

# Diagnostik och uppföljning av förstämningssyndrom

---

En systematisk litteraturöversikt

Oktober 2012



---

SBU • Statens beredning för medicinsk utvärdering  
*Swedish Council on Health Technology Assessment*

# SBU utvärderar sjukvårdens metoder

SBU, Statens beredning för medicinsk utvärdering, är en statlig myndighet som utvärderar hälso- och sjukvårdens metoder.

SBU analyserar metodernas nytta, risker och kostnader och jämför vetenskapliga fakta med svensk vårdpraxis. Målet är att ge ett bättre beslutsunderlag för alla som avgör hur vården ska utformas.

SBU ger ut flera rapportserier. I ”SBU Utvärderar” har SBU:s expertgrupper själva gjort den systematiska utvärderingen. Serien omfattar både etablerade metoder (gula rapporter) och nya metoder (Alert). ”SBU Kommenterar” sammanfattar och kommenterar utländska medicinska kunskapsöversikter. SBU svarar också på frågor direkt från beslutsfattare i vården via SBU:s Upplysningstjänst.

Välkommen att läsa mer om SBU:s rapporter och verksamhet på [www.sbu.se](http://www.sbu.se).

*Denna utvärdering publicerades år 2012. Resultat som bygger på ett starkt vetenskapligt underlag fortsätter vanligen att gälla under en lång tid framåt. Andra resultat kan ha hunnit bli inaktuella. Det gäller främst områden där det vetenskapliga underlaget är otillräckligt eller begränsat.*

## **Denna rapport (nr 212) kan beställas från:**

SBU, Box 3657, 103 59 Stockholm  
Besöksadress: Olof Palmes Gata 17  
Telefon: 08-412 32 00 • Fax: 08-411 32 60  
[www.sbu.se](http://www.sbu.se) • E-post: [registrator@sbu.se](mailto:registrator@sbu.se)

Grafisk produktion av Anna Edling, SBU  
Tryckt av Elanders, 2012

Rapportnr: 212 • ISBN 978-91-85413-52-2 • ISSN 1400-1403

# Diagnostik och uppföljning av förstämningssyndrom

---

En systematisk litteraturöversikt

## Projektgrupp

Lisa Ekselius (ordförande)	Petter Gustavsson Kerstin Hagenfeldt	Agneta Pettersson (projektledare)
Derya Akcan (informationsspecialist)	Susanna Kjellander (biträdande projektledare)	Mia Ramklint Charlotta Sunnqvist
Per Allard	Ingvar Krakau	Cecilia Svanborg
Kristina Bengtsson Boström	Anders Norlund (hälsoekonom)	Anneth Syversson (projektassistent)
Thomas Davidson (hälsoekonom)		Maria Tillfors

## Externa granskare

Malin André	Ann-Charlotte Smedler
Ata Ghaderi	Ingmar Skoog
Mikael Landén	Hans Ågren



# Innehåll

---

<b>SBU:s sammanfattning och slutsatser</b>	<b>13</b>
<b>1. Inledning</b>	<b>45</b>
Syfte	45
Bakgrund	46
Klassificering av förstämningssyndromen	46
Förekomst och samsjuklighet	49
<i>Barn och unga</i>	49
<i>Äldre</i>	50
<i>Riskgrupper</i>	50
Den diagnostiska processen	50
Användning av formulär och skalor som stöd för den diagnostiska processen	51
Bedömning av nyttan av en diagnostisk metod	54
Val av referensstandard	54
Mått på diagnostisk tillförlitlighet	55
Bedömning av svårighetsgrad och förändring	57
Rapportens innehåll	57
Referenser	58
<b>2. Metodbeskrivning</b>	<b>61</b>
Inklusionskriterier	61
Urval av studier	62
Bedömning av studiernas vetenskapliga kvalitet	62
Sammanvägning av resultaten	64
Styrka på det vetenskapliga underlaget	65
Referenser	67
<b>3. Systematisk litteraturgenomgång</b>	<b>69</b>
3.1 Frågor och avgränsningar	71
Frågor	71
Inklusionskriterier	72

Kvalitetskriterier	75
Urval av studier	77
Referenser	81
3.2 Påverkan på omhändertagande och depressionsförlopp	81
Evidensgraderade resultat	81
Introduktion	81
Beskrivning av underlaget	82
Bedömning av evidensstyrka	83
Evidensgraderade resultat	85
Diskussion	85
Tabeller	88
Referenser	92
3.3 Klinisk bedömning med och utan stöd av intervjuformulär	93
Evidensgraderade resultat	93
Introduktion	94
Klinisk diagnos av psykiatrer jämfört med LEAD	94
<i>Beskrivning av underlaget</i>	94
<i>Sammanvägda resultat</i>	95
<i>Bedömning av evidensstyrka</i>	96
<i>Evidensgraderade resultat</i>	98
Klinisk diagnos av psykiatrer jämfört med SCID-I	98
<i>Beskrivning av underlaget</i>	98
<i>Bedömning av evidensstyrka</i>	99
Klinisk diagnos av distriktsläkare jämfört med intervjubaserad diagnostik	99
<i>Beskrivning av underlaget</i>	99
<i>Sammanvägda resultat</i>	100
<i>Bedömning av evidensstyrka</i>	100
<i>Evidensgraderade resultat</i>	101
Diskussion	101
Tabeller	104
Referenser	112
3.4 Bedömningsinstrument för barn och ungdomar	113
Evidensgraderade resultat	113
<i>Intervjuer</i>	113
<i>Formulär för riktad screening</i>	113
<i>Formulär för att bedöma svårighetsgrad</i>	116

Introduktion	117
<i>Urval av studier</i>	120
<i>Studiekvalitet</i>	120
Semistrukturerade diagnostiska intervjuer	121
<i>Schedule for Affective Disorders and Schizophrenia for School Aged Children (Kiddie-SADS/K-SADS) jämfört med LEAD</i>	121
<i>Diagnostic Interview for Children and Adolescents (DICA)</i>	122
<i>Structured Clinical Interview for DSM-IV, Childhood Version (KID-SCID)</i>	123
Strukturerade diagnostiska intervjuer	123
<i>Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID)</i>	123
<i>Development and Well-Being Assessment (DAWBA)</i>	124
<i>Brief Child and Family Phone Interview (BCPFI)</i>	124
Formulär för riktad screening av depression hos barn och ungdomar	125
<i>Achenbach System of Empirically Based Assessment (ASEBA), Child Behaviour Checklist (CBCL), Youth Self Report (YSR), Teacher Report Form (TRF)</i>	125
<i>Child Behaviour Checklist (CBCL), 6–18 år</i>	126
<i>Youth Self Report (YSR)</i>	128
<i>Teacher Report Form (TRF)</i>	129
<i>Becks ungdomsskalor (BUS)</i>	129
<i>Beck Depression Inventory for Children (BDI-I)</i>	130
<i>Birleson Depression Self Rating Scale (DSRS)</i>	134
<i>Children's Depression Inventory (CDI)</i>	135
<i>Center of Epidemiological Studies Depression Scale for Children (CES-DC)</i>	137
<i>Patient Health Questionnaire (PHQ-9)</i>	138
<i>Strength and Difficulties Questionnaire (SDQ)</i>	138
Formulär för riktad screening av bipolärt syndrom	139
<i>Child Behaviour Checklist (CBCL), 6–18 år</i>	139
<i>General Behavior Inventory (GBI)</i>	140
<i>Child Bipolar Questionnaire (CBQ)</i>	143
<i>Mood Disorder Questionnaire – Adolescent (MDQ-A)</i>	143

<i>Child Mania Rating Scale – Parent (CMRS-P)</i>	144
<i>Young Mania Rating Scale (YMRS)</i>	145
Formulär för att bedöma svårighetsgrad av förstämningssyndrom	148
Diskussion	148
<i>Intervjuer</i>	148
<i>Formulär för riktad screening</i>	149
Tabeller	152
Referenser	200
3.5 Bedömningsformulär för vuxna	209
Evidensgraderade resultat	209
<i>Intervjuformulär</i>	209
<i>Riktad screening av depression</i>	210
<i>Riktad screening för bipolära syndrom</i>	212
<i>Bedömning av svårighetsgrad</i>	214
<i>Uppföljning av behandlingseffekter</i>	215
<i>Samstämmighet av skattning av depression     mellan kliniker och patient</i>	215
Introduktion	216
<i>Urval av studier</i>	220
<i>Studiekvalitet</i>	221
Semistrukturerade diagnostiska intervjuer	222
<i>Structured Clinical Interview for DSM-Axis I     psychiatric disorders (SCID-I)</i>	222
<i>Schedule for Affective Disorders and Schizophrenia     (SADS)</i>	227
<i>Schedules for Clinical Assessment in Neuropsychiatry     (SCAN)</i>	228
Strukturerade diagnostiska intervjuer	228
<i>Composite International Diagnostic Interview     (CIDI)</i>	228
<i>Diagnostic Interview Schedule (DIS)</i>	229
<i>Mini International Neuropsychiatric Interview     (MINI)</i>	230
<i>The Primary Care Evaluation of Mental Disorders     (PRIME-MD)</i>	234
<i>Structured Psychiatric Interview for General Practice     (SPIFA)</i>	237



Riktad screening av depression	238
<i>Beck Depression Inventory-II (BDI-II)</i>	238
<i>Calgary Depression Scale for Schizophrenia (CDSS)</i>	241
<i>Center for Epidemiologic Studies of Depression (CES-D)</i>	242
<i>Edinburgh Postnatal Depression Scale (EPDS)</i>	244
<i>Hospital Anxiety and Depression Scale (HADS-D)</i>	250
<i>Patient Health Questionnaire (PHQ-9)</i>	254
<i>Major Depression Inventory (MDI)</i>	259
<i>Seasonal Health Questionnaire (SHQ)</i>	260
<i>Seasonal Pattern Assessment Questionnaire (SPAQ)</i>	261
Riktad screening av bipolärt syndrom	262
<i>Hypomania Checklist-32 (HCL-32)</i>	262
<i>Mood Disorder Questionnaire (MDQ)</i>	265
Formulär för bedömning av svårighetsgrad av depression	272
<i>Beck Depression Inventory-I och -II (BDI-I, BDI-II)</i>	272
<i>Calgary Depression Scale for Schizophrenia (CDSS)</i>	273
<i>Hamilton Rating Scale for Depression (HDRS-17)</i>	274
<i>Inventory of Depressive Symptomatology (IDS och QIDS)</i>	275
<i>Montgomery-Åsberg Depression Rating Scale (MADRS)</i>	276
<i>Zung Self-Rated Depression Scale (SDS)</i>	277
Formulär för bedömning av svårighetsgrad av bipolärt syndrom	278
<i>Affective Self Rating Scale-18 (AS-18)</i>	278
<i>Bipolar Depression Rating Scale (BDRS)</i>	279
<i>Bech-Rafaelsen Mania Scale (MAS)</i>	280
<i>Young Mania Rating Scale (YMRS)</i>	280
<i>Life Chart Methodology (LCM)</i>	281
Uppföljning av behandlingseffekter	282
Samstämmighet mellan kliniker och patient	282
<i>Inventory of Depressive Symptomatology (IDS och QIDS)</i>	283
<i>Montgomery-Åsberg Depression Rating Scale (MADRS)</i>	285
Diskussion	287

Tabeller	294
Referenser	440
3.6 Bedömningsformulär för äldre	455
Evidensgraderade resultat	455
<i>Riktad screening av depression</i>	455
Introduktion	456
Urval av studier	457
Studiekvalitet	465
Semistrukturerade och strukturerade diagnostiska intervjuer	460
Formulär för riktad screening av depression hos äldre	460
<i>Geriatric Depression Scale (GDS-30)</i>	460
<i>Geriatric Depression Scale (GDS-15)</i>	463
<i>Geriatric Depression Scale (GDS-20)</i>	466
<i>Hospital Anxiety and Depression Scale (HADS)</i>	467
<i>Patient Health Questionnaire-9 (PHQ-9)</i>	469
<i>Cornell Scale for Depression in Dementia (CSDD)</i>	471
<i>Beck Depression Inventory (BDI-II)</i>	472
<i>Center for Epidemiologic Studies Depression Scale (CES-D)</i>	473
<i>Stroke Inpatient Depression Inventory (SIDI)</i>	473
Formulär för bedömning av svårighetsgrad av depression	474
<i>Montgomery-Åsberg Depression Rating Scale (MADRS)</i>	474
<i>Övriga formulär för att bedöma svårighetsgrad</i>	474
Diskussion	474
Tabeller	480
Referenser	512
3.7 Hälsoekonomi	517
Slutsatser	517
Bakgrund	517
Metod	517
Resultat	517
Tabell	518
Referenser	520

<b>4. Inventering</b>	<b>521</b>
Tabeller	524
Referenser	538
<b>5. Klinisk betydelse av prestanda hos olika tester</b>	<b>543</b>
Sensitivitet och specificitet – vad betyder de i kliniken?	543
Diagnos av depression i primärvården	544
Diagnostik av depression med PHQ-9	545
Riktad screening av bipolära syndrom	546
Referenser	548
<b>6. Etiska och sociala aspekter</b>	<b>549</b>
Frågor	549
Urval av litteratur	550
Vilka är de etiska konsekvenserna av resultaten av granskningen?	550
<i>Göra gott och inte skada</i>	550
Vilka är de etiska aspekterna på att använda bedömningsformulär på ett sätt som inte är <i>lege artis</i> ?	551
Etiska risker med riktad screening	552
Vilka är de etiska aspekterna på att alla patienter inte kan bedömas med formulär?	553
<i>Konsekvenser av språk- och läsvårigheter</i>	553
<i>Konsekvenser av kulturella skillnader</i>	554
Finns det etiska risker med att utbildningen om bedömningsformulär varierar?	554
Referenser	556
<b>7. Användning av bedömningsformulär i Sverige</b>	<b>559</b>
Vetenskapligt underlag	560
Hur vanligt är det att använda bedömningsformulär vid förstämningssyndrom?	561
<i>Strukturerade och semistrukturerade intervjuer</i>	561
<i>Riktad screening</i>	562
<i>Uppföljning av behandling av depression och mani</i>	563
Rekommenderas bedömningsformulär för förstämningssyndrom i regionala och lokala vårdprogram?	563

Hur ser läkare och annan vårdpersonal på att använda bedömningsformulär?	564
<i>Data från publicerade studier</i>	564
<i>Uppfattningar från svenska fokusgrupper och enkäter</i>	566
<i>Hur ser patienterna på att bedömningsformulär används?</i>	568
Diskussion	569
Referenser	570
<b>8. Kunskapsluckor</b>	<b>573</b>
<b>9. Konsekvenser av rapportens resultat</b>	<b>575</b>
Önskvärda förändringar	575
Konsekvenser av att införa en strukturerad diagnostisk process	577
<i>För patienterna</i>	577
<i>För sjukvården</i>	577
Referenser	579
<b>10. Ord- och förkortningslista</b>	<b>581</b>
Förkortningslista – Bedömningsformulär	
<b>11. Projektgrupp, externa granskare, bindningar och jäv</b>	<b>597</b>
<b>Bilaga 1. Sökstrategier</b>	
Publicerade på <a href="http://www.sbu.se/212_bilagor">www.sbu.se/212_bilagor</a>	
<b>Bilaga 2. Checklistor och mallar</b>	
Publicerade på <a href="http://www.sbu.se/212_bilagor">www.sbu.se/212_bilagor</a>	
<b>Bilaga 3. Exkluderade studier</b>	
Publicerade på <a href="http://www.sbu.se/212_bilagor">www.sbu.se/212_bilagor</a>	

# SBU:s sammanfattning och slutsatser

---



---

SBU • Statens beredning för medicinsk utvärdering  
*Swedish Council on Health Technology Assessment*



# SBU:s sammanfattning och slutsatser

---

Förstämningssyndrom är vanliga och allvarliga sjukdomstillstånd. Eftersom det finns effektiv behandling för förstämningssyndromen är det viktigt att diagnostisera dem på ett säkert sätt.

Detta är en systematisk översikt om bedömningsformulär som stöd för diagnostik och uppföljning av förstämningssyndrom. Översikten avser de cirka 60 formulär som finns tillgängliga i Sverige.

Syftet med granskningen var att klargöra nyttan med att använda bedömningsformulär samt att bedöma vilka av dem som har vetenskapligt stöd för en tillräcklig diagnostisk tillförlitlighet. Vi har också sammanställt litteratur om hur patienter och kliniker uppfattar användning av diagnostiska hjälpmedel och kompletterat med egna praxisundersökningar.

Rapporten är en del i ett regeringsuppdrag om utvärdering av diagnostik och behandling för psykiatriska tillstånd.

## Slutsatser

- ❑ Det kliniska samtalet räcker inte för att diagnostisera depression och bipolära syndrom. Tillstånden upptäcks ofta inte. När bedömningen kompletteras med hjälpmedel i form av strukturerade eller semistrukturerade intervjuformulär och tillgång till journalhandlingar ökar den diagnostiska känsligheten (sensitivitet) och träffsäkerheten (specificitet).
- ❑ Två intervjuformulär, som täcker en hel rad psykiatriska diagnoser, har vetenskapligt stöd för en tillräckligt hög känslighet och träffsäkerhet för depression och mani hos vuxna. Det är den semi-

strukturerade *Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I)* och den strukturerade *Mini International Neuropsychiatric Interview (MINI)*. Den strukturerade intervjun *Primary Care Evaluation of Mental Disorders (PRIME-MD)* har för låg känslighet.

- ❑ Vid riktad screening av depression har följande formulär vetenskapligt stöd för tillräcklig känslighet: *Child Behaviour Check List (CBCL)* och *Beck Depression Inventory-I (BDI-I)* för barn och ungdomar, *Beck Depression Inventory-II (BDI-II)* för vuxna samt *Geriatric Depression Scale (GDS-15 och GDS-30)* för äldre utan påtaglig kognitiv svikt. Screening av nyförlösta mödrar med *Edinburgh Postnatal Depression Scale (EPDS)* identifierar cirka två tredjedelar av dem som har depression, vilket är i underkant. Formulären *Patient Health Questionnaire-9 (PHQ-9)* och *Hospital Anxiety and Depression Scale (HADS)* har för odefinierbar känslighet för att vara användbara.

*Young Mania Rating Scale (YMRS)* och *General Behavior Inventory (GBI)* för barn och ungdomar samt *Hypomania Check List-32 (HCL-32)* respektive *Mood Disorder Questionnaire (MDQ med modifierade krav på funktionsnedsättning)* för vuxna har vetenskapligt stöd för riktad screening av bipolärt syndrom.

- ❑ Det går inte att bedöma tillförlitligheten hos formulär som mäter svårighetsgrad av depression och mani. Det finns för få studier med tillräcklig kvalitet.
- ❑ Det går inte att bedöma tillförlitligheten hos formulär som är avsedda att följa förlopp av depression och mani. Det saknas studier.
- ❑ Det saknas studier som utvärderar kostnadseffektiviteten av att använda bedömningsformulär för förstämningssyndrom.



- ❑ Det finns många kunskapsluckor. Det viktigaste forskningsbehovet är att utvärdera hur strukturerad diagnostik som del i en systematisk handläggning påverkar tillfrisknande, relation mellan patient och behandlare samt kostnadseffektivitet. Utvärdering av intervjuformulär för barn och ungdomar samt formulär för bedömning av depression hos äldre med kognitiv svikt är också angeläget. Det behövs ny forskningsmetodik för att utvärdera formulär som syftar till att följa sjukdomsförlopp.
- ❑ Användning av bedömningsformulär för förstämningssyndrom förutsätter att patienten har tillräckliga språkkunskaper samt, vid självbedömning, tillräcklig läs- och skrivförmåga. Det innebär att personer från andra kulturer, med läs- och skrivsvårigheter eller nedsatt kognitiv förmåga kan bli feltolkade eller ha svårt att fylla i formulären.

## **Bakgrund**

Förstämningssyndrom kännetecknas av en varaktig förskjutning i stämningensläget och innefattar såväl sjuklig nedstämdhet, dvs depression, som sjuklig upprymdhet, mani. Förstämningssyndromen delas in i unipolära och bipolära syndrom beroende på typ av förändring i sinnesstämning, allvarlighetsgrad och sjukdomsförlopp över tid (Tabell 1). Diagnoserna ställs utifrån definierade kriterier i klassifikationssystemen International Classification of Diseases (ICD) respektive Diagnostic and Statistical Manual of Mental Disorders (DSM). Kriterierna är gemensamma för alla åldersgrupper.

**Tabell 1** Förstämmningssyndrom som ingår i rapporten  
(modifierat från APA, *American Journal of Psychiatry*, 2002).

<b>Diagnos</b>	<b>Maniska symtom</b>	<b>Depressiva symtom</b>
Egentlig depression	Aldrig någon manisk eller hypoman episod	Enstaka eller återkommande depressiva episoder som varar minst två veckor
Dystymi	Aldrig någon manisk eller hypoman episod	Sänkt stämningsläge de flesta dagar under minst två år
Bipolärt syndrom typ I	Episoder med mani eller en blandning av mani och depression (blandepisod)	Typiskt med depressiva episoder men det krävs inte för diagnos
Bipolärt syndrom typ II	En eller flera episoder med hypomana episoder men aldrig någon manisk eller blandepisod	Återkommande episoder med depressiva episoder
Schizoaffektivt syndrom*	Betydande varaktighet av maniska symtom eller blandtillstånd samtidigt med symtom på schizofreni samt period om minst två veckor med psykotiska symtom utan mani	Betydande varaktighet av depressiva symtom samtidigt med symtom på schizofreni samt period på minst två veckor med psykotiska symtom utan depression

\* Schizoaffektivt syndrom hör till de psykotiska syndromen. Behandling av schizoaffektivt syndrom har många likheter med behandling av bipolära syndrom och det kan vara en svår differentialdiagnos.

Det vanligaste förstämmningssyndromet är egentlig depression. Några speciella former av depression som tas upp i rapporten är årstidsrelaterad depression (seasonal affective disorder) och depression efter förlossning (post partum-depression). Den senare bör särskilt betonas eftersom post partum-depression inte bara är förenat med problem för modern utan även kan påverka det nyfödda barnet negativt.

### **Förstämmningssyndrom är vanligt och allvarligt**

Globalt tillhör förstämmningssyndrom de vanligaste orsakerna till ohälsa, förtida död, produktionsbortfall och arbetsförmåga. Studier från USA och Europa fann relativt samstämmigt att cirka 9 procent av den vuxna befolkningen led av något förstämmningssyndrom under ett år. Drygt

80 procent utgjordes av egentlig depression. I en stor svensk undersökning, den så kallade Lundby-studien, var sannolikheten för att insjukna i en depressiv episod före 70 års ålder 23 procent för män och drygt 30 procent för kvinnor.

Det är angeläget att diagnostisera och behandla förstämningssyndrom eftersom de är förknippade med hög samsjuklighet med såväl kroppslig som psykisk ohälsa och med för tidig död. Risken för att en person med förstämningssyndrom avlider i förtid är ungefär dubbelt så stor som för befolkningen i helhet och risken för självmord är 15–20 gånger högre.

### **Diagnostik är en integrerad del av behandlingen**

Det finns effektiva behandlingar både mot depression och mani. En väl grundad diagnos är en viktig förutsättning för ett gott behandlingsarbete. Den ideala diagnostiska processen följer en metodik som brukar benämnas *LEAD (Longitudinal, Experts, All Data procedure)*. LEAD innebär att information från samtal med patient, och ibland även närstående, tillsammans med resultat från genomförda undersökningar, diagnostiska intervjuer och frågeformulär sammanställs och diskuteras i ett team av erfarna bedömare. När diagnosen fastställts bedöms svårighetsgrad av depression eller mani inför beslut om behandling. Effekterna av den insatta behandlingen utvärderas kontinuerligt, gärna med stöd av formulär.

### **Bedömningsformulär kan stödja hela den diagnostiska processen**

Formulär kan stödja den diagnostiska processen i alla led, från den första misstanken om ett förstämningssyndrom till systematisk uppföljning.

Benämningar på formulär är diffusa, ofta överlappande och ibland missvisande. Begreppet ”diagnostiska formulär” kan t ex användas för såväl diagnos som för riktad screening. ”Symtomskattning” eller ”skattningsskala” kan avse både formulär för riktad screening, för bedömning av svårighetsgrad och för uppföljning av behandling.

Screening som term för preliminär bedömning av symtom är det vedertagna begreppet i psykiatrisk litteratur och används t ex i vårdprogram

och riktlinjer. Vi har därför valt att använda begreppet screening även om det inte överensstämmer med WHO:s definition eftersom vi inte vill skapa onödig begreppsförvirring.

I rapporten har vi klassificerat bedömningsformulären enligt Tabell 2.

**Tabell 2** Indelning av bedömningsformulär för förstämningssyndrom.

<b>Benämning</b>	<b>Används för</b>	<b>Generellt eller specifikt</b>	<b>Bedömare</b>
Formulär för screening	Hitta individer i riskpopulation, t ex nyförlösta mödrar	Sjukdomsspecifikt	Ofta självskattning
Formulär för riktad screening	Stärka eller avfärda misstanke hos individ med symtom	Sjukdomsspecifikt	Ofta självskattning
Diagnostisk intervju	Stöd för att ställa diagnos	Inte sjukdomsspecifikt, många psykiatriska diagnoser och kan ge besked om samsjuklighet	Expert: Semistrukturerade intervjuer. Vårdpersonal med kortare träning: Strukturerade intervjuer
Formulär för svårighetsgrad	Stöd för att bedöma svårighetsgrad	Sjukdomsspecifikt	Expert- eller självskattning
Formulär för att följa effekter av behandling	Upprepade mätningar med hjälp av formulär för svårighetsgrad	Sjukdomsspecifikt	Expert- eller självskattning

Det kan inte nog betonas att bedömningsformulär är *hjälpmedel* och att de alla har olika styrkor och begränsningar. Användaren måste vara förtrogen med hur formulären utvecklats, deras syfte och tillförlitlighet.

## **Diagnostisk tillförlitlighet mäts helst med LEAD som referens**

Nyttan av att använda olika formulär avgörs av om de påverkar behandlingsstrategier och patienternas hälsa. Men få studier utvärderar konsekvenserna av diagnostiska åtgärder. Formulären bedöms då istället utifrån deras diagnostiska tillförlitlighet, som beskriver hur väl formulären kan skilja sjuka från friska. Gränsen mellan ”sjuk” och ”frisk” utgör det så kallade *tröskelvärdet*.

Den diagnostiska tillförlitligheten relateras alltid till en referensstandard. En ideal referensstandard, dvs en gold standard, kan identifiera sjukdom eller frånvaro av sjukdom på ett helt säkert sätt. I praktiken är sällan referensstandarderna perfekt, vilket i sig introducerar en källa till osäkerhet. LEAD-metodiken anses vara den optimala referensstandarderna inom psykiatrin, men metodiken är tidskrävande och få studier har utförts med LEAD som referens. Helt eller delvis strukturerade intervjuformulär, framför allt SCID-I, är acceptabla alternativ som används oftare.

Den diagnostiska tillförlitligheten beror på i vilken grad index- och referenstesterna klassificerar tillståndet på samma sätt (sant positiva, dvs sjuka, respektive sant negativa, dvs friska) vid ett givet tröskelvärde. Utifrån de värdena kan man sedan beräkna formulärets känslighet och träffsäkerhet (Faktaruta 1).

### **Faktaruta 1** Begreppen känslighet och träffsäkerhet.

Känslighet (sensitivitet): Andelen sjuka som identifieras av ett test.

Träffsäkerhet (specificitet): Andelen friska som identifieras av ett test.

## Rapportens frågor

Vi utvärderar:

- om bedömningsformulär förbättrar diagnostiken
- om diagnostik med stöd av bedömningsformulär förbättrar omhändertagande och minskar symtombördan
- vilka formulär som har tillräcklig diagnostisk tillförlitlighet
- om det finns formulär som är kostnadseffektiva
- om det finns några etiska aspekter att ta hänsyn till.

Vi definierade att tillräcklig diagnostisk tillförlitlighet för intervjuformulär var 80 procent känslighet och 80 procent träffsäkerhet. Tillräcklig diagnostisk tillförlitlighet för formulär för riktad screening definierades som minst 80 procent känslighet.

Vi sökte litteratur som avgränsades på det sätt som beskrivs i Faktaruta 2.

### Faktaruta 2 Avgränsningar (inklusionskriterier).

**Population:** Barn, vuxna och äldre med symptom som tydde på förstämningssyndrom eller som utgjorde en riskgrupp för förstämningssyndrom, t ex nyblivna mödrar. Studierna skulle ha utförts i Europa, Nordamerika och Australien/Nya Zeeland. Undantaget var studier om äldre där vi inte hade någon geografisk begränsning.

**Indextest:** Bedömningsformulären utvärderades enbart för det ursprungligen tänkta användningsområdet och tröskelvärdet.

**Referensstandard:** Strukturerade eller semistrukturerade intervjuer. Undantaget var formulär som mäter svårighetsgrad, vilka skulle jämföras mot SCID-I, Clinical Global Impression-Severity Scale (CGI-S eller CGI-BP).

**Utfallsmått:** Sensitivitet (känslighet) och specificitet (träffsäkerhet). För formulär för att bedöma svårighetsgrad använde vi olika sambandsmått som t ex intraklasskorrelation.

**Tid mellan index- och referenstesterna:** En vecka för depression, fyra dagar för mani och ett dygn för hypomani. För studier som undersökte formulär för svårighetsgrad eller frågan om samskattning satte vi gränsen ett dygn.

### Faktaruta 3 Studiekvalitet och evidensstyrka.

**Studiekvalitet** avser den vetenskapliga kvaliteten hos en enskild studie och dess förmåga att besvara en viss fråga på ett tillförlitligt sätt.

**Evidensstyrkan** är en bedömning av hur starkt det sammanlagda vetenskapliga underlaget är för att besvara en viss fråga på ett tillförlitligt sätt. SBU tillämpar det internationellt utarbetade evidensgraderingssystemet GRADE. För varje effektmått utgår man i den sammanlagda bedömningen från studiernas design. Därefter kan evidensstyrkan påverkas av förekomsten av försvagande eller förstärkande faktorer som studiekvalitet, relevans, samstämmighet, överförbarhet, effektstorlek, precision i data, risk för publikationsbias och andra aspekter, t ex dos-responssamband.

Evidensstyrka graderas i fyra nivåer:

**Starkt vetenskapligt underlag (⊕⊕⊕⊕)**

Bygger på studier med hög eller medelhög kvalitet utan försvagande faktorer vid en samlad bedömning.

