005;46:417-27.

140. McIntyre RS, Konarski JZ, Mancini DA, Fulton KA, Parikh SV, Grigoriadis S, et al. Measuring the severity of depression and remission in primary care: validation of the HAMD-7 scale. CMAJ 2005;173: 1327-34.

141. Muller MJ, Himmerich H, Kienzle B, Szegedi A. Differentiating moderate and severe depression using the Montgomery-Asberg depression rating scale (MADRS). J Affect Disord 2003;77:255-60.

142. Korner A, Lauritzen L, Abelskov K, Gulmann NC, Brodersen AM, Wedervang-Jensen T, et al. Rating scales for depression in the elderly: external and internal validity. J Clin Psychiatry 2007;68:384-9.

143. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger J, Burns C. The Inventory for Depressive Symptomatology (IDS): preliminary findings. Psychiatry Res 1986;18:65-87.

144. Dunlop BW, Li T, Kornstein SG, Friedman ES, Rothschild AJ, Pedersen R, et al. Correlation between patient and clinician assessments of depression severity in the PREVENT study. Psychiatry Res 2010;177:177-83.

145. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382-9.

146. Asberg M, Montgomery SA, Perris C, Schalling D, Sedvall G. A comprehensive psychopathological rating scale. Acta Psychiatr Scand Suppl 1978:5-27.

147. Bandelow B, Baldwin DS, Dolberg OT, Andersen HF, Stein DJ. What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder? J Clin Psychiatry 2006;67:1428-34.

148. Zung WW. A Self-Rating Depression Scale. Arch Gen Psychiatry 1965;12:63-70.

149. Spearing MK, Post RM, Leverich GS, Brandt D, Nolen W. Modification of the Clinical Global Impressions (CGI) Scale for use in bipolar illness (BP): the CGI-BP. Psychiatry Res 1997;73:159-71.

150. Adler M, Liberg B, Andersson S, Isacsson G, Hetta J. Development and validation of the Affective Self Rating Scale for manic, depressive, and mixed affective states. Nord J Psychiatry 2008; 62:130-5.

151. Berk M, Malhi GS, Cahill C, Carman AC, Hadzi-Pavlovic D, Hawkins MT, et al. The Bipolar Depression Rating Scale (BDRS): its development, validation and utility. Bipolar Disord 2007;9:571-9.

152. Berk M, Dodd S, Dean OM, Kohlmann K, Berk L, Malhi GS. The validity and internal structure of the Bipolar Depression Rating Scale: Data from a clinical trial of N-acetylcysteine as adjunctive therapy in bipolar disorder. Acta Neuropsychiatr 2010;22:237-42.

153. Bech P, Rafaelsen OJ, Kramp P, Bolwig TG. The mania rating scale: scale construction and inter-observer agreement. Neuropharmacology 1978;17:430-1.

154. Bech P. The Bech, Hamilton and Zung scales for mood disorders:screening and listening. A twenty years update with

reference to DSM-IV and ICD-10. 2nd edition. Berlin: Springer-Verlag; 1996.

155. Bech P. The Bech-Rafaelsen Mania Scale in clinical trials of therapies for bipolar disorder: a 20-year review of its use as an outcome measure. CNS Drugs 2002;16:47-63.

156. Licht RW, Jensen J. Validation of the Bech-Rafaelsen Mania Scale using latent structure analysis. Acta Psychiatr Scand 1997;96:367-72.

157. Bech P, Baastrup PC, de Bleeker E, Ropert R. Dimensionality, responsiveness and standardization of the Bech-Rafaelsen Mania Scale in the ultra-short therapy with antipsychotics in patients with severe manic episodes. Acta Psychiatr Scand 2001;104: 25-30.

158. Vieta E, Bobes J, Ballesteros J, Gonzalez-Pinto A, Luque A, Ibarra N, et al. Validity and reliability of the Spanish versions of the Bech-Rafaelsen's mania and melancholia scales for bipolar disorders. Acta Psychiatr Scand 2008;117:207-15.

159. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978;133:429-35.

160. Berk M, Ng F, Wang WV, Calabrese JR, Mitchell PB, Malhi GS, et al. The empirical redefinition of the psychometric criteria for remission in bipolar disorder. J Affect Disord 2008;106:153-8.

161. Favre S, Aubry JM, Gex-Fabry M, Ragama-Pardos E, McQuillan A, Bertschy G. [Translation and validation of a French version of the Young Mania Rating Scale (YMRS)]. Encephale 2003;29:499-505.

162. Roy-Byrne P, Post RM, Uhde TW, Porcu T, Davis D. The longitudinal course of recurrent affective illness: life chart data from research patients at the NIMH. Acta Psychiatr Scand Suppl 1985;317:1-34.

163. Post RM, Roy-Byrne PP, Uhde TW. Graphic representation of the life course of illness in patients with affective disorder. Am J Psychiatry 1988;145:844-8.

164. Meaden PM, Daniels RE, Zajecka J. Construct validity of life chart functioning scales for use in naturalistic studies of bipolar disorder. J Psychiatr Res 2000;34: 187-92.

165. Denicoff KD, Smith-Jackson EE, Disney ER, Suddath RL, Leverich GS, Post RM. Preliminary evidence of the reliability and validity of the prospective life-chart methodology (LCM-p). J Psychiatr Res 1997;31:593-603.

166. Denicoff KD, Leverich GS, Nolen WA, Rush AJ, McElroy SL, Keck PE, et al. Validation of the prospective NIMH-Life-Chart Method (NIMH-LCM-p) for longitudinal assessment of bipolar illness. Psychol Med 2000;30:1391-7.

167. Bernstein IH, Rush AJ, Thomas CJ, Woo A, Trivedi MH. Item Response Analysis of the Inventory of Depressive Symptomatology. Neuropsychiatr Dis Treat 2006;2:557-64.

168. Drieling T, Scharer LO, Langosch JM. The Inventory of Depressive Symptomatology: German translation and psychometric validation. Int J Methods Psychiatr Res 2007;16:230-6.

169. Doraiswamy PM, Bernstein IH, Rush AJ, Kyutoku Y, Carmody TJ,

Macleod L, et al. Diagnostic utility of the Quick Inventory of Depressive Symptomatology (QIDS-C(16) and QIDS-SR(16)) in the elderly. Acta Psychiatr Scand 2010;122:226-34.

170. Rush AJ, Carmody TJ, Ibrahim HM, Trivedi MH, Biggs MM, Shores-Wilson K, et al. Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. Psychiatr Serv 2006;57:829-37.

171. Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC, et al. An evaluation of the quick inventory of depressive symptomatology and the Hamilton rating scale for depression: A sequenced treatment alternatives to relieve depression trial report. Biol Psychiatry 2006;59:493-501.

172. Biggs MM, Shores-Wilson K, Rush AJ, Carmody TJ, Trivedi MH, Crismon ML, et al. A comparison of alternative assessments of depressive symptom severity: a pilot study. Psychiatry Res 2000; 95:55-65.

173. Domken M, Scott J, Kelly P. What factors predict discrepancies between self and observer ratings of depression? J Affect Disord 1994;31:253-9.

174. Corruble E, Legrand JM, Zvenigorowski H, Duret C, Guelfi JD. Concordance between self-report and clinician's assessment of depression. J Psychiatr Res 1999;33:457-65.

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

176. Bondolfi G, Jermann F, Rouget BW, Gex-Fabry M, McQuillan A, Dupont-Willemin A, et al. Self- and clinician-rated Montgomery-Asberg Depression Rating Scale: evaluation in clinical practice. J Affect Disord 2010;121:268-72.

177. Cunningham JL, Wernroth L, von Knorring L, Berglund L, Ekselius L. Agreement between physicians' and patients' ratings on the Montgomery-Asberg Depression Rating Scale. J Affect Disord 2011;135:148-53.

178. Kobak KA, Skodol AE, Bender DS. Diagnostic measures for adults. In: Rush AJ, First MB, Blacker D, editors. Handbook of psychiatric measures. 2nd ed. Arlington: American Psychiatric Publishing; 2008. p 39.

179. Brennan C, Worrall-Davies A, McMillan D, Gilbody S, House A. The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. J Psychosom Res 2010;69:371-8.

180. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry 2007;29:388-95.

181. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007;22: 1596-602.

182. Gibson J, McKenzie-McHarg K, Shakespeare J, Price J, Gray R. A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum

and postpartum women. Acta Psychiatr Scand 2009;119:350-64.

183. Lundh W, Gyllang C. Use of the Edinburgh Postnatal Depression Scale in some Swedish child health care centres. Scand J Caring Sci 1993;7:149-54.

184. Wickberg B, Tjus T, Hwang P. Using the EPDS in routine antenatal care in Sweden: A naturalistic study. J Reprod Infant Psychol 2005;23:33-41.

185. Rubertsson C, Borjesson K, Berglund A, Josefsson A, Sydsjo G. The Swedish validation of Edinburgh Postnatal Depression Scale (EPDS) during pregnancy. Nord J Psychiatry 2011;65:414-8. Epub 2011 Jul 5.

186. Kiejna A, Pawlowski T, Dudek D, Lojko D, Siwek M, Roczen R, et al. The utility of Mood Disorder Questionnaire for the detection of bipolar diathesis in treatment-resistant depression. J Affect Disord 2010;124:270-4. Epub 2010 Jan 13.

187. Das AK, Olfson M, Gameroff MJ, Pilowsky DJ, Blanco C, Feder A, et al. Screening for bipolar disorder in a primary care practice. JAMA 2005;293:956-63.

188. Rouillon F, Gasquet I, Garay RP, Lancrenon S. Screening for bipolar disorder in patients consulting general practitioners in France. J Affect Disord 2011;130:492-5.

189. Tohen M, Frank E, Bowden CL, Colom F, Ghaemi SN, Yatham LN, et al. The International Society for Bipolar Disorders (ISBD) Task Force report on the nomenclature of course and outcome in bipolar disorders. Bipolar Disord 2009;11:453-73.

190. Fantino B, Moore N. The self-reported Montgomery-Asberg Depression Rating Scale is a useful evaluative tool in Major Depressive Disorder. BMC Psychiatry 2009;9:26.

191. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001; 16:606-13.

192. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. Appl Psychol Meas 1977;1:385-401.

193. Bech P, Rasmussen NA, Olsen LR, Noerholm V, Abildgaard W. The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. J Affect Disord 2001;66:159-64.

3.6 Bedömningsformulär för äldre

Evidensgraderade resultat

Riktad screening av depression

- GDS-30 med tröskelvärde 11 för riktad screening av depression hos äldre i icke-psykiatrisk öppenvård och äldre personer med hjärtsjukdom har en sensitivitet på 87 procent (95 % KI, 82 till 92) med SCID-I som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕◯) och en specificitet på 76 procent (95 % KI, 72 till 80) (måttligt starkt vetenskapligt underlag ⊕⊕⊕◯).

- GDS-15 för riktad screening av depression hos personer i äldreboenden samt äldre i primärvård, med Parkinsons sjukdom eller med hjärtsjukdom har en sensitivitet på 87 procent (95 % KI, 79 till 92) vid tröskelvärdet 6 poäng och med SCID-I som referensstandard (begränsat vetenskapligt underlag ⊕⊕◯◯) och en specificitet på 71 procent (95 % KI, 66 till 75) (begränsat vetenskapligt underlag ⊕⊕◯◯).

- Det går inte att bedöma sensitivitet och specificitet för HADS som formulär för riktad screening av depression hos äldre personer pga att det finns för få studier med tillräcklig kvalitet (otillräckligt vetenskapligt underlag ⊕◯◯◯).

- Det går inte att bedöma sensitivitet för PHQ-9 som formulär för riktad screening av depression hos äldre med strukturerad eller semistrukturerad intervju som referensstandard pga att det finns för få studier med tillräcklig kvalitet (otillräckligt vetenskapligt underlag ⊕◯◯◯). Specificiteten för depression är 91 procent (95 % KI, 88 till 94) för äldre i primärvård eller med Parkinsons sjukdom med SCID-I som referensstandard (begränsat vetenskapligt underlag ⊕⊕◯◯).

- Det finns för få studier av tillräcklig kvalitet för att kunna bedöma sensitivitet och specificitet för GDS-20, CSDD, BDI-II, CES-D och SIDI som formulär för riktad screening av depression hos äldre med strukturerade eller semistrukturerade intervjuer som referensstandard (otillräckligt vetenskapligt underlag ⊕◯◯◯).

- Det går inte att bedöma om SCL, GHQ, BCDRS och CPRS-D kan användas som formulär för riktad screening av depression hos äldre med strukturerad eller semistrukturerad intervju som referensstandard eftersom det saknas studier.

Introduktion

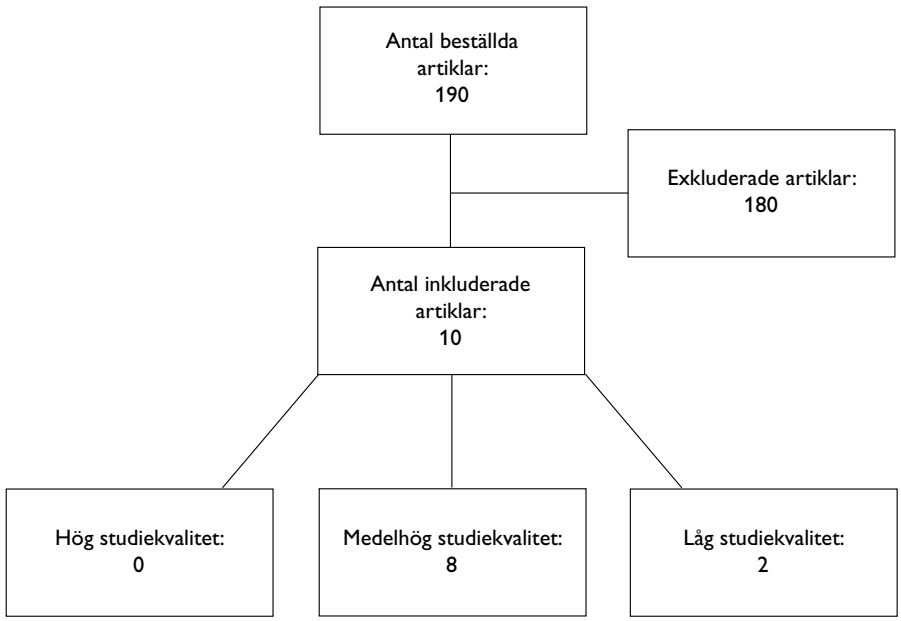
Detta avsnitt granskar det vetenskapliga underlaget för bedömningsformulär som används för äldre. Med äldre avser vi personer som är 65 år och däröver. Mellan 10 och 15 procent av äldre i västvärlden har en depression [1,2]. Depression hos äldre medför en ökad mortalitetsrisk [3–5] även om man tar hänsyn till samsjuklighet. Självmord är nästan dubbelt så vanligt hos äldre som hos befolkningen i övrigt [6]. En tillförlitlig diagnostik är därför angelägen för att äldre med depression ska få adekvat omhändertagande.

Diagnostiken försvåras av att de diagnostiska kriterierna är utvecklade för vuxna. Äldres symtom faller dock ofta i varierande grad utanför DSM- och ICD-kriterierna [7]. Symtom som t ex förändringar i stämningsläge och depressivt tankeinnehåll är inte alltid lika tydligt framträdande som hos vuxna [8,9]. Det finns en, jämfört med för vuxna, ökad samsjuklighet mellan depression och ångest hos äldre [10] och det kan ibland vara svårt att skilja mellan tillstånden. Kroppsliga och kognitiva symtom är vanliga [11,12] och kan vara mera framträdande än affektiva symtom [13].

Depression hos äldre har klara beröringspunkter med kognitiv nedsättning och demens både som riskfaktor och tidig manifestation [14–16]. Organiska förändringar och kognitiva symtom är vanliga framför allt vid sent debuterande depressioner (>60 år). Sådana depressioner kan vara uttryck för degenerativ hjärnsjukdom [17]. En konsekvens av att kriterierna inte är optimala för äldre är att det är osäkert hur tillförlitliga våra referensstandarder är. Det saknas studier som undersöker t ex SCID-I mot LEAD för äldre.

Depression hos äldre kan ha heterogena symtombilder. Det är inte ovanligt att symtomen vid depression hos äldre överensstämmer bäst med minor depression. Vi har därför valt att även inkludera studier där depressionsdiagnoserna omfattar minor depression.

Urval av studier

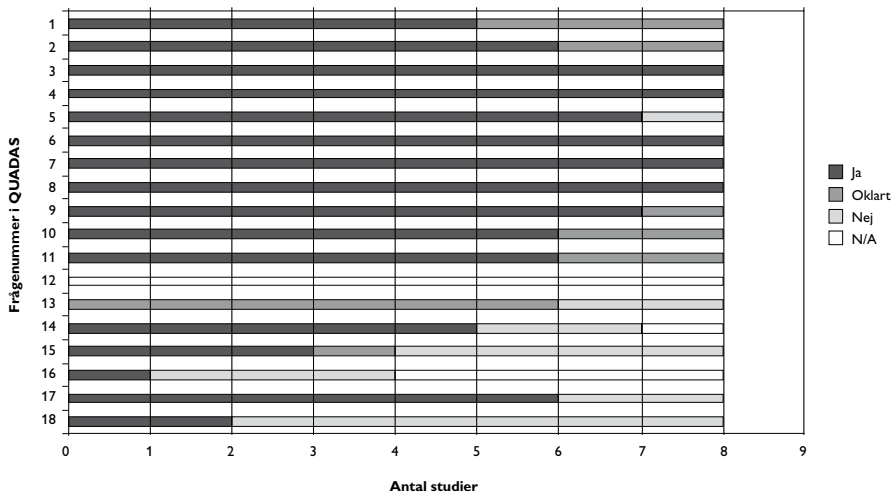


Figur 3.6.1 *Flödesschema äldre.*

Studiekvalitet

Figur 3.6.2 visar översiktligt vilka metodproblem som fanns i de åtta studierna. I figuren representeras varje rad av ett påstående (item) i vår QUADAS-mall (Kapitel 2). Svarta staplar åskådliggör hur många av de ingående studierna som uppfyllde respektive påstående. På samma sätt motsvarar längden på de grå staplarna i hur många studier det var oklart om påståendet uppfylldes. Ljusgrå staplar visar antalet studier där det inte framgick att kriteriet uppfylldes. ”NA” innebär att påståendet inte är relevant för studien. Det kan t ex handla om blindning i studier där

patienten fyllde i ett självskattningsformulär efter intervjun. Som framgår av figuren håller underlaget som helhet en god kvalitet. De största svagheterna rör frågorna 13, 15 och 16. Inga studier redovisade hur stor andel av påståendena i formulären som fylldes i korrekt av kliniker, patient eller närstående. I över hälften av studierna framgick det inte hur väl utbildade de som utförde referensintervjun var, och bara en av de granskade studierna rapporterade interbedömarreliabilitet för referensstandarden.



Figur 3.6.2 *Sammanställning av hur kvaliteten på studier bedömdes med formuläret QUADAS.*

De formulär som ingick i granskningen sammanfattas i Tabell 3.6.1.

Tabell 3.6.1 *Sammanfattning av resultat av granskningen.*

Formulär	Syfte	Referensstandard	Evidensgraderat resultat
GDS-30	Riktad screening, tröskelvärde 11	SCID-I	Sensitivitet: 87% (82; 92) Specificitet: 76% (72; 80)
GDS-15	Riktad screening, tröskelvärde 6	SCID-I	Sensitivitet: 87% (79; 92) Specificitet: 71% (66; 75)
GDS-20	Riktad screening	Semistrukturerad eller strukturerad intervju	Studier saknas
HADS	Riktad screening, tröskelvärde 6	SCID-I	Otillräckligt stöd
PHQ-9	Riktad screening, tröskelvärde 10	SCID-I	Sensitivitet: Otillräckligt stöd Specificitet: 91% (88; 94)
Cornellskalan (CSDD)	Riktad screening	Semistrukturerad eller strukturerad intervju	Otillräckligt stöd
BDI-II	Riktad screening	Semistrukturerad eller strukturerad intervju	Otillräckligt stöd
CES-D	Riktad screening	Semistrukturerad eller strukturerad intervju	Otillräckligt stöd
SIDI	Riktad screening vid stroke	Semistrukturerad eller strukturerad intervju	Otillräckligt stöd
MADRS	Bedömning av svårighetsgrad	CGI-S eller SCID-I	Studier saknas
HDRS	Bedömning av svårighetsgrad	CGI-S eller SCID-I	Otillräckligt stöd

Semistrukturerade och strukturerade diagnostiska intervjuer

Vi kunde inte identifiera några studier som undersökte om semistrukturerade eller strukturerade intervjuer är tillförlitliga för äldre, mätt som sensitivitet och specificitet, och med SCID eller LEAD som referensstandard.

Formulär för riktad screening av depression hos äldre

Syftet med granskningen var att bedöma den diagnostiska tillförlitligheten, mätt som sensitivitet och specificitet. Strukturerade och semistrukturerade intervjuer skulle utgöra referensstandard.

Geriatric Depression Scale (GDS-30)

GDS-30 är ett formulär för självbedömning [18]. Frågor om kroppsliga symtom som kan förekomma vid åldrande ingår inte. Formuläret är relativt enkelt att besvara. Det administreras skriftligt eller muntligt och består av 30 frågor som besvaras med ja eller nej. Enligt upphovsmännen talar poäng från 11 till och med 20 för mild depression och poäng däröver för medelsvår till svår depression [19].

Beskrivning av underlaget

Av 39 artiklar som lästes i fulltext exkluderades 33. En artikel kunde uteslutas av flera orsaker samtidigt. De vanligaste anledningarna till att artiklar exkluderades var att de använt annan referensstandard än strukturerade/semistrukturerade intervjuer, att det förflutit för lång tid mellan index- och referenstest eller att tidsintervallet inte redovisades. För artiklar utan tidsangivelse som var högst tio år gamla och som i övrigt var metodologiskt acceptabla kontaktade vi försteförfattaren för att få besked om tidsintervall mellan index- och referenstest. Två artiklar uteslöts då den studerade populationen inte var relevant och fyra pga att det indextest som använts var modifierat. En av de inkluderade artiklarna bedömdes ha låg studiekvalitet pga brister i beskrivningen av urvalsprocess [20].

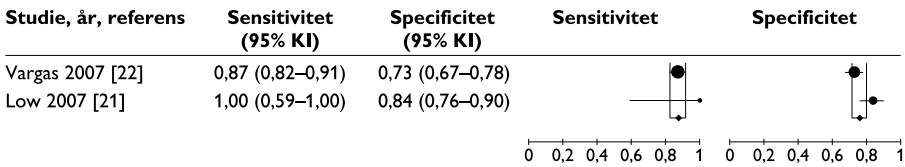
Det vetenskapliga underlaget bestod därmed av två studier som bedömdes ha medelhög kvalitet [21,22]. SCID-I [23] användes som referensstandard i båda studierna. Se Tabell 3.6.6.

I studien av Low och medarbetare ingick 119 personer med ischemisk hjärtsjukdom vid två kardiologkliniker i Kanada [21]. Medelåldern var 63 år. Personer med akut delirium, demens eller annan kognitiv nedsättning exkluderades. Studien bedömde formulärets förmåga att korrekt identifiera och utesluta personer med egentlig depression eller dubbeldiagnosen egentlig depression och dystymi.

Vargas och medarbetare rekryterade 484 personer på icke-psykiatriska öppenvårdskliniker i Brasilien [22]. Lägsta ålder för att ingå i studien var 60 år, således lägre än vårt inklusionskriterium. Medelåldern var 70 år. För icke läskunniga var tröskelvärdet för MMSE (Mini-Mental State Examination) 13, för låg- till medelutbildade 18 och för högutbildade 26 [24]. Studien bedömde formulärets förmåga att identifiera och utesluta egentlig depression. Tröskelvärdet för GDS-30 baserades på ROC-analys och var 12 poäng. Då antalet inkluderbara studier gällande GDS-30 är litet har vi valt att inkludera studien trots att den avviker något från våra inklusionskriterier eftersom den metodologiska kvaliteten är god.

Sammanvägda resultat

I de två studierna var sensitiviteten 87 respektive 100 procent och specificiteten 73 respektive 84 procent. Det sammanvägda estimatet för sensitivitet blev 87 procent (95 % KI, 82 till 92). För specificiteten blev det sammanvägda estimatet 76 procent (95 % KI, 72 till 80). Spridningen visas i nedanstående forest plot (Figur 3.6.3).



Figur 3.6.3 *Forest plot för sensitivitet och specificitet för GDS-30 vid tröskelvärde 11 och med SCID-I som referens.*

Bedömning av evidensstyrka

Det fanns brister i det vetenskapliga underlaget som påverkade evidensstyrkan. Det var oklart hur representativ populationen var i och med att en stor andel av inkluderbara personer inte deltog i en av studierna [21] och att en av studierna härrör från annan kulturkrets [22].

Tabell 3.6.2 *Resultatsammanställning för GDS-30.*

Utfallsmått	Antal studier Antal personer	Sammanvägt resultat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	2 studier n=603	87% (82; 92)	Måttligt starkt ⊕⊕⊕◯	Bristande överförbarhet (–1)
Specificitet	2 studier n=603	76% (72; 80)	Måttligt starkt ⊕⊕⊕◯	Bristande överförbarhet (–1)

KI = Konfidensintervall

Evidensgraderat resultat

GDS-30 med tröskelvärde 11 som formulär för riktad screening av depression hos äldre i icke-psykiatrisk öppenvård och äldre personer med hjärtsjukdom har en sensitivitet på 87 procent (95 % KI, 82 till 92) för att identifiera depression med SCID som referensstandard (måttligt starkt vetenskapligt underlag, ⊕⊕⊕◯) och en specificitet på 76 procent (95 % KI, 72 till 80) (måttligt starkt vetenskapligt underlag ⊕⊕⊕◯).

Geriatric Depression Scale (GDS-15)

GDS-15 är en förkortad version av GDS-30 [25]. Faktorer som trötthet och försämrad koncentration kan enligt upphovsmännen påverka resultatet, vilket ledde till att man ville ta fram ett formulär som är lättare att besvara med bibehållen fokusering. De 15 frågor som hade högst samband med depressiva symtom i tidigare valideringsstudier valdes ut för detta förkortade formulär. Administrationssättet är detsamma som för den längre versionen. Någon ROC-analys för att bestämma tröskelvärde gjordes inte i originalstudien. I de inkluderade studierna angavs tröskelvärde 5, dock utan stöd av referens.