**Måttligt starkt vetenskapligt underlag (⊕⊕⊕○)**

Bygger på studier med hög eller medelhög kvalitet med förekomst av enstaka försvagande faktorer vid en samlad bedömning.

**Begränsat vetenskapligt underlag (⊕⊕○○)**

Bygger på studier med hög eller medelhög kvalitet med försvagande faktorer vid en samlad bedömning.

**Otillräckligt vetenskapligt underlag (⊕○○○)**

När vetenskapligt underlag saknas, tillgängliga studier har låg kvalitet eller där studier av likartad kvalitet visar motsägande resultat, anges det vetenskapliga underlaget som otillräckligt.

Ju starkare evidens desto mindre sannolikt är det att redovisade resultat kommer att påverkas av nya forskningsrön inom överblickbar framtid.

**Slutsatser**

I SBU:s slutsatser görs en sammanfattande bedömning av nytta, risker och kostnadseffektivitet.

## Metod

Den systematiska översikten genomfördes i enlighet med SBU:s procedur. Styrkan på det vetenskapliga underlaget bedömdes med hjälp av GRADE (Faktaruta 3).

## Resultat av den systematiska litteraturgranskningen

Utvärderingen omfattade de drygt 60 formulär enligt en inledande praxisundersökning som används i Sverige idag.

För många av formulären saknades studier som uppfyllde våra krav på undersökningsgrupp, referensstandard eller tid mellan testerna, och de finns förtecknade i Tabell 11. För andra formulär fanns enstaka studier men inte tillräckligt många av god kvalitet för att kunna dra slutsatser om deras diagnostiska tillförlitlighet. Dessa formulär finns sammanställda i Tabell 10.

Fullständiga namn på de formulär som kunde granskas redovisas i Tabell 9.

## Riktad screening som del i förstärkt handläggning kan minska symtom på depression

Vår fråga var om omhändertagandet och patienternas symtombörda påverkas av att bedömningsformulär används i samband med diagnos eller uppföljning.

Vårt underlag bestod av fem systematiska översikter varav två var fokuserade på primärvården, en rörde depression efter förlossning, en depression hos barn och ungdom och en depression i samband med hjärt- och kärlsjukdom. Samtliga utvärderade effekten av riktad screening. I några av de ingående studierna var en systematisk diagnostik del av en intervention med förstärkta resurser för behandling.

Översikterna visade att screening enbart av vuxna, oselektade patienter i väntrummet inte påverkade symtombördan hos dem som faktiskt hade



depression (starkt vetenskapligt underlag ⊕⊕⊕⊕). Om screeningen begränsades till patienter i primärvården som tillhörde någon riskpopulation för depression gav studierna inga entydiga resultat.

Om screening av riskgrupper ingick som del i en förstärkt handläggning i primärvården påverkades tillfrisknande och symtombörda i liten men signifikant utsträckning (begränsat vetenskapligt underlag ⊕⊕○○) men det gick inte att avgöra hur stor del av förbättringen som kunde hänföras till diagnostiken. Detsamma gällde för screening av nyförlösta kvinnor med EPDS där den genomsnittliga symtombördan minskade med cirka 40 procent (OR = 0,61; 0,95 % KI, 0,48 till 0,76) (begränsat vetenskapligt underlag ⊕⊕○○).

Däremot gick det inte att bedöma om vårdrutinerna i primärvården påverkades eftersom studierna inte var samstämmiga (otillräckligt vetenskapligt underlag ⊕○○○).

Vi kunde inte finna någon studie som undersökte värdet av att följa svårighetsgrad av förstämningssyndrom över tid med hjälp av symptomskattningsskalor. Men en systematisk översikt som gällde många olika psykiatriska diagnoser tydde på att det kan finnas ett värde med systematisk uppföljning med stöd av formulär.

## **Klinisk bedömning är inte tillräcklig vid misstanke om förstämningssyndrom**

En av våra frågor var om diagnostiskt stöd ökar möjligheten till en korrekt diagnos av förstämningssyndrom. Vi undersökte den diagnostiska tillförlitligheten för psykiatrikers respektive distriktsläkares kliniska bedömning utan hjälpmedel. Referensstandard kunde antingen vara LEAD eller intervjuer som var strukturerade eller semistrukturerade.

### **Distriktsläkarens kliniska diagnos**

Vårt underlag bestod av en systematisk översikt av drygt 40 studier med över 50 000 patienter. Ingen av studierna var genomförd i Sverige. Referensstandarder var intervjuer eller annan kriteriebaserad diagnostik.

I genomsnitt kunde distriktsläkarna identifiera depression hos 47 procent av dem som enligt referensstandarderna hade depression (begränsat vetenskapligt underlag ⊕⊕○○) men andelen varierade från 7 till 79 procent. Distriktsläkarnas förmåga att ställa diagnosen egentlig depression varierade mellan länderna. Den var högst i Italien och lägst i Storbritannien.

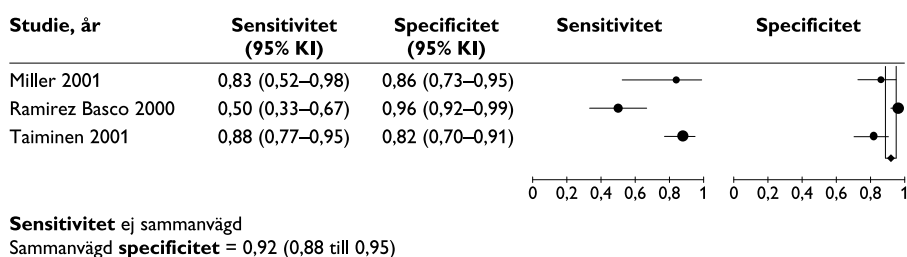
Liksom psykiatrerna kunde distriktsläkarna i högre grad utesluta dem som var friska, och träffsäkerheten uppgick till 81 procent (begränsat vetenskapligt underlag ⊕⊕○○).

## Psykiatrers kliniska diagnos

Tre studier med cirka 400 patienter jämförde klinisk diagnos med LEAD. I alla studierna var överensstämmelsen mellan klinisk diagnos ställd av psykiatrer och LEAD ganska låg.

I de två studier som genomfördes i slutenvården kunde psykiatrerna identifiera över 80 procent av de deprimerade och hälften av patienterna med bipolära syndrom (Figur 1). Öppenvårdsstudien gav samma resultat som i primärvårdsstudierna, dvs läkarna upptäckte endast cirka hälften av de deprimerade patienterna.

Däremot uteslöt psykiatrerna med hög säkerhet de som inte led av förstämningssjukdom. Träffsäkerheten låg över 90 procent både för depression och bipolära syndrom (starkt vetenskapligt underlag ⊕⊕⊕⊕).



**Figur 1** Metaanalys av känslighet (sensitivitet) och träffsäkerhet (specificitet) för klinisk diagnos av egentlig depression i psykiatrin med LEAD som referensstandard.

## Andelen vuxna med förstämningssyndrom som identifieras kan öka om SCID-I eller MINI används som komplement

Vi undersökte vilken känslighet och träffsäkerhet för förstämningssyndrom de strukturerade och semistrukturerade intervjuer som används i Sverige har. Det fanns för få studier för att kunna bedöma intervjuer för barn och ungdomar och inga studier alls med äldre. Det fanns tillräckligt med välgjorda studier för att kunna bedöma tre intervjuformulär som används för vuxna: den semistrukturerade SCID-I samt de strukturerade MINI och PRIME-MD. Tabell 3 sammanställer uppgifter om känslighet, träffsäkerhet och evidensstyrka för dem.

SCID-I och MINI har tillräcklig diagnostisk tillförlitlighet för depression och bipolära syndrom medan PRIME-MD har för låg känslighet för depression för att vara användbar. Tillförlitligheten för SCID-I vid bipolära och schizoaffektiva syndrom ökar om intervjun utförs med tillgång till journalhandlingar.

**Tabell 3** Intervjuformulär, med strukturerade eller semistrukturerade intervjuer som referensstandard.

Formulär	Syfte	Sensitivitet Vetenskapligt stöd	Specificitet Vetenskapligt stöd
SCID-I	Depression (mot LEAD)	86% (95% KI, 73; 94) Begränsat ⊕⊕○○	92% (95% KI, 88; 95) Starkt ⊕⊕⊕⊕
SCID-I	Bipolärt syndrom (mot LEAD)	79% (95% KI, 66; 88) Begränsat ⊕⊕○○	94% (95% KI, 90; 97) Starkt ⊕⊕⊕⊕
SCID-I	Schizoaffektivt syndrom (mot LEAD)	61% (95% KI, 46; 74) Begränsat ⊕⊕○○	97% (95% KI, 94; 99) Starkt ⊕⊕⊕⊕
MINI	Depression	95% (95% KI, 93; 97) Måttligt starkt ⊕⊕⊕○	84% (95% KI, 80; 87) Måttligt starkt ⊕⊕⊕○
MINI	Pågående mani	83% (95% KI, 71; 92) Begränsat ⊕⊕○○	95% (95% KI, 93; 97) Måttligt starkt ⊕⊕⊕○
MINI	Mani någon gång under livet	79% (95% KI, 71; 86) Begränsat ⊕⊕○○	93% (95% KI, 90; 95) Måttligt starkt ⊕⊕⊕○

Tabellen fortsätter på nästa sida

**Tabell 3** fortsättning

Formulär	Syfte	Sensitivitet Vetenskapligt stöd	Specificitet Vetenskapligt stöd
PRIME-MD	Depression	<70% Begränsat ⊕⊕○○	85% (95%KI, 82; 88) Begränsat ⊕⊕○○

KI = Konfidensintervall; MINI = Mini international Neuropsychiatric interview; PRIME-MD = The primary care evaluation of mental disorders; SCID-I = Structured clinical interview for DSM-IV disorders

Vad betyder nu dessa värden för känslighet (sensitivitet) och träffsäkerhet (specificitet) i praktiken? Ett illustrativt exempel är MINI. Intervjun är i och för sig enbart utvärderad för öppenvårdspsykiatri, men eftersom formuläret är strukturerat och kan användas av tränad lekman bör känslighet och träffsäkerhet vara ungefär desamma i primärvården.

Ett rimligt antagande är att 20 av 100 patienter i primärvården har depression. Om läkarens diagnostiska tillförlitlighet är cirka 50 procent känslighet och 80 procent träffsäkerhet kommer 10 av 20 deprimerade upptäckas samtidigt som 10 deprimerade inte upptäcks. Om MINI används som komplement kommer antalet som identifieras öka till 19 av 20. Däremot ändras inte antalet patienter utan depression som kan uteslutas eftersom läkaren och MINI har ungefär samma träffsäkerhet.

### **Några formulär för riktad screening har tillräcklig känslighet för depression**

Utvärderingen omfattade 21 bedömningsformulär. Tabell 4 sammanställer resultaten för de sju formulär som hade tillräckligt många studier för att vi skulle kunna bedöma den diagnostiska tillförlitligheten.

Vi kom fram till att BDI har tillräcklig känslighet för både barn, ungdomar och vuxna. En svaghet med formuläret är att träffsäkerheten är låg eller osäker, vilket medför en risk för att även personer som inte är deprimerade klassificeras som sjuka.

Ett alternativ för barn och ungdomar är CBCL som också har högre träffsäkerhet jämfört med BDI-I.

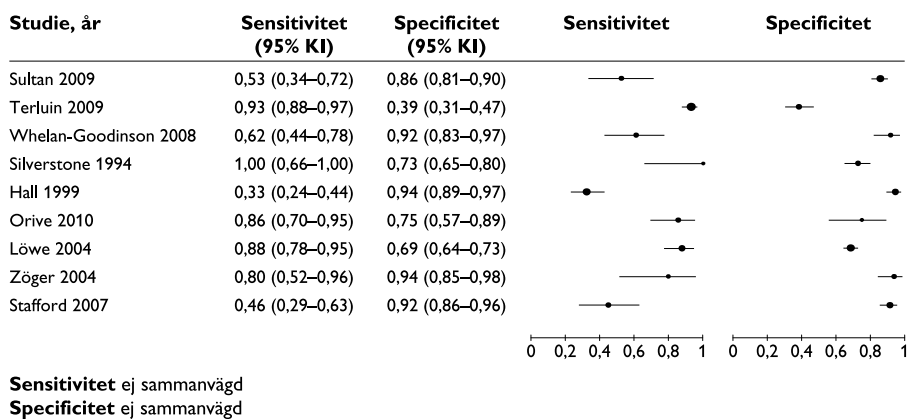
Två versioner av GDS som är avsedda för äldre har också tillräcklig känslighet. Studierna på GDS hade oftast exkluderat deltagare med demens, vilket gör att vi inte kan uttala oss om värdet av formuläret för personer med kognitiv funktionsnedsättning.

**Tabell 4** Formulär för riktad screening av depression.

Formulär tröskelvärde	Ålder	Sensitivitet Vetenskapligt stöd	Specificitet Vetenskapligt stöd
BDI-I Tröskelvärde 11	Barn och ungdomar	81% (95% KI, 75; 87) Begränsat ⊕⊕○○	59% (95% KI, 53; 65) Begränsat ⊕⊕○○
CBCL Tröskelvärde 65	Barn och ungdomar	88% (95% KI, 78; 94) Begränsat ⊕⊕○○	76% (95% KI, 67; 83) Begränsat ⊕⊕○○
CDI Tröskelvärde 12–13	Barn och ungdomar	69% (95% KI, 52; 84) Begränsat ⊕⊕○○	70% (95% KI, 63; 76) Måttligt starkt ⊕⊕⊕○
BDI-II Tröskelvärde 14	Vuxna	90% (95% KI, 83; 95) Begränsat ⊕⊕○○	50 till 84% Ottillräckligt ⊕○○○
EPDS Tröskelvärde 12–13	Nyförlost kvinnor	72% (95% KI, 69; 76) Begränsat ⊕⊕○○	88% (95% KI, 86; 89) Begränsat ⊕⊕○○
HADS Tröskelvärde 8	Vuxna	Minst 33 till max 100%, dvs varierande Måttligt starkt ⊕⊕⊕○	Minst 39 till max 94%, dvs varierande Måttligt starkt ⊕⊕⊕○
PHQ-9 Algoritm-baserat	Vuxna	34 till 84%, dvs varierande Måttligt starkt ⊕⊕⊕○	93% (95% KI, 92; 93) Måttligt starkt ⊕⊕⊕○
PHQ-9 Tröskelvärde 10	Vuxna	54 till 100%, dvs osäker Måttligt starkt ⊕⊕⊕○	84% (95% KI, 83; 86) Måttligt starkt ⊕⊕⊕○
GDS-15 Tröskelvärde 5	Äldre	87% (95% KI, 79; 92) Måttligt starkt ⊕⊕⊕○	71% (95% KI, 66; 75) Måttligt starkt ⊕⊕⊕○
GDS-30 Tröskelvärde 11	Äldre	87% (95% KI, 82; 92) Begränsat ⊕⊕○○	76% (95% KI, 72; 80) Begränsat ⊕⊕○○

KI = Konfidensintervall

Vi fann att HADS och PHQ-9 inte uppfyller krav på tillräcklig känslighet. HADS undersöktes i nio studier med nära 1 700 patienter och PHQ-9 i fjorton studier med drygt 3 000 vuxna och äldre patienter från olika typer av somatisk och psykiatrisk vård. Trots den stora mängden data gick det inte att fastställa känsligheten för vare sig HADS eller PHQ-9. Det illustreras i metaanalysen för HADS (Figur 2). Vi bedömde att ytterligare studier inte skulle tillföra mera information och att det därför finns ett måttligt starkt vetenskapligt underlag för att känsligheten inte kan definieras.



**Figur 2** Metaanalys av känslighet och träffsäkerhet för HADS.

Nyförlostna kvinnor har högre risk att drabbas av depression. Screening av mödrar i samband med besök på barnvårdscentraler genomförs numera i större delen av landet. I Sverige används EPDS där mödrarna skattar symtom på depression. EPDS har utvärderats i ett stort antal studier. Men studierna var så heterogena att det inte gick att få någon information från en metaanalys. Vi analyserade därför resultaten enbart för de åtta studier som använde SCID-I som referensstandard och fann då en känslighet på 72 procent. Det är i och för sig lägre än de 80 procent som vi anser vara en rimlig gräns, men å andra sidan är screening med EPDS en enkel procedur för att öka upptäckt av depression post partum avsevärt.

## **Några formulär för riktad screening av bipolära syndrom har tillräcklig känslighet**

Vi sökte litteratur för sju formulär och det fanns tillräckligt underlag för att granska fyra av dem (Tabell 5).

Studier på MDQ har utvärderat hur väl formuläret fungerar om funktionsnedsättningen är måttlig, men också enligt de så kallade Genève-kriterierna där funktionsnedsättningen kan vara lindrig. Vi granskade båda alternativen.

Mycket få studier utvärderade känslighet och träffsäkerhet för formulär avsedda för barn och ungdomar. Istället beräknade de måttet ROC-AUC (Receiver Operating Characteristic – Area Under Curve) som är oberoende av tröskelvärde. En nackdel med ROC-AUC är att det saknas etablerade gränser i litteraturen för när värdet är ”tillräckligt högt”. I en tidigare SBU-rapport, om riskbedömningar i psykiatri, tolkades ROC-AUC över 0,70 som modest, över 0,80 som måttlig och över 0,90 som hög överensstämmelse. Vi beslutade att göra ett undantag och godkänna studier med ROC-analys som utfallsmått och att utgå från den ovan nämnda SBU-rapportens tolkning.

Vi kom fram till att barn och ungdomsformulären General Behavior Inventory (GBI) med föräldraskattning och YMRS när kliniker skattar har tillräcklig känslighet. För vuxna är alternativen MDQ om kravet är minst lindring funktionsnedsättning eller HCL-32. En nackdel med HCL-32 är den osäkra specificiteten som riskerar att leda till att många felaktigt bedöms vara sjuka. Det saknades studier på äldre personer.

**Tabell 5** Formulär för riktad screening av bipolära syndrom.

Formulär	Ålder	Sensitivitet Vetenskapligt stöd	Specificitet Vetenskapligt stöd	ROC-AUC Vetenskapligt stöd
GBI föräldra- skattning	Barn och ungdomar			>0,80 Måttligt starkt ⊕⊕⊕○
YMRS expert- skattning	Barn och ungdomar			0,97 Måttligt starkt ⊕⊕⊕○
YMRS föräldra- skattning	Barn och ungdomar			<0,76 Begränsat ⊕⊕○○
HCL-32 ≥14	Vuxna	Sensitivitet: 82% (95% KI, 78; 85) Måttligt starkt ⊕⊕⊕○	Specificitet: 36 till 71% Otilräckligt ⊕○ ○○	
MDQ Normal- kriterier ≥7	Vuxna	Sensitivitet: 72% (95% KI, 71; 84) Begränsat ⊕⊕○○	Specificitet: 78% (95% KI, 76; 80) Måttligt starkt ⊕⊕⊕○	
MDQ Genève- kriterier ≥7	Vuxna	Sensitivitet: 78% (95% KI, 71; 84) Begränsat ⊕⊕○○	Specificitet: 79% (95% KI, 76; 82) Måttligt starkt ⊕⊕⊕○	

KI = Konfidensintervall; ROC-AUC = Receiver Operating Characteristic – Area Under Curve

## Det går inte att bedöma värdet av formulär för bestämning av svårighetsgrad

Denna kategori av formulär används dels för att bedöma svårighetsgrad i samband med diagnos, dels för att följa hur symtomen utvecklas över tid. Hit hör välkända symtomskattningsskalor som Montgomery-Åsberg Depression Rating Scale (MADRS) och Hamilton Depression Rating Scale (HDRS).



Svårighetsgrad av depression och mani klassificeras i DSM som lindrig, måttlig och svår. Kriterierna för svårighetsgrader ingår i intervjuerna SCID-I och Kiddie-Schedule for Affective Disorders and Schizophrenia (K-SADS) samt i formuläret CGI-S (CGI-BP för bipolära syndrom) som vi valde som möjliga referensstandarder.

Det fanns många studier som utvärderade formulär för bedömning av svårighetsgrad. Ett litet fåtal baserades på DSM:s kriterier. Merparten av studierna jämförde två olika symtomskattningsformulär med varandra, t ex MADRS mot HDRS. Eftersom det inte finns några brytpunkter för olika svårighetsgrad med i sådana studier exkluderades de.

Det fanns följaktligen inte tillräckligt många studier för att utvärdera formulär för bestämning av svårighetsgrad.

### **Det saknas relevanta studier om formulär för att följa upp behandling**

Formulär för att bedöma svårighetsgrad kan användas vid upprepade tillfällen för att följa effekter av behandling. Vi hittade inga studier som byggde på DSM-kriterier och som undersökte formulärens förmåga att följa förändring av svårighetsgrad över tid.

### **Expert- och patientskattningar har god överensstämmelse för IDS och MADRS**

Flera studier har undersökt hur väl skattningar som görs av en kliniker respektive patient överensstämmer med varandra. Det fanns tillräckligt med studier för att utvärdera Inventory of Depressive Symptomatology (IDS) och MADRS. Överensstämmelsen var acceptabel för MADRS (begränsat vetenskapligt underlag ⊕⊕○○) och god för IDS (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

### **Det saknas studier om kostnadseffektivitet**

Vi hittade inga studier som uppfyllde våra inklusionskriterier.

## Svenska versioner har sällan en validerad översättning

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.

Vi kompletterade den systematiska litteraturgranskningen med en sammanställning av uppgifter om översättning, omfattning, tillgänglighet och kostnader. Data hämtades från användarmanualer, webbplatser och från svenska forskare inom området.

Ändå blev sammanställningen långt ifrån komplett. Det visade sig vara svårt att hitta uppgifter om översättningar och tillgänglighet för många av formulärens. I Tabell 6 redovisas de uppgifter vi har fått för de formulär som vi kunde utvärdera.

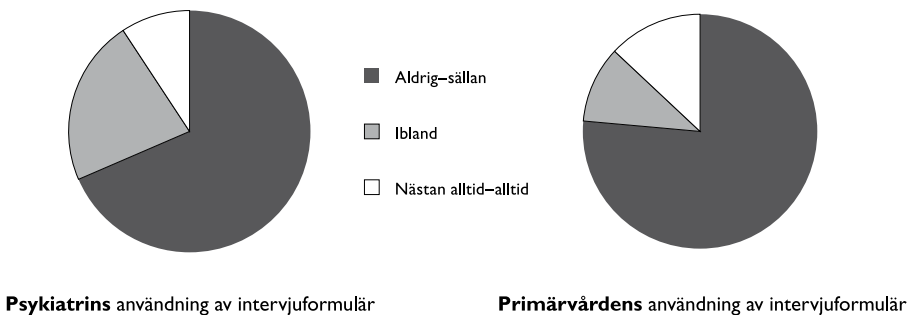
**Tabell 6** Översättning och tillgänglighet för formulär med vetenskapligt stöd för tillräcklig sensitivitet och specificitet.

Formulär	Översättning	Tillgänglig
MINI, intervju	Översatt med tillstånd	Fritt
SCID-I, intervju	Översatt med tillstånd	Avgift
BDI-II	Översatt med tillstånd	Avgift
CBCL	Översatt med tillstånd	Fritt
EPDS	Översatt	Fritt
GDS-15	Översatt med tillstånd	Fritt
HCL-32	Översatt	Fritt
MDQ	Översatt med tillstånd	Fritt

## Vetenskap och praxis överensstämmer inte alltid

Vi kompletterade litteraturgranskningen med insamling av data om dagens praxis. Vi genomförde en enkät till primärvården och vuxenpsykiatri och hämtade uppgifter från utredningar, framför allt från Socialstyrelsen. Svarsfrekvensen för enkäterna var låg; 36 procent för psykiatrer och 41 procent för läkare från primärvården, men resultaten stämde väl överens med äldre praxisundersökningar.

Enkätsvaren tydde på att knappt en fjärdedel inom såväl vuxenpsykiatri som primärvården använde intervjuformulär regelbundet medan mer än hälften aldrig använde dem (se Figur 3). Däremot uppgav var tredje läkare från primärvården att de alltid eller nästan alltid använde riktad screening. Det vanligaste formuläret var MADRS. MADRS (eller MADRS-S) utnyttjades också av cirka en tredjedel av läkarna från primärvården för att följa upp behandling.



**Figur 3** Användning av helt eller delvis strukturerade intervjuer i primärvården och vuxenpsykiatri (Källa: SBU:s enkäter januari 2011).

## Det finns olika uppfattningar om värdet av bedömningsformulär

Vi identifierade fem studier som belyste patienternas erfarenheter av bedömningsformulär och åtta som undersökte vårdpersonalens.

Två studier hade inhämtat både patienters och läkares uppfattningar, och författarna noterade att patienterna var mera positiva till bedömningsformulär än läkarna/psykologerna.

Patienterna var nöjda i samtliga fem studier. I två studier med MINI sågs intervjun som meningsfull, den täckte deras symtom och det strukturerade formatet innebar inte något problem.

Vårdpersonalen lämnade mera blandade synpunkter. I några studier fick personalen i hög grad fram den relevanta informationen med bedömningsformulär och uppnådde en bra relation till patienten. Två studier som genomförts i England, där användning av bedömningsformulär är ersättningsgrundande, redovisade mera kritiska synpunkter. Läkarna var skeptiska till formulär för egen del men ansåg att de kunde vara värdefulla ”för andra”. Andra invändningar var att formulären inte kan upptäcka subtila förändringar, att tiden inte räckte till och att infrastrukturen inte fungerade.

Spektrat av synpunkter i studierna överensstämde med vad vi fann i våra enkäter och i en serie fokusgrupper med läkare från svensk primärvård.

## **Bedömningsformulär ger en etisk vinst men det finns risker**

Det är ett etiskt problem att många personer med förstämningssyndrom inte upptäcks genom den sedvanliga diagnostiken. Användning av bedömningsformulär innebär därmed en etisk vinst. Men det förutsätter att formuläret har tillräcklig diagnostisk tillförlitlighet och används på rätt sätt. Över-, under- eller felanvändning av formulären medför etiska komplikationer.

Ett grundläggande etiskt problem är att vårdpersonal har varierande grad av utbildning när det gäller bedömningsformulär. Detta medför en risk för felanvändning av formulären. En mindre del av läkarna utbildas av sina arbetsgivare och våra enkäter tydde på att många läkare i primärvården inte fått någon utbildning alls.

Bristande kunskap om enskilda formulärs känslighet och träffsäkerhet innebär en risk att patienterna får fel diagnos. Bristande kunskap om syftet med formulären utgör också en risk. Det kan t ex finnas en överdriven tilltro till att en viss poäng på ett formulär för att bedöma svårighetsgrad kan översättas till en diagnos. Formulären bygger på förståelse av språk, vilket leder till risk för ojämlig vård. Personer med annan språktillhörighet, med skriv- och lässvårigheter eller med nedsatt kognitiv funktion kan då inte bedömas med formulären. Det är inte heller säkert att personer från andra kulturer definierar förstämningssyndrom på samma sätt som i klassificeringssystemen, och formulären kommer då inte att vara användbara.

## **Våra bedömningar**

En strukturerad diagnostik förbättrar förutsättningarna för att patienterna får en korrekt diagnos. På så sätt kan en större andel av de patienter som har förstämningssyndrom bli identifierade och få tillgång till effektiv behandling.

### **Det går att förbättra diagnostiken för förstämningssyndrom**

Det optimala vore om procedurer som bygger på LEAD används i både primärvård och specialiserad vård. Diagnosen skulle då baseras på en sammanvägningsdiskussion baserad på information från klinisk intervju, resultat från bedömningsformulär och resultat från andra undersökningar, t ex laboratorieprover och röntgen.

Ett steg på vägen är att använda strukturerade eller semistrukturerade intervjuer med tillräcklig diagnostisk tillförlitlighet som del i processen. För verksamheterna kan det innebära såväl att formulär införs som att andra, där evidensen tyder på att tillförlitligheten är osäker, fasas ut (Tabell 7 och 8 som bygger på resultaten från vår granskning).

**Tabell 7** Formulär med vetenskapligt stöd för tillräcklig diagnostisk tillförlitlighet.

Åldersintervall	Intervju	Riktad screening	Svårighetsgrad
Barn och ungdomar	–	Depression: CBCL, BDI-I Mani: YMRS, GBI	–
Vuxna	SCID, MINI	Depression: BDI-II Post partum-depression: EPDS Mani: HCL-32, MDQ med modifierade kriterier	–
Äldre än 65 år	–	Depression: GDS-15, GDS-30	–

**Tabell 8** Formulär med vetenskapligt stöd för att den diagnostiska tillförlitligheten inte är tillräcklig.

Åldersintervall	Intervju	Riktad screening
Barn och ungdomar	–	CDI
Vuxna	PRIME-MD	HADS, PHQ-9
Äldre än 65 år	–	PHQ-9

En nackdel är att diagnostiken kan komma att ta längre tid. En semi-strukturerad intervju kan pågå i cirka en timme i okomplicerade fall och en strukturerad intervju cirka en halvtimme. Å andra sidan ger intervjuerna möjlighet att tidigt få en uppfattning om samsjuklighet och kanske behov av att remittera patienterna vidare. Diagnosspecifika självbedömningsformulär behöver däremot inte ta tid från konsultationen eftersom patienterna kan fylla i dem i hemmet eller väntrummet.

En annan konsekvens av att använda bedömningsformulär är att patientjournalerna kan behöva anpassas. Om intervjuer och formulär ska kunna införlivas i klinisk rutin måste de kunna anslutas till patientjournalen på ett enkelt sätt.

Slutligen skulle det behövas rutiner för att kvalitetssäkra diagnostiken. De kan bl a omfatta utbildning om bedömningsformulär, deras användningsområden, styrkor och svagheter och hur resultaten från dem ska tolkas samt regelbundna samskattningar.

### **En webbaserad nationell sammanställning av bedömningsformulär skulle vara till hjälp**

En webbaserad nationell sammanställning över bedömningsformulär skulle underlätta för användarna att välja lämpligt formulär. Utöver information om ändamålet med formuläret och dess prestanda kunde sammanställningen innehålla uppgifter om tillgänglighet och om formuläret är översatt på vedertaget sätt.

### **Screeningen med EPDS är värdefull – men inte tillräcklig**

En önskvärd förändring är att screening för post partum-depression erbjuds samtliga kvinnor, oavsett språktillhörighet. Det är viktigt att det finns en handlingsplan kopplad till screeningen. Den bör också ta hänsyn till att många mödrar med depression inte upptäcks, erbjuder fortsatt diagnostik till dem som har depression enligt EPDS samt evidensbaserad behandling för dem som får diagnosen depression.

### **Det finns kunskapsluckor som behöver täppas till**

De mest angelägna forskningsbehoven som också är relevanta för svensk sjukvård är:

- Att undersöka hur användning av diagnostiska hjälpmedel påverkar behandlingsresultat vid handläggning av förstämningssyndrom.
- Hur patienterna och klinikerna upplever omhändertagandet med och utan diagnostiska hjälpmedel; hur tillfrisknande, livskvalitet, funktionsförmåga och relation mellan läkare, andra berörda vårdgivare och patient påverkas, samt vilka de hälsoekonomiska och organisatoriska effekterna blir.
- Att verifiera diagnostisk tillförlitlighet för den svenska versionen av formulär som har vetenskapligt stöd.

- Att undersöka diagnostisk tillförlitlighet för intervjuformulär som är avsedda för barn och ungdomar samt för de allra äldsta och för dem med kognitiv svikt.
- Att utveckla metodik för att utvärdera formulär som mäter sjukdomsförlopp.

**Tabell 9** Bedömningsformulär för förstämningssyndrom – en förkortningslista

AS-18	Affective self rating Scale 18
BCFPI	Brief Child and Family Phone Interview
BDI-II	Beck Depression Inventory II
BUS	Beck Ungdomsskalor (Beck Youth Inventories, BYI)
CBCL	Child Behaviour Checklist
CES-D	Center for Epidemiologic Studies for Depression
CDSS	Calagary Depression Scale for Schizophrenia
CIDI	Composite International Diagnostic Interview
CMRS	Child Mania Rating Scale
DAWBA	Development and Well-Being Assessment
DICA	Diagnostic Interview for Children and Adolescents
DIS	Diagnostic Interview Schedule
EPDS	Edinburgh Post Natal Depression Scale
GBI	General Behavior Inventory
GDS-15	Geriatric Depression Scale (15 items)
GDS-30	Geriatric Depression Scale (30 items)
HADS	Hospital Anxiety and Depression Scale
HCL-32	Hypomania/Mania Symptom Checklist
HDRS	Hamilton Depression Rating Scale
IDS	Inventory of Depressive Symptomatology
K-SADS	Kiddie-Schedule for Affective Disorders and Schizophrenia

*Tabellen fortsätter på nästa sida*



**Tabell 9** fortsättning

LCM	Life Chart Methodology
LEAD	Longitudinal, Experts, All Data procedure
MADRS	Montgomery-Åsberg Depression Rating Scale
MAS	Bech-Rafaelsen Mania Scale
MDI	Major Depression Inventory
MDQ	Mood Disorder Questionnaire
MINI	Mini International Neuropsychiatric Interview
MINI-KID	Mini International Neuropsychiatric Interview for Children and Adolescents
PHQ-9	Patient Health Questionnaire
PRIME-MD	Primary Care Evaluation of Mental Disorders
QIDS	Quick Inventory of Depressive Symptomatology
SADS	Schedule for Affective Disorders and Schizophrenia
SCAN	Schedules for Clinical Assessment in Neuropsychiatry
SCID	Structured Clinical Interview for DSM-IV disorders
SDQ	Strength and Difficulties Questionnaire
SHQ	Seasonal Health Questionnaire
SPAQ	Seasonal Pattern Assessment Questionnaire
SPIFA	Structured Psychiatric Interview for General Practice
TRF	Teacher Report Form
YMRS	Young Mania Rating Scale

**Tabell 10** Formulär som har för få studier med tillräcklig kvalitet för att kunna bedömas.

<b>Formulär</b>	<b>Syfte</b>
<b>För barn</b>	
K-SADS	Intervju
DICA	Intervju
MINI-KID	Intervju
TRF	Riktad screening, depression
BUS	Riktad screening, depression
CES-D	Riktad screening, depression
SDQ:s emotionella delskala	Riktad screening, depression
CBCL	Riktad screening, mani
GBI självskattning	Riktad screening, mani
Riktad screening, mani	Riktad screening, mani
MDQ	Riktad screening, mani
CMRS	Riktad screening, mani
YMRS självskattning	Riktad screening, mani
<b>För vuxna</b>	
SADS	Intervju
SCAN	Intervju
CIDI	Intervju
DIS	Intervju
SPIFA	Intervju
CDSS	Intervju
CES-D	Riktad screening, depression
EPDS	Riktad screening, depression, gravida
SPAQ	Riktad screening, årstidsbunden depression
SHQ	Riktad screening, årstidsbunden depression
HCL-32	Riktad screening, bipolärt syndrom typ II

*Tabellen fortsätter på nästa sida*

**Tabell 10** fortsättning

BDI-II	Svårighetsgrad, depression
CDSS	Svårighetsgrad, depression
HDRS-17	Svårighetsgrad, depression
IDS	Svårighetsgrad, depression
MADRS	Svårighetsgrad, depression
AS-18	Svårighetsgrad, bipolära syndrom
MAS	Svårighetsgrad, bipolära syndrom
YMRS	Svårighetsgrad, bipolära syndrom
LCM-p	Svårighetsgrad, bipolära syndrom
<b>För äldre</b>	
BDI-II	Riktad screening, depression
GDS-20	Riktad screening, depression
CES-D	Riktad screening, depression
CSDD (Cornellskalan)	Riktad screening, depression vid demens
SIDI	Riktad screening, depression

**Tabell 11** Formulär där studier som uppfyllt inklusionskriterierna saknas.

Barn	Vuxna	Äldre
DAWBA	MDI	SCL
KID-SCID	Zung SDS	BCDRS
BCFPI	BDRS (svårighetsgrad mani)	CPRS-D
Birlesons DSRS		HSCL-25 MADRS



# 1. Inledning

---

## Syfte

Förstämningssyndromen omfattar episoder med depression, mani, hypomani, dystymi och blandtillstånd av dessa. Det är angeläget att identifiera och behandla förstämningssyndrom eftersom de, utöver stort lidande, även är förknippade med ökad risk för samtidiga kroppsliga sjukdomar och för tidig död. Risken för att en person med förstämningssyndrom avlider i förtid är ungefär dubbelt så stor som för befolkningen som helhet, och risken för självmord är 15–20 gånger högre [1]. Vid egentlig depression är risken för självmord vid långtidsuppföljningar cirka 3–7 procent [2–4].

Syndromen diagnostiseras med hjälp av kriterier som beskriver symtom och deras varaktighet. World Health Organization (WHO) och amerikanska psykiaterföreningen (APA) har var sitt klassificeringssystem (International Classification of Diseases, ICD, respektive Diagnostic and Statistical Manual of Mental Disorders, DSM) för psykiska tillstånd där kriterierna revideras regelbundet.

Diagnosen grundas på observationer och information vid samtal med patienten. Samtalen kan kompletteras med bedömningsformulär, t ex strukturerade intervjuer och formulär för riktad screening. Bedömningsformulär kan även användas för att följa upp behandlingseffekter.

Det är oklart vilka bedömningsformulär som används i Sverige och också i vilken utsträckning de används.

Syftet med översikten är att bistå verksamheter och myndigheter med kunskap om vilka bedömningsformulär som har ett vetenskapligt stöd för tillräcklig diagnostisk tillförlitlighet. Ytterligare ett syfte är att identifiera forskningsluckor och utvecklingsbehov.

Vårt uppdrag omfattade en systematisk litteraturgranskning, kompletterad med praxisundersökningar. Vi avgränsade oss till formulär som används i Sverige.

Rapporten är en del i ett uppdrag från Socialdepartementet om att utvärdera metoder inom psykiatri.

## **Bakgrund**

### **Klassificering av förstämningssyndromen**

Förstämningssyndromen karakteriseras av en eller flera episoder där patientens stämningssläge, energi och aktivitetsnivå är så förändrade att de påverkar funktionsförmågan. Den kliniska bilden av förstämningssyndrom skiljer sig mellan olika kulturer, och vår aktuella diagnostik är i hög grad en västerländsk konstruktion. Det verkar dock finnas en kärna av symtom som är gemensam i olika kulturer [5].

Förstämningssyndromen delas in i unipolära och bipolära syndrom, vilka redovisas i Tabell 1.1.

Det vanligaste unipolära syndromet är egentlig depression. I DSM karakteriseras en episod av egentlig depression av nio symtom varav minst fem ska vara uppfyllda och ha pågått under minst två veckor samt orsaka signifikant lidande eller försämrad funktion. Personen ska antingen ha varit nedstämd eller känt ett minskat intresse för dagliga aktiviteter under större delen av den tiden. De övriga symtomen utgörs av viktförändring, sömnproblem, psykomotorisk agitation eller hämning, trötthet eller brist på energi, känslor av att vara värdelös, nedsatt förmåga att tänka eller koncentrera sig samt återkommande tankar på döden.

Några speciella former av depression som tas upp i rapporten är årstidsrelaterad depression (seasonal affective disorder) och post partum-depression (depression efter förlossning).

De bipolära syndromen karakteriseras av såväl episoder med förhöjt stämningstillstånd, ökad energi och aktivitet (hypomani eller mani) som episoder med sänkt stämningstillstånd, minskad energi och minskad aktivitet (depression). Hypomani skiljer sig från mani genom att episoden inte orsakar påtaglig funktionsnedsättning och att det inte finns psykotiska symtom, men förändringen av sinnestämning ska ändå vara tydligt iakttagbar för omgivningen. Enligt DSM-IV innefattar bipolära syndrom fyra kategorier som avspeglar vilken typ av episoder individen haft under sin livstid [6] (Tabell 1.1).

Såväl DSM-IV som ICD-10 använder sig utöver polaritet och förlopp också av svårighetsgrad som underlag för diagnos av förstämningssyndrom. Svårighetsgraden avgörs av antalet symtom som är uppfyllda och vilken grad av funktionsnedsättning de medför. Både egentlig depression och mani delas in i lindrig, måttlig och svår (med eller utan psykotiska symtom) svårighetsgrad. Dessutom specificeras om episoden är i delvis eller full återhämtning (remission). Indelningen är viktig för val av behandling liksom för utvärdering av behandlingseffekter.

Vi har inkluderat diagnosen schizoaffektiva syndrom, som brukar räknas till de psykotiska syndromen, i vår översikt. Anledningen är att behandlingen av schizoaffektiva syndrom har många likheter med behandling av bipolära syndrom och att det kan vara en svår differentialdiagnos [7,8].

**Tabell 1.1 Förstämningssyndromen**  
(modifierat från APA, *American Journal of Psychiatry*, 2002).

<b>Diagnos</b>	<b>Maniska symtom</b>	<b>Depressiva symtom</b>
Egentlig depression	Aldrig mani eller hypomani	Enstaka eller återkommande episoder under minst två veckor
Dystymi	Aldrig mani eller hypomani	Sänkt stämningsläge de flesta dagarna under minst två år
Depression UNS	Aldrig mani eller hypomani	Depressiva symtom som inte uppfyller kriterier för egentlig depression eller dystymi, t ex premenstruellt dysforiskt syndrom
Bipolärt syndrom typ I	Episoder med mani eller blandning av mani och depression (blandtillstånd)	Typiskt med episoder med egentlig depression men krävs inte för diagnos
Bipolärt syndrom typ II	En eller flera episoder med hypomani; aldrig mani eller blandtillstånd	Återkommande episoder med egentlig depression
Cyklotymi	Flera perioder med hypomana symtom under minst två år	Flera perioder med depressiva symtom som inte uppfyller kriterier för egentlig depression
Bipolärt syndrom UNS	Maniska symtom som inte uppfyller kriterier för bipolär I, II eller cyklotymi	Symtom som inte uppfyller kriterierna kan förekomma
Schizoaffektivt syndrom	Betydande varaktighet av maniska symtom eller blandtillstånd samtidigt som symtom på schizofreni + period på minst två veckor med psykotiska symtom utan mani (bipolär form)	Betydande varaktighet av egentlig depression innefattande nedstämdhet samtidigt som symtom på schizofreni + period på minst två veckor med psykotiska symtom utan depression (depressiv form)

UNS = Utan närmare specifikation



## Förekomst och samsjuklighet

Studier från USA och Europa har visat att upp till cirka 9 procent av den vuxna befolkningen led av förstämningssyndrom under ett år (12-månaders prevalens) [9,10]. Drygt 80 procent av dem hade egentlig depression. I den svenska Lundby-studien var sannolikheten för att insjukna i en depressiv episod före 70 års ålder 23 procent för män och drygt 30 procent för kvinnor [11]. I en norsk studie var livstidsprevalensen hos drygt 2 000 Oslobor mellan 18 och 65 år cirka 10 procent hos män och 24 procent hos kvinnor [12].

I den ovan nämnda amerikanska studien var 12-månadersprevalensen av bipolära syndrom 2,6 procent. I den europeiska översikten, som inte innefattade tonåringar eller personer över 65 år, varierade 12-månaders prevalens för bipolärt syndrom mellan 0,2 och 1,1 procent.

I den amerikanska undersökningen hade ungefär 45 procent av personerna med förstämningssyndrom allvarlig och 40 procent måttlig svårighetsgrad. Det fanns ett starkt samband mellan svårighetsgrad och förekomst av andra psykiska sjukdomar, framför allt ångestsyndrom, impulskontrollstörningar eller missbruk.

Personlighetsstörningar förekommer hos 41 till 81 procent av deprimerade patienter [13] och hos cirka 30 procent av patienter med bipolära syndrom [14]. Med personlighetsstörning menas ett genomgående oflexibelt sätt att tänka, uppleva eller bete sig som medför lidande eller försämrar funktionen. Mönstret ska ha funnits sedan ungdomstiden.

## Barn och unga

Prevalensen av depression verkar öka bland barn och unga och med tidigare sjukdomsdebut över generationerna [15]. I en finsk studie var ettårsprevalensen 5,3 procent för tonåringar mellan 15 och 19 år och 9,4 procent för unga vuxna mellan 20 och 24 år [16].

För bipolära syndrom är den vanligaste debutåldern mellan 15 och 25 år [17]. Vanligen är den första episoden depressiv. Man vet mindre om förekomsten av bipolära syndrom hos barn, men det finns en hög samsjuklighet med uppmärksamhetsstörningar hos dem som debuterar

i barndomen [18]. Avgränsningen mot diagnosen Attention Deficit Hyperactivity Disorder (ADHD) kan vara svår.

## **Äldre**

Depressioner hos äldre kan ha heterogena och atypiska symtombilder med inslag av kroppsliga och kognitiva symtom. Systematiska översikter har kommit fram till att prevalensen för depression hos äldre ligger mellan 10 och 15 procent [19,20]. I en studie från Göteborg steg prevalensen för depression från drygt 5 procent vid 70 års ålder till 13 procent vid 85 års ålder [21].

Vid depression hos äldre är samsjuklighet i ångestsyndrom vanligt [22]. Depression kan också vara ett tidigt tecken på demensutveckling.

## **Riskgrupper**

Risken att insjukna i depression är högre i vissa perioder av livet och vid vissa kroppsliga sjukdomar. Efter förlossning insjuknar 10–20 procent av alla nyförlösta mammor i så kallad post partum-depression [23].

Depression är vanligare hos personer med hjärt- och kärlsjukdomar, stroke, smärttillstånd i rygg och nacke, ledgångsreumatism, diabetes, vissa cancersjukdomar samt neurologiska sjukdomar som Parkinsons sjukdom och epilepsi (för en översikt se t ex [5]). Upp till 50 procent av patienter med Alzheimers sjukdom drabbas [24].

## **Den diagnostiska processen**

Gemensamt för både primärvård och specialiserad psykiatrisk vård är att diagnostiken ofta är komplicerad. Inom primärvård söker patienter oftast för kroppsliga symtom. Symtomen kan bero på kroppslig sjukdom men kan också vara tecken på förstämnings- eller ångestsyndrom. Inom psykiatrisk vård söker patienten oftast för psykiska symtom men som regel har patienterna en symtomrik sjukdomsbild och kan ha flera tillstånd samtidigt.

Den diagnostiska processen består i att värdera typ och svårighetsgrad av psykiska problem och att bedöma påverkan på funktion och livskvalitet

med hänsyn tagen till sammanhanget. Finns det familjemönster och yttre stressande omständigheter som spelar in? Diagnostiken baseras på patientens egen beskrivning av upplevelser och bedömarens frågor och observationer av beteendet. Insamling och bearbetning av information sker stegvis.

Kommunikationen mellan patient och kliniker är följaktligen navet i den diagnostiska processen. Dialogen kan försvåras av många skäl, t ex problemets komplexitet, språk- och kulturskillnader samt förutfattade meningar och förhastade slutsatser hos både patient och vårdgivare.

En *strukturerad* diagnostik inleds med en klinisk bedömning av patienten. Om klinikern misstänker psykiska problem kan ett formulär för riktad screening vara ett stöd för att bedöma sannolikheten för den misstänkta diagnosen. Om misstanken stärks kompletteras bedömningen med strukturerade eller semistrukturerade intervjuformulär för att säkerställa att samtliga kriterier i klassifikationssystemen (DSM-IV eller ICD-10) täcks. Det gäller både den troliga diagnosen men även andra syndrom, eftersom samsjukligheten ofta är betydande. Den slutliga diagnosen bör alltid baseras på en sammanvägning av anamnes och all tillgänglig klinisk information från diagnostiska hjälpmedel samt information från närstående. När diagnosen är ställd kan syndromets svårighetsgrad bedömas med hjälp av formulär som patienten eller klinikern fyller i.

En noggrann diagnostik ger förutsättningar att formulera en behandlingsplan och påbörja behandling. Effekterna av behandlingen utvärderas kontinuerligt med stöd av formulär för att bedöma svårighetsgrad.

## **Användning av formulär och skalor som stöd för den diagnostiska processen**

De bedömningsformulär som granskas i rapporten baseras på de bakomliggande kriterierna för förstämningssyndrom. Tanken är att formulären ska kunna fånga in eller kvantifiera ett tillstånd som är mångfasetterat. Formulären består därför av flera olika frågor (påståenden eller items).

Utveckling och utvärdering av psykologiska mätformulär följer principer från psykometrin (vetenskapen för psykologisk bedömning). Formulär och skalor måste konstrueras så att de verkligen mäter det som åsyftas, *validitet*. Frågorna ska t ex vara relevanta för tillståndet, antalet frågor ska vara tillräckligt för att täcka in olika aspekter och de ska tolkas på ett likartat sätt av användarna.

*Reliabilitet* beskriver exaktheten i mätningarna som t ex om testet visar samma resultat vid upprepade testomgångar och oberoende av vem som utför testet. I rapporten är graden av *interbedömarreliabilitet* en viktig kvalitetsaspekt. Den är ett mått på i vilken grad olika personers bedömningar överensstämmer med varandra. För utfall med två eller tre kategorier, t ex ställda diagnoser, brukar interbedömarreliabiliteten uttryckas som en kappakoefficient ( $\kappa$ ). Ett värde på  $\kappa$  som är mindre än 0,20 anses som låg, 0,21–0,40 relativt låg, 0,41–0,60 måttlig, mer än 0,61 god och mer än 0,81 mycket god [25].

Vi har delat in formulären i tre kategorier, beroende på syftet med användningen. Indelningen framgår av Faktaruta 1.1. Bedömningsformulären fyller dock flera funktioner. Förutom att mäta förekomst och grad av störningar säkerställer de att alla kriterier efterfrågas och att det görs på samma sätt. De kan också underlätta kontakt och ge tillfälle till att observera hur patienten förhåller sig till situationen.

Vissa formulär fylls i av patienten eller närstående medan andra fylls i av den som utför diagnostiken, en expertbedömare. Hit hör läkare, psykologer och specialistsjuksköterskor. I rapporten har vi använt benämningen *kliniker* för sådana expertbedömare.

## **Faktaruta 1.1** Indelning av bedömningsformulär i rapporten.

### **Formulär för riktad screening**

Riktad screening används för en preliminär bedömning av patienter vars symtom inger misstanke om förstämningssyndrom. Syftet är att stärka eller avfärda misstankarna och de kan alltså inte användas för att ställa en diagnos. Formulären fylls vanligen i av patienten eller närstående.

### **Intervjuformulär**

Intervjuformulär innehåller moduler för flera psykiska sjukdomar och används som stöd för att ställa diagnos. Formulären består av manualer för frågor och tolkningar av svaren vilka är uppbyggda på kriterier enligt ett diagnostiskt system som t ex DSM eller ICD. Intervjuerna kan vara strukturerade eller semistrukturerade. En strukturerad intervju kan utföras av vårdpersonal efter kortare tids träning. En semistrukturerad intervju utförs av en tränad expert inom området som utöver de färdigformulerade frågorna kan ställa följdfrågor för att säkrare utvärdera kriterierna. Semistrukturerade intervjuer är mer omfattande och tar längre tid att genomföra än de strukturerade.

### **Symtom/sjukdomsspecifika formulär för att bedöma svårighetsgrad**

Formulären används som stöd för att kvantifiera svårighetsgraden av ett tillstånd när diagnosen har ställts. De är vanligen uppbyggda som frågeformulär där patienten eller en kliniker skattar graden av symtom. De kan användas för att vägleda behandlingsplanering samt för att följa effekter av behandlingen.

Det kan inte nog betonas att formulären är hjälpmedel och att de alla har olika styrkor och begränsningar. Användaren måste vara förtrogen med hur formulären utvecklats, deras syfte och tillförlitlighet. Det bör vidare finnas en dokumentation över hur formulären ska administreras och tolkas samt om och hur de validerats för relevanta grupper [26]. Ofta är formulären utvecklade i andra länder än Sverige och bör därför ha en godkänd översättning med angiven källa (huvudreferens och översättare). Den enskilda användaren är därmed beroende av tillgång till uppdaterade manualer och systematiska utprövningar av formulären.

## Bedömning av nyttan av en diagnostisk metod

Ett diagnostiskt test är en integrerad del i handläggningen av en patient. Värdet av testet avgörs av om det påverkar behandlingsstrategier, den enskilde patientens hälsa och, ytterst, ger nytta för hela samhället. Få studier utvärderar dock konsekvenser av diagnostiska åtgärder. Metoden bedöms då istället utifrån dess diagnostiska tillförlitlighet (accuracy). Den diagnostiska tillförlitligheten beskriver hur väl formulären kan skilja sjuka från friska.

Gränsen mellan ”sjuk” och ”frisk” utgör det så kallade tröskelvärdet (cut off). Ju lägre tröskelvärdet sätts desto fler kommer att testa positivt.

## Val av referensstandard

Den diagnostiska tillförlitligheten relateras alltid till en referensstandard. En ideal referensstandard, dvs en gold standard, kan identifiera sjukdom eller frånvaro av sjukdom på ett säkert sätt. I praktiken är sällan referensstandarderna perfekta, vilket i sig introducerar en källa till osäkerhet.

LEAD (Longitudinal, Experts, All Data procedure) och Best Estimate Diagnosis anses vara optimala referensstandarder inom psykiatri [27]. Dessa metoder innebär att information från samtal med patienten, resultat från diagnostiska intervjuer, frågeformulär och laboratorieundersökningar samt information från närstående och observationer över tid sammanställs och diskuteras i team av erfarna bedömare. Begreppen används ofta överlappande och innebörden är ungefär densamma. För att öka läsbarheten har vi valt att genomgående använda begreppet LEAD.

LEAD är en krävande procedur, och få studier utförs med LEAD som referens. Näst bästa alternativ är att använda väletablerade semistrukturerade intervjuer som referensstandard. För vuxna är den vanligaste referensstandarderna DSM-systemets officiella intervju, Structured Clinical Interview for DSM-Axis I (SCID-I).

## Mått på diagnostisk tillförlitlighet

Studierna undersöker i vilken utsträckning referensstandard och det test som ska undersökas (indextestet) klassificerar tillståndet på samma sätt (sant positiva, dvs sjuka, respektive sant negativa, dvs friska). För en andel patienter ger referensstandard och indextest inte samma resultat (falskt positiva respektive falskt negativa). De olika utfallen åskådliggörs i den så kallade fyrfältstabellen i Tabell 1.2.

Studierna uttrycker ofta den diagnostiska tillförlitligheten som sensitivitet (känslighet) och specificitet (träffsäkerhet). Sensitiviteten anger hur stor del av dem som är sjuka som ligger över tröskelvärdet för indextestet, dvs

$$\text{Sensitivitet} = \frac{TP}{TP + FN}$$

På motsvarande sätt är specificiteten den andel av de friska som ligger under tröskelvärdet för indextestet, dvs

$$\text{Specificitet} = \frac{TN}{TN + FP}$$

**Tabell 1.2** Fyrfältstabell.

	<b>Sjuk enligt referensstandard</b>	<b>Frisk enligt referensstandard</b>
Test +	TP (sant positiv)	FP (falskt positiv)
Test –	FN (falskt negativ)	TN (sant negativ)

I allmänhet medför en ökad sensitivitet lägre specificitet och vice versa. Beroende på tillståndet, sammansättning av populationen och syftet med mätningen får man göra en avvägning av vilken balans mellan sensitivitet och specificitet som är optimal.

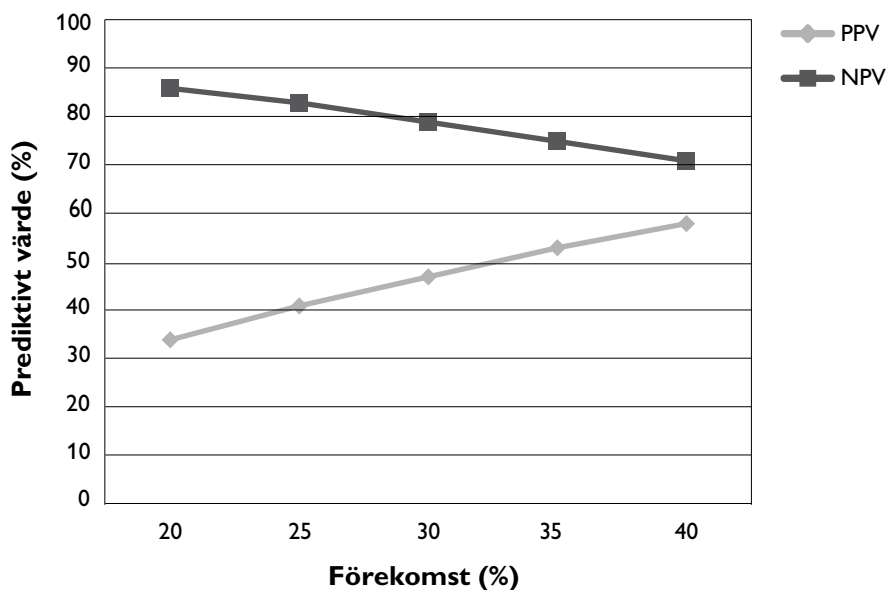
Från fyrfältstabellen går det också att beräkna positivt prediktionsvärde (PPV) och negativt prediktionsvärde (NPV). PPV anger andelen av dem som testar positivt som faktiskt är sjuka dvs

$$PPV = \frac{TP}{TP + FP}$$

NPV är ett uttryck för hur stor andel av dem som testar negativt som är friska, dvs

$$NPV = \frac{TN}{TN + FN}$$

Prediktionsvärdena kan vara värdefulla för planering av de resurser som ska avsättas för vidare diagnostik av den enskilda patienten. Till skillnad från sensitivitet och specificitet är PPV och NPV beroende av prevalensen, vilket framgår av Figur 1.1.



**Figur 1.1** Ett tests prediktiva värde varierar med prevalensen.



För en mera detaljerad genomgång av diagnostiska statistiska mått hänvisar vi till SBU:s metodbok ([www.sbu.se/sv/Evidensbaserad-varld-Utvardering-av-metoder-i-halso-och-sjukvarden--En-handbok/](http://www.sbu.se/sv/Evidensbaserad-varld-Utvardering-av-metoder-i-halso-och-sjukvarden--En-handbok/)) eller läroböcker om diagnostik.

## **Bedömning av svårighetsgrad och förändring**

Effekten av behandling kan värderas med hjälp av formulär som kan mäta svårighetsgrad av tillståndet inför, under och efter behandling. Bedömningen av svårighetsgrad försvåras av att depression och mani är komplexa tillstånd som inte kan observeras direkt. Både DSM- och ICD-kriterierna betonar antalet symtom som ett mått på svårighetsgrad av syndromen, men kliniskt är sambandet mellan antal symtom och grad av funktionsnedsättning relativt svagt. Olika skalor viktar symtomen olika, t ex lägger tonvikt på kroppsliga symtom eller på tankeinnehåll som pessimism och skuld.

Det är svårt att anse att något formulär kan utgöra gold standard eftersom de delvis mäter olika aspekter av förstämningssyndrom [28]. En lösning, som vi har valt i denna rapport är att korrelera skalan gentemot en oberoende referensstandard som är kliniskt relevant, Clinical Global Impression (CGI).

## **Rapportens innehåll**

Tyngdpunkten i rapporten är den systematiska översikten i Kapitel 3. Metodiken för systematisk granskning redovisas i Kapitel 2. Kapitel 4 innehåller en inventering av vilka formulär som används i Sverige samt data om dem.

Kapitel 6 diskuterar viktiga etiska och sociala aspekter på diagnos och uppföljning av förstämningssyndrom. Kapitel 7 beskriver de kvantitativa och kvalitativa praxisundersökningar som ingått i projektet. Kapitel 8 analyserar vilka som är de mest angelägna forskningsbehoven. Kapitel 9 diskuterar vilka konsekvenser resultaten av rapporten skulle kunna få för svensk praxis. Rapporten avslutas med en ordlista.