Beskrivning av underlaget

Av 37 artiklar som lästes i fulltext exkluderades 30. De vanligaste anledningarna till att studier exkluderades var att referensstandarden inte ingick i våra inklusionskriterier (22 artiklar) och att det förflutit för lång tid mellan index- och referenstest (4 artiklar) samt att tidsintervallet inte redovisades (12 artiklar). Femton artiklar saknade såväl accepterad referensstandard som tidsangivelse mellan index och referenstest. Två artiklar uteslöts då den studerade populationen inte var relevant och fem artiklar uteslöts av andra orsaker.

Två av de sju inkluderade artiklarna bedömdes ha låg kvalitet pga brister i referensstandard och testreproducerbarhet [26] samt oklarheter om blindning respektive tidsintervall mellan index- och referenstest [27]. Det vetenskapliga underlaget bestod därmed av fem studier som bedömdes ha medelhög kvalitet. SCID-I användes som referensstandard i samtliga [23] (Tabell 3.6.7).

Davison och medarbetare rekryterade 168 personer från slumpvis utvalda australiensiska äldreboenden med medelålder 84,7 år [28]. Personer med mindre än 24 poäng på MMSE exkluderades [24]. Studien utvärderade tillförlitlighet för egentlig depression.

Haworth och medarbetare inkluderade konsekutivt 88 personer med hjärtsjukdom från öppenvårdskliniker i England med medelåldern 70 år [29]. Personer med MMSE under 24 poäng exkluderades. Studien utvärderade tillförlitlighet för någon depressionsdiagnos; vanligen egentlig depression, dystymi eller minor depression.

Phelan och medarbetare rekryterade konsekutivt 69 personer från två primärvårdsenheter i USA med medelålder 78 år [30]. Svårt dementa personer exkluderades. Studien utvärderade tillförlitlighet för egentlig depression och minor depression.

I studien av Thompson och medarbetare ingick 214 amerikanska personer med Parkinsons sjukdom med medelålder 72,5 år [31]. Personerna ingick i ett större forskningsprojekt om Parkinsons sjukdom. Personer med MMSE under 15 exkluderades. Studien utvärderade tillförlitlighet för egentlig depression och minor depression.

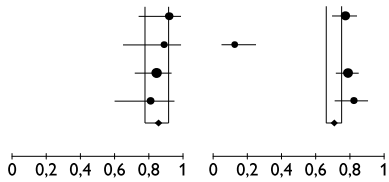
En av studierna, av Roger och medarbetare, undersökte 67 personer med stroke vid en rehabiliteringsklinik i USA med medelålder 72 år [32]. Personer med allvarliga sensoriska störningar exkluderades. Studien utvärderade tillförlitlighet för egentlig depression och minor depression. Eftersom skattningarna gjordes i genomsnitt 8 dagar efter stroke, vilket är kortare än tidskriteriet i DSM, bedömdes studien som inte relevant och ingår inte i vårt underlag.

Sammanvägda resultat

I de fyra studierna låg sensitiviteten mellan 82 och 93 procent och specificiteten mellan 13 och 83 procent [28–31]. Spridningen visas i nedanstående forest plot (Figur 3.6.4).

Den sammanvägda sensitiviteten blev 87 procent (95 % KI, 79 till 92) och den sammanvägda specificiteten 71 procent (95 % KI, 66 till 75).

Studie, år, referens	Sensitivitet (95% KI)	Specificitet (95% KI)
Davison 2009 [28]	0,93 (0,76–0,99)	0,77 (0,69–0,84)
Phelan 2010 [30]	0,89 (0,65–0,99)	0,13 (0,05–0,25)
Thompson 2011 [31]	0,85 (0,72–0,93)	0,79 (0,72–0,85)
Haworth 2007 [29]	0,82 (0,60–0,95)	0,83 (0,72–0,91)



Sammanvägd **sensitivitet** = 0,87 (0,79 till 0,92)
Sammanvägd **specificitet** = 0,71 (0,66 till 0,75)

Figur 3.6.4 *Forest plot för sensitivitet och specificitet för GDS-15 med SCID-I som referens.*

Bedömning av evidensstyrka

Våra överväganden finns sammanställda i Tabell 3.6.3. Evidensstyrkan sänktes ett steg som följd av brister i studiekvalitet. Tre studier hade en oklar redovisning av urvalsprocessen [28,29,31], en stor andel av inkluderbara personer deltog inte i tre av studierna [28,30,31], interbedömarreliabilitet redovisades inte i två studier [28,29] och blindningen var oklart beskriven i en.

För sensitivitet var konfidensintervallet för det sammanvägda estimatet relativt brett, vilket medförde att evidensstyrkan sänktes ytterligare ett steg.

Tabell 3.6.3 *Resultatsammanställning för GDS-15.*

Utfallsmått	Antal studier Antal personer	Sammanvägt resultat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	4 studier n=539	87% (79; 92)	Begränsat ⊕⊕◯◯	Bristande studiekvalitet (–1) Bristande precision (–1)
Specificitet	4 studier n=539	71% (66; 75)	Begränsat ⊕⊕◯◯	Bristande studiekvalitet (–1) Bristande samstämmighet (–1)

KI = Konfidensintervall

Evidensgraderat resultat

GDS-15 som formulär för riktad screening av depression hos personer i äldreboenden, äldre i primärvård samt personer med Parkinsons sjukdom eller hjärtsjukdom har en sensitivitet på 87 procent (95 % KI, 79 till 92) vid tröskelvärdet 6 med SCID-I som referensstandard (begränsat vetenskapligt underlag ⊕⊕◯◯) och en specificitet på 71 procent (95 % KI, 66 och 75) (begränsat vetenskapligt underlag ⊕⊕◯◯).

Geriatric Depression Scale (GDS-20)

GDS-20 är en svensk version av GDS för användning främst bland äldre i primärvård. Den har, utöver de frågor som ingår i GDS-15, fem tilläggsfrågor om sömnproblem, ångest, paniksymtom, smärta och hypokondri. Denna version utvärderades i en studie av Gottfries och medarbetare [33]. Studien exkluderades pga valet av referensstandard och att för lång tid förflutit mellan indextest och referenstest.

Evidensgraderat resultat

Det finns för få studier för att kunna bedöma om GDS-20 kan användas som formulär för riktad screening av depression hos äldre med strukturerade eller semistrukturerade intervjuer som referensstandard (otillräckligt vetenskapligt underlag ⊕◯◯◯).

Hospital Anxiety and Depression Scale (HADS)

HADS är ett formulär för självbedömning som kan användas för att identifiera depression och ångest [34]. Formuläret innehåller 14 påståenden, varav sju rör depression. Patienten anger på en fyrgradig skala hur väl påståendet överensstämmer med hur personen mått under den senaste veckan.

Formuläret togs fram för att användas i första hand inom icke-psykiatrisk sjukhusvård. Det låga antalet frågor gör att formuläret är relativt enkelt att besvara. HADS är därför användbart för äldre personer. Det etablerade tröskelvärdet är 8.

Beskrivning av underlaget

Av nio artiklar som lästes i fulltext uppfyllde två inklusionskriterierna. Övriga exkluderades. I fem av dem var tidsintervallet för långt eller inte redovisat. Tre artiklar använde inte accepterad referensstandard och en testade en modifierad form av HADS.

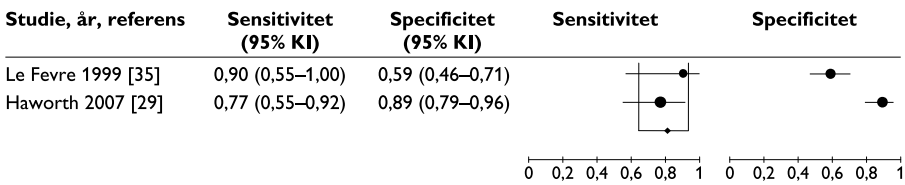
Båda studier hade medelhög studiekvalitet [29,35] (Tabell 3.6.8). Den ena var den ovan nämnda studien av Haworth och medarbetare [29]. Den andra studien, av Le Fevre och medarbetare, omfattade 79 konsekutivt rekryterade personer [35]. De vårdades palliativt vid ett hospice i England, de flesta pga cancer. Medelåldern var 68 år. Som referensstandard användes Clinical Interview Schedule [36]; en semistrukturerad intervju som modifierats för att generera DSM-III-diagnoser.

Sammanvägda resultat

Det är inte säkert att tröskelvärdet 8 är optimalt för äldre. I studien av Haworth och medarbetare var överensstämmelsen mellan HADS och SCID-I dålig [29]. Prevalensen, mätt med HADS-I, var 13 procent jämfört med 25 procent med SCID-I. Vi beslutade därför att använda tröskelvärdet 6 som föreföll vara optimalt i Haworths ROC-analys.

Sensitiviteten var 77 respektive 90 procent i de två studierna, och det sammanvägda estimatet blev 81 procent (95 % KI, 64 till 93). Specificiteten var 59 respektive 89 procent. I och med att det bara fanns två

studier bedömde vi att en sammanvägning skulle ge ett otillförlitligt resultat. Spridningen visas i sammanställningen i Figur 3.6.5.



Figur 3.6.5 *Forest plot för sensitivitet och specificitet för HADS med strukturerade/semistrukturerade intervjuer som referensstandard, vid tröskelvärde 6.*

Bedömning av evidensstyrka

Det fanns flera brister i underlaget som påverkade evidensstyrkan för sensitiviteten. Det fanns viss risk för bias pga att blindningen var oklar i en av studierna, men vi bedömde att den inte minskade tillförlitligheten påtagligt.

Överförbarheten påverkades av att det var oklart om patientspektrum var representativt (många personer som uppfyllde inklusionskriterierna deltog inte i studierna) och att studierna använde olika referensstandards. Dessutom var det få och små studier, vilket ytterligare ökade osäkerheten. Konfidensintervallet var brett, vilket drar ner evidensstyrkan ytterligare två steg.

Bedömningarna sammanfattas i nedanstående resultatsammanställning (Tabell 3.6.4).

Tabell 3.6.4 *Resultatsammanställning för HADS.*

Utfallsmått	Antal studier Antal personer	Sammanvägt estimat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poäng-avdrag
Sensitivitet	2 studier n=167	81% (64; 93%)	Otillräckligt ⊕○○○	Bristande över-förbarhet (–2) Bristande preci-sion (–1)
Specificitet	2 studier n=167	Går inte att väga samman	Otillräckligt ⊕○○○	Bristande över-förbarhet (–2) Bristande samstämmighet (–1)

Evidensgraderat resultat

Det går inte att bedöma sensitivitet och specificitet för HADS som formulär för riktad screening av depression hos äldre personer med hjärtsjukdom vid öppenvårdskliniker eller i palliativ vård pga kvalitetsbrister i underlaget (otillräckligt vetenskapligt underlag ⊕○○○).

Patient Health Questionnaire-9 (PHQ-9)

PHQ-9 är ett formulär för självbedömning med nio frågor om depression baserat på DSM-IV-kriterierna. Formuläret kan också besvaras genom att frågorna läses upp. PHQ-9 som utvecklats av Spitzer och medarbetare [37] har tagits fram och validerats främst för användning i primärvård (Kapitel 3.5 avsnitt ”Riktad screening av depression”). Det etablerade tröskelvärdet för depression är 10 poäng [38].

Beskrivning av underlaget.

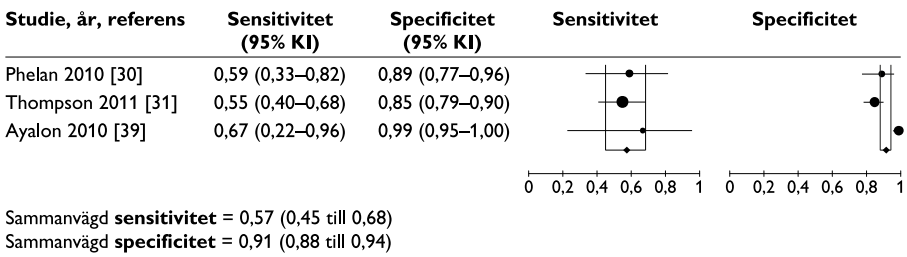
Av sju artiklar som lästes i fulltext exkluderades fyra, främst pga irrelevant referensstandard, att det förflutit för lång tid mellan index- och referenstest samt att tidsintervallet inte redovisades.

Tre studier med medelhög kvalitet utgjorde det vetenskapliga underlaget [30,31,39] (Tabell 3.6.9). De omfattade nära 450 äldre personer, varav

hälften i primärvård [30,39] och hälften inom ett forskningsprojekt om Parkinsons sjukdom [31]. I en av studierna ingick personer från två primärvårdsenheter i Israel. Inklusions- och exklusionskriterier angavs inte [39]. Medelåldern i studierna varierade mellan 72 och 78 år. Två av studierna exkluderade personer med svår kognitiv funktionsnedsättning (MMSE <15 respektive ”svår demens”). SCID-I användes som referensstandard i samtliga studier [23]. Syftet med studierna var att bedöma tillförlitligheten hos formuläret vad gällde egentlig depression och minor depression. Två av studierna undersökte även GDS-15 och resultaten av den utvärderingen finns beskrivet i avsnittet om GDS-15 [30,31].

Sammanvägda resultat

Sensitiviteten låg mellan 55 och 67 procent och specificiteten mellan 85 och 99 procent i de tre studierna. Spridningen illustreras i Figur 3.6.6.



Figur 3.6.6 *Forest plot för sensitivitet och specificitet för PHQ-9.*

Bedömning av evidensstyrka

Säkerheten i våra estimat påverkades av brister i det vetenskapliga underlaget. Det fanns en risk för bias genom att redovisningen av urvalet var oklar i en studie [31], att tiden mellan de två testerna överskred tre dygn för en okänd andel av personerna [31] (personlig kommunikation) och att blindningen var oklar i en studie [39].