## Referenser

1. Osby U, Brandt L, Correia N, Ekblom A, Sparen P. Excess mortality in bipolar and unipolar disorder in Sweden. *Arch Gen Psychiatry* 2001;58:844-50.
2. Inskip HM, Harris EC, Barraclough B. Lifetime risk of suicide for affective disorder, alcoholism and schizophrenia. *Br J Psychiatry* 1998;172:35-7.
3. Bostwick JM, Pankratz VS. Affective disorders and suicide risk: a reexamination. *Am J Psychiatry* 2000;157:1925-32.
4. Nordentoft M, Mortensen PB, Pedersen CB. Absolute risk of suicide after first hospital contact in mental disorder. *Arch Gen Psychiatry* 2011;68:1058-64.
5. SBU. Behandling av depressionssjukdomar. En systematisk litteraturoversikt. Volym 1. Statens beredning för medicinsk utvärdering (SBU); 2004. SBU-rapport nr 166/1. ISBN 91-87890-87-9.
6. APA. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fourth edition. Washington; 1994.
7. Vieta E. Developing an individualized treatment plan for patients with schizoaffective disorder: from pharmacotherapy to psychoeducation. *J Clin Psychiatry* 2010;71:14-9.
8. Kane JM. Strategies for making an accurate differential diagnosis of schizoaffective disorder. *J Clin Psychiatry* 2010;71:4-7.
9. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005;62:617-27.
10. Wittchen HU, Jacobi F. Size and burden of mental disorders in Europe – a critical review and appraisal of 27 studies. *Eur Neuropsychopharmacol* 2005;15:357-76.
11. Mattisson C, Bogren M, Nettelbladt P, Munk-Jorgensen P, Bhugra D. First incidence depression in the Lundby Study: a comparison of the two time periods 1947-1972 and 1972-1997. *J Affect Disord* 2005;87:151-60.
12. Kringlen E, Torgersen S, Cramer V. A Norwegian psychiatric epidemiological study. *Am J Psychiatry* 2001;158:1091-8.
13. Hirschfeld RM. Personality disorders and depression: comorbidity. *Depress Anxiety* 1999;10:142-6.
14. George EL, Miklowitz DJ, Richards JA, Simoneau TL, Taylor DO. The comorbidity of bipolar disorder and axis II personality disorders: prevalence and clinical correlates. *Bipolar Disord* 2003;5:115-22.
15. Klerman GL, Weissman MM. Increasing rates of depression. *JAMA* 1989;261:2229-35.
16. Haarasilta L, Marttunen M, Kaprio J, Aro H. The 12-month prevalence and characteristics of major depressive episode in a representative nationwide sample of adolescents and young adults. *Psychol Med* 2001;31:1169-79.

17. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005;62:593-602.
18. Chang KD, Steiner H, Ketter TA. Psychiatric phenomenology of child and adolescent bipolar offspring. *J Am Acad Child Adolesc Psychiatry* 2000;39:453-60.
19. Beekman AT, Copeland JR, Prince MJ. Review of community prevalence of depression in later life. *Br J Psychiatry* 1999;174:307-11.
20. 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.
21. Palsson SP, Ostling S, Skoog I. The incidence of first-onset depression in a population followed from the age of 70 to 85. *Psychol Med* 2001;31:1159-68.
22. Beekman AT, de Beurs E, van Balkom AJ, Deeg DJ, 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.
23. 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.
24. Modrego PJ. Depression in Alzheimer's disease. Pathophysiology, diagnosis, and treatment. *J Alzheimers Dis* 2010;21:1077-87.
25. Altman DG. Practical statistics for medical research. London, Chapman & Hall/CRC; 1991.
26. AERA: The standards for educational and psychological testing. American Educational Research Association. Washington DC; 1999.
27. Spitzer RL. Psychiatric diagnosis: are clinicians still necessary? *Compr Psychiatry* 1983;24:399-411.
28. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry* 2004;161:2163-77.



## 2. Metodbeskrivning

---

Syftet med en systematisk översikt är att få en objektiv kartläggning av kunskapsläget. En systematisk översikt innebär att samtliga vetenskapliga studier som är relevanta för rapportens frågor identifieras och kvalitetsgranskas med avseende på tillförlitlighet. Studier som bedöms som tillförlitliga blir sedan underlag för evidensgraderade resultat.

Detta avsnitt beskriver i generella termer de olika steg som leder fram till en systematisk översikt. Kapitel 3.1 redovisar i detalj urvals- och kvalitetskriterier. För mera information om metodik hänvisar vi till SBU:s metodbok ([www.sbu.se/sv/Evidensbaserad-varld/Utvardering-av-metoder-i-halso-och-sjukvarden--En-handbok/](http://www.sbu.se/sv/Evidensbaserad-varld/Utvardering-av-metoder-i-halso-och-sjukvarden--En-handbok/)).

### Inklusionskriterier

Först formulerade vi frågorna för projektet så att de lämpade sig för systematiskt arbete. Frågorna var fokuserade och definierade diagnos, ålder, typ av bedömningsformulär, referensstandard och utfallsmått (Kapitel 3.1).

Därefter fastställde vi inklusions- och exklusionskriterier för de studier som skulle granskas samt kvalitetskriterier (Kapitel 3.1). Ett avgörande inklusionskriterium var vilka typer av studiedesign som kan accepteras. Valet av studiedesign beror på vilka frågor som ska besvaras. Randomiserade kontrollerade studier inkluderades för alla frågorna. För studier om diagnostisk tillförlitlighet accepterade vi även tvärsnittsstudier. Studier med etiska frågor hade inga begränsningar i studietyp.

Vi inkluderade studier på skandinaviska språk och engelska.

## Urval av studier

Ämnesexperterna tog fram sökstrategier för de olika frågorna i samarbete med en av SBU:s informatiker. Databassökningarna utfördes vid tre tillfällen. Den sista sökningen genomfördes i april 2011.

Vi sökte litteratur i databaserna Cochrane Database of Systematic Reviews, Medline, PsycInfo och Embase. Förutom databaser sökte vi litteratur i referenslistor för originalstudier och systematiska översikter. En mer detaljerad beskrivning av sökstrategierna redovisas i Bilaga 1.

Resultatet av litteratursökningen för respektive fråga, presenterat som abstraktlistor (korta sammanfattningar av studier), granskades först oberoende av två personer på SBU:s kansli (projektledaren samt antingen biträdande projektledare eller informatiker). Dessa två sållade bort uppenbart ovidkommande studier. Den förkortade listan skickades därefter till respektive läspar av ämnesexperter. Alla studier som åtminstone en i läsparet bedömde kunde tänkas uppfylla våra inklusionskriterier beställdes i fulltext.

## Bedömning av studiernas vetenskapliga kvalitet

De beställda studierna granskades av de två ämnesexperterna i läsparet, oberoende av varandra. Studier som vid närmare betraktande inte uppfyllde inklusionskriterierna eller inte var relevanta för svenska förhållanden exkluderades. Samtliga exkluderade studier finns förtecknade i Bilaga 3 på SBU:s webbplats ([www.sbu.se/212](http://www.sbu.se/212)).

Bedömningen av studiernas kvalitet och tillförlitlighet baserades på en systematisk granskning av de inkluderade studiernas design, utförande, resultatredovisning och slutsatser. För att minska risken för inter- och intrabedömarvariation användes granskningsmallar (checklistor). Syften med mallarna var att uppnå en gemensam norm för vad som är god kvalitet, minimera riskerna för glidning i värderingarna samt öka reproducerbarheten i bedömningarna.

Granskarna använde flera olika checklistor. För randomiserade studier användes SBU:s granskningsmall för randomiserade studier (se Bilaga 2). Kvaliteten på systematiska översikter bedömdes med hjälp av formuläret AMSTAR [1]. Majoriteten av studierna var dock tvärsnittsstudier om diagnostisk tillförlitlighet. Dessa granskades med hjälp av mallar som byggde på formuläret QUADAS [2]. QUADAS modifierades för att passa projektets frågor, vilket beskrivs närmare i Kapitel 3.1.

Vi testade interbedömarreliabiliteten för bedömning med QUADAS vid fyra tillfällen. Vid varje tillfälle läste ämnesexperterna en utvald studie och fyllde i QUADAS individuellt. Vi beräknade Fleiss  $\kappa$  [3] för överensstämmelse mellan experternas enskilda bedömningar av QUADAS-kriterierna. Vid den sista mätningen uppnåddes ett  $\kappa$  över 0,7 vilket anses motsvara en god (substantial) överensstämmelse.

Kvalitetsgranskningen av hälsoekonomiska artiklar gjordes av två hälsoekonomer med hjälp av SBU:s checklista för hälsoekonomiska studier (Bilaga 2 på SBU:s webbplats [www.sbu.se/212](http://www.sbu.se/212)).

Granskarna i respektive läspar diskuterade sina individuella bedömningar av studierna med varandra. Eventuell oenighet löstes med konsensus. Om inte ämnesexperterna kunde komma överens granskades studien av hela expertgruppen. Studierna klassificerades som *hög*, *medelhög* och *låg* kvalitet.

En person i varje läspar extraherade demografiska data, beskrivning av index- och referensstandard samt resultaten från de studier som hade minst medelhög studiekvalitet, och förde in uppgifterna i tabeller. Uppgifterna kontrollerades av ytterligare minst en person. Studier med låg studiekvalitet tabellerades enbart om det saknades studier med högre kvalitet.

## Sammanvägning av resultaten

Endast studier med minst medelhög studiekvalitet användes för att bedöma diagnostisk tillförlitlighet, patientnytta och kostnadseffektivitet.

Det finns flera metoder för att uppskatta eller beräkna estimat för den diagnostiska tillförlitligheten. Tekniken med så kallade forest plots som ofta används i metaanalyser är en sådan metod. För diagnostiska studier behövs då två forest plots, som speglar sensitivitet respektive specificitet. Forest plots innebär en förenkling och medför att de sammanvägda resultaten inte är helt tillförlitliga. Problemen ökar om studierna bygger på olika tröskelvärden. Det kan då bli nödvändigt att använda mer avancerade statistiska metoder med bivariat analys. Våra frågor var dock avgränsade till att endast inkludera studier med ett enda tröskelvärde, vanligen det som användes i den ursprungliga valideringen för ett givet instrument (Kapitel 3.1). Vi beslutade därför att i ett första skede ta fram forest plots för instrument med minst två välgjorda studier. Data från fyrfältstabeller (sant positiv, sant negativ, falskt positiv och falskt negativ) matades in i Microsoft Office Excel och överfördes därefter till programmet Meta Disc [4] för statistisk bearbetning till forest plots. Om resultaten från enskilda studier var likartade kunde det sammanvägda estimatet användas för att ge en uppskattning av sensitivitet och specificitet.

För sammanvägning av resultat från studier om överensstämmelse mellan klinikers och patienters/närståendes skattningar gjorde vi metaanalyser enligt metodik utvecklad av Hedges och Olkin [5]. Data fördes in i dataprogrammet CMA.



## Styrka på det vetenskapliga underlaget

Evidensstyrkan anger hur tillförlitligt det sammanvägda estimatet är. Ju starkare evidensen är, desto mindre sannolikt är det att redovisade resultat kommer att påverkas av nya forskningsrön inom en överblickbar framtid.

Evidensstyrkan bedöms med hjälp av GRADE-systemet [6]. Det kategoriserar det vetenskapliga stödet på fyra nivåer, som starkt (⊕⊕⊕⊕), måttligt starkt (⊕⊕⊕○), begränsat (⊕⊕○○) och otillräckligt (⊕○○○). En preliminär evidensstyrka sätts som beror på vilken studietyp som används i huvuddelen av det vetenskapliga underlaget. För interventionsstudier innebär det att randomiserade, kontrollerade studier leder till en preliminär evidensstyrka på ⊕⊕⊕⊕. För diagnostisk tillförlitlighet anser GRADE Working Group att välgjorda tvärsnittsstudier medför en preliminär evidensstyrka på ⊕⊕⊕⊕ [7].

Därefter bedöms i vilken utsträckning det vetenskapliga underlaget har svagheter som beror på fem faktorer: metodologiska brister (dvs studie-kvalitet, risk of bias), bristande samstämmighet (heterogeneity), bristande relevans (indirectness), bristande precision i data och risk för publikationsbias. Den preliminära evidensstyrkan dras ned med ett eller flera steg om det finns påtagliga svagheter i underlaget med avseende på dessa faktorer.

För att bedöma risken för bias som följd av systematiska metodologiska svagheter i underlaget använde vi den så kallade risk of bias table som utvecklats av Cochrane Collaboration [8]. Tabellen ger en överskådlig sammanställning över vilka metodbrister som finns för respektive studie. Items i QUADAS som besvarats med ”ja” markerades med svart och de som besvarats med ”nej” med ljusgrått. Om det inte framgick klart av artikeln huruvida ett kvalitetskriterium uppfylldes markerades detta med grått. Resultatet blir en matris där graden av grå och ljusgrå fält speglar hur stor osäkerheten är.

Vi beslutade att tillämpa följande riktlinjer vid arbetet med GRADE:

- Graden av samstämmighet mellan studier bedömdes med hjälp av de tidigare beskrivna forest plots.
- Överförbarheten (indirectness) påverkas av bl a population, val av kontrollåtgärd (referensstandard), val av utfallsmått och antal studier. I och med att studier med uppenbart irrelevanta populationer sållats bort i ett tidigt skede kunde avdrag för population högst bli ett steg. Om studierna använt olika referensstandards kunde det bli aktuellt att dra ner evidensstyrkan. Evidensstyrkan kunde också sänkas om det bara fanns en eller två studier och dessa omfattade få patienter.
- Evidensstyrkan kunde justeras ner pga bristande precision om konfidensintervallet för estimatet översteg 0,1 (-1) respektive 0,2 (-2) i forest plot.

## Referenser

1. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
2. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;6:9.
3. Fleiss J. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378-82.
4. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006;6:31.
5. Borenstein MH, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Chichester, West Sussex, UK, John Wileys & Sons Ltd; 2009.
6. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-6.
7. Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *Evid Based Med* 2008;13:162-3.
8. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).



### 3. Systematisk litteraturgenomgång

---

I detta avsnitt redovisas resultaten av den systematiska översikten av de formulär som kan användas som stöd för diagnostik och uppföljning vid depression och bipolärt syndrom.

Projektet avgränsades till att gälla sådana formulär som används i klinisk verksamhet i Sverige idag. Vi kontaktade därför representanter för de professioner inom hälso- och sjukvård som, enligt vår bedömning, ställer sådana diagnoser och följer upp resultat av behandling. Enkäten gick till Svensk förening för allmänmedicin, Svenska psykiatriska föreningen, Svenska föreningen för barn- och ungdomspsykiatri, Psykologförbundet, valda företrädare för äldrepsykiatri och obstetrik samt för olika somatiska specialiteter som t ex internmedicin. Vi frågade vilka formulär som används i Sverige.

De som besvarade förfrågan tog upp sammanlagt 64 olika formulär, vilka omfattade intervjuer, formulär för riktad screening och formulär för bedömning av svårighetsgrad. Det är dessa formulär som ingår i granskningen. De förkortningar som används i avsnittet finns förklarade i Kapitel 10 i rapporten.

Vi har utvärderat formulären uppdelat på åldersgrupperna barn och ungdomar, vuxna samt äldre. Redovisningen bygger på denna uppdelning.



## 3.1 Frågor och avgränsningar

### Frågor

De frågor som projektet omfattade var:

1. Förbättras omhändertagandet och minskar symtombördan då den kliniska bedömningen kompletteras med formulär för diagnos och symtomskattning, jämfört med en klinisk bedömning utan sådana hjälpmedel?
2. Har klinisk bedömning av förstämningssyndrom med stöd av diagnostiska hjälpmedel högre sensitivitet och specificitet än klinisk bedömning utan sådana hjälpmedel?
3. Vilken sensitivitet och specificitet har strukturerade och semistrukturerade intervjuer jämfört med LEAD för att ställa diagnoser på förstämningssyndrom?
4. Vilken sensitivitet och specificitet har strukturerade/semistrukturerade intervjuer för vuxna och äldre jämfört med Structured Clinical Interview for DSM-Axis I (SCID-I) och för barn jämfört med Schedule for Affective Disorders and Schizophrenia for School Aged Children (K-SADS) när det gäller att ställa diagnoserna depression och bipolärt syndrom?
5. Vilken sensitivitet och specificitet har formulär för riktad screening av depression och bipolärt syndrom jämfört med strukturerade eller semistrukturerade intervjuer?
6. Finns det några formulär för att bedöma svårighetsgrad av depression och bipolärt syndrom som är likvärdiga med svårighetsgraderingar enligt SCID-I eller Clinical Global Impression Severity Scale (CGI-S, CGI-BP)?
7. Finns det formulär som kan följa förlopp av svårighetsgrad av depression eller bipolärt syndrom och som är likvärdiga med svårighetsgraderingar enligt SCID eller Clinical Global Impression Severity Scale (CGI-S)?

8. Överensstämmer expertskattning och skattning utförd av patienter eller deras närstående när formuläret finns i två versioner?
9. Är det kostnadseffektivt att använda formulär i den diagnostiska processen respektive för att bedöma svårighetsgrad jämfört med att inte göra det?

## Inklusionskriterier

Vi inkluderade originalstudier i fulltext samt systematiska översikter, publicerade från år 1966 på engelska eller på skandinaviska språk.

Kriterier för *studiedesign* bestämdes av frågans art. Vi accepterade bara randomiserade studier (och översikter av randomiserade studier) för frågan om huruvida omhändertagande, behandling och utfall förbättras om den kliniska bedömningen kompletteras med diagnostiska formulär och skattningsformulär samt för frågan om kostnadseffektivitet. För bedömning av diagnostisk tillförlitlighet och överensstämmelse inkluderade vi även tvärsnittstudier. För studier om etiska aspekter hade vi inga begränsningar i studiedesign.

*Populationen* omfattade barn, vuxna och äldre som hade symtom som tydde på förstämningssyndrom (depression, bipolärt syndrom och dystymi) eller som utgjorde en riskgrupp, t ex nyblivna mödrar. Studierna skulle ha utförts i Europa, Nordamerika och Australien/Nya Zeeland. För studier om äldre inkluderade vi även studier om minor depression och studier utan geografisk begränsning.

Studier om allmän screening, som t ex hälsoundersökningar på arbetsplatser, eller epidemiologiska undersökningar ingick inte.

*Bedömningsformulären (indextestet)* omfattade strukturerade eller semi-strukturerade diagnostiska intervjuer, formulär för riktad screening samt skattningsskalor för att bedöma svårighetsgrad eller för att följa behandling över tid.



Formulären skulle enbart utvärderas för det ursprungligen tänkta användningsområdet och formatet, dvs olika former av ”indikationsglidning” skulle exkluderas. Undantaget var om upphovsmannen/kvinnan själv vidareutvecklat formuläret och validerat det. Studier som undersökte vilket tröskelvärde som var optimalt, utan att testa det ursprungligen validerade tröskelvärdet ingick inte.

Referensstandarderna varierade mellan de olika frågorna (Tabell 3.1.1).

**Tabell 3.1.1** Referensstandarder för de olika indextesten för diagnos och uppföljning av förstämningssyndrom.

Indextest	Referensstandard
SCID-I, K-SADS	LEAD
Klinisk bedömning utan diagnostiskt stöd	LEAD, SCID-I, K-SADS
Övriga strukturerade och semi-strukturerade intervjuer	Strukturerade eller semi-strukturerade intervjuer
Formulär för riktad screening	Strukturerade eller semi-strukturerade intervjuer
Formulär för bedömning av svårighetsgrad av depression och mani vid diagnos och uppföljning, t ex symtomskattningsskalor	CGI-S, CGI-BP, SCID-I eller K-SADS

CGI-BP = Clinical Global Impressions Scale, Bipolar Version; CGI-S = Clinical Global Impressions, Severity Scale; K-SADS = Schedule for Affective Disorders and Schizophrenia for School-Age Children; LEAD = Longitudinal, Experts, All Data procedure; SCID-I = Structured Clinical Interview for DSM-IV Axis I disorders

*Utfallsmått:* För bedömning av patientnyttan med diagnostiska hjälpmedel skulle resultatet mätas som förbättring, tillfrisknande eller som förändrat omhändertagande. För studier som undersökte den diagnostiska tillförlitligheten för intervjuformulär och formulär för riktad screening var sensitivitet och specificitet primärt utfallsmått.

Forskare har diskuterat vilken sensitivitet och specificitet som är tillräcklig vid depression och mani [1]. Vi beslutade att ett riktmärke för acceptabel sensitivitet och specificitet var 80 procent för intervjuer medan riktmärket för riktad screening var en sensitivitet på minst 80 procent [1].

I några fall accepterade vi också ROC-AUC som utfallsmått (Kapitel 3.4). ROC-kurvan är en grafisk representation av sambandet mellan sensitivitet och specificitet vid alla tröskelvärden. Ytan under kurvan (dvs AUC, Area Under the Curve) är ett mått på hur väl testet kan skilja mellan sjuk och frisk. Ju större ytan är, desto bättre prestanda har testet. En diagonal linje från origo, dvs ett AUC-värde på 0,50, motsvarar vad som skulle åstadkommas av slumpen medan ett AUC-värde på 1,0 avspeglar en perfekt diagnostik. Vilket AUC-värde som krävs beror på sammanhanget och det saknas etablerade gränser. Sjöstedt och Grann föreslår t ex att för prognosinstrument för riskbedömningar inom psykiatrin ska ett AUC på 0,70–0,80 ses som modest, 0,81–0,90 som måttlig och >0,90 som hög prediktiv förmåga [2]. Vi beslutade att använda deras gränsdragningar.

Olika sambandsmått som t ex intraklasskorrelation (ICC) eller Cohen's kappa ( $\kappa$ ) var utfallsmått för studier som bedömde om formulären kan användas för att bedöma svårighetsgrad av sjukdom respektive för samstämmighet mellan expert- och patientskattat depressionsdjup.

*Övriga inklusionskriterier* avsåg tid och studiestorlek. Maximal tid mellan referens- och indextest fick högst vara

- 1 vecka för depression
- 1 månad för dystymi
- 4 dagar för mani (undantag: livstidsmätningar som t ex Mood Disorder Questionnaire)
- 1 dygn för hypomani
- 24 timmar för jämförelse mellan expert- och självskattning eller för bedömning av depressionsdjup.

Det skulle ingå minst 30 deltagare i studier om skalor för bedömning av depressionsdjup och minst 30 par deltagare i studier som jämförde expert- och självskattning.

## Kvalitetskriterier

Studiernas kvalitet bedömdes med olika checklistor. Systematiska översikter värderades med AMSTAR [3]. Majoriteten av studier var tvärsnittsstudier om diagnostisk tillförlitlighet. Dessa granskades med hjälp av mallar som byggde på formuläret QUADAS [4]. Vi utgick från QUADAS i den version som består av 14 parametrar. Vi lade till ytterligare fyra parametrar som vi bedömde vara viktiga för studiekvaliteten (nummer 15–18).

### Faktaruta 3.1.2 Kriterier i QUADAS modifierat av SBU [4].

1. Var patientspektrum representativt?
2. Var selektionskriterier väl beskrivna?
3. Klassificerar referensstandardens tillståndet korrekt?
4. Adekvat tid mellan index och referenstest?
5. Verifiering av alla eller slumpmässigt utvalda patienter?
6. Samma referenstest oberoende av resultat från indextest?
7. Referenstest oberoende av indextest?
8. Beskrivning av indextestets utförande?
9. Beskrivning av referenstestets utförande?
10. Tolkades indextestet utan kännedom om resultat från referenstestet?
11. Tolkades referenstestet utan kännedom om resultat från indextestet?
12. Var tillgång till kliniska data samma som vid användning i praktiken?
13. Redovisning av ofullständiga data?
14. Redovisning av bortfall?
15. Adekvat utbildning av personer som utförde testen?
16. Beskrivning av metod för bedömning av tests reproducerbarhet
17. Rapporteras diagnostisk tillförlitlighet och statistisk spridning?
18. Rapporteras eventuella intressekonflikter?

QUADAS vidareutvecklades därefter till tre olika checklistor, anpassade för att bedöma studier om tillförlitlighet av intervjuer och diagnostiska formulär, samstämmighet mellan symtomskattningsskalor och CGI-S eller SCID-I (depressionsdjup) respektive samstämmighet mellan expert- och patientskattningar (Bilaga 2).

Vi ställde också upp kriterier för vilka parametrar som var avgörande för att studien skulle anses hålla minst medelhög studiekvalitet respektive hög studiekvalitet (Tabell 3.1.2).

**Tabell 3.1.2** *Krav för medelhög och hög studiekvalitet i studier om diagnostisk tillförlitlighet.*

Item	Intervjuer och diagnostiska formulär	Överensstämmelse expert-patient	Överensstämmelse CGI-S och symtomskattningsskala
Selektionskriterier (item 2)	Inte avgörande	Inte avgörande	Låg kvalitet om patienter selekteras på grundval av CGI-S (CGI-BP) eller skattningspoäng
Val av referensstandard (item 3)	Strukturerade intervjuer: medelhög Semistrukturerade intervjuer: hög	Inte avgörande	Inklusionskriterium: CGI-S eller CGI-BP. SCID, K-SADS eller DSM-kriterier tillsammans med GAF
Blindning (item 10 och 11)	Medelhög kvalitet	Medelhög kvalitet	Medelhög kvalitet
Redovisning av ofullständiga data (item 13)	Hög kvalitet	Hög kvalitet	Hög kvalitet
Redovisning av bortfall (item 14)	Hög kvalitet om bortfall <10 procent eller välgjord bortfallsanalys	Medelhög kvalitet	Medelhög kvalitet

*Tabellen fortsätter på nästa sida*

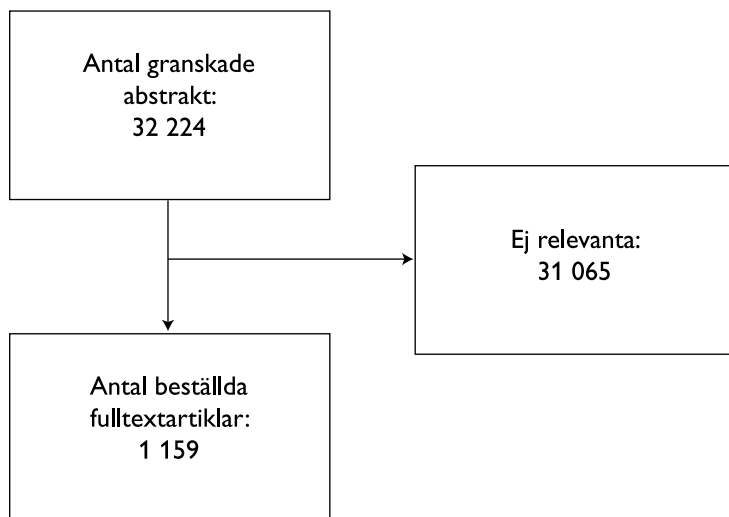
**Tabell 3.1.2** fortsättning

<b>Item</b>	<b>Intervjuer och diagnostiska formulär</b>	<b>Överensstämmelse expert-patient</b>	<b>Överensstämmelse CGI-S och symtomskattningsskala</b>
Adekvat utbildning i referensstandard (item 15)	Hög kvalitet	Hög kvalitet	Hög kvalitet
Bedömarreliabilitet >0,70 (item 16)	Hög kvalitet	Hög kvalitet	Hög kvalitet

## Urval av studier

Vi sökte litteratur i databaserna Cochrane Database of Systematic Reviews, Medline, PsycInfo och Embase. Databassökningarna utfördes vid tre tillfällen. Den sista sökningen genomfördes i april 2011. Vi kompletterade med litteratur från referenslistor i originalstudier och systematiska översikter.

Sökningarna som redovisas i Bilaga 1 på SBU:s webbplats ([www.sbu.se/212](http://www.sbu.se/212)) genererade drygt 32 000 abstrakt. Det visade sig vara mycket svårt att utifrån de beställda listorna dela upp abstrakt per fråga pga indexeringen. Flödesschemat i Figur 3.1.1 sammanställer därför det totala antalet titlar, oberoende av fråga och antalet artiklar som beställts i fulltext. Hur artiklarna fördelade sig mellan de olika åldersgrupperna framgår under respektive underavsnitt.



**Figur 3.1.1** Flödesschema.

För många av de intervjuformulär och skattningsskalor som ingick i litteratursökningen fanns inga studier som uppfyllde inklusionskriterierna. Tabell 3.1.3 sammanställer de formulär där vi har kunnat inkludera studier och som granskas i påföljande avsnitt.

**Tabell 3.1.3** Formulär som ingår i granskningen.

<b>Syfte</b>	<b>Formulär</b>	<b>Avsedd för åldrar</b>
Strukturerad intervju	CIDI, DIS, MINI, PRIME-MD, SPIFA	Vuxna
	MINI-KID, DAWBA, KID-SCID, BCPFI	Barn och ungdom
Semistrukturerad intervju	SCID-I, SADS, SCAN	Vuxna
	K-SADS, DICA	Barn och ungdom
Bedömningsformulär som stöd för diagnos av depression ("riktad screening"), expert- eller självbedömning	BDI-II, CDSS, PHQ-9, CES-D, EPDS, HADS, MDI, SDS	Vuxna
	ASEBA, Becks ungdomsskalor, BDI, SDQ, DSRs, CDI, CES-DC, PHQ-9	Barn och ungdom
	GDS, Cornellskalan, HADS, PHQ-9	Äldre
Bedömningsformulär som stöd för diagnos av mani ("riktad screening"), expert- eller självbedömning	HCL-32, MDQ	Vuxna
	GBI, CBQ, MDQ-A, YMRS, CMRS-P	Barn
Bedömningsformulär som stöd för diagnos av årstidsbunden depression ("riktad screening")	SHQ, SPAQ	Vuxna
Formulär för bedömning av djup vid diagnos och uppföljning av depression, t ex symtomskattning	BDI, CDSS, HDRS-17, MADRS, PHQ-9	Vuxna
Formulär för bedömning av svårighetsgrad vid diagnos och uppföljning av bipolära syndrom, t ex symtomskattning	AS-18, BDRS, MAS, YMRS, LCM-p	Vuxna
Överensstämmelse mellan patient- och klinikerskattning	IDS/QIDS, MADRS	Vuxna

## Referenser

1. Gilbody S, Sheldon T, Wessely S. Should we screen for depression? *BMJ* 2006;332:1027-30.
2. Sjöstedt G, Grann M. Risk assessment: What is being predicted by actuarial "prediction instruments"? *Int J Forensic Ment Health* 2002;1:179-83.
3. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
4. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;6:9.



## 3.2 Påverkan på omhändertagande och depressionsförlopp

### Evidensgraderade resultat

- Riktad screening som del i en förstärkt handläggning påverkar tillfrisknande eller symtombörda av depression hos vuxna patienter i primärvården i liten men signifikant utsträckning (begränsat vetenskapligt underlag ⊕⊕○○).
- Riktad screening som enda intervention har ingen effekt på symtombörda av depression (SMD  $-0,02$  (95 % KI,  $-0,25$  till  $0,20$ )) hos vuxna, oselektade patienter i primärvården (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- Det går inte att bedöma om riktad screening för depression hos vuxna i primärvård påverkar vårdrutinerna pga att studierna har bristande samstämmighet och precision (otillräckligt vetenskapligt underlag ⊕○○○).
- En förstärkt handläggning som inkluderar screening med EPDS minskar symtombördan, OR  $0,61$  (95 % KI,  $0,48$  till  $0,76$ ), av post partum-depression hos nyförlösta (begränsat vetenskapligt underlag ⊕⊕○○).
- Det saknas studier som utvärderar effekten på patientens symtombörda, eller på förloppet av att följa patienternas depressionsdjup med hjälp av symtomskattningsskala.

### Introduktion

Vår fråga var om omhändertagandet av patienterna påverkades av att diagnostiska hjälpmedel användes i samband med diagnos eller uppföljning.

Sökningarna genererade 2 777 abstrakt; 31 artiklar lästes i fulltext. Av dessa var fem systematiska översikter. Vi bedömde att de systematiska översikterna hade hög studiekvalitet. Vi kunde inte identifiera några

relevanta originalstudier förutom dem som ingick i översikterna. Samtliga studier handlade om effekten av att genomföra screening för depression. I några av dem var en systematisk diagnostik del av en intervention med förstärkta resurser för behandling. Vi kunde inte finna någon studie som berörde värdet av att följa depressionens svårighetsgrad över tid med hjälp av symtomskattningsskalor.

## Beskrivning av underlaget

De ingående översikterna finns sammanfattade i Tabell 3.2.2. Tre av de systematiska översikterna handlade om screening i primärvården [1–3].

Gilbody och medarbetare [1] inkluderade 16 randomiserade, kontrollerade studier varav sju avsåg patienter med förhöjd risk för depression medan resterande studier inte selekterade patienter efter deras risk för depression. Gilbody exkluderade studier där screening var en del av förstärkt intervention. Vanligen skattade patienten sina besvär på ett depressionsformulär före läkarbesöket. Läkaren fick dels information om hur patienten hade skattat sig, dels om att en depression var sannolik vid höga poäng. Kontrollåtgärden var oftast att patienterna fyllde i formuläret utan att läkaren fick tillgång till resultaten. Författarna vägde samman resultaten i metaanalyser. Läkarnas handläggning förändrades inte för ”oselekterade” patienter i fem studier, riskkvot 0,97 (95 % KI, 0,81 till 1,18). Huruvida handläggningen för högriskpatienter förändrades är mera svårbedömt. Sammanvägning av resultaten från fem studier gav en riskkvot på 1,50 (95 % KI, 0,89 till 2,53). De fem studier som även undersökte effekten på depressionsutfall såg ingen effekt av screeningen, SMD -0,02 (95 % KI, -0,25 till 0,20).

O'Connor och medarbetare undersökte om screening för depression i primärvården ledde till ökad grad av förbättring och tillfrisknande [2]. Åtta studier som författarna bedömde hålla god eller måttligt god kvalitet ingick. Till skillnad från översikten av Gilbody och medarbetare inkluderade O'Connor och medarbetare även studier där diagnostiken var del av ett förbättrat patientomhändertagande, t ex case managementmetodik eller konsultation med psykiatrer. Studierna analyserades efter åldersgrupp (vuxna respektive äldre) men inte efter om patienterna till-

hörde någon riskpopulation. Författarna bedömde att studierna var för heterogena för att metaanalyser skulle vara meningsfulla. De kom fram till att screening av vuxna personer, med tillägg av förbättrat patientomhändertagande, kan reducera depressiva symtom jämfört med rutinsjukvård. Fyra studier rörde äldre deltagare och i tre av dem uteblev positiva effekter.

Williams och medarbetare skulle undersöka det vetenskapliga underlaget för screening av barn och ungdom men fann inga studier [3].

Två andra översikter fokuserade på effekten av screening på selekterade riskgrupper [4,6]. Thombs och medarbetare avsåg att analysera om systematisk screening för depression jämfört med rutinhandläggning påverkade förekomsten av depressiva symtom vid uppföljning av patienter med hjärt- och kärlsjukdom [6]. Författarna hittade inga studier som besvarade frågan.

Hewitt och medarbetare studerade om screening av nyförlösta kvinnor med symtom på depression, separat eller tillsammans med annan förbättrad handläggning, leder till en lägre andel kvinnor med post partum-depression [4]. Resultatet från tre inkluderade studier visade att screening med formuläret EPDS tillsammans med ett förbättrat patientomhändertagande ledde till en minskning av depressionspoäng vid uppföljning.

## **Bedömning av evidensstyrka**

Tabell 3.2.1 sammanställer det vetenskapliga underlaget för att bedöma patientens nytta av att formulär för riktad screening används som ett hjälpmedel för att upptäcka depression. Evidensstyrkan bedömdes med hjälp av kriterierna i GRADE.

Samtliga översikter är välgjorda. De brister som finns i underlaget gäller O'Connor och medarbetares översikt. O'Connor och medarbetare fann att studierna var för heterogena för att medge en metaanalys. Slutsatsen är därför endast kvalitativ, vilket leder till en bristande precision.

**Tabell 3.2.1** Resultatsammanställning. Studier som undersöker patientnyttan med screening för depression.

Utfallsmått	Antal studier Referens Antal patienter	Sammanvägd effekt (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Depressionspoäng, oselekterade vuxna och äldre i primärvård när screening inte kopplas till andra åtgärder	5 [1] n=1 842	SMD -0,02 (95% KI, -0,25; 0,20)	Starkt (⊕⊕⊕⊕)	
Depressionspoäng, vuxna och äldre i primärvård när screening ingår i en större intervention	8 [2]	Liten effekt när det ingår i en större intervention	Begränsat (⊕⊕○○)	Bristande precision: -2
Depressionspoäng, barn och ungdom i primärvård	0 [2,3]	-	-	Studier saknas
Depressionspoäng, vuxna med hjärt- och kärlsjukdom	0 [6]	-	-	Studier saknas
Depressionspoäng, nyblivna mödrar	4 [1,4] n= >1 400	OR 0,61 (95% KI, 0,48; 0,76) när screening med EPDS är del av intervention	Begränsat (⊕⊕○○)	Bristande överförbarhet: -1 Bristande precision: -1
Förändrad vårdprocess	10 [1] n=2 333	RR 1,30 (95% KI, 0,97; 1,76)	Otillräckligt (⊕○○○)	Bristande samstämmighet: -1 Bristande precision: -2

KI = Konfidensintervall; OR = Oddsquot; RR = Relativ risk; SMD = Standard Mean Difference

## Evidensgraderade resultat

- Screening som enda intervention har ingen effekt på symtombörda av depression, SMD  $-0,02$  (95 % KI,  $-0,25$  till  $0,20$ ) hos vuxna, oselekerade patienter i primärvården (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- Screening som del i en förstärkt handläggning påverkar tillfrisknande eller symtombörda av depression hos vuxna patienter i primärvården i liten utsträckning (begränsat vetenskapligt underlag ⊕⊕○○).
- Det går inte att bedöma om screening för depression hos vuxna i primärvård påverkar vårdrutinerna pga att studierna har bristande samstämmighet och precision (otillräckligt vetenskapligt underlag ⊕○○○).
- En förstärkt handläggning som inkluderar screening med EPDS minskar symtombördan, OR  $0,61$  (95 % KI,  $0,48$  till  $0,76$ ) av post partum-depression hos nyförlösta (begränsat vetenskapligt underlag ⊕⊕○○).
- Det saknas studier som utvärderar effekter på tillfrisknande, symtombörda eller handläggning av att följa patienternas depressionsdjup med hjälp av symtomskattningsskala.

## Diskussion

Vår granskning visar att det finns anmärkningsvärt få studier som har undersökt om hälsa och omhändertagande av patienterna påverkas av att diagnostiska hjälpmedel används i samband med diagnos eller uppföljning. Om man ser till hela diagnostiska processen kan man konstatera att det endast är screening som är utvärderad. Det är också värt att notera att de få studier som finns uteslutande avser vuxna och äldre inom primärvården eller kvinnor som är gravida eller nyförlösta.

Resultatet av de översikter som undersökt hur screening som enda åtgärd påverkar depressionsutfallet är föga uppmuntrande med en effektstorlek, SMD på  $-0,02$ . Med samma aktivitet, i kombination med förstärkt om-

händertagande, kunde symtombördan reduceras jämfört med rutinsjukvård, men utfallet stod sig inte i gruppen av äldre patienter då resultaten särredovisades [2].

Det kan finnas flera orsaker till att screening och återkoppling som enda åtgärd inte resulterar i ett positivt utfall. En sådan möjlig förklaring är att depression som identifierats med hjälp av screening skiljer sig från depression som lätt går att identifiera kliniskt. Patienter vars depression förblivit oupptäckt i primärvård har visat sig ha mindre funktionspåverkan och lägre svårighetsgrad jämfört med depression som identifierats utan screening. Det kan inte uteslutas att patienter med sådana lättare former av depression inte har lika stor nytta av aktiv behandling alternativt inte svarar lika bra på behandling som de med uppenbar klinisk depression [5].

Att screening i kombination med förstärkt omhändertagande resulterar i bättre utfall är förväntat. Men, som visades i O'Connors och medarbetares översikt [2], var förbättringen relativt blygsam sett över en längre tidsperiod. Betydligt mer övertygande var resultaten från den översikt som utvärderat effekten av screening och systematiskt omhändertagande av kvinnor med post partum-depression. Begränsningen i dessa översikter utgörs av det faktum att det inte går att separera effekten av enskilda åtgärder.

Studier som undersöker om en sekventiell diagnostik med strukturerade diagnostiska intervjuer och symtomskattning påverkar omhändertagande och symtombörda saknas helt. Likaså saknas studier som följer upp behandlingseffekter med hjälp av bedömningsformulär. En systematisk översikt av Knaup och medarbetare antyder dock att det kan finnas ett värde med systematisk uppföljning [7]. Översikten utvärderade effekter av att återkoppla resultat av behandling inom specialistpsykiatri. Tolv studier inkluderades. Patienterna hade olika psykiska sjukdomar, däribland depression. Effekter av behandling mättes med standardiserade formulär, vanligen OQ-45 (Outcome Questionnaire). Studierna visade en liten men signifikant effekt på patienternas psykiska hälsa.



**Table 3.2.2** Does improved rate of recognition of depression result in better management of depression and better outcomes for the patient?

Author Year Reference Country	Design Setting Population	Intervention Control Drop-out rate	Process outcome	Patient outcome	Study quality Comments
Gilbody et al 2008 [1] United Kingdom	<p><u>Design</u> Systematic review based on RCT published before January 2008</p> <p>n=5 studies (1 842 patients) were included for review of patient outcome</p> <p>n=10 studies (2 333 patients) were included for review of management of depression</p> <p><u>Setting</u> Primary care in Europe and the USA</p> <p><u>Population</u> Unselected patients in the waiting room or high risk population</p>	<p><u>Intervention</u> Screening with various rating scales for depression in the waiting room without feedback to the GP</p> <p><u>Control</u> Screening for depression in the waiting room without feedback to the GP</p>	<p><u>Management of depression, unselected patients</u> Risk ratio: 0.97 (95% CI, 0.81; 1.18)</p> <p><u>Management of depression, high-risk patients</u> Risk ratio: 1.50 (95% CI, 0.89; 2.53)</p>	<p><u>Depression at follow-up</u> SMD -0.02 (95% CI, -0.25; 0.20), favouring control</p>	High
Hewitt et al 2009 [4] United Kingdom	<p><u>Design</u> Systematic review based on RCT and controlled trials published before March 1, 2007</p> <p>n=4 studies were included</p> <p><u>Setting</u> GP clinics and hospitals in United Kingdom, China, USA and Australia</p> <p><u>Population</u> Women with newborns</p> <p><u>Follow-up</u> 6 to 16 weeks</p>	<p><u>Intervention</u> Screening with EPDS, with (2 studies) or without (2 studies) enhanced care n=1 802 (EPDS)</p> <p><u>Control</u> Usual care n=1 659</p> <p><u>Drop-out rates</u> 26 to 63%</p>	Not investigated in the review	<p><u>EPDS score above cut off at follow-up</u> OR 0.61 (95% CI 0.48; 0.77)</p> <p>Note: As screening was part of other changes in two studies out of three, it is difficult to assess the impact of screening component itself</p>	High

The table continues on the next page



Table 3.2.2 continued

Author Year Reference Country	Design Setting Population	Intervention Control Drop-out rate	Process outcome	Patient outcome	Study quality Comments
Thombs et al 2008 [6] Canada	<u>Design</u> Systematic review based on RCT and controlled trials up to May 2008  <u>Setting</u> Cardiovascular care  <u>Population</u> Patients with cardiovascular disease	<u>Intervention</u> Screening for depression (various rating scales)  <u>Control</u> No screening	No studies found	No studies found	High
O'Connor et al 2009 [2] USA	<u>Design</u> Systematic review based on RCT and controlled trials up to December 31, 2007  n=6 studies on adults n=4 on elderly fulfilled inclusion and quality criteria  <u>Setting</u> Primary care  <u>Population</u> Adults	<u>Intervention</u> Various screening methods in the primary care waiting room with or without enhanced care  <u>Control</u> No screening	Not investigated in the review	Screening programs are likely to reduce morbidity and mortality when other staff provided part of the depression care, such as assessment and monitoring in coordination with the primary care provider's treatment (9 studies)  No data identifying harm with screening	High
Williams et al 2009 [3] USA	<u>Design</u> Systematic review  <u>Setting</u> Primary care  <u>Population</u> Children and adolescents, 7–18 years	<u>Intervention</u> Routine screening  <u>Control</u> No screening	No studies found	No studies found	High

CI = Confidence interval; EPDS = Edinburgh post natal depression scale; GP = General practitioner; OR = Odds ratio; RCT = Randomised controlled trial; SMD = Standard mean difference

## Referenser

1. Gilbody S, Sheldon T, House A. Screening and case-finding instruments for depression: a meta-analysis. *CMAJ* 2008;178:997-1003.
2. O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Ann Intern Med* 2009;151:793-803.
3. Williams SB, O'Connor EA, Eder M, Whitlock EP. Screening for child and adolescent depression in primary care settings: a Systematic Evidence Review for the US Preventive Services Task Force Screening for Child and Adolescent Depression in Primary Care Settings. *Pediatrics* 2009; 123:e716-35.
4. 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.
5. Oxman TE, Dietrich AJ, Williams JW, Jr., Kroenke K. A three-component model for reengineering systems for the treatment of depression in primary care. *Psychosomatics* 2002; 43:441-50.
6. Thombs BD, de Jonge P, Coyne JC, Whooley MA, Frasure-Smith N, Mitchell AJ, et al. Depression screening and patient outcomes in cardiovascular care: a systematic review. *JAMA* 2008; 300:2161-71.
7. Knaup C, Koesters M, Schoefer D, Becker T, Puschner B. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *Br J Psychiatry* 2009;195:15-22.

### 3.3 Klinisk bedömning med och utan stöd av intervjuformulär

#### Evidensgraderade resultat

- Det går inte att bedöma sensitiviteten för klinisk diagnos utan stöd av formulär utförd av psykiatrer för egentlig depression med LEAD som referensstandard pga för få studier och stor spridning i resultaten mellan öppenvård och slutenvård (otillräckligt vetenskapligt underlag ⊕○○○). Specificiteten är 92 procent (95 % KI, 88 till 95) (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- Klinisk diagnos utförd av psykiatrer, utan stöd av formulär, har en sensitivitet på 52 procent (95 % KI, 40 till 64) (begränsat vetenskapligt underlag ⊕⊕○○) och specificitet på 94 procent för bipolärt syndrom, med LEAD som referensstandard (95 % KI, 91 till 97) (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- Det går inte att bedöma sensitiviteten för klinisk diagnos utan stöd av formulär utförd av psykiatrer för schizoaffektivt syndrom med LEAD som referensstandard pga för få studier och för stor spridning i resultaten (otillräckligt vetenskapligt underlag ⊕○○○). Specificiteten är 90 procent (95 % KI, 85 till 94) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Det går inte att bedöma sensitivitet och specificitet för klinisk diagnos utan stöd av formulär utförd av psykiatrer med SCID-I som referensstandard för att diagnostisera förstämningssyndrom i kliniska populationer pga att antalet studier är för litet (otillräckligt vetenskapligt underlag ⊕○○○).
- Klinisk diagnos av egentlig depression utförd av distriktsläkare utan stöd av formulär har en sensitivitet på 47,3 procent (95 % KI, 41,7 till 53,0) (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 81,3 procent (95 % KI, 74,5 till 87,3) (begränsat vetenskapligt underlag ⊕⊕○○), med intervjubaserad diagnostik som referensstandard.

## Introduktion

En av våra grundläggande frågor var om diagnostiskt stöd i form av intervjuer och skattningar av besvär ökar möjligheten till en korrekt diagnos av förstämningssyndrom. Men hur träffsäker är läkarens/ klinikerens bedömning utan diagnostiskt stöd?

I detta avsnitt undersöker vi den diagnostiska tillförlitligheten för psykiatrers respektive distriktsläkares kliniska bedömning utan hjälpmedel. Utfallsmåtten var sensitivitet och specificitet. Referensstandarderna kunde vara antingen LEAD (Longitudinal, Experts, All Data procedure) eller strukturerade/semistrukturerade intervjuer.

## Klinisk diagnos av psykiatrer jämfört med LEAD

Litteratursökningen resulterade i 1 420 abstrakt som avsåg klinisk diagnos med antingen LEAD eller strukturerade/semistrukturerade intervjuer.

### Beskrivning av underlaget

Fyra studier lästes i fulltext, varav en exkluderades eftersom den inte använde sedvanlig LEAD [1]. Det vetenskapliga underlaget bestod därmed av tre studier varav två hade hög och en hade medelhög studie-kvalitet [2–4]. Två studier utfördes i USA och jämförde även den semistrukturerade intervjun SCID med LEAD (Kapitel 3.5). Totalt ingick 372 patienter i de tre studierna.

Studien av Ramirez Basco och medarbetare omfattade 200 patienter från en psykiatrisk öppenvårdsmottagning [2]. Miller och medarbetare inkluderade 56 patienter från universitetsanknuten psykiatrisk slutenvård [3]. I studien av Taiminen och medarbetare deltog 116 patienter, varav mer än 80 procent vårdades för allvarlig psykisk sjukdom för första gången, på en finsk universitetsanknuten psykiatrisk klinik inom offentlig vård [4]. I denna studie ingick bl a intervju med SCAN som del i LEAD-förfarandet.

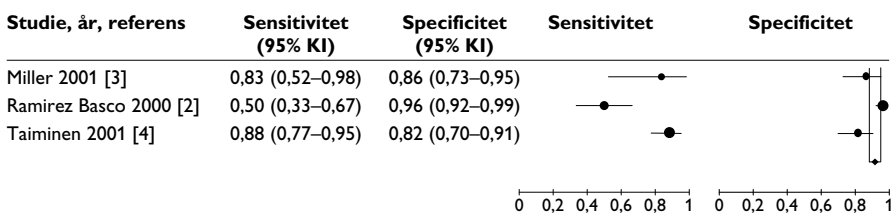
## Sammanvägda resultat

Den generella överensstämmelsen för samtliga diagnoser mellan klinisk diagnostik och LEAD, mätt med Cohen's kappas, var låg: 0,43, 0,45 respektive 0,51.

För klinisk diagnos av *egentlig depression* var sensitiviteten mellan 50 och 88 procent i de tre studierna medan specificiteten låg mellan 82 och 96 procent (Figur 3.3.1). Vi bedömde att en sammanvägning av resultaten för sensitivitet inte var meningsfull. Det sammanvägda estimatet för specificitet blev 92 procent (95 % KI, 88 till 95).

Vid *bipolärt syndrom* var sensitiviteten lägre, mellan 46 och 54 procent medan specificiteten låg mellan 92 och 98 procent (Figur 3.3.2). De sammanvägda estimaten blev 52 procent (95 % KI, 40 till 64) för sensitivitet och 94 procent (95 % KI, 91 till 97) för specificitet.

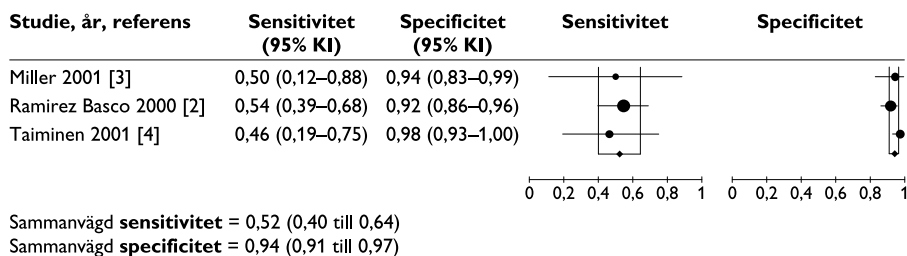
Två av studierna undersökte också tillförlitligheten vid schizo-affektivt syndrom och där var sensitiviteten 33 respektive 53 procent och specificiteten 89 respektive 91 procent [2,3] (Figur 3.3.3). Vi bedömde att en sammanvägning av resultaten för sensitivitet inte var meningsfull. Den sammanvägda specificiteten var 90 procent (95 % KI, 85 till 94).



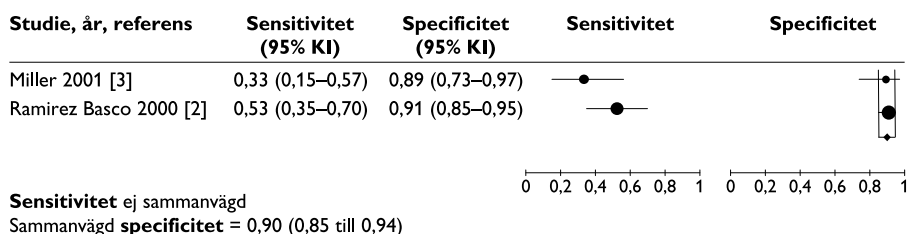
Sensitivitet ej sammanvägd

Sammanvägd **specificitet** = 0,92 (0,88 till 0,95)

**Figur 3.3.1** Forest plot. Spridning av sensitivitet och specificitet för klinisk diagnos av *egentlig depression* med LEAD som referensstandard.



**Figur 3.3.2** Forest plot. Spridning av sensitivitet och specificitet för klinisk diagnos av bipolärt syndrom med LEAD som referensstandard.



**Figur 3.3.3** Forest plot. Spridning av sensitivitet och specificitet för klinisk diagnos av schizoaftektivt syndrom med LEAD som referensstandard.

## Bedömning av evidensstyrka

Underlaget medgav följaktligen en bedömning med GRADE av sensitivitet för bipolära syndrom och för specificitet för såväl depression som bipolära syndrom och schizoaftektivt syndrom.

Två av de ingående studierna hade hög studiekvalitet [2,3] och den tredje medelhög [4]. I alla studierna genomfördes en mycket noggrann LEAD med konsensusförfarande vilket gör att frågan om interbedömarreliabilitet av referenstestet inte var tillämplig. I studien av Taiminen och medarbetare framgick inte tiden klart mellan LEAD och klinisk diagnos, men man hade på ett systematiskt sätt bedömt vilka diagnoser som hade satts under vårdperioden. Det innebär att vi bedömde att risken för bias var låg.

Överförbarheten var god. De ingående populationerna var relevanta för svenska kliniska förhållanden och representerade både psykiatrisk öppenvård och slutenvård.

Precisionen var låg för sensitivitet för bipolära syndrom vilket medförde att evidensstyrkan sänktes två steg.

Våra bedömningar finns sammanställda i Tabell 3.3.1 nedan.

**Tabell 3.3.1** Sammanvägda resultat avseende diagnoserna egentlig depression, bipolärt syndrom och schizoaffektivt syndrom för klinisk intervju jämfört med LEAD.

Utfallsmått	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
<b>Sensitivitet</b>				
Egentlig depression	3 studier n=372	Bör inte vägas samman	Otillräckligt ⊕○○○	
Bipolärt syndrom	3 studier n=372	52% (95% KI, 40; 64)	Begränsat ⊕⊕○○	Bristande precision (-2)
Schizoaffektivt syndrom	2 studier n=256	Bör inte vägas samman	Otillräckligt ⊕○○○	
<b>Specificitet</b>				
Egentlig depression	3 studier n=372	92% (95% KI, 88; 95)	Starkt ⊕⊕⊕⊕	
Bipolärt syndrom	3 studier n=372	94% (95% KI, 91; 97)	Starkt ⊕⊕⊕⊕	
Schizoaffektivt syndrom	2 studier n=256	90% (95% KI, 85; 94)	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (-1)

KI = Konfidensintervall

## Evidensgraderade resultat

- Det går inte att bedöma sensitiviteten för klinisk diagnos utförd av psykiatrer för att identifiera egentlig depression med LEAD som referensmetod pga stor spridning i resultaten mellan öppen- och slutenvård (otillräckligt vetenskapligt underlag ⊕○○○).
- Klinisk diagnos utförd av psykiatrer för att identifiera egentlig depression med LEAD som referensstandard har en specificitet på 92 procent (95 % KI, 88 till 95), (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- Klinisk diagnos utförd av psykiatrer för att identifiera bipolärt syndrom, med LEAD som referensstandard har en sensitivitet på 52 procent (95 % KI, 40 till 64) (begränsat vetenskapligt underlag ⊕⊕○○) och specificitet på 94 procent (95 % KI, 91 till 97) (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- Det går inte att bedöma sensitiviteten för klinisk diagnos utförd av psykiatrer för schizoaffektivt syndrom med LEAD som referensstandard pga för få studier och för stor spridning i resultaten (otillräckligt vetenskapligt underlag ⊕○○○). Specificiteten är 90 procent (95 % KI, 85 till 94) för schizoaffektivt syndrom, med LEAD som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

## Klinisk diagnos av psykiatrer jämfört med SCID

En annan kliniskt relevant fråga är om sedvanlig klinisk diagnostik utförd av psykiatrer har samma sensitivitet och specificitet som det semi-strukturerade formuläret SCID som vi valt som referensstandard. SCID utvecklades för att öka tillförlitligheten av psykiatriska diagnoser [5].

### Beskrivning av underlaget

Av 26 studier som lästes i fulltext berörde fyra studier frågeställningen. Tre av dem exkluderades. Orsaker var för lång tid mellan klinisk diagnos och SCID i två studier [6,7]. Den tredje studien innehöll så många metodologiska oklarheter att den exkluderades [8]. En studie hade medelhög kvalitet och utgjorde därmed det vetenskapliga underlaget [9].



Steiner och medarbetare [9] rekryterade 100 patienter som var aktuella för inläggning eller dagvård vid en psykiatrisk klinik i Connecticut, USA. Den kliniska diagnostiken utfördes av erfarna psykiatrer, och SCID-intervjuarna var välutbildad forskningspersonal. Psykiatrerna var omedvetna om att deras diagnoser skulle ingå i undersökningen. Generell överensstämmelse mellan bedömningarna med och utan intervjustöd för alla diagnoser, mätt med viktad kappa, var 0,25, dvs låg. Sensitiviteten var 32 procent för egentlig depression, 15 procent för dystymi, 37 procent för bipolärt syndrom och 12 procent för schizoaffektivt syndrom. Specificiteten var 44 procent för egentlig depression, 88 procent för dystymi, 86 procent för bipolärt syndrom och 35 procent för schizoaffektivt syndrom.

### **Bedömning av evidensstyrka**

Det går inte att bedöma sensitivitet och specificitet för klinisk diagnos utförd av psykiatrer med SCID som referensstandard för att diagnostisera förstämningssyndrom i kliniska populationer pga att antalet studier är för litet (otillräckligt vetenskapligt underlag ⊕○○○).

### **Klinisk diagnos av distriktsläkare jämfört med intervjubaserad diagnostik**

Vi gjorde ingen egen systematisk litteratursökning för frågan om sensitivitet och specificitet för klinisk diagnos ställd av distriktsläkare jämfört med strukturerade/semistrukturerade intervjuer eftersom det visade sig finnas en nyligen publicerad systematisk översikt [10]. Vi gjorde en kompletterande sökning efter nytillkommen litteratur med samma söktermer som i översikten. Inga ytterligare studier återfanns.

### **Beskrivning av underlaget**

Mitchell och medarbetare genomförde en systematisk litteraturöversikt med hög kvalitet om sensitivitet och specificitet för klinisk diagnos utförd av distriktsläkare jämförd med intervjubaserad diagnostik för egentlig depression [10]. Översikten omfattade studier som publicerats före maj 2009 i ett brett urval databaser. Av 118 studier som berörde distriktsläkares kliniska diagnos utan hjälpmedel kunde 41 studier inkluderas. Anledningar till att studier exkluderades var att de använde annan

referensstandard som t ex symtomskattningsskalor (61 studier), dubbelpublikationer (5 studier) eller otillräckliga data (7 studier). Nitton studier rapporterade både sensitivitet och specificitet, varav 13 studier berörde vuxna patienter och 6 studier äldre patienter. Övriga 22 studier rapporterade bara sensitivitet.

Totalt ingick över 50 000 patienter med i genomsnitt 1 228 patienter per studie. Tjugoåtta studier använde semistrukturerad eller strukturerad intervju och 12 studier annan expertintervju som referenstest. Vårdcentralerna var i huvudsak belägna i städer i tio olika länder. Den sammanvägda prevalensen för depression var 19,5 procent (95 % KI, 15,7 till 23,7) och varierade från 10,2 procent i Australien till 22,6 procent i Nederländerna.

### **Sammanvägda resultat**

Samtliga 41 studier redovisade sensitivitet. Sett över alla studier kunde distriktsläkarna identifiera egentlig depression hos 45,4 procent av dem som enligt den diagnostiska intervjun hade depression. Sensitiviteten varierade mycket mellan enskilda studier (från 6,6 till 78,8 %). Det sammanvägda estimatet var 47,3 procent (95 % KI, 41,7 till 53). Den sammanvägda specificiteten för de 19 studier som redovisade sådana uppgifter blev 81,3 procent (95 % KI, 74,5 till 87,3).

Distriktsläkarnas förmåga att ställa diagnosen egentlig depression varierade mellan länderna. Den var högst i Italien och lägst i Storbritannien (andel korrekt identifierade 82,5 jämfört med 50 %) [11].

### **Bedömning av evidensstyrka**

Kvaliteten på den systematiska översikten, granskad med AMSTAR, bedömdes som hög. Materialet innefattade ett mycket stort antal patienter i primärvård från många länder. Det medförde att överförbarheten var god för svenska förhållanden även om ingen svensk studie ingick. Däremot användes flera olika referensstandarder, varav några inte var strukturerade intervjuer, vilket drog ned evidensgraden något. Vidare var det en stor spridning mellan de individuella studierna både för sensitivitet och för specificitet vilket medförde ett visst avdrag pga bristande

samstämmighet. Precisionen var otillräcklig vilket ledde till att evidensgraden sänktes ytterligare ett steg.