Överförbarheten minskade av att det var oklart hur representativ patientgruppen var i två av studierna [30,31]. Precisionen för sensitivitet påverkades av det breda konfidensintervallet.

Tabell 3.6.5 *Resultatsammanställning för PHQ-9.*

Utfallsmått	Antal studier Antal personer	Samman-vägt estimat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	3 studier n=436	57% (45; 68%)	Otillräckligt ⊕○○○	Bristande studie-kvalitet (–1) Bristande precision (–2) Bristande över-förbarhet (–1)
Specificitet	3 studier n=436	91% (88; 94%)	Begränsat ⊕⊕○○	Bristande studie-kvalitet (–1) Bristande över-förbarhet (–1)

Evidensgraderade resultat

Det går inte att bedöma sensitivitet för PHQ-9 som formulär för riktad screening av depression hos äldre med strukturerad eller semistrukturerad intervju som referensstandard pga kvalitetsbrister i underlaget (otillräckligt vetenskapligt underlag ⊕○○○). Specificiteten för depression är 91 procent (95 % KI, 88 och 94) för äldre personer i primärvård eller med Parkinsons sjukdom med SCID-I som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○).

Cornell Scale for Depression in Dementia (CSDD)

CSDD utvecklades för att möta svårigheterna med att identifiera depression hos personer med demens [40]. Både patient och närmaste vårdgivare genomgår en semistrukturerad intervju som omfattar 19 symtomområden. Om det är skillnader i svar upprepas intervjun med vårdgivaren. Varje område skattas med 0 till 2 poäng. Formuläret är validerat även på personer utan demens [41]. Originalartikeln redovisade inte något tröskelvärde. I en senare manual angav författaren att poäng över 10 tyder på möjlig egentlig depression och att poäng över 18 tyder på egentlig depression. Poäng under 6 anses tyda på avsaknad av depression.

Beskrivning av underlaget

Av de 20 artiklar som lästes i fulltext exkluderades 19. De vanligaste orsakerna var irrelevant referensstandard, att det förflutit för lång tid mellan index- och referenstest samt att tidsintervallet inte redovisades.

En studie av Barca och medarbetare uppfyllde inklusionskriterierna [42]. Där ingick 231 personer med eller utan demenssjukdom. Personer från sjukhem och strokeenheter utvaldes slumpvis och personer från sjukhusavdelningar inkluderades konsekutivt. Medelåldern för personer på sjukhem var 86 år och för personer på sjukhus 78 år. Tre referensstandarder användes, nämligen ICD-10 research criteria, DSM-IV TR och Provisional diagnostic criteria for depression of Alzheimer's disease (PCD-dAD) [43]) men diagnoserna ställdes enbart utifrån checklistor och diagnostiska kriterier, vilket gjorde att studien inte kunde få mer än låg kvalitet.

Evidensgraderat resultat

Det finns för få studier med tillräcklig kvalitet för att kunna bedöma sensitivitet och specificitet för CSDD som formulär för riktad screening av depression hos äldre med eller utan demens (otillräckligt vetenskapligt underlag ⊕○○○).

Beck Depression Inventory (BDI-II)

Av de 10 artiklar som lästes i fulltext exkluderades nio. En studie med medelhög kvalitet av Low och medarbetare inkluderades [21] (Tabell 3.6.11). Studien har beskrivits i avsnittet om GDS-30. Low fann för BDI-II vid tröskelvärdet 14 en sensitivitet på 86 procent och en specificitet på 89 procent med SCID-I som referens.

Evidensgraderat resultat

Det finns för få studier med tillräcklig kvalitet för att kunna bedöma sensitivitet och specificitet för BDI-II som formulär för riktad screening av depression hos äldre, med SCID-I som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

Center for Epidemiologic Studies Depression Scale (CES-D)

Av de 25 artiklar som lästes i fulltext exkluderades 24. Studien av Roger och medarbetare som har beskrivits i avsnittet om GDS-15 uppfyllde inklusionskriterierna [32]. Vid det etablerade tröskelvärdet 16 var sensitiviteten 60 procent och specificiteten 76 procent. Studien bedömdes inte som relevant eftersom skattningarna gjordes i genomsnitt åtta dagar efter stroke, vilket är kortare tid än vad som anges i DSM-kriterierna (Tabell 3.6.12).

Evidensgraderat resultat

Det finns för få studier med tillräcklig kvalitet för att kunna bedöma sensitivitet och specificitet för CES-D som formulär för riktad screening av depression hos äldre med strukturerade eller semistrukturerade intervjuer som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

Stroke Inpatient Depression Inventory (SIDI)

I studien av Roger och medarbetare ingick även SIDI som konstruerades för att mäta depression efter stroke [32]. I studien var sensitiviteten endast 19 procent medan specificiteten låg på 95 procent vid det etablerade tröskelvärdet 17 (Tabell 3.6.13). Studien uteslöts av samma anledning som redovisats under avsnitten om GDS-15 och CES-D.

Evidensgraderat resultat

Det finns för få studier med tillräcklig kvalitet för att kunna bedöma sensitivitet och specificitet för SIDI som formulär för riktad screening av depression hos äldre efter stroke, med strukturerad eller semistrukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

Övriga formulär för riktad screening

För bedömningsformulären HSCL-25, SCL, GHQ, BCDRS och CPRS-D fanns det studier, men de uppfyllde inte våra inklusionskriterier. Det går därmed inte att uttala sig om sensitivitet och specificitet för dem.

Formulär för bedömning av svårighetsgrad av depression

Vår fråga var vilka formulär som är tillförlitliga för att bedöma svårighetsgrad av depression. Clinical Global Impression Scale – Severity (CGI-S) eller depressionsdjup enligt DSM-IV:s kriterier skulle vara referensstandard (Kapitel 3.1).

Montgomery-Åsberg Depression Rating Scale (MADRS)

Vi identifierade en studie som utvärderade MADRS av Hawley och medarbetare [44]. Studien exkluderades då den använde annan referensstandard.

Evidensgraderat resultat

Det saknas studier som undersöker om MADRS kan användas för att bedöma svårighetsgrad av depression hos äldre med CGI-S eller SCID som referensmetod.

Övriga formulär för att bedöma svårighetsgrad

Vi kunde inte identifiera några studier på andra formulär som uppfyllde våra inklusionskriterier

Diskussion

Förmågan hos ett formulär för riktad screening att fastställa eller utesluta en diagnos mäts i relation till en referensstandard. De diagnostiska kriterierna är inte anpassade till äldre och följaktligen inte heller referensstandard som SCID-I. Resultatet av diagnostiska tillförlitlighetsstudier hos äldre kan därför vara svårvärderade. Även om GDS-30 är utvecklat för äldre är formuläret ett exempel på svårigheter som kan finnas i depressionsdiagnostik hos äldre. Frågor saknas om kroppsligt präglade symtom som försämrad sömn och nedsatt sexuell funktion.

Det vetenskapliga underlaget för formulär för riktad screening av depression hos äldre är svagt. Endast tio av de nästan 200 artiklar som lästes i fulltext uppfyllde inklusionskriterierna.

Av de bedömda formulären kan GDS-30 och GDS-15 enligt vår litteraturgenomgång relativt tillförlitligt identifiera en depression hos en äldre

person och något mindre tillförlitligt identifiera en äldre person som inte är deprimerad.

Det finns två översikter, av Wancata och medarbetare respektive Mitchell och medarbetare, som gått igenom litteraturen för GDS-30 [45,46]. Översikterna har inte tillämpat samma inklusionskriterier som vi och har inte heller haft samma kvalitetskrav och begränsningar i val av tröskelvärde. Wancata och medarbetare inkluderade 33 studier av GDS-30 i varierande äldrepopulationer och fann en sammanvägd sensitivitet på 75 procent med variation mellan 34 och 100 procent och specificitet på 77 procent med variation mellan 63 och 96 procent för olika depressionsdiagnoser [45]. I en metaanalys av GDS i primärvård av Mitchell och medarbetare ingick sju studier av GDS-30 där den sammanvägda sensitiviteten var 77 procent med variation mellan 66 och 87 procent och specificiteten 65 procent med variation mellan 44 och 84 procent [46]. I vår litteraturgenomgång har vi kommit fram till likartade resultat.

De fyra studier av GDS-15 som vi inkluderat, är trots att de sinsemellan är heterogena, relativt samstämmiga. Både Wancata och medarbetare samt Mitchell och medarbetare kom fram till resultat som står i samklang med våra i sina respektive översikter. Wancata och medarbetare tog upp 21 studier av GDS-15. Den sammanvägda sensitiviteten var 81 procent med variation mellan 60 och 94 procent och specificiteten 75 procent med variation mellan 57 och 87 procent för någon depressionssjukdom [45]. Mitchell och medarbetare inkluderade 10 studier av GDS-15 i sin metaanalys. Sensitiviteten var 81 procent med variation mellan 77 och 85 procent och specificiteten 78 procent med variation mellan 71 och 84 procent [46].

Få studier av GDS-15 har utförts på personer med Parkinsons sjukdom. I en studie av personer mellan 65 och 75 år samt 75 år och däröver var sensitiviteten 89 respektive 90 procent, specificiteten 82 respektive 85 procent [27].

Av övriga formulär för riktad screening skulle PHQ-9 kunna vara ett värdefullt hjälpmedel för att identifiera äldre som inte är deprimerade. Låg sensitivitet för PHQ-9 har konstaterats även i andra äldrestudier

och populationer, exempelvis 54 procent i en studie av hjärtsjuka äldre [47] samt i Kapitel 3.5 om vuxna. Däremot går det inte att uttala sig om värdet av att använda HADS. Vi har inte kunnat identifiera några systematiska översikter eller metaanalyser av PHQ-9 och HADS med äldre.

När det gäller formulär som är anpassade för personer med kognitiva nedsättningar är Cornellskalan, CSDD, den mest kända och använda. Bristen på studier av god kvalitet gör att det saknas underlag för bedömning av formuläret. Detta är beklagligt och ytterligare studier är angelägna.

En begränsning i vår granskning är kravet på ett tidsintervall på maximalt en vecka mellan indextest och referenstest. I och med att depression hos äldre inte sällan kan ha ett kroniskt förlopp kan möjligen kravet vara för strängt [48]. Totalt exkluderade vi sex studier enbart pga tidskravet. Två av dem utvärderade GDS-15 och GDS-30 [49,50] och ingår i Wancata och medarbetares översikt [45]. I en senare studie av GDS-15 i sjukhemspopulationer var sensitiviteten 72 procent och specificiteten 78 procent vid tröskelvärdet 5 [51].

En annan begränsning ligger i att några av studierna avsåg specifika patientgrupper, t ex personer med Parkinsons sjukdom. I en majoritet av de redovisade studierna med medelhög kvalitet ingår dock personer i äldreboenden, primärvård och andra öppna vårdformer. Generaliserbarheten får därför ändå anses vara god.

Vi har i några fall accepterat studier med andra tröskelvärden än de etablerade eller de av författarna rekommenderade. För GDS och CSDD utfördes ingen ROC-analys för att få fram tröskelvärden i originalartiklarna om formulären. I senare arbeten är de vanligast förekommande tröskelvärdena 10 och 11 för GDS-30 (8 respektive 13 studier i Wancata och medarbetares översikt) medan fem och sex var vanligast för GDS-15 (sex respektive sju studier) [45]. I ursprungsartikeln för HADS föreslås ett tröskelvärde på 8–10 poäng för tveksamma fall, men gränsen har visat sig kunna medföra falskt negativa svar hos äldre [52,53]. För övrigt gav sällan de tröskelvärden som är rekommenderade eller anses etablerade den bästa avvägningen mellan sensitivitet och specificitet i de genom-

gångna studierna. Behovet av att modifiera tröskelvärden för äldre-populationer har påpekats av Blank och medarbetare samt Lamers och medarbetare [54,55].

När det gäller deltagarnas kognitiva förmåga tillämpade studierna olika exklusionskriterier. Ytterligheterna är de studier som krävde minst 24 poäng på MMSE [28,29] och ”svår demens” [30], eller mindre än 15 poäng på MMSE [31]. Två studier hade inte kognitiv förmåga som exklusionskriterium [35] HADS, [39] PHQ-9. En majoritet av de inkluderade studierna omfattade således personer med kognitiva begränsningar mellan lätt till medelsvår demens.

Sambandet mellan MMSE-poäng och poäng på GDS-15 respektive GDS-30 har studerats av Gerety och medarbetare [50]. I denna studie påverkades inte validiteten hos GDS-15 och GDS-30 av MMSE för personer på sjukhem, när de med svår demens exkluderats. Den lägsta sensitiviteten för GDS-30 i Wancata och medarbetares översikt [45] härrörde från en studie av personer med Alzheimers sjukdom [56]. Lägre sensitivitet och oförändrad specificitet för GDS-30 hos kognitivt nedsatta eller dementa personer jämfört med kognitivt intakta har noterats även av Kafonek och medarbetare [57].

MMSE anses otillräckligt för bedömning av kognitiv nedsättning vid depressioner hos äldre [16]. I de studier som ingår i vår litteraturgenomgång saknas dessutom närmare uppgifter om vilka kriterier utöver MMSE-poäng som användes vid diagnostik av kognitiv nedsättning eller demens. Detta gör att säkra slutsatser om formulärens användbarhet vid kognitiv nedsättning inte kan dras. Eftersom kognitiva symtom ofta är uttryck för depression hos äldre är det viktigt att en noggrann undersökning och beskrivning av deltagarnas kognitiva funktioner görs vid studier av diagnostiska formulär i denna åldersgrupp.

Kunskapsbristerna vad gäller bedömning av äldre med depression är framför allt kopplade till att det saknas en adekvat referensstandard. DSM och SCID-I är inte anpassade till depressioner hos äldre där kognitiva bortfall och samtidiga kroppsliga sjukdomar kan ge överlappande symtom samtidigt som symtombilden ofta är annorlunda än för yngre.

Med bristen på tillförlitlig referensstandard följer att det är svårt att utveckla valida bedömningsformulär. Detta är inte en kunskapslucka i konventionell mening utan snarare ett uttryck för en fundamental svårighet. Det är en vetenskaplig utmaning att utveckla en valid referensstandard och att försöka beskriva hur en sådan standard ska utformas ligger utanför ramarna för vårt uppdrag. En diagnostik som utgår från sammansatta mått i olika symtomdimensioner och med flera bedömare skulle kanske kunna vara en framkomlig väg.

Sammanfattningsvis pekar vår granskning på att det ändå finns formulär för riktad screening som kan vara ett stöd i samband med diagnostik hos äldre. Det är viktigt att diagnostik av depression hos äldre alltid baseras på en noggrann klinisk undersökning som omfattar psykiatriska, kognitiva, kroppsliga och psykosociala faktorer. Intensifierad forskning avseende utveckling av diagnostiska instrument för äldre är angelägen.

Table 3.6.6 *GDS-30 for screening of depression with semistructured or structured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Vargas et al 2007 [22] Brazil	*Design* Cross sectional *Setting* General non-psychiatric outpatient clinics *Population* Non-randomised sample, predominantly women Mean age: 70 years (range 60–96) Caucasians: 78% *Inclusion criteria* Age: ≥60 years *Exclusion criteria* MMSE >13 for illiterate, ≥18 for low to medium-level education, ≥26 for high educational level *Prevalence MDD* 43% (210/484)	*Index test* GDS-30, orally *Reference test* SCID-CV *Number of patients* n=484 *Drop-out rate* Not reported	*Rater of index test* 5 trained interviewers *Training of index rater* Eight hour program, conducted by a pedagogue *Rater of reference test* Psychiatrist *Training of reference test* Trained in the use of SCID-SV *Interobserver reliability* Not specified	Sensitivity: 87% (95% CI, 81.3; 91.0) Specificity: 73% (95% CI, 66.9; 77.8)	Moderate *Comments* *Blinding* Adequate *Handling of missing data* Not reported *Other comments* Cut off point = 12 Index and reference test were completed within one day (personal communication with authors)

The table continues on the next page

Table 3.6.6 *continued*

Author Year Reference Country	Study design Setting Population Inclusion & exclusion criteria Prevalence	Index test Reference test Number of patients Drop-out rate	Rater Rater training Interobserver reliability Outcome measure	Sensitivity Specificity	Study quality Comments
Low et al 2007 [21] Canada	*Design* Cross sectional *Setting* Two coronary care units *Population* n=224 consecutively *Inclusion criteria* Acute (recent or evolving) MI or unstable angina pectoris. Ability to speak and understand English. Medical stablity, freedom of angina pain and of other life threatening medical conditions *Exclusion criteria* Acute delirium, diagnosed dementia, cognitive impairment, MMSE 23 *Prevalence* MDD: 5% MDD+DD: 5.9% Any depressive disorder: 11.8%	*Index test* GDS-30 *Reference test* SCID-I *Number of patients* n=119 (30 female) Mean age: 62.97 years (SD 11.61) 16.7% of the women and 7.9% of the men were currently on anti-depressant medication *Drop-out rate* Not reported	*Rater of index test* Self rating *Training of index rater* Not reported *Rater of reference test* First author *Training of reference test* Not reported *Interobserver reliability* Not specified	*MDD+DD* Sensitivity: 100% Specificity: 84%	Moderate *Comments* *Sampling method* Adequate *Blinding* Adequate *Handling of missing data* No missing data *Other comments* CI not reported The index and reference tests were given in the same session (personal communication with authors)

CI = Confidence interval; DD = Dysthymia; GDS-30 = Geriatric depression scale 30 items; MDD = Major depressive disorder; MI = Myocardial infarction; MMSE = Mini-mental state examination; SCID-CV = Structured clinical interview for DSM disorders, clinical version; SD = Standard deviation

Table 3.6.7 *GDS-15 for screening of depression with semistructured or structured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Davison et al 2009 [28] Australia	*Design* Cross sectional *Setting* Randomly selected Low level care facilities, "homes for the elderly" in Melbourne *Population* n=282 eligible *Inclusion criteria* MMSE: ≥24/30 >65 years *Exclusion criteria* Psychotic disorder or intellectual disability, inability to communicate in English, severe hearing impairment *Prevalence MDD* 16.1% According to senior staff: 22.0%	*Index test* GDS-15 *Reference test* SCID-I elderly SCID-I informant *Number of patients* n=177 (76.8% female) Mean age: 84.7 years (SD 6.15) *Drop-out rate* 5.1%	*Rater of index test* A research assistant read the GDS-15 aloud *Training of index rater* Not reported *Rater of reference test* A clinical geropsychologist interviewed the elderly A psychologist interviewed a senior staff member who knew the participant well *Training of reference test* Not reported *Interobserver reliability* Overall concordance between diagnostic procedures=94%, κ 0.81 between the two reference standards	*MDD at cut off 5* *Self rating* Sensitivity: 93% (95% CI, 76; 99) Specificity: 77% (95% CI, 69; 84) *Informant rating* Sensitivity: 70% (95% CI, 53; 84) Specificity: 76% (95% CI, 68; 83)	Moderate *Comments* *Sampling method* Adequate *Blinding* Order of the interviews varied randomly to avoid order bias *Handling of missing data* No missing data were reported *Other comments* Medium delay between GDS-15 and SCID was 4.82 days SD 2.76 days

The table continues on the next page

Table 3.6.7 *continued*

Author Year Reference Country	Study design Setting Population Inclusion & exclusion criteria Prevalence	Index test Reference test Number of patients Drop-out rate	Rater Rater training Interobserver reliability Outcome measure	Sensitivity Specificity	Study quality Comments
Roger et al 2099 [32] USA	*Design* Cross sectional *Setting* An inpatient based rehabilitation unit *Population* Consecutive recruitment of persons with stroke (52.2% women) Mean age: 71.9±9.5 years 83% consented to participate Mean number of post-stroke-days at test was 8 *Exclusion criteria* Severe sensory deficits, cognitive impairment, aphasia, abuse problems *Prevalence (SCID)* Minor depression: 28.4% MDD: 15%	*Index test* GDS-15 *Reference test* SCID-CV *Number of patients* n=67 *Drop-out rate* None	*Rater of index test* Self rating *Training of index rater* Not reported *Rater of reference test* A trained doctoral-level provider *Training of reference test* Not reported *Interobserver reliability* Not specified	*Minor + major depression at standard cut off ≥5* Sensitivity: 46% Specificity: 90%	Moderate *Comments* Selection of persons not described *Blinding* Relevant *Handling of missing data* Not reported *Other comments* CI not reported All patients did not fulfill DSM-IV criteria for duration of symptoms of depression

The table continues on the next page

Table 3.6.7 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Phelan et al 2010 [30] USA	*Design* Cross sectional *Setting* Two primary care clinics affiliated with a university *Population* n=502 consecutively recruited persons (62% females) n=227 fulfilled inclusion criteria Mean age: 78±7 years Comorbidity: 2 chronic medical conditions on average *Inclusion criteria* Age: ≥65 *Exclusion criteria* Severe dementia, unstable medical conditions, not fluent in English *Prevalence* MDD: 12% Minor depression: 13%	*Index test* GDS-15 *Reference test* SCID-I (DSM-IV) *Number of patients* n=71 *Drop-out rate* 2/71	*Rater of index test* "Administered by a research assistant" *Rater of reference test* A geriatric psychiatrist or gerontologic psychiatric nurse practitioner *Rater training* No more information than "trained in SCID administration" *Interobserver reliability* Not specified	*Minor and major depression at cut off ≥5* Sensitivity: 88% (95% CI, 71; 96) Specificity: 14% (95% CI, 8; 16)	Moderate *Comments* *Sampling method* Adequate but only 1/3 of eligible persons consented *Blinding* Adequate *Handling of missing data* Not reported

The table continues on the next page

Table 3.6.7 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Thompson et al 2011 [31] USA	*Design* Cross sectional *Setting* University clinic *Population* n=254 persons with Parkinson disease participating in a clinical trial *Inclusion criteria* Presenting for clinical follow-up during a two year period *Exclusion criteria* MMSE: <15 *Prevalence* MDD: 12% Minor depression: 13%	*Index test* GDS-15 *Reference test* SCID-I (DSM-IV) *Number of patients* n=214 (42% female) Mean age: 72.5 years (SD 9.5) *Drop-out rate* None	*Index rater* A research assistant could assist those who had difficulty marking responses *Rater of reference test* Research assistants trained and supervised by a psychiatrist *Interobserver reliability* Not reported	*Minor or major depression* Sensitivity: 84.8% (95% CI, 71.1; 93.7) Specificity: 78.6% (95% CI, 71.6; 84.5)	Moderate *Comments* – *Blinding* Satisfactory *Handling of missing data* Not reported *Other comments* All instruments were administered on the same day (personal communication with authors)

The table continues on the next page

Table 3.6.7 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Haworth 2007 [29] United Kingdom	*Design* Cross sectional *Setting* Hospital outpatient cardiology clinic serving a local community *Population* n=203 consecutive persons *Inclusion criteria* Symptoms of heart failure ≥3 months despite loop diuretics left ventricular systolic dysfunction *Exclusion criteria* Implantable Cardiac Defibrillator MMSE score <24 *Prevalence* Depressive disorders: 25% (total) MDD: 14.8% Dysthymia: 2.3% Minor depression: 2.3%	*Index test* GDS-15 *Reference test* SCID-I *Number of patients* n=88 (17% females) Mean age: 69.9 years Comorbidity: 89% *Drop-out rate* Not reported	*Rater of index test* GDS-15 was administered as a structured interview by lead researcher *Training of index rater* Not reported *Rater of reference test* Lead researcher *Training of reference test* Not reported *Interobserver reliability* Not reported	*Depressive disorders cut off 5* Sensitivity: 81.8% Specificity: 83.3%	Moderate *Comments* *Sampling method* Adequate but representativity not ascertained *Blinding* Risk for evaluation bias *Other comments* Index and reference test were completed within one day (personal communication with authors)

CI = Confidence interval; DSM = Diagnostic and statistical manual of mental disorders; DSM-IV = Diagnostic and statistical manual of mental disorders, fourth edition; GDS-15 = Geriatric depression scale 15 items; MDD = Major depressive disorder; MMSE = Mini-mental scale examination; SCID = Structured clinical interview for DSM-disorders; SCID-I = Structured clinical interviews for DSM-IV axis I disorder; SCID-CV = Structured clinical interview for DSM-disorders, clinical version; SD = Standard deviation

Table 3.6.8 *Hospital Anxiety and Depression Scale (HADS) for screening of depression with semistructured or structured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Le Fevre et al 1999 [35] United Kingdom	*Design* Cross sectional *Setting* Hospice *Population* n=148 consecutive persons in palliative care were eligible *Exclusion criteria* Seriously ill or close to death *Prevalence* MDD: 18% 13% were classified as depressed using the Endicott criteria	*Index test* HADS-D *Reference test* CIS-R, diagnoses according to ICD-10 *Number of patients* n=79 Mean age: 68 years *Drop-out rate* None	*Index test rater* Self rating *Index test training* NA *Reference test rater* One clinician *Rater training* Not reported *Interobserver reliability* Not reported	*MDD cut off ≥8* Sensitivity: 91% Specificity: 59%	Moderate *Comments* *Sampling method* Adequate *Blinding* Satisfactory *Handling of missing data* Not reported

The table continues on the next page

Table 3.6.8 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Haworth et al 2007 [29] United Kingdom	*Design* Cross sectional *Setting* Hospital outpatient cardiology clinic serving a local community *Population* n=203 consecutive persons approached by mail *Inclusion criteria* Symptoms of heart failure ≥3 months, despite the use of loop diuretics left ventricular systolic dysfunction *Exclusion criteria* Implantable Cardiac Defibrillator MMSE score <24 *Prevalence* Depressive disorders 25% (total) MDD: 14.8% Dysthymia: 2.3% Minor depression: 2.3%	*Index test* HADS *Reference test* SCID *Number of patients* n=88 (17% female) Mean age: 69.9 years *Drop-out rate* Not reported	*Rater of index test* HADS administered as an interview by lead researcher *Training of index rater* Not reported *Rater of reference test* Lead researcher *Training of reference test* Not reported *Interobserver reliability* Not reported	*Depressive disorders cut off 7* Sensitivity: 93.8% Specificity: 84.7%	Moderate *Comments* Cut off 7 accepted in spite of recommendation *Sampling method* Adequate but but representativity not ascertained *Blinding* Risk for evaluation bias *Other comments* Index and reference test were completed within one day (personal communication with authors)

CIS-R = Clinical interview schedule-revised; HADS = Hospital anxiety and depression scale; HADS-D = Hospital anxiety and depression scale, depression subscale; ICD-10 = International statistical classification of diseases and related health problems, tenth revision; MDD = Major depressive disorder; MMSE = Mini-mental scale examination; NA = Not applicable; SCID = Structured clinical interview for DSM-disorders