**Tabell 3.3.2** Sammanvägda resultat för klinisk diagnos av egentlig depression utförd av distriktsläkare jämfört med intervju-baserad diagnostik.

Utfallsmått	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	41 studier n= >50 000	47,3% (95% KI, 41,7; 53,0)	Begränsat ⊕⊕○○	Smärre brister i överförbarhet och samstämmighet: -1 Precision: -1
Specificitet	19 studier Antal patienter framgår inte	81,3% (95% KI, 74,5; 87,3)	Begränsat ⊕⊕○○	Smärre brister i överförbarhet och samstämmighet: -1 Precision: -1

### Evidensgraderade resultat

Klinisk diagnos utförd av distriktsläkare har en sensitivitet på 47,3 procent (95 % KI, 41,7 till 53) (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 81,3 procent (95 % KI, 74,5 till 87,3) (begränsat vetenskapligt underlag ⊕⊕○○) för egentlig depression, med intervju-baserad diagnostik som referensstandard.

### Diskussion

En grundläggande fråga för vår utvärdering var om diagnosen kan bli mer träffsäker om den kliniska proceduren kompletteras med olika bedömningsformulär. Vi kunde konstatera att det finns en påfallande skillnad vad gäller antal och storlek på studier. För primärvård omfattade underlaget drygt 40 studier med över 50 000 patienter, medan de tre studierna från psykiatrisk vård inkluderade knappt 400 patienter.

Ett genomgående drag i studierna från primärvård och psykiatri är en hög specificitet, dvs att läkaren rätt väl kan identifiera de som är friska. De två studierna i psykiatrisk slutenvård visade även att läkaren hade en god förmåga att identifiera de som hade depression. Däremot missade läkarna i primärvård och psykiatrisk öppenvård cirka hälften av patienterna med depression. Detsamma gällde bipolära syndrom inom den psykiatriska öppenvården.

Användning av strukturerade intervjuer ökar antalet patienter med förstämningssyndrom som upptäcks. De kan även motverka en annan risk, den att vissa diagnoser blir ”modediagnoser”. En modediagnos innebär att förekomsten ökar som följd av uppmärksamhet från t ex media eller forskarsamhället kring diagnosen. Det leder i sin tur till en selektiv uppmärksamhet från klinikerns sida. Här har strukturerade eller semistrukturerade intervjuer en viktig funktion att fylla för att motverka sådana influenser.



**Table 3.3.3** Are clinical evaluation and best estimate (LEAD) equal for diagnosing affective disorders?

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
Miller et al 2001 [3] USA	<p><u>Design</u> Cross sectional and longitudinal</p> <p><u>Setting</u> University-affiliated public psychiatric hospital</p> <p><u>Population</u> First patient each week was screened n=75 inpatients eligible</p> <p><u>Inclusion criteria</u> &gt;18 years, enough English fluency and cognitive function to participate in interviews</p> <p><u>Exclusion criteria</u> Too acutely ill</p> <p><u>Prevalence</u> Mania: 10.7% (6/56) Depression: 21.4% (12/56) Schizoaffective: 37.5% (21/56)</p>	<p><u>Index test</u> TDA (traditional diagnostic assessment). At discharge comprising unstructured interview and medical records</p> <p><u>Reference test</u> LEAD Consensus by raters using discharge diagnoses, SCID-I, structured interview, all data 1–10 days after discharge</p> <p><u>Number of patients</u> n=56 (45% females) Mean age: 36.9 years (SD 12.4, range 19–59)</p> <p><u>Drop-out rate</u> 0</p>	<p><u>Rater of index test</u> 6 experienced clinicians (5 psychiatrists, 1 PhD psychologist)</p> <p><u>Rater training</u> NA</p> <p><u>Rater of reference test</u> Three to four of the index test raters</p> <p><u>Training of reference test</u> Extensive training</p> <p><u>Interobserver reliability</u> NA for LEAD</p>	<p><u>Bipolar disorder</u> Sensitivity: 50% Specificity: 94%</p> <p><u>Depression</u> Sensitivity: 83% Specificity: 86%</p> <p><u>Schizoaffective disorder</u> Sensitivity: 33% Specificity: 89%</p>	<p>High</p> <p><u>Comments</u> Sampling method Adequate</p> <p>Blinding Satisfactory</p> <p>Handling of missing data NA</p> <p>Other comments Withdrawals explained</p>

The table continues on the next page

Table 3.3.3 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
Ramirez Basco et al 2000 [2] USA	<p><u>Design</u> Cross sectional and longitudinal</p> <p><u>Setting</u> Community mental health center</p> <p><u>Population</u> n=210 outpatients recruited by in-clinic advertisements and clinical referrals</p> <p><u>Inclusion criteria</u> Not reported</p> <p><u>Exclusion criteria</u> Not reported</p> <p><u>Prevalence</u> BP: 25% (50/200) MDD: 19% (38/200) Schizoaffective: 18% (36/200)</p>	<p><u>Index test</u> Routine diagnoses Interview by psychiatrists 40–45 minutes without access to prior medical records</p> <p><u>Reference test</u> LEAD Interview after a week with one of prime investigators (psychiatrist or psychologist) after reviewing all data</p> <p><u>Number of patients</u> n=200 (58% females) Mean age: 37.6 years (SD 10)</p> <p><u>Drop-out rate</u> 0</p>	<p><u>Rater of index test</u> Three psychiatrists</p> <p><u>Training of index test</u> Yes</p> <p><u>Rater of reference test</u> Psychiatrist or psychologist</p> <p><u>Training of reference test</u> NA</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>Bipolar I disorder</u> Sensitivity: 54% Specificity: 92%</p> <p><u>Major depressive disorder</u> Sensitivity: 50% Specificity: 96%</p> <p><u>Schizoaffective disorder</u> Sensitivity: 53% Specificity: 91%</p>	<p>High</p> <p><u>Comments</u> <u>Sampling method</u> Adequate</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> Adequate</p> <p><u>Other comments</u> Analysis of time requirements for the diagnostic steps, and effects on treatment of diagnostic feedback, adds further to clinical relevance</p>

The table continues on the next page

Table 3.3.3 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
Taiminen et al 2001 [4] Finland	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Turku University Central Hospital and Turku City Mental Health Centre, Finland</p> <p><u>Population</u> All patients admitted during study period n=116 53% females, 84% inpatients</p> <p>Mean age: 34.5 years (SD 11.4, range 16–62)</p> <p><u>Inclusion criteria</u> 1. Schizophrenia spectrum disorder, bipolar disorder, severe major depression 2. First-ever psychiatric treatment in the public sector for the disorder 3. Age 16–64 years 4. Residence in the catchment area</p> <p><u>Exclusion criteria</u> Not reported</p> <p><u>Prevalence</u> BP: 11% (13/116) MDD: 51% (59/116)</p>	<p><u>Index test</u> Clinical discharge diagnoses according to DSM-IV and ICD-10, retrospectively obtained by a researcher</p> <p><u>Reference test</u> LEAD Interview with structured anamnestic information, SCAN-interview and assessment of the severity of psychiatric symptoms</p> <p>Consensus meetings, with three psychiatrists among one had performed the interview, decided Best-estimate diagnoses according to DSM-IV using all data</p> <p><u>Number of patients</u> n=116</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of index test</u> Psychiatrists</p> <p><u>Training of index rater</u> NA</p> <p><u>Rater of reference test</u> Five research psychiatrists</p> <p><u>Training of reference test</u> Not reported</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>Depressive disorders</u> Sensitivity: 88% Specificity: 82%</p> <p><u>Bipolar disorder</u> Sensitivity: 46% Specificity: 98%</p>	<p>Moderate</p> <p><u>Comments</u> <u>Sampling method</u> Yes, but inclusion criteria might imply some bias</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> NA</p> <p><u>Other comments</u> Training and interrater reliability not reported but careful Best estimate process with consensus diagnoses. Period between clinical evaluation and LEAD unclear</p>

BP = Bipolar disorder; DSM-IV = Diagnostic and statistical manual of mental disorder, fourth edition; LEAD = Longitudinal, Experts, All Data procedure; MDD = Major depressive disorder; NA = Not applicable; SCAN = Schedules for Clinical Assessment in Neuropsychiatry; SCID-I = Structured Clinical Interview for DSM-IV Axis I disorders; SD = Standard deviation



**Table 3.3.4** Is clinical evaluation by psychiatrists and SCID equal for diagnosing affective disorder?

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
Steiner et al 1995 [9] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Entry and Crisis Service at a Mental Health Center serving an urban, poor clientele</p> <p><u>Population</u> Patients in need of hospitalisation, n=100</p> <p><u>Inclusion criteria</u> Active psychiatric disorder, severe or moderate disturbance of functioning, dangerous to self or others, not responding to earlier treatment, no other mental services available</p> <p><u>Exclusion criteria</u> Not reported</p> <p><u>Prevalence</u> MDD: 38% (38/100) BP: 19% (19/100) Schizoaffective disorder: 34% Dysthymia: 13% (13/100)</p>	<p><u>Index test</u> Clinical interviews, sometimes supplemented by medical records</p> <p><u>Reference test</u> SCID-P (DSM-III-R) with limited chart information</p> <p><u>Number of patients</u> n=100 (51% females) Mean age: 33.6 years (SD 10.4)</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of index test</u> Senior psychiatrists</p> <p><u>Training of index rater</u> Psychiatrists were unaware that their assessments should be scrutinized</p> <p><u>Rater of reference test</u> One clinical psychologist and three experienced clinical research interviewers</p> <p><u>Training of reference test</u> Extensive training including SCID-P videos and in vivo observations</p> <p><u>Interobserver reliability</u> SCID-P protocols were reviewed by the psychologist with SCID-P interviewers, arriving at consensus</p>	<p><u>MDD</u> Sensitivity: 32% Specificity: 44%</p> <p><u>Bipolar disorder</u> Sensitivity: 37% Specificity: 86%</p> <p><u>Dysthymia</u> Sensitivity: 15% Specificity: 88%</p> <p><u>Schizoaffective disorder</u> Sensitivity: 12% Specificity: 35%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Yes, except sharing some chart information</p> <p><u>Handling of missing data</u> NA</p> <p><u>Other comments</u> No interrater reliability reported of the SCID-P but there were consensus ratings</p>

BP = Bipolar disorder; MDD = Major depressive disorder; NA = Not applicable;  
SCID-P = Structured clinical interview for DSM-IV axis I disorders, patient edition;  
SD = Standard deviation

## Referenser

1. 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.
2. 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.
3. 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.
4. Taiminen T, Ranta K, Karlsson H, Lauerma H, Leinonen KM, Wallenius E, et al. Comparison of clinical and best-estimate research DSM-IV diagnoses in a Finnish sample of first-admission psychosis and severe affective disorder. *Nord J Psychiatry* 2001;55:107-11.
5. Williams JB, Gibbon M, First MB, Spitzer RL, Davies M, Borus J, et al. The Structured Clinical Interview for DSM-III-R (SCID). II. Multisite test-retest reliability. *Arch Gen Psychiatry* 1992;49:630-6.
6. Kashner TM, Rush AJ, Suris A, Biggs MM, Gajewski VL, Hooker DJ, et al. Impact of structured clinical interviews on physicians' practices in community mental health settings. *Psychiatr Serv* 2003;54:712-8.
7. Shear MK, Greeno C, Kang J, Ludewig D, Frank E, Swartz HA, et al. Diagnosis of nonpsychotic patients in community clinics. *Am J Psychiatry* 2000;157:581-7.
8. Dunner DL, Tay LK. Diagnostic reliability of the history of hypomania in bipolar II patients and patients with major depression. *Compr Psychiatry* 1993;34:303-7.
9. Steiner JL, Tebes JK, Sledge WH, Walker ML. A comparison of the structured clinical interview for DSM-III-R and clinical diagnoses. *J Nerv Ment Dis* 1995;183:365-9.
10. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009;374:609-19.
11. Mitchell AJ, Rao S, Vaze A. International comparison of clinicians' ability to identify depression in primary care: meta-analysis and meta-regression of predictors. *Br J Gen Pract* 2011;61:e72-80.

## 3.4 Bedömningsinstrument för barn och ungdomar

### Evidensgraderade resultat

#### Intervjuer

- Det finns för få studier för att göra en säker uppskattning av sensitivitet och specificitet för att identifiera förstämningssyndrom hos barn eller ungdomar för de semistrukturerade intervjuerna K-SADS och DICA jämfört med LEAD (otillräckligt vetenskapligt underlag ⊕○○○).
- Det saknas studier som undersöker sensitivitet och specificitet för den semistrukturerade intervjun DAWBA för att identifiera förstämningssyndrom hos barn och ungdomar jämfört med semistrukturerad intervju.
- Det saknas studier som undersöker sensitivitet och specificitet för den semistrukturerade intervjun KID-SCID för att identifiera förstämningssyndrom hos barn och ungdomar jämfört med semistrukturerad intervju. Det finns för få studier för att kunna bedöma sensitivitet och specificitet för den strukturerade intervjun MINI-KID för att identifiera förstämningssyndrom hos barn och ungdomar jämfört med semistrukturerade intervjuer (otillräckligt vetenskapligt underlag ⊕○○○).
- Det saknas studier som undersöker sensitivitet och specificitet för telefonintervjun BCFPI för att identifiera förstämningssyndrom hos barn eller ungdomar jämfört med semistrukturerad intervju för att identifiera förstämningssyndrom.

#### Formulär för riktad screening

- Sensitiviteten för CBCL:s DSM-orienterade affektiva skala med tröskelvärde 65 med semistrukturerad intervju som referensstandard, är 88 procent (95 % KI, 78 till 94) för att identifiera depression hos barn och ungdomar (begränsat vetenskapligt underlag ⊕⊕○○). Specificiteten är 76 procent (95 % KI, 67 till 83) (begränsat vetenskapligt underlag ⊕⊕○○).

- Det går inte att bedöma sensitivitet och specificitet för YRS som formulär för riktad screening av depression hos barn och ungdomar med en strukturerad eller semistrukturerad intervju som referensstandard eftersom studierna har bristande kvalitet (otillräckligt vetenskapligt underlag ⊕○○○).
- Det går inte att bedöma sensitivitet och specificitet för TRF som formulär för riktad screening av depression hos barn och ungdomar med en strukturerad eller semistrukturerad intervju som referensstandard eftersom det saknas studier som redovisar dessa utfallsmått.
- Det går inte att bedöma sensitivitet och specificitet för BUS som formulär för riktad screening av depression hos barn och ungdomar med en strukturerad eller semistrukturerad intervju som referensstandard eftersom det saknas studier.
- För barn och ungdomar i slutet respektive öppen psykiatrisk vård har BDI-I som formulär för riktad screening av depression med tröskelvärdet 11 poäng en sensitivitet på 81 procent (95 % KI, 75 till 87) och en specificitet på 59 procent (95 % KI, 53 till 65) med semistrukturerade intervjuer som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○). Vid tröskelvärdet 16 är sensitiviteten 59 procent (95 % KI, 51 till 68) (begränsat vetenskapligt underlag ⊕⊕○○) och specificiteten 80 procent (95 % KI, 75 till 85) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Det går inte att bedöma sensitivitet och specificitet för DSRS som formulär för riktad screening av barn och ungdomar med strukturerad eller semistrukturerad intervju som referensstandard eftersom det saknas studier.
- CDI som formulär för riktad screening för depression hos barn och ungdomar har en sensitivitet på 69 procent (95 % KI, 52 till 84) med K-SADS som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 70 procent (95 % KI, 63 till 76) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

- Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för CES-D som formulär för riktad screening för depression hos barn och ungdomar med strukturerade eller semistrukturerade intervjuer som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för SDQ:s emotionella delskala som formulär för riktad screening av depression hos barn och ungdomar med en strukturerad eller semistrukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för CBCL vid riktad screening av bipolära syndrom hos barn och ungdomar med strukturerade eller semistrukturerade intervjuer som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att kunna bedöma sensitivitet och specificitet för GBI som formulär för riktad screening av bipolära syndrom hos barn och ungdomar med strukturerade eller semistrukturerade intervjuer som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- GBI:s föräldraskala har en tillräcklig förmåga (ROC-AUC >0,80) att identifiera mani hos barn och ungdomar med K-SADS som referensstandard (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- Det går inte att bedöma ROC-AUC för GBI:s självskattningsskala för att identifiera mani hos barn och ungdomar med strukturerad eller semistrukturerad intervju som referensstandard pga bristande samstämmighet mellan studierna (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för CBQ som formulär för riktad screening för bipolärt syndrom hos barn och ungdomar med strukturerad eller

semistrukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

- Det går inte att bedöma sensitivitet och specificitet för MDQ som formulär för riktad screening av bipolärt syndrom hos barn och ungdomar med semistrukturerad intervju som referensstandard. Det finns inte tillräckligt många studier med adekvat kvalitet (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för fullängdsversionen respektive kortversionen av CMRS-P som formulär för riktad screening för bipolärt syndrom hos barn och ungdomar med strukturerad eller semistrukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det går inte att bedöma hur tillförlitlig självskattning med YMRS är som formulär för riktad screening av bipolära syndrom hos ungdomar med semistrukturerade eller strukturerade intervjuer som referensstandard eftersom det finns för få studier med tillräcklig kvalitet (otillräckligt vetenskapligt underlag ⊕○○○).
- Expertskattning med YMRS har en hög tillförlitlighet (ROC-AUC  $\geq 0,97$ ) som formulär för riktad screening av bipolära syndrom hos barn och ungdomar med K-SADS som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Föräldraskattning med YMRS har en god tillförlitlighet (ROC-AUC  $\geq 0,76$ ) som formulär för riktad screening av bipolära syndrom hos barn och ungdomar med K-SADS som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○).

### **Formulär för att bedöma svårighetsgrad**

- Det saknas studier som utvärderar formulär för att mäta svårighetsgrad av förstämningssyndrom jämfört med K-SADS eller CGI-S.

## Introduktion

Syftet var att undersöka den diagnostiska tillförlitligheten för intervjuformulär, för formulär för riktad screening och för formulär för att bedöma svårighetsgrad av förstämningssyndrom.

Den optimala diagnostiska proceduren är LEAD (Longitudinal, Experts, All Data procedure) (Kapitel 3.1), som därmed kan ses som referensstandard. Alternativa referensstandarder är semistrukturerade eller strukturerade intervjuformulär.

Avsnittet inleds med en granskning av studier som jämför den vanligast använda referensstandard, K-SADS, med LEAD. Därefter jämförs övriga semistrukturerade och strukturerade intervjuer som används i Sverige (enkätsvar, se Kapitel 3.1) med LEAD eller K-SADS.

Därefter följer en utvärdering av formulär för riktad screening av depression respektive bipolära syndrom. Utfallsmåtten är i första hand sensitivitet och specificitet. Vissa välgjorda studier har emellertid valt att enbart redovisa resultaten i form av ROC-AUC (Kapitel 1). Vi har kontaktat författarna, men de har inte kunnat återskapa de data som vi behövt för att kunna beräkna sensitivitet och specificitet. Vi har därför valt att även utvärdera ROC-AUC-data för dessa studier.

De formulär som ingick i granskningen sammanfattas i Tabell 3.4.1.

**Tabell 3.4.1** Sammanfattning av resultat av granskningen.

<b>Formulär</b>	<b>Syfte</b>	<b>Referensstandard</b>	<b>Evidensgraderat resultat</b>
K-SADS	Semistrukturerad intervju	LEAD/Best estimate	Otillräckligt
DICA	Semistrukturerad intervju	LEAD/Best estimate	Otillräckligt
KID-SCID	Semistrukturerad intervju	Semistrukturerad intervju	Saknas
MINI-KID	Strukturerad intervju	Semistrukturerad intervju	Otillräckligt
DAWBA	Strukturerad intervju	Semistrukturerad intervju	Saknas
BCPFI	Strukturerad intervju	Semistrukturerad intervju	Saknas
CBCL, DSM-IV orienterade affektiva skalan, tröskelvärde 65 (subklinisk)	Riktad screening av depression	Semistrukturerad intervju	Sensitivitet: 88% Specificitet: 76%
YSR	Riktad screening av depression	Semistrukturerad intervju	Otillräckligt
TRF	Riktad screening av depression	Semistrukturerad intervju	Otillräckligt
BUS	Riktad screening av depression	Semistrukturerad intervju	Saknas
BDI-I, tröskelvärde 11 respektive 16	Riktad screening av depression	Semistrukturerad intervju	Tröskelvärde 11 Sensitivitet: 81% Specificitet: 59%  Tröskelvärde 16 Sensitivitet: 59% Specificitet: 80%
DSRS	Riktad screening av depression	Semistrukturerad intervju	Saknas
CDI, tröskelvärden 12, 12,5, 13	Riktad screening av depression	Semistrukturerad intervju	Sensitivitet: 69% Specificitet: 70%

Tabellen fortsätter på nästa sida

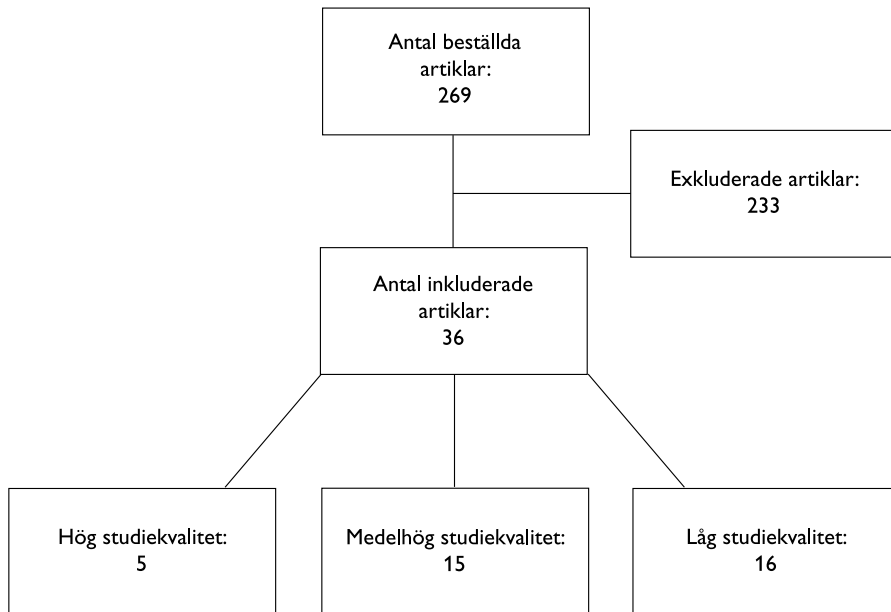


**Tabell 3.4.1** fortsättning

<b>Formulär</b>	<b>Syfte</b>	<b>Referensstandard</b>	<b>Evidensgraderat resultat</b>
CES-DC	Riktad screening av depression	Semistrukturerad intervju	Otillräckligt
PHQ-9	Riktad screening av depression	Semistrukturerad intervju	Otillräckligt
SDQ: emotionella skala	Riktad screening av förstämningssyndrom	Semistrukturerad intervju	Saknas
GBI	Riktad screening av bipolära syndrom	Semistrukturerad intervju	Sensitivitet och specificitet: Saknas ROC-AUC föräldrar: >0,80 ROC-AUC barn-ungdom: Otillräckligt
CBQ	Riktad screening av bipolära syndrom	Semistrukturerad intervju	Otillräckligt
MDQ-A	Riktad screening av bipolära syndrom	Semistrukturerad intervju	Otillräckligt
CMRS-P	Riktad screening av bipolära syndrom	Semistrukturerad intervju	Otillräckligt
YMRS	Riktad screening av bipolära syndrom	Semistrukturerad intervju	Sensitivitet och specificitet: Saknas ROC-AUC expert: $\geq 0,97$ ROC-AUC föräldrar: $\geq 0,76$ ROC-AUC barn-ungdom: Otillräckligt
BDI, CMRS m fl	Bedömning av svårighetsgrad vid ett eller flera tillfällen	K-SADS eller CGI-S	Saknas

ROC-AUC = Receiving characteristics - area under curve

## Urval av studier

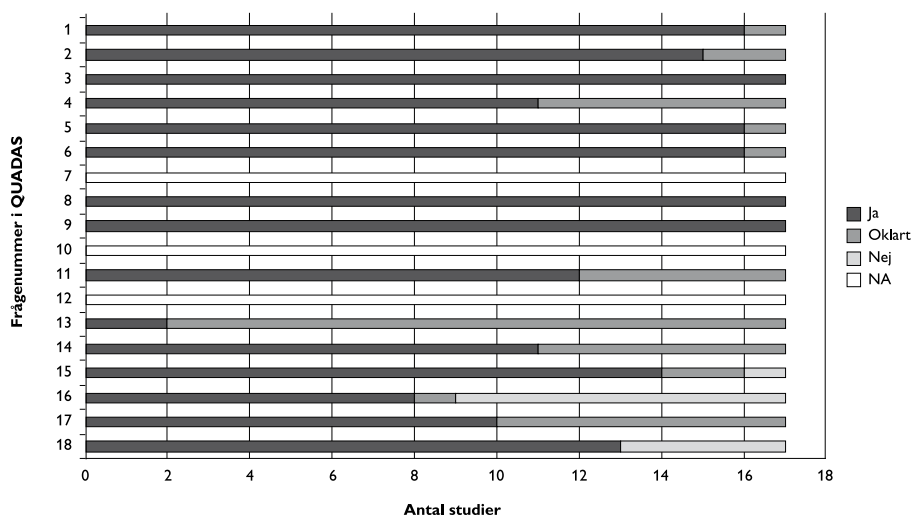


**Figur 3.4.1** Flödesschema av beställda och inkluderade studier om bedömningsformulär för barn och ungdomar.

## Studiekvalitet

Figur 3.4.2 visar hur stor andel av de olika frågorna (items) i QUADAS (Kapitel 3.1 för en förteckning över frågorna) som bedömdes uppfylla kraven på kvalitet (svart del av staplarna) för de studier som ingick i granskningen. Grå markering motsvarar att det inte framgick av artikeln om kraven var uppfyllda och ljusgrå att det framgick klart att kraven inte var uppfyllda. Vit färg användes om frågorna inte var relevanta (not applicable, NA).

Som helhet uppfyllde det vetenskapliga underlaget kriterier för studiekvalitet. Den största svagheten rörde fråga 16 om interbedömarreliabilitet, där nära hälften av studierna inte redovisade några uppgifter.



**Figur 3.4.2** Sammanställning av bedömning av studiekvalitet med QUADAS för det vetenskapliga underlaget.

## Semistrukturerade diagnostiska intervjuer

### Schedule for Affective Disorders and Schizophrenia for School Aged Children (Kiddie-SADS/K-SADS) jämfört med LEAD

De semistrukturerade formuläret K-SADS används ofta som referensstandard vid utvärdering av diagnostiska formulär inom barnpsykiatri [1]. Intervjun är en barnversion av vuxenintervjun SADS, framtagen av Chambers och Puig-Antich i USA. Den kan användas för barn mellan 6 och 18 år. Intervjun omfattar de flesta barnpsykiatriska symtom och tar minst två timmar att genomföra. Vår första fråga var om K-SADS har tillräckligt hög sensitivitet och specificitet jämfört med LEAD för att kunna användas som referensstandard.

#### *Beskrivning av underlaget*

Sex studier lästes i fulltext varav en inkluderades och bedömdes ha medelhög studiekvalitet [2]. I den inkluderade studien av Carlson och medarbetare som genomfördes i USA genomgick 30 barn och deras föräldrar K-SADS-intervjun i samband med att de lades in på en barn-

psykiatrisk avdelning. Vid utskrivningen ställdes så kallade LEAD-diagnoser baserade på all tillgänglig information förutom den från K-SADS-intervjuerna. K-SADS hade då 81 procents sensitivitet och 79 procents specificitet för att identifiera något förstämningssyndrom.

### ***Evidensgraderade resultat***

Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för den semistrukturerade intervjun K-SADS jämfört med LEAD för att korrekt klassificera något förstämningssyndrom hos barn eller ungdomar (otillräckligt vetenskapligt underlag ⊕○○○).

### **Diagnostic Interview for Children and Adolescents (DICA)**

DICA är en semistrukturerad diagnostisk intervju, konstruerad av Herjanic och Reich 1982 [3]. Den senaste versionen är från 1997 [4] och den finns i tre versioner: för barn, tonåringar och föräldrar. Intervjun omfattar de flesta barnpsykiatriska symtom och tar cirka två timmar att genomföra.

### ***Beskrivning av underlaget***

Fyra artiklar lästes i fullformat [2,5–7]. Tre av dem exkluderades [5–7]. Boyle och medarbetare [5] undersökte flera egenskaper hos DICA men inte sensitivitet och specificitet jämfört med en referensstandard. De båda andra studierna som exkluderades använde symtomskala respektive intervjuer med barn och föräldrar som referensstandard [6,7].

En studie, den ovan nämnda studien av medelhög kvalitet av Carlson och medarbetare, inkluderades [2]. De ställde LEAD-diagnoser baserat på all information utom den från de semistrukturerade intervjuerna. DICA hade då 81 procents sensitivitet och 68 procent specificitet för att identifiera något förstämningssyndrom.

### ***Evidensgraderat resultat***

Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för den semistrukturerade intervjun DICA jämfört med LEAD för att korrekt klassificera förstämningssyndrom hos barn eller ungdomar (otillräckligt vetenskapligt underlag ⊕○○○).

## **Structured Clinical Interview for DSM-IV, Childhood Version (KID-SCID)**

Amerikanska psykiatriska föreningens officiella diagnostiska intervju för vuxna, SCID, finns även i en barnversion, KID-SCID, som utvecklades 1993 av Fred Matzner och medarbetare [8].

### ***Beskrivning av underlaget***

Vi kunde inte identifiera någon studie som jämförde KID-SCID med en semistrukturerad intervju för att diagnostisera förstämningssyndrom.

### ***Evidensgraderat resultat***

Det saknas studier som undersöker sensitivitet och specificitet för den semistrukturerade intervjun KID-SCID jämfört med semistrukturerad intervju för att ställa diagnoser på förstämningssyndrom hos barn eller ungdomar.

## **Strukturerade diagnostiska intervjuer**

### **Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID)**

MINI-KID är en strukturerad diagnostisk intervju för barn mellan 6 och 17 år, konstruerad 1989 av Sheehan och medarbetare i USA och Europa [9]. Den täcker 24 DSM-IV- och ICD-10-diagnoser och har utvecklats i samma format som vuxenvarianten MINI.