Table 3.6.9 *PHQ-9 for screening of depression with semistructured or structured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Phelan et al 2010 [30] USA	*Design* Cross sectional *Setting* Two primary care clinics affiliated with a university *Population* n=502 consecutively recruited persons (62% females) n=227 fulfilled inclusion criteria Mean age: 78±7 years Comorbidity: 2 chronic medical conditions on average *Inclusion criteria* Age: ≥65 years *Exclusion criteria* Severe dementia, unstable medical conditions, not fluent in English language *Prevalence* MDD: 12% Minor depression: 13%	*Index test* PHQ-9 *Reference test* SCID (DSM-IV) *Number of patients* n=71 *Drop-out rate* 2/71	*Index test rater* "Administered by a research assistant" *Reference test rater* A geriatric psychiatrist or gerontologic psychiatric nurse practitioner *Rater training* No more information than "trained in SCID administration" *Interobserver reliability* Not specified	*Minor and major depression at cut off ≥10* Sensitivity: 59% (95% CI, 40; 74) Specificity: 89% (95% CI, 82; 93)	Moderate *Comments* *Sampling method* Adequate but only 1/3 of eligible persons consented *Blinding* Adequate *Handling of missing data* Not reported

The table continues on the next page

Table 3.6.9 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Thompson et al 2011 [31] USA	*Design* Cross sectional *Setting* University clinic *Population* n=254 Persons with Parkinson disease participating in a clinical trial *Inclusion criteria* Presenting for clinical follow-up during a two year period *Exclusion criteria* MMSE: <15 *Prevalence* MDD: 12% Minor depression: 13%	*Index test* PHQ-9 *Reference test* SCID-I (DSM-IV) *Number of patients* n=214 (42% female) Mean age: 72.5 years (SD 9.5) *Drop-out rate* None	*Rater of index test* Self rating. "A research assistant could assist those who had difficulty marking responses" *Training of index rater* NA *Rater of reference test* Research assistants trained and supervised by a psychiatrist *Interobserver reliability* Not specified	*Minor or major depression* Sensitivity:54% (95% CI, 39.0; 69.1) Specificity: 84.5% (95% CI, 78; 89)	Moderate *Comments* *Sampling method* adequate *Blinding* Satisfactory *Handling of missing data* Not reported *Other comments* All instruments were administered on the same day (personal communication with authors)

The table continues on the next page

Table 3.6.9 *continued*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Ayalon et al 2010 [39] Israel	*Design* Cross sectional *Setting* Two health centers in Israel *Population* Recruitment from the waiting rooms by research assistants *Inclusion criteria* Older primary care patients *Exclusion criteria* Not specified *Prevalence* MDD and/or dysthymia: 3.9%	*Index test* PHQ-9 *Reference test* SCID 1 (DSM-IV) *Number of patients* n=153 (40.5% female) Mean age: 75 years (SD 8.1) *Drop-out rate* Not reported	*Rater of index test* Self rating *Training of index test rater* *Rater of reference test* Two trained mental health graduate research assistants *Training of reference test rater* 3 day training course *Interobserver reliability* Adequate interobserver reliability was established against an experienced clinical psychologist	*MDD and/or dysthymia* Sensitivity: 67% Specificity: 99%	*Study quality* Moderate *Comments* *Sampling method* Adequate *Blinding* Not satisfactory *Handling of missing data* Not reported *Other comments* Instruments were administered ”on the same time” (personal communication with authors)

CI = Confidence interval; DSM-IV = Diagnostic and statistical manual of mental disorders, fourth edition; MDD = Major depressive disorder; MMSE = Mini-mental scale examination; NA = Not applicable; PHQ-9 = Patient health questionnaire, nine items; SCID = Structured clinical interview for DSM-disorders; SD = Standard deviation

Table 3.6.10 *The Cornell Scale for Depression in Dementia for screening of depression with semistructured or structured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Barca et al 2010 [42] Norway	*Design* Cross sectional reliability and validity study (the latter referred here) *Setting* 2 hospital departments of geriatric psychiatry, 5 nursing homes and one stroke unit *Population* *Hospital sample:* Consecutive recruitment Mean age: 78 years (females 61.2%) *Nursing homes, stroke unit:* Randomly addressed for participation Mean age: 86 years (females 75.3%) *Inclusion criteria* Not specified *Exclusion criteria* Bipolar syndrome *Prevalence* ICD-10: 45.5% DSM-IV-TR: 29.9% In persons with dementia: 53.3%	*Index test* CSDD *Reference test* A template with research criteria of ICD-10 and DSM-IV-TR PCD-dAD was used to diagnose Alzheimer depression *Number of patients* n=231 *Drop-out rate* Not reported	*Rater of index test* Trained nurses (after inteview with patients primary caregiver and the patient) *Training of index rater* Not specified *Rater of reference test* Geriatric psychiatrist *Training of reference test* Not specified *Interobserver reliability* Not specified in validity study. In reliability study, κ was 0.91 (range 0.8–1.0) the correlation between 2 raters was 0.97	*All patients* *Cut off 8/9 produced highest accuracy against ICD-10* Sensitivity: 63% Specificity: 79% *Cut off 10/11 produced highest accuracy against DSM-IV-TR* Sensitivity: 63% Specificity: 86%	Low Diagnoses were made solely on the basis of checklists and diagnostic criteria *Comments* *Sampling method* Adequate *Blinding* Adequate *Handling of missing data* Not reported

CSDD = Cornell scale for depression in dementia; DSM-IV-TR = Diagnostic and statistical manual of mental disorder (fourth edition, text revision); ICD-10 = International statistical classification of diseases and health related problems, tenth revision; PCD-dAD = Provisional criteria for depression in Alzheimer's disease

Table 3.6.11 *BDI-II for screening of depression with semistructured or structured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Low et al 2007 [21] Canada	*Design* Cross sectional *Setting* Two coronary care units *Population* n=224 consecutively *Inclusion criteria* Acute (recent or evolving) MI or unstable angina pectoris. Ability to speak and understand english. Medical stablity, freedom of angina pain and of other life threatening medical conditions *Exclusion criteria* Acute delirium, cognitive impairment MMSE: ≤23 Ongoing kidney dialysis *Prevalence* MDD: 5% MDD + dystymic disorder: 5.9% Any depressive disorder: 11.8%	*Index test* BDI-II *Reference test* SCID *Number of patients* n=119 (30 females) Mean age: 62.97 years (SD 11.61). 16.7% of the women and 7.9% of the men were currently on anti-depressant medication *Drop-out rate* Not reported	*Rater of index test* Self rating *Training of index rater* Not reported *Rater of reference test* First author *Training of reference test* Not reported *Interobserver reliability* Not reported	*MDD+ dystymic disorder* Sensitivity: 86% Specificity: 89%	Moderate *Comments* *Sampling method* Adequate *Blinding* Adequate *Handling of missing data* Not reported *Other comments* CI not reported The index and reference tests were given in the same session (personal communication with authors)

BDI-II = Beck depression inventory, second edition; CI = Confidence interval; MDD = Major depressive disorder; MI = Myocardial infarction; MMSE = Mini-mental scale examination; SCID = Structured clinical interview for DSM-disorders; SD = Standard deviation

Table 3.6.12 *CES-D for screening of depression with semistructured or structured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Roger et al 2009 [32] USA	*Design* Cross sectional *Setting* An inpatient based rehabilitation unit *Population* Consecutive recruitment of persons with stroke (52.2% women) Mean age: 71.9±9.5 years 83% consented to participate Mean number of post-stroke-days at test was 8 *Exclusion criteria* Severe sensory deficits, cognitive impairment, significant aphasia, abuse problems *Prevalence (SCID)* Minor depression: 28.4% MDD: 15%	*Index test* CES-D *Reference test* SCID-CV *Number of patients* n=67 *Drop-out rate* 0%	*Rater of index test* Self rating *Training of index rater* Not reported *Rater of reference test* A trained doctoral-level provider *Training of reference test* Not reported *Interobserver reliability* Not specified	*Minor and major depression at cut off ≥16* Sensitivity: 60% Specificity: 76%	Moderate *Comments* *Sampling method* Not reported *Blinding* Relevant *Handling of missing data* Not reported *Other comments* CI not reported All patients did not fulfill DSM-IV criteria for duration of depression

CES-D = Center for epidemiological studies depression scale; CI = Confidence interval; DSM-IV = Diagnostic and statistical manual of mental disorder, fourth edition; MDD = Major depressive disorder; SCID = Structured clinical interview for DSM-disorders; SCID-CV = Structured clinical interview for DSM-disorders, clinical version

Table 3.6.13 *SIDI for screening of depression with semistructured or structured interview as reference standard.*

Author **Year** **Reference** **Country**	**Study design** **Setting** **Population** **Inclusion & exclusion criteria** **Prevalence**	**Index test** **Reference test** **Number of patients** **Drop-out rate**	**Rater** **Rater training** **Interobserver reliability** **Outcome measure**	**Sensitivity** **Specificity**	**Study quality** **Comments**
Roger et al 2009 [32] USA	*Design* Cross sectional *Setting* An inpatient based rehabilitation unit *Population* Consecutive recruitment of persons with stroke (52.2% women) Mean age: 71.9±9.5 years 83% consented to participate Mean number of post-stroke-days at test was 8 *Exclusion criteria* Severe sensory deficits, cognitive impairment, significant aphasia and abuse problems *Prevalence (SCID)* Minor depression: 28.4% MDD: 15%	*Index test* SIDI *Reference test* SCID-CV *Number of patients* n=67 Mean MMSE: 23.8 (SD 4.0) *Drop-out rate* 0%	*Rater of index test* Self rating *Training of index rater* Not reported *Rater of reference test* A trained doctoral-level provider *Training of reference test* Not reported *Interobserver reliability* Not specified	*Minor and major depression at cut off ≥17* Sensitivity: 19% Specificity: 95%	*Study quality* Moderate *Comments* Selection of persons not described *Blinding* Satisfactory *Handling of missing data* Not reported *Other comments* CI not reported All patients did not fulfill DSM-IV criteria for duration of depression

CI = Confidence interval; DSM-IV = Diagnostic and statistical manual of mental disorder, fourth edition; MDD = Major depressive disorder; MMSE = Mini-mental scale examination; SCID-CV = Structured clinical interview for DSM-disorders, clinical version; SD = Standard deviation; SIDI = Stroke inpatient depression inventor

Referenser

1. Beekman AT, Copeland JR, Prince MJ. Review of community prevalence of depression in later life. Br J Psychiatry 1999;174:307-11.

2. Copeland JR, Beekman AT, Braam AW, Dewey ME, Delespaul P, Fuhrer R, et al. Depression among older people in Europe: the EURODEP studies. World Psychiatry 2004;3:45-9.

3. Lesher EL, Berryhill JS. Validation of the Geriatric Depression Scale – Short Form among inpatients. J Clin Psychol 1994;50:256-60.

4. Covinsky KE, Kahana E, Chin MH, Palmer RM, Fortinsky RH, Landefeld CS. Depressive symptoms and 3-year mortality in older hospitalized medical patients. Ann Intern Med 1999;130:563-9.

5. Adamson JA, Price GM, Breeze E, Bulpitt CJ, Fletcher AE. Are older people dying of depression? Findings from the Medical Research Council trial of the assessment and management of older people in the community. J Am Geriatr Soc 2005; 53:1128-32.

6. Minino AM, Arias E, Kochanek KD, Murphy SL, Smith BL. Deaths: final data for 2000. Natl Vital Stat Rep 2002;50: 1-119.

7. Alexopoulos GS. Depression in the elderly. Lancet 2005;365:1961-70.

8. Girling DM, Barkley C, Paykel ES, Gehlhaar E, Brayne C, Gill C, et al. The prevalence of depression in a cohort of the very elderly. J Affect Disord 1995; 34:319-29.

9. Corruble E, Gorwood P, Falissard B. Association between age of onset and symptom profiles of late-life depression. Acta Psychiatr Scand 2008;118:389-94.

10. Beekman ATF, De Beurs E, Van Balkom AJLM, Deeg DJH, Van Dyck R, Van Tilburg W. Anxiety and depression in later life: Co-occurrence and communality of risk factors. Am J Psychiatry 2000;157:89-95.

11. Lockwood KA, Alexopoulos GS, van Gorp WG. Executive dysfunction in geriatric depression. Am J Psychiatry 2002;159:1119-26.

12. Elderkin-Thompson V, Kumar A, Bilker WB, Dunkin JJ, Mintz J, Moberg PJ, et al. Neuropsychological deficits among patients with late-onset minor and major depression. Arch Clin Neuropsychol 2003; 18:529-49.

13. Alexopoulos GS, Borson S, Cuthbert BN, Devanand DP, Mulsant BH, Olin JT, et al. Assessment of late life depression. Biol Psychiatry 2002;52:164-74.

14. Devanand DP, Sano M, Tang MX, Taylor S, Gurland BJ, Wilder D, et al. Depressed mood and the incidence of Alzheimer's disease in the elderly living in the community. Arch Gen Psychiatry 1996;53:175-82.

15. Yaffe K, Blackwell T, Gore R, Sands L, Reus V, Browner WS. Depressive symptoms

and cognitive decline in nondemented elderly women: a prospective study. Arch Gen Psychiatry 1999;56:425-30.

16. Steffens DC, Otey E, Alexopoulos GS, Butters MA, Cuthbert B, Ganguli M, et al. Perspectives on depression, mild cognitive impairment, and cognitive decline. Arch Gen Psychiatry 2006;63:130-8.

17. Dillon C, Allegri RF, Serrano CM, Iturry M, Salgado P, Glaser FB, et al. Late- versus early-onset geriatric depression in a memory research center. Neuropsychiatr Dis Treat 2009;5:517-26.

18. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17:37-49.

19. Brink TL. Screening tests for geriatric depression. Clinical Gerontologist: The Journal of Aging and Mental Health 1982; 1:37-43.

20. Shah A, Phongsathorn V, George C, Bielawska C. Psychiatric morbidity among continuing care geriatric inpatients. Int J Geriatr Psychiatry 1992;7:517-25.

21. Low GD, Hubley AM. Screening for depression after cardiac events using the Beck Depression Inventory-II and the Geriatric Depression Scale. Soc Indic Res 2007;82:527-48.

22. Vargas HO, Matsuo T, Blay SL. Validity of the Geriatric Depression Scale for patients seen at general outpatient clinics. Clinical Gerontologist: The Journal of Aging and Mental Health 2007;30:65-78.

23. First M, Spitzer R, Gibbon M, Williams JBW. Structured Clinical Interview for DSM-IV Axis I disorders (SCID I). New York: Biometric Research Department; 1997.

24. Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12: 189-98.

25. Sheikh JI, Yesavage JA. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontologist: The Journal of Aging and Mental Health 1986;5:165-73.

26. McCabe MP, Davison T, Mellor D, George K, Moore K, Ski C. Depression among older people with cognitive impairment: prevalence and detection. Int J Geriatr Psychiatry 2006;21:633-44.

27. Weintraub D, Saboe K, Stern MB. Effect of age on geriatric depression scale performance in Parkinson's disease. Mov Disord 2007;22:1331-5.

28. Davison TE, McCabe MP, Mellor D. An examination of the "gold standard" diagnosis of major depression in aged-care settings. Am J Geriatr Psychiatry 2009;17:359-67.

29. Haworth JE, Moniz-Cook E, Clark AL, Wang M, Cleland JG. An evaluation of two self-report screening measures for mood in an out-patient chronic heart failure population. Int J Geriatr Psychiatry 2007;22: 1147-53.

30. Phelan E, Williams B, Meeker K, Bonn K, Frederick J, Logerfo J, et al. A study of

the diagnostic accuracy of the PHQ-9 in primary care elderly. BMC Fam Pract 2010;11:63.

31. Thompson AW, Liu H, Hays RD, Katon WJ, Rausch R, Diaz N, et al. Diagnostic accuracy and agreement across three depression assessment measures for Parkinson's disease. Parkinsonism Relat Disord 2011;17:40-5. Epub 2010 Nov 17.

32. Roger PR, Johnson-Greene D. Comparison of assessment measures for post-stroke depression. Clin Neuropsychol 2009;23:780-93.

33. Gottfries GG, Noltorp S, Norgaard N. Experience with a Swedish version of the Geriatric Depression Scale in primary care centres. Int J Geriatr Psychiatry 1997;12: 1029-34.

34. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361-70.

35. Le Fevre P, Devereux J, Smith S, Lawrie SM, Cornbleet M. Screening for psychiatric illness in the palliative care inpatient setting: a comparison between the Hospital Anxiety and Depression Scale and the General Health Questionnaire-12. Palliat Med 1999;13:399-407.

36. Goldberg DP, Blackwell B. Psychiatric illness in general practice. A detailed study using a new method of case identification. Br Med J 1970;1:439-43.

37. Spitzer RL, Wakefield JC. DSM-IV diagnostic criterion for clinical significance: does it help solve the false positives problem? Am J Psychiatry 1999;156:1856-64.

38. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001; 16:606-13.

39. Ayalon L, Goldfracht M, Bech P. 'Do you think you suffer from depression?' Reevaluating the use of a single item question for the screening of depression in older primary care patients. Int J Geriatr Psychiatry 2010;25:497-502.

40. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell Scale for Depression in Dementia. Biol Psychiatry 1988;23:271-84.

41. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Use of the Cornell scale in nondemented patients. J Am Geriatr Soc 1988;36:230-6.

42. Barca ML, Engedal K, Selbaek G. A reliability and validity study of the cornell scale among elderly inpatients, using various clinical criteria. Dement Geriatr Cogn Disord 2010;29:438-47.

43. Olin JT, Schneider LS, Katz IR, Meyers BS, Alexopoulos GS, Breitner JC, et al. Provisional diagnostic criteria for depression of Alzheimer disease. Am J Geriatr Psychiatry 2002;10:125-8.

44. Hawley CJ, Gale TM, Sivakumaran T, Hertfordshire Neuroscience Research g. Defining remission by cut off score on the MADRS: Selecting the optimal value. J Affect Disord 2002;72:177-84.

45. Wancata J, Alexandrowicz R, Marquart B, Weiss M, Friedrich F. The criterion validity of the Geriatric Depression Scale:

a systematic review. Acta Psychiatr Scand 2006;114:398-410.

46. Mitchell AJ, Bird V, Rizzo M, Meader N. Diagnostic validity and added value of the geriatric depression scale for depression in primary care: a meta-analysis of GDS30 and GDS15. J Affect Disord 2010;125: 10-7. Epub 2009 Oct 2.

47. McManus D, Pipkin SS, Whooley MA. Screening for depression in patients with coronary heart disease (data from the Heart and Soul Study). Am J Cardiol 2005;96:1076-81.

48. Beekman AT, Geerlings SW, Deeg DJ, Smit JH, Schoevers RS, de Beurs E, et al. The natural history of late-life depression: a 6-year prospective study in the community. Arch Gen Psychiatry 2002;59:605-11.

49. Arthur A, Jagger C, Lindesay J, Graham C, Clarke M. Using an annual over-75 health check to screen for depression: validation of the short Geriatric Depression Scale (GDS15) within general practice. Int J Geriatr Psychiatry 1999;14:431-9.

50. Gerety MB, Williams JW, Jr., Mulrow CD, Cornell JE, Kadri AA, Rosenberg J, et al. Performance of case-finding tools for depression in the nursing home: influence of clinical and functional characteristics and selection of optimal threshold scores. J Am Geriatr Soc 1994;42:1103-9.

51. Marc LG, Raue PJ, Bruce ML. Screening performance of the 15-item geriatric depression scale in a diverse elderly home care population. Am J Geriatr Psychiatry 2008;16:914-21.

52. Johnson G, Burvill PW, Anderson CS, Jamrozik K, Stewart-Wynne EG, Chakera TM. Screening instruments for depression and anxiety following stroke: experience in the Perth community stroke study. Acta Psychiatr Scand 1995;91:252-7.

53. Lam CLK, Pa P-C, Chan AWT, Chan S-Y. Can the Hospital and Anxiety Depression (HAD) Scale be used on Chinese elderly in general practice? Fam Pract 1995;12:149-54.

54. Blank K, Gruman C, Robison JT. Case-finding for depression in elderly people: balancing ease of administration with validity in varied treatment settings. J Gerontol A Biol Sci Med Sci 2004;59:78-84.

55. Lamers F, Jonkers CC, Bosma H, Penninx BW, Knottnerus JA, van Eijk JT. Summed score of the Patient Health Questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. J Clin Epidemiol 2008;61:679-87.

56. Gilley DW, Wilson RS. Criterion-related validity of the Geriatric Depression Scale in Alzheimer's disease. J Clin Exp Neuropsychol 1997;19:489-99.

57. Kafonek S, Ettinger WH, Roca R, Kittner S, Taylor N, German PS. Instruments for screening for depression and dementia in a long-term care facility. J Am Geriatr Soc 1989;37:29-34.

3.7 Hälsoekonomi

Slutsatser

Det går inte att bedöma om det är kostnadseffektivt att komplettera klinisk diagnos med bedömningsformulär för förstämningssyndrom eftersom det saknas studier.

Bakgrund

Mätt enligt WHO:s DALY (disability adjusted life years) svarar depressioner och andra psykiska sjukdomar för en femtedel av den totala sjukdomsbördan i Sverige [1]. Omkring 40 procent av alla fall av sjuk- och aktivitetsersättningar (”förtidspensioneringar”) orsakas av psykisk ohälsa [2].

Samhällets kostnader för depressionssjukdomar fördubblades mellan åren 1997 och 2005, från cirka 16 till 32 miljarder kronor [3]. Sjukskrivningar svarade för en stor andel av kostnadsökningen. Omkring två tredjedelar av den beräknade samhällskostnaden utgörs av produktionsförluster till följd av kort- och långvarig sjukfrånvaro (inklusive förtidspensioneringar). Personer med psykisk funktionsnedsättning är arbetslösa i högre grad än befolkningen i övrigt [4].

Metod

Litteratursökning med tillämpning av söktermer använda för de medicinska frågeställningarna, med tillägg för ekonomiska söktermer (costs and cost analyses), angav endast ett abstrakt, som därefter beställdes i fulltext och inkluderades.

Resultat

Hewitt och medarbetare har gjort en systematisk översikt avseende screening för postnatal depression där de utöver kliniska aspekter även har sökt efter hälsoekonomiska studier [5]. I studien hittades ingen fullständig hälsoekonomisk analys och inga slutsatser om ekonomiska aspekter kan därför dras i översikten.

Table 3.7.1 *Health economy outcome.*

Author Year Reference Country	Study design	Population characteristics	Intervention Control	Follow-up period Drop-out rate	Results	Study quality and relevance	Comments
Hewitt et al 2009 [5] United Kingdom	Systematic review	Women with potential postnatal depression	*Intervention* Formal methods to identify postnatal depression *Control* Not using formal methods to identify postnatal depression or usual care	Not applicable	No full economic evaluation of methods to identify postnatal depression was found	Moderate	

Referenser

1. Moradi T, Allebeck P, Jacobsson A, Mathers C. [The burden of disease in Sweden measured with DALY. Neuropsychiatric diseases and cardiovascular diseases dominate]. Lakartidningen 2006;103:137-41.

2. Socialstyrelsen. Folkhälsorapport 2009. Artikelnr: 2009-126-71.

3. Sobocki P, Ekman M, Agren H, Krakau I, Runeson B, Martensson B, et al. Resource use and costs associated with patients treated for depression in primary care. Eur J Health Econ 2007; 8:67-76.

4. SCB. Situationen på arbetsmarknaden för personer med funktionsnedsättning – 4:e kvartalet 2000. Information om utbildning och arbetsmarknad 2001:3. Statistiska Centralbyrån.

5. Hewitt CE, Gilbody SM. Is it clinically and cost effective to screen for postnatal depression: a systematic review of controlled clinical trials and economic evidence. BJOG 2009;116:1019-27.

4. **Inventering**

Det finns mängder av formulär som är avsedda för att bedöma förekomst och djup av förstämningssyndrom eller för att följa sjukdomsförloppet under behandling. Som framgick av Kapitel 3 och 7 förefaller minst 60 formulär för screening, diagnostik eller symtomskattning användas mer eller mindre rutinmässigt. Några formulär används av många, andra av enstaka kliniker. Enkätsvaren vi fick in visade påtagliga skillnader i vilket syfte formulären användes. Symtomskattningsformuläret MADRS används t ex även för att ställa diagnos (Kapitel 7).

Vetenskapligt stöd för en god diagnostisk tillförlitlighet är en viktig komponent vid val av bedömningsformulär. Men det finns också andra aspekter som bör tas i beaktande. Hit hör hur översättningen av formuläret till svenska har gått till. Har den gjorts med godkännande av upphovsmannen? Är översättningen validerad genom återöversättning? Om översättningen inte är gjord ”enligt konstens alla regler” finns det en risk för att frågorna tolkas annorlunda än vad som är tänkt.

Dessutom finns det praktiska aspekter att ta hänsyn till. Hur omfattande är formuläret? Tar det alltför lång tid att gå igenom för att fungera? Vilket är att föredra: att patienterna fyller i formuläret eller att vårdgivaren gör det? Är formuläret lätt tillgängligt? Är det kostnadsfritt att använda?

Detta kapitel kompletterar den systematiska litteraturgranskningen med en sammanställning av uppgifter om översättning, omfattning, tillgänglighet och kostnader. De formulär som har vetenskapligt stöd är markerade med fetstil för att underlätta läsningen.

Vi har samlat data från flera källor. Vi gick igenom användarmanualer och översiktslitteratur [1]. Via internet fick vi information från dels webbsidor som är specifika för vissa formulär och dels från webbplatser som sammanställer uppgifter för många instrument (t ex

www.vardverktyget.se). Flera svenska forskare inom området har också bidragit med information.

Ändå är sammanställningen långt ifrån komplett. Det visade sig vara svårt att hitta uppgifter om översättningar och tillgänglighet för många av formulären. För somliga formulär cirkulerar det flera olika versioner och där vi hittade information kunde det vara svårt att bedöma trovärdigheten. Det är ofta oklart om alla frågor är med i den aktuella versionen, eller om det har lagts till några extra jämfört med originalet. Var översättningen till svenska korrekt gjord och godkänd av upphovsmannen? Ofta saknades information om användande (av vem, i vilket syfte och för vilka indikationer) och tolkning av resultaten.

Den bristande tillgången och tveksamma trovärdigheten speglar sannolikt den kliniska verkligheten och visar på en stor svårighet för de kliniker och psykologer som faktiskt vill använda strukturerad diagnostik idag. Det finns ett tydligt behov av kvalitetssäkrade översättningar. Det finns även ett behov av en central ”nod” för information och tillgänglighet av rätt version av formulären.