### ***Beskrivning av underlaget***

En studie lästes i fulltext [10]. Den uppfyllde inklusionskriterierna och bedömdes ha hög studiekvalitet. Studien genomfördes inom olika psykiatriska verksamheter i Florida. Totalt rekryterades 195 deltagare. K-SADS användes som referensstandard. Sensitiviteten var 54 procent och specificiteten 87 procent för depression. För manisk/hypoman episod var sensitiviteten 96 procent och specificiteten 73 procent.

### ***Evidensgraderat resultat***

Det finns för få studier för att kunna bedöma sensitivitet och specificitet för den strukturerade intervjun MINI-KID jämfört med semistrukturerade intervjuer för att diagnostisera depression eller pågående mani/

hypomani hos barn eller ungdomar (otillräckligt vetenskapligt underlag ⊕○○○).

### **Development and Well-Being Assessment (DAWBA)**

DAWBA är en utvärderingsmetod som omfattar såväl frågeformulär (SDQ) som intervjuer och skattningsmetodik [11], vilka tillsammans används för att ställa diagnoser enligt ICD-10 och DSM-IV. DAWBA är utvecklat av Goodman och medarbetare i England. DAWBA är avsett för barn mellan 5 och 16 år [11].

#### ***Beskrivning av underlaget***

Vi kunde inte identifiera någon studie där DAWBA jämfördes med semistrukturerad intervju för att ställa diagnoser på förstämningssyndrom hos barn och ungdomar.

#### ***Evidensgraderat resultat***

Det saknas studier som undersöker sensitivitet och specificitet för den strukturerade intervjun DAWBA jämfört med semistrukturerad intervju för att diagnostisera förstämningssyndrom hos barn eller ungdomar.

### **Brief Child and Family Phone Interview (BCPFI)**

BCPFI är en datorbaserad telefonintervju för föräldrar om psykiska symtom hos deras barn i åldrarna 3 till 18 år. Den utvecklades i Kanada i början av 1990-talet, men den första publikationen kom först 2009 [12]. Syftet med intervjun är screening, inte diagnostik, och den avser att inhämta information från föräldrar som söker psykiatrisk hjälp för sina barn.

#### ***Beskrivning av underlaget***

Två studier lästes i fulltext och båda exkluderades. Den första undersökte inte den diagnostiska tillförlitligheten utan beskrev hur BCFPI konstruerats [12]. Den andra hade för lång tid mellan indextest och referenstest [13].

### *Evidensgraderat resultat*

Det saknas studier som undersöker sensitivitet och specificitet för telefonintervjun BCFPI jämfört med semistrukturerad intervju för att ställa diagnoser på förstämningssyndrom hos barn eller ungdomar.

## **Formulär för riktad screening av depression hos barn och ungdomar**

Syftet var att bedöma den diagnostiska tillförlitligheten, mätt som sensitivitet och specificitet, för formulär som är avsedda att användas för riktad screening för depression hos barn och ungdomar med strukturerade eller semistrukturerade intervjuer som referensstandard.

### **Achenbach System of Empirically Based Assessment (ASEBA), Child Behaviour Checklist (CBCL), Youth Self Report (YSR), Teacher Report Form (TRF)**

Achenbach System of Empirically Based Assessment (ASEBA) ([www.aseba.org/](http://www.aseba.org/)) utvecklades av Achenbach i USA på 1960-talet för att göra det möjligt att standardisera beskrivningar av barns psykiska problem [14]. CBCL, YSR och TRF är tre frågeformulär som ingår i ASEBA. Formulären används för att ge en integrerad bild av ett barns eller en ungdoms förmågor, anpassning och psykiska problem.

CBCL fylls i av föräldrar för att beskriva problem hos barn i åldrarna 6 till 18 år. Det finns även en version för barn mellan 1,5 och 5 år. YSR besvaras av ungdomar mellan 11 och 18 år och TRF av lärare till barn och ungdomar i åldrarna mellan 6 och 18 år (se nedan).

ASEBA innehåller nio syndromskalor som kan kategoriseras ytterligare i ”internaliserande” (inåtvända) eller ”externaliserande” (utåtriktade) problem samt en total problemskala. Skalorna kan relateras till diagnoser i DSM-systemet och Achenbach har utvecklat sex DSM-orienterade skalor, varav en rör affektiva problem. Skalorna syftar inte till att vara diagnostiska men kan användas för screening, t ex av samsjuklighet.

För syndromskalor och DSM-orienterade skalor gäller ett tröskelvärde på 65 för subklinisk nivå och på 70 för klinisk nivå. För de sammanslagna skalorna (internaliserande, externaliserande och total problemskala) gäller ett subkliniskt tröskelvärde på 60 och ett kliniskt på 64.

Det finns flera diagnostiska skalor, baserade på CBCL och YSR, som konstruerats av andra än Achenbach. De är vetenskapligt prövade, t ex för att screena för bipolärt syndrom, men Achenbach stödjer inte användningen av dem (personlig kommunikation år 2011). Därför ingår de inte i denna översikt.

ASEBA-skalorna är normerade för olika åldrar, kön, och i olika länder. CBCL från 6 år och YSR är normerade i Sverige [16,17]. De CBCL-skalor som finns för förskolebarn (1,5–5 år) saknar däremot svensk normering. För dem finns inte heller DSM-orienterade skalor. CBCL för åldrarna 1,5–5 år ingår därför inte i vår granskning.

## **Child Behaviour Checklist (CBCL) 6–18 år**

### *Beskrivning av underlaget*

Av 37 studier som lästes i fulltext undersökte 18 diagnostisk tillförlitlighet för depression, 12 för bipolärt syndrom, tre för någon diagnos över huvud taget och två ställde inga diagnoser. Av de 18 depressionsstudierna exkluderades 13 [18–30]. Av de fem depressionsstudierna som inkluderades [31–35] bedömdes tre ha låg studiekvalitet [31,32,35] pga oklar tid mellan referens och indextest och att ingen interbedömarreliabilitet presenterades. Dessutom var bortfallet oklart i två av dem [32,35].

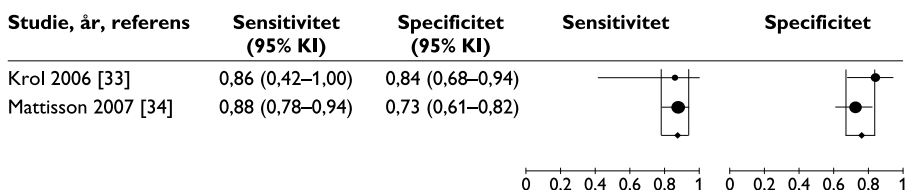
Två studier med medelhög studiekvalitet, av Krol och medarbetare respektive Mattison och medarbetare, rapporterade sensitivitet och specificitet [33,34] (se Tabell 3.4.8). De omfattade totalt 197 barn. Studien av Krol och medarbetare genomfördes i barnpsykiatrisk öppenvård i Nederländerna [33]. Föräldrarna till 44 konsekutivt remitterade barnpsykiatriska patienter tillfrågades och samtliga valde att delta. Gruppen bestod av 25 pojkar och 19 flickor med en genomsnittlig ålder på 10,6 år. Föräldrarna fyllde i CBCL. Den DSM-orienterade CBCL-skalen för affektiva problem (tröskelvärde 65) jämfördes med den strukturerade intervjun DISC.



Mattison och medarbetare rekryterade 153 barn med en genomsnittlig ålder på 11 år på en forskningsklinik i USA [34]. Ungdomarna valdes för att de hade antingen depression eller ADHD eller ingendera av diagnoserna. Det innebar att blindningen för diagnostillhörighet inte var optimal. CBCL fylldes i av föräldrarna. Den DSM-orienterade skalan för affektiva problem (tröskelvärde 65) jämfördes med K-SADS.

### *Sammanvägda resultat*

De två studierna hade en sensitivitet på 86 respektive 88 procent och en specificitet på 84 respektive 73 procent. Den sammanvägda sensitiviteten blev 88 procent (95 % KI, 78 till 94) och specificiteten 76 procent (95 % KI, 67 till 83) (Figur 3.4.3).



Sammanvägd **sensitivitet** = 0,88 (0,78 till 0,94)

Sammanvägd **specificitet** = 0,76 (0,67 till 0,83)

**Figur 3.4.3** Forest plots. Spridning av sensitivitet och specificitet för CBCL:s DSM-orienterade skala för affektiva problem.

### *Bedömning av evidensstyrka*

Vi bedömde evidensstyrkan med hjälp av GRADE-kriterierna. Vi bedömde risken för bias som följd av bristande studiekvalitet som liten. Studien av Krol och medarbetare var väl genomförd men vissa uppgifter saknades [33]. I studien av Mattison och medarbetare var det oklart hur väl blindningen upprätthållits [34].

Däremot fanns det brister i överförbarheten. Gruppen med depression i Mattison och medarbetares studie var inte homogen eftersom en tredjedel hade diagnosen dystymi. Studierna använde också olika referensintervjuer (K-SADS respektive DISC). Detta medförde att evidensstyrkan

minskade ett steg. Även precisionen var bristfällig. Det innebar att evidensstyrkan sänktes ytterligare ett steg för både sensitivitet och specificitet (Tabell 3.4.2).

**Tabell 3.4.2** Sammanvägda resultat avseende CBCL:s DSM-orienterade skala för affektiva problems förmåga att identifiera depression hos barn.

Utfallsmått	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	2 studier n=197	88% (78; 94)	Begränsat ⊕⊕○○	Bristande överförbarhet (-1)  Bristande precision (-1)
Specificitet	2 studier n=197	76% (67; 83)	Begränsat ⊕⊕○○	Bristande överförbarhet (-1)  Bristande precision (-1)

KI = Konfidensintervall

### *Evidensgraderat resultat*

Sensitiviteten för CBCL:s DSM-orienterade affektiva skala för riktad screening av depression hos barn och ungdomar vid tröskelvärde 65 är 88 procent (95 % KI, 78 till 94) med semistrukturerade intervjuer som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○). Specificiteten är 76 procent (95 % KI, 67 till 83) (begränsat vetenskapligt underlag ⊕⊕○○).

### **Youth Self Report (YSR)**

#### *Beskrivning av underlaget*

Av 13 studier som lästes i fulltext [26–28,35–44] undersökte nio diagnostisk tillförlitlighet för depression [26–28,35,38,39,42–44], tre för bipolärt syndrom och tre ställde inga diagnoser (alla sex exkluderade). Av depressionsstudierna inkluderades fyra [35,39,42,43]. Samtliga

bedömdes ha låg studiekvalitet pga oklar tid mellan referens- och indextest [35,42,43], att de inte presenterade interbedömarreliabilitet [35,42,43] och att referensstandarderna var bristfälliga [39].

### ***Evidensgraderade resultat***

Det går inte att bedöma sensitivitet och specificitet för YSR för riktad screening av depression hos barn och ungdomar med en strukturerad eller semistrukturerad intervju som referensstandard eftersom det inte finns tillräckligt med studier av god kvalitet (otillräckligt vetenskapligt underlag ⊕○○○).

## **Teacher Report Form (TRF)**

### ***Beskrivning av underlaget***

Fyra studier lästes i fulltext och samtliga exkluderades [34,41,45,46]. Orsakerna var att det var för lång tid mellan referenstest och indextest [34], att referensstandarderna inte var en strukturerad eller semistrukturerad intervju [41], att ett modifierat formulär användes [45] eller att studien inte redovisade relevanta utfallsmått [46].

### ***Evidensgraderat resultat***

Det går inte att bedöma sensitivitet och specificitet för TRF för riktad screening av depression hos barn och ungdomar med en semistrukturerad eller strukturerad intervju som referensstandard eftersom det saknas studier.

## **Becks ungdomsskalor (BUS)**

Becks ungdomsskalor (Beck Youth Inventories, BYI) består av fem självbedömningsformulär rörande emotionella och sociala problem för barn och ungdomar mellan 7 och 18 år. Formulären utvecklades i USA 2001 [47]. Ett av formulären avser depression och har totalpoängen 60. Varje fråga har fyra svarsalternativ, från aldrig till alltid, motsvarande 0–3 poäng. Tröskelvärde för depressionsskalan för flickor anges till 23 poäng och för pojkar till 17 poäng.

### ***Beskrivning av underlaget***

En studie lästes i fulltext men den exkluderades eftersom uppgifter om sensitivitet och specificitet mot en diagnostisk intervju saknades [48].

### *Evidensgraderat resultat*

Det går inte att bedöma sensitivitet och specificitet för Becks ungdomsskalor för riktad screening av depression med en semistrukturerad eller strukturerad intervju som referensstandard eftersom det saknas studier.

### **Beck Depression Inventory for Children (BDI-I)**

BDI utvecklades av Beck och medarbetare som ett formulär för att mäta svårighetsgrad, och originalartikeln publicerades år 1961 [49]. Formuläret har reviderats, senast i mitten på 1990-talet, och anpassades då till DSM-IV-kriterierna. Nuvarande version, BDI-II, kan användas för såväl riktad screening av egentlig depression som för att mäta svårighetsgrad av depressiva symtom [50]. BDI-II är avsedd för att användas från 13 år och uppåt. För barn och ungdomar är de två etablerade tröskelvärdena 11 och 16 för BDI-I. Motsvarande tröskelvärden finns inte presenterade för BDI-II.

BDI-C är en modifiering av BDI-I och används för tonåringar (13–18 år). Maximalt kan man få 63 poäng. Det finns även en särskild barnvariant för barn från sju års ålder som del i Becks ungdomsskalor, se ovan.

### *Beskrivning av underlaget*

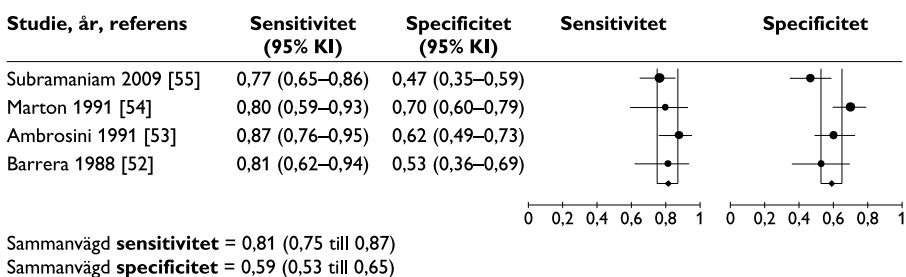
Vi sökte artiklar som utvärderade någon av versionerna av BDI för barn och ungdomar. Av 15 artiklar som lästes i fulltext uppfyllde fem inklusionskriterierna. Samtliga utvärderade BDI-I och hade medelhög kvalitet [51–55] (Tabell 3.4.9). Av dessa prövade Subramaniam och medarbetare enbart tröskelvärde 11 [55] och Kashani och medarbetare endast tröskelvärde 16 [51]. Totalt innefattade studierna 551 barn och ungdomar som vårdades på barnpsykiatrisk öppen- eller slutenvård. Prevalensen låg mellan 21 och 50 procent i de olika studierna.

Barrera och medarbetare inkluderade 65 barn och ungdomar som vårdades på barnpsykiatrisk avdelning [52]. Studien av Ambrosini och medarbetare omfattade 122 tonåringar som var patienter i öppen psykiatrisk vård [53]. Marton och medarbetare undersökte 119 tonåringar som vårdades inom öppenvården (n=93) eller på en barnpsykiatrisk avdelning [54]. Subramaniam och medarbetare inkluderade 145 tonåringar med

alkohol- eller cannabisberoende som deltog i behandlingsprogram för beroende [55]. I studien av Kashani och medarbetare inkluderades 100 öppenvårdspatienter [51].

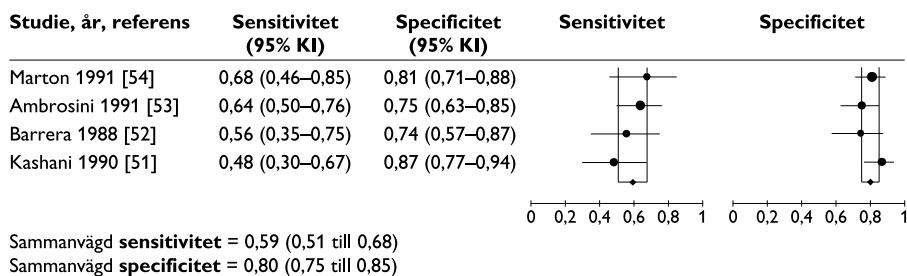
### *Sammanvägning av resultaten*

Sensitiviteten i de fyra studier som undersökte tröskelvärde 11 låg mellan 77 och 87 procent. Specificiteten låg mellan 47 och 70 procent. Det sammanvägda estimatet för sensitivitet blev 81 procent (95 % KI, 75 till 87). Det sammanvägda estimatet för specificitet blev 59 procent (95 % KI, 53 till 65) (Figur 3.4.4).



**Figur 3.4.4** Forest plot för sensitivitet och specificitet för BDI-I med tröskelvärde 11.

Fyra studier undersökte även tröskelvärde 16 [51–54]. Här blev spridningen i sensitivitet större mellan studierna. Den låg mellan 48 och 68 procent. Specificiteten låg mellan 74 och 87 procent. Det sammanvägda estimatet för sensitivitet blev 59 procent (95 % KI, 51 till 68). För specificiteten var det sammanvägda estimatet 80 procent (95 % KI, 75 till 85) (Figur 3.4.5).



**Figur 3.4.5** Forest plot för sensitivitet och specificitet för BDI-I med tröskelvärde 16.

### *Bedömning av evidensstyrka*

Vi bedömde tillförlitligheten med hjälp av GRADE-kriterierna. Det fanns vissa brister i studiekvalitet, t ex att endast en studie uppgav hur stor andel av formulären som fyllts i fullständigt. En annan brist var att studierna använde olika referensstandarder, vilket påverkade överförbarheten. Svagheter i studiekvalitet och överförbarhet innebar att evidensstyrkan sänktes ett steg.

Precisionen var bristfällig för sensitivitet och specificitet vid tröskelvärdet 11 och för sensitivitet vid tröskelvärdet 16, vilket också påverkade evidensstyrkan. Se Tabell 3.4.3.

**Tabell 3.4.3** Resultatsammanställning för BDI-I som formulär för riktad screening av depression hos barn och ungdomar vid tröskelvärde 11 och 16.

Utfallsmått	Antal studier Antal patienter	Sammanvägt estimat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
<b>Tröskelvärde 11</b>				
Sensitivitet	4 studier n=451	81% (75; 87)	Begränsat ⊕⊕○○	Bristande studiekvalitet och överförbarhet (-1) Bristande precision (-1)
Specificitet	4 studier n=451	59% (53; 65)	Begränsat ⊕⊕○○	Bristande studiekvalitet och överförbarhet (-1) Bristande precision (-1)
<b>Tröskelvärde 16</b>				
Sensitivitet	4 studier n=406	59% (51; 68)	Begränsat ⊕⊕○○	Bristande studiekvalitet och överförbarhet (-1) Bristande precision (-1)
Specificitet	4 studier n=406	80% (75; 85)	Måttligt starkt ⊕⊕⊕○	Bristande studiekvalitet och överförbarhet (-1)

KI = Konfidensintervall

### ***Evidensgraderade resultat***

- Sensitiviteten för BDI-I vid riktad screening av depression hos barn och ungdomar i sluten respektive öppen psykiatrisk vård är 81 procent (95 % KI, 75 till 87) vid tröskelvärde 11 med semistrukturerade intervjuer som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○).
- Specificiteten för BDI-I vid riktad screening av depression hos barn och ungdomar i sluten respektive öppen psykiatrisk vård är 59 procent (95 % KI, 53 till 65) vid tröskelvärde 11 och med semistrukturerade intervjuer som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○).
- Sensitiviteten för BDI-I vid riktad screening av depression hos barn och ungdomar i sluten respektive öppen psykiatrisk vård är 59 procent (95 % KI, 51 till 68) vid tröskelvärde 16 poäng och med semistrukturerade intervjuer som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○).
- Specificiteten för BDI-I vid riktad screening av depression hos barn och ungdomar i sluten respektive öppen psykiatrisk vård är 80 procent (95 % KI, 75 till 85) vid tröskelvärde 16 poäng och med semistrukturerade intervjuer som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

### **Birleson Depression Self Rating Scale (DSRS)**

DSRS är ett frågeformulär med 18 frågor för riktad screening av depression hos barn. Formuläret utvecklades år 1981 i Skottland [56]. Det översattes till svenska av Gillberg samma år och har senare validerats i Sverige [57]. Varje fråga skattas på en tregradig skala (0–2). I originalarbetet föreslås ett tröskelvärde på 15.

### ***Beskrivning av underlaget***

Tre artiklar lästes i fulltext men alla exkluderades. Den svenska valideringen av Ivarsson och medarbetare använde en klinisk intervju som referensstandard [57], och den andra svenska studien av samma författare var en normeringsstudie i ett icke kliniskt material [58]. Den tredje



artikeln av Fundudis och medarbetare angav diagnostisk tillförlitlighet för tröskelvärde 11 [59]. Dessutom uppgavs inte tiden mellan index- och referenstest.

### ***Evidensgraderat resultat***

Det går inte att bedöma sensitivitet och specificitet för Birlerson Depression Rating Scale (DSRS) med strukturerad eller semistrukturerad intervju som referensstandard eftersom det saknas studier.

### **Children's Depression Inventory (CDI)**

CDI är ett självbedömningsformulär för barn och ungdomar mellan 7 och 17 år som utvecklades av Kovacs i början av 1990-talet [60]. CDI har BDI som förebild. Formuläret innehåller 27 påståenden med svar som poängsätts 0–2 vilket ger maximalt 54 poäng. Etablerade tröskelvärden är 12–13.

### ***Beskrivning av underlaget***

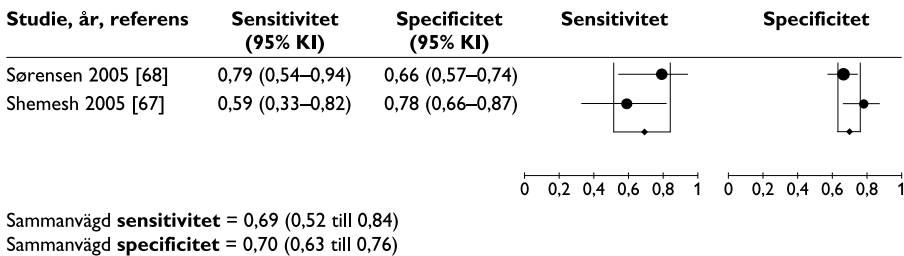
Av 29 studier som lästes i fulltext exkluderades 20 studier pga irrelevant population, att de saknade godkänd referensstandard, att de använde en modifierad metod, att tiden var för lång mellan index- och referenstest samt irrelevanta utfallsmått. Av de nio studier som inkluderades bedömdes sju ha låg kvalitet [18,61–66] och två medelhög kvalitet [67,68]. Studierna som bedömdes ha låg kvalitet hade för lång tid mellan referenstest och index- och test [61,63], fel utfallsmått [18,62], eller en tveksam referensstandard (ADIS-intervjun [59], checklista [64]). Båda använde K-SADS som referensstandard (Tabell 3.4.10).

Shemesh och medarbetare inkluderade 81 barn och ungdomar som sökte primärvård pga somatisk sjukdom och som samtidigt verkade ha någon form av emotionell störning [67]. Av dessa hade 16 procent någon form av depression, i huvudsak egentlig depression. Både föräldrar och barn intervjuades med K-SADS, och den sammantagna bedömningen utgjorde referensstandard.

Studien av Sørensen och medarbetare omfattade 149 barn och ungdomar med egentlig depression i psykiatrisk öppen- och slutenvård [68]. Barnen fyllde i CDI. Både föräldrar och barn intervjuades med K-SADS och resultaten redovisades separat.

### Sammanvägt resultat

Sensitiviteten i de två studierna, baserat på barnens bedömningar, var 79 respektive 59 procent och specificiteten 66 procent respektive 78 procent (Figur 3.4.6). Det sammanvägda estimatet för sensitivitet blev 69 procent (95 % KI, 52 till 84). En sammanvägning för specificitet gav ett estimat på 70 procent (95 % KI, 63 till 76).



**Figur 3.4.6** Forest plot för sensitivitet och specificitet för CDI som formulär för riktad screening av depression.

### Bedömning av evidensstyrka

Vi bedömde tillförlitligheten av de sammanvägda estimaten med hjälp av GRADE-kriterierna. Risken för bias som följd av metodproblem föreföll vara liten. För sensitiviteten var precisionen mycket bristfällig och för specificiteten bristfällig, vilket sänkte evidensstyrkan med två respektive ett steg.

**Tabell 3.4.4** Resultatsammanställning för CDI som formulär för riktad screening.

Utfallsmått	Antal studier Antal patienter	Sammanvägt estimat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	2 studier n=230	69% (52; 84)	Begränsat ⊕⊕○○	Bristande precision (–2)
Specificitet	2 studier n=230	70% (63; 76)	Måttligt starkt ⊕⊕⊕○	Bristande precision (–1)

### ***Evidensgraderade resultat***

- CDI som formulär för riktad screening av depression hos barn och ungdomar har en sensitivitet på 69 procent (95 % KI, 52 till 84) med K-SADS som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○).
- CDI som formulär för riktad screening av depression hos barn och ungdomar har en specificitet på 70 procent (95 % KI, 63 till 76) med K-SADS som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

### **Center of Epidemiological Studies Depression Scale for Children (CES-DC)**

CES-DC är en för barn modifierad version [69] av vuxenvarianten CES-D (Kapitel 3.5). Den är avsedd för barn mellan 12 och 18 år och består av 20 frågor som skattas mellan 0 och 3. Tröskelvärde för egentlig depression är 16. CES-DC finns översatt till svenska och har använts i svensk forskning [70].

### ***Beskrivning av underlaget***

Av 14 artiklar som lästes i fulltext uppfyllde en inklusionskriterierna och bedömdes ha medelhög kvalitet [71] (se Tabell 3.4.11). Studien av Fendrich och medarbetare undersökte 166 barn till deprimerade föräldrar och en kontrollgrupp [71]. Prevalensen av depression var 17 procent. Studien gjordes på en forskningsklinik i USA. Referensstandard var LEAD inkluderande K-SADS-intervju. För ett tröskelvärde på 16 var sensitiviteten 71 procent och specificiteten 62 procent.

### ***Evidensgraderat resultat***

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

## **Patient Health Questionnaire (PHQ-9)**

PHQ-9 är ett frågeformulär med nio påståenden som är avsett för riktad screening av depression i primärvård [72]. Det har utvecklats för vuxna (Kapitel 3.5), men har utvärderats även för tonåringar. Frågeformuläret bygger på frågorna i den strukturerade diagnostiska intervjun PRIME-MD.

### ***Beskrivning av underlaget***

Av tre artiklar som lästes i fulltext exkluderades två, en pga att endast två av frågorna i PHQ-9 ingick [73] och den andra pga att referensstandard inte var relevant (kliniska diagnoser) [74]. En studie uppfyllde inklusionskriterierna [75]. Vi bedömde att den hade låg kvalitet eftersom den hade bristfällig blindning (Tabell 3.4.16a).

### ***Evidensgraderat resultat***

Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för PHQ-9 som formulär för riktad screening för depression hos barn och ungdomar med strukturerade eller semistrukturerade intervjuer som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## **Strength and Difficulties Questionnaire (SDQ)**

SDQ heter på svenska ”Styrkor och svårigheter” och avser barn i åldrarna 3 till 16 år. SDQ finns i en version för lärare och en för föräldrar. Dessutom finns en självbedömning för barn i åldrarna 11 till 17 år.

SDQ:s frågor grupperas i fem problemområden: emotionella problem, normbrytande beteende, hyperaktivitet och koncentrationssvårigheter, kamratproblem och social kompetens. Dessutom kan man få ett totalvärde och ett mått på symtomens funktionspåverkan. Den emotionella delskalan kan användas som screening för förstämningssyndrom, och poängen avgör om man går vidare med intervjuundersökningen DAWBA [76,77].

### ***Beskrivning av underlaget***

Vi läste fem artiklar i fulltext [78-82], men samtliga exkluderades pga felaktigt utfallsmått, felaktig referensstandard, DAWBA [81] alternativt

olika formulär [79], eller irrelevant population [82]. Den svenska föräldra-versionen av SDQ har prövats i två studier men de använde inte någon semistrukturerad diagnostisk intervju som referensstandard [76,77].

### *Evidensgraderat resultat*

Det går inte att bedöma sensitivitet och specificitet för SDQ:s emotionella delskala som formulär för riktad screening av depression hos barn och ungdomar med en strukturerad eller semistrukturerad intervju som referensstandard eftersom det saknas studier.

## **Formulär för riktad screening av bipolärt syndrom**

Syftet var att granska den diagnostiska tillförlitligheten, mätt som specificitet och sensitivitet för formulär som är avsedda att användas för riktad screening av bipolärt syndrom med strukturerade eller semistrukturerade intervjuer som referensstandard.

### **Child Behaviour Checklist (CBCL), 6–18 år**

Av de 37 CBCL-studier som inkluderades (se depressionsavsnittet) var det tolv studier som avsåg diagnostisk tillförlitlighet för bipolärt syndrom. Tio av dem exkluderades [46,83–91]. De båda som inkluderades hade hög studiekvalitet [36,92]. Ingen av dem presenterade dock sensitivitet och specificitet, utan bara AUC-ROC. Då studierna var välgjorda bedömde vi att det var intressant att analysera dem.

### *Beskrivning av det vetenskapliga underlaget*

Youngstrom och medarbetare undersökte den diagnostiska tillförlitligheten hos åtta olika formulär, däribland CBCL, med K-SADS som referensstandard [93]. Populationen bestod av 500 barn och ungdomar från barnpsykiatrisk öppenvård. Prevalensen bipolära syndrom var över 40 procent. ROC-AUC för CBCL:s delskala för utagerande problem var 0,80.

Samma forskargrupp undersökte om CBCL kan diskriminera mellan bipolära syndrom och andra psykiska tillstånd [36]. Det var 439 barn och ungdomar från en psykiatrisk forskningsklinik som deltog. Av dem

hade 38 procent bipolära syndrom. ROC-AUC för CBCL:s aggressiva delskala var 0,81.

### ***Sammanvägda resultat***

Studierna använde olika delskalor av CBCL, vilket medför att resultaten inte kan vägas samman.