Tabell 4.1 *Strukturerade och semistrukturerade intervjuer.*

Formulär	Typ, åldersgrupp	Originalpublikation	Översättning	Omfattning	Kostnad
BCFPI Brief Child and Family Phone Interview	Strukturerad intervju anpassad för telefon-intervju 3–18 år	Cunningham et al 2009 [2]	Version på officiell hemsida: (C) BCFPI Inc. Mars 2002. Översatt av Jens Dymling. Oklar metod	Max 30 min	Ja
CIDI Composite International Diagnostic Interview	Strukturerad Från 18 år	Robins et al 1988 [3]	Ja Metod och tillstånd oklart	1–2 timmar	Oklart
DAWBA Development and Wellbeing Assessment	Strukturerad 5–16 år	Goodman et al 2000 [4]	Ja Officiell hemsida: ”Jan-Olov Larsson and Hans Smedje have made major contributions to the translation, back-translation or validation”	30– 50 minuter	Fritt tillgänglig
DICA Diagnostic Interview for Children and Adolescents	Semistrukturerad Barn och ungdomar	Herjanic et al 1982 [5]	Ja Gunilla Olsson. Metod oklar	5–20 minuter per diagnos	Oklart
DIS Diagnostic Interview Schedule	Strukturerad Vuxna	Robins et al 1981 [6]	Oklart	90–120 minuter	Ja, se webbplats
KID-SCID KID-Structured Clinical Interview for DSM	Semistrukturerad Barn	Matzner et al 1997 [7]	Oklart	Oklart	Oklart
K-SADS Kiddie-Schedule of Affective Disorders and Schizophrenia for school age children	Semistrukturerad Föräldrar och barn (6–18 år) intervjuas var för sig	Kaufman et al 1997 [8]	Ja Tord Ivarsson. Översättningen är återöversatt till engelska och granskad och godkänd av upp-hovsman	Cirka 75 minuter för barn/ungdom respektive föräld-rar (150 minuter totalt)	Fritt tillgänglig
MINI Mini International Neuropsychiatric Interview	Strukturerad Från 18 år	Sheehan et al 1998 [9]	Ja Svensk version 6.0.0b. Allgulander C, Wærn M, Humble M, Andersch S, Ågren H	20 minuter	Fritt tillgänglig i pappersversion
MINI-KID Mini International Neuropsychiatric Interview for Children and Adolescents	Strukturerad 6–17 år	Sheehan et al 1998 [9]	Ja	20 minuter	Fritt tillgänglig i pappersversion

Tabellen fortsätter på nästa sida

Tabell 4.1 *fortsättning*

Formulär	Typ, åldersgrupp	Originalpublikation	Översättning	Omfattning	Kostnad
PRIME-MD Primary Care Evaluation of Mental Disorders	Strukturerad Från 18 år	Spitzer et al 1994 [10]	Ja Rättigheter till den svenska utgåvan: Pfizer AB och Pär Svanborg, Karolinska Universitetssjukhuset, Stockholm	15 minuter	Fritt tillgänglig
SADS Schedule for Affective Disorders and Schizophrenia	Semistrukturerad, baserad på RDC-kriterier Från 18 år	Endicott et al 1978 [11]	Ej identifierad	1–4 timmar	Oklart
SCAN Schedules for Clinical Assessment in Neuropsychiatry	Semistrukturerad Från 18 år	Wing et al 1990 [12]	Ej identifierad	2 timmar	Ja
SCID-I Structured Clinical Interview for DSM-IV Axis I Disorders	Semistrukturerad Från 18 år	Spitzer et al 1992 [13]	Ja Både papper och elektroniskt	1–1,5 timme	Ja
SPIFA Structured Psychiatric Interview for General Practitioners	Strukturerad Från 18 år	Dahl et al 2009 [14]	Ja Metod och tillstånd oklart (nordiskt original)	15–20 minuter	Oklart

Tabell 4.2 *Formulär för riktad screening av depression.*

Formulär	Ålder och bedömare	Originalpublikation	Översättning	Omfattning	Kostnad
BDI-II Beck Depression Inventory II	Från 13 år Självbedömning	Beck et al 1961 [15] Beck et al 1996 [16]	Ja Svensk version av BDI-II expertgranskad av Niklas Pålsson. Manual är översatt och anpassad till svenska med tillstånd. Copyright 2005 by Harcourt Assessment, Inc USA	BDI-II 21 frågor, 5–10 minuter	Ja
BYI/BUS	7–18 år Självbedömning		Ja	5–10 minuter	Ja
CBCL	Se Tabell 4.3				
CDI Childrens' Depression Inventory	6–17 år Självbedömning	Kovacs 1985 [17]	Ja Bo Larsson, BUP UAS, med tillåtelse från upphovsman. Ingen återöversättning utfördes	27 frågor, 10–15 minuter	Fri version cirkulerar
CDSS Calgary Depression Scale for Schizophrenia	Vuxna Självbedömning	Addington et al 1992 [18]	Ja Metod och tillstånd oklart	9 frågor	Fritt tillgänglig för studier och icke vinstdrivande organisationer
CES-D Center for epidemiologic studies of depression	Vuxna Självbedömning	Radloff et al 1977 [19]	Ja Metod och tillstånd oklart	20 frågor, 10 minuter	
CES-DC Center for Epidemiological Studies Depression Scale for Children	12–18 år Självbedömning	Weissman et al 1980 [20]	Ja Gunilla Olsson. Oklart om rättigheter och återöversättning	20 frågor	Fritt tillgänglig
Cornell/CSDD Cornell Scale for depression in dementia	Äldre med demens Semistrukturerad intervju	Alexopoulos et al 1988 [21]	Ja Metod och tillstånd oklart	19 frågor	Oklart
DSRS Birleson Depression Rating Scale	Barn Självbedömning	Birleson 1981 [22]	Ja Christopher Gillberg. Metod och rättigheter oklara	18 frågor	Fritt tillgänglig
EPDS Edinburgh Post Natal Depression Scale	Kvinnor, gravida och post partum Självbedömning	Cox et al 1987 [23]	Ja	10 frågor, 5 minuter	Fritt tillgänglig

Tabellen fortsätter på nästa sida

Tabell 4.2 *fortsättning*

Formulär	Ålder och bedömare	Originalpublikation	Översättning	Omfattning	Kostnad
GDS-15 Geriatric Depression Scale	Äldre Självbedömning	Brink 1982 [24]	Ja Carl-Gerhard Gottfries med tillstånd av upphovsmannen Yesavage, men inte återöversatt	15 frågor	Oklart
GDS-30 Geriatric Depression Scale	Äldre Självbedömning	Se GDS-15	Ja Metod och tillstånd oklart	30 frågor	Oklart
GDS-20	Äldre Självbedömning	Gottfries et al 1997 [25]	Som GDS-15 samt komplettering med 5 frågor	20 frågor	Oklart
HADS-D Hospital Anxiety and Depression Scale (depressionsdelen)	Vuxna Självbedömning	Zigmond et al 1983 [26]	Ja Rättigheten tillhör Marianne Sullivan, HRQL-gruppen	7 frågor, 5 minuter	Fritt för landsting och kommuner
HSCL-25 Hopkins symtom checklist	Vuxna Självbedömning	Hesbacher et al 1980 [49]	Ja Metod och tillstånd oklart	25 frågor, 20 minuter	Fritt tillgängligt
MDI Major Depression Inventory	Vuxna Självbedömning	Bech et al 2001 [27]	Ja	12 frågor	Kostnadsfritt
PHQ-9 Patient Health Questionnaire	Tonåringar och vuxna Självbedömning	Spitzer et al 1999 [28]	Ja Det krävs inget tillstånd från upphovsmännen för att översätta	9 frågor	Kostnadsfritt
SDQ Strength and Difficulties Questionnaire, emotionell delskala	Barn och ungdomar, 3–16 år föräldra- och lärarskattning 11–17 år Självbedömning	Goodman et al 2000 [29]	Ja	25 frågor	Kostnadsfritt
SHQ Seasonal Health Questionnaire	Från 18 år Självbedömning av säsongsbunden depression	Thompson et al 2001 [30]	Ja Cecilia Rastad	27 frågor	Oklart
SIDI Stroke Inpatient Depression Inventory	Efter stroke Självbedömning	Rybarczyk et al 1996 [31]	Ej identifierad	Oklart	Oklart

Tabellen fortsätter på nästa sida

Tabell 4.2 *fortsättning*

Formulär	Ålder och bedömare	Originalpublikation	Översättning	Omfattning	Kostnad
SPAQ Seasonal Pattern Assessment Questionnaire	Från 18 år Självbedömning av säsongsbunden depression	Rosenthal et al 1987 [32]	Ja Jerker Hetta och Lennart Wetterberg. Metod oklar	6+1 frågor	Oklar för svensk version
TRF Teacher Report Form (del av ASEBA)	Barn och ungdomar, 6–18 år Lärarskattning av adaptivt beteende och beteende/känslomässiga svårigheter	Achenbach 2009 [33]	Ja	118 frågor	Ja

Tabell 4.3 *Formulär för riktad screening av mani.*

Formulär	Ålder och bedömare	Originalpublikation	Översättning	Omfattning	Kostnad
AS-18 Affective Self Rating Scale -18	Vuxna Självbedömning	Adler et al 2008 [34]	Ja Mats Adler, Karolinska Universitetssjukhuset, Huddinge	18 frågor, 5 minuter	Fritt tillgänglig för sjukvårdspersonal
CBCL Children Behaviour check List	6–18 år Föräldrabedömning	Manual: Achenbach T & Rescorla L. Manual fot the ASEBA School-age forms & profiles. Burlington, VT: University of Vermon, Research center for children, youth and families. 2001	Ja Information via Birgitta Bäcklund eller Bruno Hägglöf, Umeå universitet	120 frågor	Ja
CBQ Children Bipolar Questionnaire	Barn och ungdom Föräldrabedömning, klinikerbedömning	Papolos et al 2006 [35]	Ej identifierad	65 frågor	Oklart
CMRS-P Child Mania Rating Scale, Parent version	Barn 7–13 år Föräldrabedömning	Pavuluri et al 2006 [36]	Ja Copyright för den svenska versionen innehas av Håkan Jarbin	21 frågor, 10 minuter	Forskare och kliniker inom offentlig verksamhet får kopiera den svenska versionen för eget bruk inom klinik eller forskning
GBI General Behavior Inventory	Barn Versioner finns för både föräldra- och självbedömningar	Klein et al 1989 [37]	Ej identifierad	73 frågor	Oklar
HCL-32 Hypomania Checklist	Från 18 år Självbedömning	Angst et al 2005 [38]	Ja	9 frågor	Kostnadsfritt
LCM-p Life Chart Methodology (p=prospektiv)	Från 13 år Stämningsdagböcker	Roy-Byrne et al 1985 [39]	Flera svenska varianter finns	Inte tillämpligt	Oklart
MDQ Mood Disorder Questionnaire Samt MDQ-A	Från 18 år Självbedömning MDQ-A för tonåringar Föräldrabedömning	Hirschfeld et al 2000 [40]	Ja Lars Häggström med tillstånd från upphovs-mannen (vuxenversionen)	13 frågor + ytterligare 2 frågor	Kostnadsfritt (vuxenversionen)
YMRS Young Mania Rating Scale	5–17 år Föräldrabedömning samt klinikerbedömning Vuxna Klinikerbedömning	Young et al 1978 [41]	Ja För vuxenformuläret av Peter Skeppar 1999. Översättning kräver inget till-stånd av upphovsmannen. Ingen återöversättning har gjorts	11 frågor, 15–20 minuter	Kostnadsfritt

Tabell 4.4 *Formulär för bedömning av svårighetsgrad av depression.*

Formulär	Ålder och bedömare	Originalpublikation	Översättning	Omfattning	Kostnad
BDI-I och BDI-II	Se Tabell 4.2				
HDRS-17 (SIGH-D) Hamilton Rating Scale for Depression	Vuxna Expertbedömning	Hamilton 1960 [42]	Ja	Finns i flera versioner men grundversionen innehåller 17 frågor	Kostnadsfri version cirkulerar
IDS-30/QIDS-16 Inventory of Depressive Symtomatology	Vuxna Självbedömning eller expertbedömning	Rush et al 1986 [43]	Ej identifierad för IDS Ja för QIDS	IDS 30 frågor, QIDS 16 frågor	QIDS kostnadsfritt
MADRS och MADRS-S Montgomery Åsberg Depression Rating Scale	Vuxna Självbedömning eller expertbedömning	Montgomery et al 1979 [44] Svanborg et al 1994 [45]	Svenska är originalspråk	Expertbedömning cirka 1 timme. Självskattning 10 frågor, cirka 5–10 minuter	Kostnadsfritt
SDS Sung self rating depression scale	Vuxna Självbedömning	Zung 1965 [46]	Oklart	20 frågor, 5 minuter	Oklart

Tabell 4.5 *Formulär för bedömning av svårighetsgrad av bipolära syndrom.*

Formulär	Ålder och bedömare	Originalpublikation	Översättning	Omfattning	Kostnad
BDRS Bipolar Depression Rating Scale	Från 18 år	Berk et al 2007 [47]	Ja Översatt av Lars Häggström i samarbete med AstraZeneca AB med tillstånd av Berk och medarbetare	20 frågor	Kostnadsfritt
MAS Bech-Rafaelsen Mania Rating Scale	Från 18 år Semistrukturerad intervju	Bech et al 1978 [48]	Ja	11 frågor	Ja