### ***Evidensgraderat resultat***

Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för CBCL vid riktad screening av bipolära syndrom hos barn och ungdomar med strukturerade eller semistrukturerade intervjuer som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **General Behavior Inventory (GBI)**

GBI är ett formulär med 73 frågor avseende symtom på depression, hypomani, mani och blandepisoder. Varje fråga skattas på en Likertskala mellan ett och fyra. De vanligaste angivna tröskelvärdena för hypomani/mani är 16/17 och för depression 23/24 [94]. Dessutom finns två kortversioner för manisymtom [95].

### ***Beskrivning av underlaget***

Av sex artiklar som lästes i fulltext uppfyllde tre inklusionskriterierna [93,95,96]. En av dem bedömdes ha hög studiekvalitet [93] och två medelhög [95,96]. Av de exkluderade artiklarna hade en fel utfallsmått [97], en undersökte utfall i sjukdom fem år senare [98] och en hade modifierat metoden [45].

Youngstrom och medarbetare undersökte fullängdsversionen av manisalan (28 frågor) på 642 ungdomar som deltog i läkemedelsstudier på en forskningsklinik i USA [93]. Studien redovisade ROC-AUC för föräldraskattning på 0,81 för barn upp till 10 år och 0,84 för barn och ungdomar mellan 11 och 17 år. När ungdomarna skattade själva var ROC-AUC 0,67.

I en andra studie undersökte Youngstrom och medarbetare diagnostisk tillförlitlighet för fullängds- respektive kortversionen med tio frågor [95]. Föräldrarna till drygt 600 barn och ungdomar från psykiatrisk öppenvård gjorde skattningarna. Youngstrom och medarbetare redovisade ROC-AUC 0,83 för fullängdsversionen av maniskalan och 0,86 för kortversionen.

Findling och medarbetare undersökte 196 barn- och ungdomspsykiatriska öppenvårdspatienter mellan 5 och 17 år [96]. För föräldraskattning av mani var ROC-AUC 0,88 medan ungdomarnas självskattning uppnådde ett ROC-AUC på 0,82. Findlings studie är den enda som också redovisar sensitivitet och specificitet. Vid tröskelvärde 17 på maniskalan var sensitiviteten 91 procent och specificiteten 68 procent för föräldraskattning.

### ***Sammanvägda resultat***

För föräldraskattning låg ROC-AUC över 0,80 i tre studier. Två av dem redovisade också ungdomarnas egen skattning. I och med att ROC-AUC i den ena studien var 0,67 och i den andra 0,82 bedömde vi att det inte var meningsfullt att väga samman resultaten.

### ***Bedömning av evidensstyrka***

Vi tillämpade GRADE-kriterierna för att bedöma säkerheten i data för ROC-AUC (Tabell 3.4.13). Studierna hade en god kvalitet och risken för bias var liten. Överförbarheten var god. Studierna var stora och genomförda i barnpsykiatrisk öppenvård. Alla studier använde samma referensstandard. För självskattning var resultaten inte samstämmiga.

**Tabell 3.4.5** Resultatsammanställning för GBI, riktad screening för mani hos barn och ungdomar.

Utfallsmått	Antal studier Antal patienter	Sammanvägt estimat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet, föräldraskattning	1 studie n=196	91%	Otillräcklig ⊕○○○	En studie
Specificitet, föräldraskattning	2 studier n=196	68%	Otillräckligt ⊕○○○	En studie
ROC-AUC, föräldraskattning	3 studier n=1 475	>80%	Starkt ⊕⊕⊕⊕	
ROC-AUC, självskattning	2 studier n=838	Bör inte vägas samman	Otillräckligt ⊕○○○	

KI = Konfidensintervall

### ***Evidensgraderade resultat***

Det finns för få studier med tillräcklig kvalitet för att kunna bedöma sensitivitet och specificitet för GBI för att identifiera bipolära syndrom hos barn och ungdomar med strukturerade eller semistrukturerade intervjuer som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

Mätt som ROC-AUC har GBI:s föräldraskala en tillräcklig förmåga (ROC-AUC >0,80) att identifiera mani hos barn och ungdomar med K-SADS som referensstandard (starkt vetenskapligt underlag ⊕⊕⊕⊕).

Det går inte att bedöma ROC-AUC för GBI:s självskattning för att identifiera mani hos barn- och ungdomar med strukturerad eller semistrukturerad intervju som referensstandard pga bristande samstämmighet mellan studierna (otillräckligt vetenskapligt underlag ⊕○○○).



## **Child Bipolar Questionnaire (CBQ)**

CBQ består av 65 frågor och fylls i av förälder eller kliniker. Frågorna graderas utifrån en Likertskala 1 till 4 [99]. Tröskelvärde för bipolärt syndrom enligt DSM-IV baseras på en algoritm. Den bygger på värden för utvalda frågor och inte på en totalpoäng för alla frågor. Vi fann ingen formell svensk översättning.

### ***Beskrivning av underlaget***

Endast en studie lästes i fulltext och den uppfyllde inklusionskriterierna [99]. Den bedömdes ha låg studiekvalitet eftersom uppgift om tiden mellan indextest och referenstest var oklar och det saknades uppgifter om blindning. Det framgick inte heller om interbedömarreliabiliteten hade undersökts.

### ***Evidensgraderat resultat***

Det går inte att bedöma sensitivitet och specificitet för CBQ som formulär för riktad screening av bipolärt syndrom hos barn och ungdomar med strukturerad eller semistrukturerad intervju som referensstandard. Det finns inte tillräckligt många studier med adekvat kvalitet (otillräckligt vetenskapligt underlag ⊕○○○).

## **Mood Disorder Questionnaire – Adolescent (MDQ-A)**

MDQ-A finns i en version för tonåringar och en för föräldrar. MDQ-A är identiskt med formuläret MDQ för vuxna [100].

### ***Beskrivning av underlaget***

Endast en studie lästes i fulltext och den uppfyllde inklusionskriterierna, men den bedömdes ha låg studiekvalitet eftersom tiden mellan index- och referenstest inte redovisades [101] (Tabell 3.4.16b).

### ***Evidensgraderat resultat***

Det går inte att bedöma sensitivitet och specificitet för MDQ-A som formulär för riktad screening av bipolärt syndrom hos barn och ungdomar med strukturerade eller semistrukturerade intervjuer som referensstandard. Det finns för få studier med adekvat studiekvalitet (otillräckligt vetenskapligt underlag ⊕○○○).

## **Child Mania Rating Scale – Parent (CMRS-P)**

CMRS-P fylls i av barnets förälder. Formuläret finns i en fullängdsversion med 21 frågor och en kortversion med 11 frågor, båda utvärderade av konstruktörerna i två originalarbeten [102,103]. Varje fråga graderas på en fyrgradig skala (0–3), mellan ”aldrig/sällan” och ”mycket ofta”. Tröskelvärdet för något bipolärt syndrom är för fullängdsversionen 20 poäng och för kortversionen 10 poäng. CMRS-P (fullängdsversionen) finns i svensk översättning.

### ***Beskrivning av underlaget***

Två studier, publicerade i tre artiklar, lästes i fulltext. En av studierna, som undersökte förmågan hos CMRS-P att mäta förändring exkluderades [104]. Den andra, som redovisades i två artiklar bedömdes ha hög studiekvalitet [102,103] (Tabell 3.4.14). Den ena artikeln utvärderade fullängdsversionen och den andra kortversionen av CMRS-P.

Studien genomfördes på en barnpsykiatrisk forskningsklinik i USA [102,103]. Totalt rekryterades 150 deltagare; 50 med ADHD, 50 med bipolärt syndrom (BPI, BPII och BP-UNS) och 50 kontroller. Referensstandard var en något utökad K-SADS (WASH-U-KADS).

CMRS-P i fullängdsversion hade en relativt god förmåga att skilja patienter med bipolärt syndrom från patienter med ADHD. Sensitiviteten var 82 procent (95 % KI, 69 till 91) och specificiteten var 94 procent (95 % KI, 83 till 99). Motsvarande sensitivitet och specificitet för kortversionen av CMRS-P var 84 procent (95 % KI, 71 till 93) respektive 83 procent (95 % KI, 71 till 93).

### ***Evidensgraderat resultat***

Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för fullängdsversionen respektive kortversionen av CMRS-P som formulär för riktad screening av bipolärt syndrom hos barn och ungdomar med strukturerad eller semistrukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## Young Mania Rating Scale (YMRS)

YMRS är en klinikerskattad skala som konstruerades år 1978 av Young och medarbetare i USA [105]. Skalan finns i tre versioner; en för kliniker, en för tonåringar och en för föräldrar. Skattningen baseras på patientens tillstånd de två senaste veckorna [105]. Formuläret består av 11 påståenden, och poängen är mellan 0 och 56. Högre värden indikerar mer maniska symtom, men det finns inte något etablerat tröskelvärde.

### *Beskrivning av underlaget*

Fyra studier i fem publikationer lästes i fulltext och samtliga uppfyllde inklusionskriterierna [93,106–109]. Två studier i två publikationer bedömdes ha hög kvalitet [106,107] och tre medelhög [93,108,109] (Tabell 3.4.15).

I samtliga studier utom en rekryterades deltagarna från forskningskliniker. En av Youngstrom och medarbetares studier inkluderade deltagarna från en vanlig klinisk barnpsykiatrisk mottagning [109].

I Youngstrom och medarbetares studie undersöktes föräldraskattning med YMRS [93]. ROC-AUC var 0,83 för barn mellan 5 och 10 år och 0,80 för barn och ungdomar mellan 11 och 17 år. Prevalensen av bipolärt syndrom i den yngre åldersgruppen var 50 procent och i den äldre 41 procent.

I Gracious och medarbetares samt Youngstrom och medarbetares studie presenterad i två publikationer rekryterades 117 patienter från en forskningsklinik med en prevalens av bipolärt syndrom på 50 procent [106,107]. ROC-AUC för föräldraskattningarna var 0,85 och för expertskattningen cirka 1,0. Denna studie redovisade även en sensitivitet på 79 procent och specificitet på 78 procent för tröskelvärdet 17.

I studierna av medelhög kvalitet hade Frazier rekryterat 1 014 barn och ungdomar från en forskningsklinik [108]. Prevalensen av bipolärt syndrom var 53 procent. ROC-AUC för skattning gjord av psykiatrer/psykolog var 0,97.

Slutligen rekryterade Youngstrom och medarbetare 262 barn och ungdomar från barnpsykiatrisk öppenvård med en prevalens av bipolärt syndrom på 34 procent [109]. Han redovisade ROC-AUC för föräldraskattning för dem under 10 år på 0,66, för dem mellan 11 och 17 år på 0,70 och slutligen självskattning i den äldre gruppen med ROC-AUC på 0,50.

### ***Sammanvägt resultat***

Två studier har använt ROC-AUC för expertskattning som utfallsmått. Båda studierna visar en hög AUC; 1,0 respektive 0,97. Tre studier redovisade ROC-AUC för föräldraskattning, och resultaten var 0,76, 0,80 respektive 0,85.

### ***Bedömning av evidensstyrka***

Vi tillämpade GRADE-kriterierna för utfallsmåtten expert- och föräldraskattning med YMRS. Studierna är välgjorda och vi bedömde att risken för bias var liten. Det fanns en brist i överförbarhet eftersom större delen av studierna var genomförda på forskningsklinik med en prevalens för bipolära syndrom på cirka 50 procent. Samstämmigheten var god för expertskattning men bristfällig för föräldraskattning. Samtliga studier har genomförts av samma forskningsgrupp, men gruppen är oberoende av upphovsmännen vilket innebär att vi bedömer att risken för publikationsbias är liten.

**Tabell 3.4.6** Resultatsammanställning för YMRS som formulär för riktad screening av mani hos barn och ungdomar.

<b>Utfallsmått</b>	<b>Antal studier Antal patienter</b>	<b>Sammanvägt estimat (95% KI)</b>	<b>Bedömning av vetenskapligt underlag</b>	<b>Kommentarer till poängavdrag</b>
Sensitivitet, föräldraskattning	1 studie n=117	79%	Otillräcklig ⊕○○○	En studie
Specificitet, föräldraskattning	1 studie n=117	78%	Otillräckligt ⊕○○○	En studie

*Tabellen fortsätter på nästa sida*

**Tabell 3.4.5** fortsättning

Utfallsmått	Antal studier Antal patienter	Sammanvägt estimat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
ROC-AUC, föräldraskattning	3 studier n=1 016	76–85%	Begränsat ⊕⊕○○	Bristande överförbarhet (–1) Bristande samstämmighet (–1)
ROC-AUC, expertskattning	2 studier n=1 131	97–100%	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (–1)
ROC-AUC, självskattning	1 studie n=262	50%	Otillräckligt ⊕○○○	En studie

KI = Konfidensintervall

### *Evidensgraderade resultat*

- Det finns för få studier av tillräcklig kvalitet för att bedöma sensitivitet och specificitet för YMRS vid riktad screening av mani hos barn och ungdomar vid tröskelvärde 17 och med semistrukturerad eller strukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Expertskattning med YMRS har en hög tillförlitlighet (ROC-AUC  $\geq 0,97$ ) som formulär för riktad screening av bipolära syndrom hos barn och ungdomar inom barnpsykiatri med K-SADS som referensstandard (måttligt starkt vetenskapligt stöd ⊕⊕⊕○).
- Föräldraskattning med YMRS har en tillräcklig tillförlitlighet (ROC-AUC  $\geq 0,76$ ) som formulär för riktad screening av bipolära syndrom hos barn och ungdomar inom barnpsykiatri med K-SADS som referensmetod (begränsat vetenskapligt stöd ⊕⊕○○).
- Det går inte att bedöma hur tillförlitlig självskattning med YMRS är som formulär för riktad screening av bipolära syndrom hos ungdomar med semistrukturerade eller strukturerade intervjuer som

referensstandard eftersom det finns för få studier med tillräcklig kvalitet (otillräckligt vetenskapligt stöd ⊕○○○).

## **Formulär för att bedöma svårighetsgrad av förstämningssyndrom**

Vår fråga var vilka formulär som är tillförlitliga för att bedöma svårighetsgraden av förstämningssyndrom hos barn och ungdomar vid ett eller flera tillfällen. Vi accepterade svårighetsgrad mätt enligt K-SADS eller CGI-S som referensstandarder (se Kapitel 3.1).

### *Beskrivning av underlaget*

Vi identifierade endast en studie som utvärderade formulär för bedömning av svårighetsgrad. Studien av West och medarbetare syftade till att utvärdera förändring av maniska symtom över tid [104]. I studien följdes 66 barn och ungdomar med bipolärt syndrom en gång i veckan under sex veckor. Studien jämförde CMRS-P med CGI-S. Studien redovisade inte interbedömarreliabilitet för CGI-S och exkluderades därför.

### *Evidensgraderade resultat*

Det saknas studier som utvärderar formulär för att mäta svårighetsgrad av förstämningssyndrom jämfört med K-SADS eller CGI-S.

## **Diskussion**

### **Intervjuer**

Mot bakgrund av att LEAD ofta betraktas som referensstandard i valideringsstudier [110] är det anmärkningsvärt att det bara finns en studie som jämför K-SADS med LEAD. Även för DICA och MINI-KID finns det bara en studie vardera, mätt mot K-SADS eller LEAD. Även om underlaget är knapphändigt tyder resultaten på att intervjuerna har tillräcklig sensitivitet och specificitet för att vara värdefulla i klinisk praxis [111,112]. Generellt sett råder det dock stor brist på studier som utvärderar de semi-strukturerade och strukturerade intervjuformulär som används i Sverige mot en adekvat referensstandard.

Användning av semistrukturerade intervjuer är inte regel inom svensk barnpsykiatri [113]. Tidsbrist har angetts som ett skäl. Två nordiska studier har utvärderat K-SADS i rutinsjukvård. Lauth och medarbetare följde implementering av K-SADS inom slutna barnpsykiatrisk vård på Island och visade att flera barn med en diagnos identifierades med formuläret än utan [114]. En dansk studie undersökte erfarenheter av intervjun hos föräldrar, barn och personal inom barnpsykiatri [115]. Generellt sett var erfarenheterna goda. Som nackdelar nämndes att intervjun tog lång tid i anspråk och att den möjligen kunde vara för krävande för yngre, utagerande barn.

Det amerikanska projektet Improving the assessment of bipolar disorder jämförde bl a hur diagnostiska formulär fungerade på en barnpsykiatrisk öppenvårdsmottagning respektive en specialistklinik [116]. Såväl forskare som behandlande läkare fick utbildning i att använda K-SADS. Läkarna upplevde att utbildningen bidrog till en kvalitativ förbättring av deras kompetens i diagnostik. När studien avslutats valde klinikerna att fortsätta använda K-SADS trots den större tidsåtgången.

Vi har inte identifierat några systematiska översikter om diagnostiska intervjuer för barn. Däremot finns det två äldre, icke systematiska, översikter av Hodges [112] respektive Renou och medarbetare [111]. Båda översikterna framhöll värdet av K-SADS men också att det behövs mera forskning.

Våra resultat måste värderas i ljuset av våra avgränsningar. Vi inkluderade enbart studier som utförts i västvärlden. Det är möjligt att vi därigenom har exkluderat studier som skulle kunnat stärka det vetenskapliga underlaget.

### **Formulär för riktad screening**

De flesta studier har använt andra referensstandarder än intervjuer. Det fanns tillräckligt med studier för att kunna bedöma tre formulär, BDI-I, CDI och CBCL. CDI hade en låg sensitivitet medan sensitiviteten för BDI-I var 81 procent och CBCL-affektiva skalan 88 procent (95 % KI, 78 till 94). Detta kan jämföras med resultaten från en systematisk över-

sikt av Warnick och medarbetare där sensitiviteten för delskalorna Mood/Anxiety beräknades till 61 procent (95 % KI, 47 till 75) [29]. Warnick och medarbetare hade inkluderat sex studier utan hänsyn till tröskelvärde och de flesta studierna hade ett tröskelvärde som var lägre än Aschenbachs rekommenderade,  $\geq 65$ , vilket gör Warnick och medarbetares studie svårtolkad.

Formulären kan inte användas för att säkert utesluta eller fastställa en diagnos på depression. Deras funktion är att stärka eller försvaga en misstanke om förstämningssyndrom. Diagnostiken måste alltid baseras på bred information och även omfatta anamnes, diagnostisk intervju och gradering av svårighetsgrad.

En stor svaghet i underlaget är bristen på studier som utförts inom primärvården. En enda studie på depression var utförd inom barnmedicinsk öppenvård [67]. Vi kunde inte identifiera någon studie om bipolära syndrom gjord i primärvård. Vi saknar därmed kunskap om hur formulären fungerar i första linjens barn- och ungdomspsykiatri, dvs inom barn- och skolhälsovård och inom primärvård. Detta är en brist eftersom man har visat att bipolära syndrom ofta förblir oidentifierade om man inte använder formulär för riktad screening för att upptäcka dem [117].





**Table 3.4.7** Sensitivity and specificity for structured and semistructured interviews.

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
Carlson et al 1987 [2] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> One Mental Health Center Children's inpatient services</p> <p><u>Population</u> n=51 consecutive children with mean duration 12 weeks at the hospital</p> <p><u>Inclusion criteria</u> Age: ≥7 years Lived with a parent or foster parent ≥2 years</p> <p><u>Prevalence</u> MDD and dysthymia: 43%</p>	<p><u>Index test</u> K-SADS</p> <p><u>Reference test</u> LEAD diagnosis</p> <p><u>Number of patients</u> n=30 (10 girls) Mean age: 10.1±1.5 years</p> <p><u>Drop-out rate</u> 5/30</p>	<p><u>Index test rater</u> A professor in psychiatry</p> <p><u>Reference test rater</u> Two child psychiatrists</p> <p><u>Rater training</u> Not reported</p> <p><u>Interobserver reliability</u> κ 0.58</p>	<p><u>MDD + dysthymia</u> Sensitivity: 81% Specificity: 79%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> NA</p>
Carlson et al 1987 [2] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> One Mental Health Center Children's inpatient services</p> <p><u>Population</u> n=51 consecutive children with mean duration 12 weeks at the hospital</p> <p><u>Inclusion criteria</u> Age: ≥7 years Lived with a parent or foster parent ≥2 years</p> <p><u>Prevalence</u> MDD and dysthymia: 43%</p>	<p><u>Index test</u> DICA</p> <p><u>Reference test</u> LEAD diagnosis</p> <p><u>Number of patients</u> n=30 (10 girls) Mean age: 10.1±1.5 years</p> <p><u>Drop-out rate</u> 5/30</p>	<p><u>Index test rater</u> A child psychiatrist or the clinical psychology graduate student</p> <p><u>Reference test rater</u> Two child psychiatrists</p> <p><u>Rater training</u> Not reported</p> <p><u>Interobserver reliability</u> κ 0.48</p>	<p><u>MDD + dysthymia</u> Sensitivity: 81% Specificity: 68%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> NA</p>

The table continues on the next page

Table 3.4.7 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	Sensitivity Specificity	Study quality Comments
Sheehan et al 2010 [10] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Child and adolescent psychiatric research setting</p> <p><u>Population</u> 231 children and adolescents patients volunteers being seen for evaluation and treatment and 36 recruited controls</p> <p><u>Inclusion criteria</u> Age: 12–17 years and 11 months</p> <p><u>Exclusion criteria</u> Mental retardation, significant developmental disorders, congenital abnormalities, delirium, dementia, a language problem, serious medical illness, organic brain damage</p> <p><u>Prevalence</u> MDD: 18% Mania/hypomania: 41%</p>	<p><u>Index test</u> MINI-KID</p> <p><u>Reference test</u> K-SADS</p> <p><u>Number of patients</u> n=195 (94 girls) Mean age: 12.8±3.5 years</p> <p><u>Drop-out rate</u> 2.5%</p>	<p><u>Rater</u> Three board-certified faculty psychiatrists, 6 psychiatry residents or child fellows, 4 doctoral-level faculty psychologists, 1 RN/MS and 5 BA and MS research coordinators</p> <p><u>Rater training</u> All raters had received training on the MINI-KID and the K-SADS</p> <p><u>Interobserver reliability</u> κ 0.93 for any mood disorders, 1.00 for MDD episode and 1.00 for mania/hypomania</p>	<p><u>Major depressive disorder</u> Sensitivity: 54% Specificity: 87%</p> <p><u>Mania/hypomania</u> Sensitivity: 96% Specificity: 73%</p>	<p>High</p> <p><u>Comments</u> <u>Sampling method</u> Adequate</p> <p><u>Blinding</u> The raters were blind to the results of the other interview</p> <p><u>Handling of missing data</u> NA</p>

BA = Bachelor of arts; DICA = Diagnostic interview for children and adolescents; K-SADS = Kiddie-schedule for affective disorders and schizophrenia; LEAD = Longitudinal, Experts, All Data procedure; MDD = Major depressive disorder; MINI-KID = Mini international neuropsychiatric interview for children and adolescents; MS = Master of science; NA = Not applicable; RN/MS = Master of science in nursing

**Table 3.4.8** Sensitivity and specificity for Child Behaviour Check List (CBCL) with semistructured or structured 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
Krol et al 2006 [33] The Netherlands	<u>Design</u> Cohort, clinical  <u>Setting</u> Outpatient, mental health clinic  <u>Population</u> n=44 consecutively admitted  <u>Inclusion criteria</u> Referred patients, parent agreement  <u>Exclusion criteria</u> None described  <u>Prevalence</u> MDD: 16%	<u>Index test</u> CBCL affective problems scale  <u>Reference test</u> DISC-IV  <u>Number of patients</u> n=44 (43% girls) Mean age: 10.6±2.7 years  <u>Drop-out rate</u> Not reported	<u>Index test rater</u> Parent  <u>Reference test rater</u> Four graduate students and a psychologist employed by the clinic  <u>Rater training, reference test</u> Rater by the translator  <u>Interobserver reliability</u> During each interview, a co-student was present to check the reliability by simultaneously scoring the interview, $\kappa$ 0.88–1.00	<u>MDD cut off score 64</u> Sensitivity: 86% Specificity: 83%	Moderate  <u>Comments</u> <u>Sampling method</u> Consecutively included, all invited parents participated, relatively small sample  <u>Blinding</u> Not explicitly explained, but probably blind  <u>Handling of missing data</u> Unclear  <u>Other comments</u> The order in which the CBCL and the DISC were administered was random. The time between the reference test and index test is not presented but probably they were done at the same time

The table continues on the next page

Table 3.4.8 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	Sensitivity Specificity	Study quality Comments
Mattison et al 2007 [34] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Research setting</p> <p><u>Population</u> Participators in a university project, or recruited through several sources</p> <p><u>Inclusion criteria</u> Age: 6–18 years Index test completed by both parents and teachers Current diagnosis of 1) MDD, dysthymia or both 2) ADHD 3) Community control</p> <p><u>Exclusion criteria</u> None described</p> <p><u>Prevalence</u> The depressed group: MDD 48.1%, dysthymia 32.7%, both 19.2%</p> <p>The ADHD and control group: MDD &amp; dysthymia 0%</p>	<p><u>Index test</u> CBCL affective problems scale</p> <p><u>Reference test</u> K-SADS-E, parents- and child- (if &gt;8 years old) interviews</p> <p><u>Number of patients</u> n=153 (38% girls) Mean age: 11.1±2.9 years</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Index test rater</u> Parents or self rating</p> <p><u>Reference test rater</u> Experienced interviewers</p> <p><u>Rater training</u> Not described</p> <p><u>Interobserver reliability</u> κ 0.77–0.86 for the major diagnostic categories</p>	<p><u>MDD cut off score 65</u> Sensitivity: 88% Specificity: 73%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Not reported</p> <p><u>Handling of missing data</u> Items where imputed if missing in 13.7% of the CBCL questionnaires</p> <p><u>Other comments</u> 'Current', is the last 60 days. Time between index and reference test OK, communicated by author</p>

ADHD = Attention deficit/hyperactivity disorder; CBCL = Child behaviour check list; DISC-IV = Diagnostic interview schedule for children, fourth edition; K-SADS-E = Kiddie-schedule for affective disorders and schizophrenia, epidemiological edition; MDD = Major depressive disorder

**Table 3.4.9** Sensitivity and specificity for Beck Depression Inventory (BDI) with semistructured or structured 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
Subramaniam et al 2009 [55] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Clinical, substance abuse treatment program</p> <p><u>Population</u> 184 adolescents with either opioid use or cannabis/alcohol abuse disorder</p> <p><u>Inclusion criteria</u> Less than 2 weeks of abstinence and treatment seeking</p> <p><u>Exclusion criteria</u> Not reported</p> <p><u>Prevalence MDD</u> 50% (73/145)</p>	<p><u>Index test</u> BDI-I</p> <p><u>Reference test</u> DICA-IV</p> <p><u>Number of patients</u> n=145</p> <p><u>Drop-out rate</u> None</p>	<p><u>Rater, index test</u> Self rating</p> <p><u>Rater training</u> NA</p> <p><u>Rater, reference test</u> The first author (MD) and two other psychiatrists</p> <p><u>Rater training</u> The two psychiatrists were trained by the first psychiatrist</p> <p><u>Interobserver reliability</u> 100% agreement on observed interviews prior to the study, but no information about how many interviews that were included or any reported <math>\kappa</math> values</p>	<p><u>Cut off score 11</u> Sensitivity: 77% Specificity: 47%</p>	<p>Moderate</p> <p><u>Comments</u> <u>Sampling method</u> Probably consecutively recruited, but not clearly described</p> <p><u>Blinding</u> Unclear</p> <p><u>Handling of missing data</u> Patients with missing data were excluded</p> <p><u>Other comments</u> The time between reference and index time is unclear</p>

The table continues on the next page

Table 3.4.9 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	Sensitivity Specificity	Study quality Comments
Marton et al 1991 [54] Canada	<u>Design</u> Cross sectional  <u>Setting</u> Clinical, child-psychiatric  <u>Population</u> n=119 consecutively admitted adolescents to either an out-patient clinic (n=93) or to a psychiatric ward  <u>Inclusion criteria</u> —  <u>Exclusion criteria</u> Not described  <u>Prevalence</u> 21% (25/119)	<u>Index test</u> BDI-I  <u>Reference test</u> DICA  <u>Number of patients</u> n=119 (57% girls)  <u>Drop-out rate</u> Not reported	<u>Rater, index test</u> Self rating  <u>Rater training, index test</u> NA  <u>Rater, reference test</u> Psychiatrists, psychologists and social workers, number not reported  <u>Rater training, reference test</u> Not described  <u>Interobserver reliability</u> Not reported	<u>Cut off score 11</u> Sensitivity: 80% Specificity: 70%  <u>Cut off score 16</u> Sensitivity: 68% Specificity: 81%	Moderate  <u>Comments</u> Sampling method Adequate, consecutively admitted adolescents were recruited  <u>Blinding</u> Probably  <u>Handling of missing data</u> Unclear
Ambrosini et al 1991 [53] USA	<u>Design</u> Cross sectional  <u>Setting</u> Child psychiatric outpatient clinic  <u>Population</u> n=122 consecutively admitted adolescents  <u>Inclusion criteria</u> Referred to a child depression clinic  <u>Exclusion criteria</u> Physically ill, mentally retarded, symptoms clearly suggestive of anorexia/bulimia nervosa or autism  <u>Prevalence MDD</u> 45% (55/122)	<u>Index test</u> BDI-I  <u>Reference test</u> K-SADS  <u>Number of patients</u> n=122 (57% girls) Mean age: 15.5 years (range 12.6 to 19.8)  <u>Drop-out rate</u> Not reported	<u>Rater, index test</u> Self report  <u>Rater training, index test</u> NA  <u>Rater, reference test</u> Not reported  <u>Rater training, reference test</u> Not reported  <u>Interobserver reliability</u> κ for rater agreement ranged from 0.72 to 0.92	<u>Cut off scores &gt;11</u> Sensitivity: 48% Specificity: 40%  <u>Cut off scores &gt;16</u> Sensitivity: 35% Specificity: 49%	Moderate  <u>Comments</u> Sampling method Adequate  <u>Blinding</u> Unclear  <u>Handling of missing data</u> Not reported

The table continues on the next page

Table 3.4.9 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	Sensitivity Specificity	Study quality Comments
Kashani et al 1990 [51] USA	<u>Design</u> Cross sectional  <u>Setting</u> Clinical, counselling centre  <u>Population</u> n=102 consecutively admitted Mean age: 14.8 years, range 13–18 High level of comorbidity  <u>Inclusion criteria</u> –  <u>Exclusion criteria</u> Not reported  <u>Prevalence MDD</u> 31%	<u>Index test</u> BDI-I  <u>Reference test</u> DICA  <u>Number of patients</u> n=100 (53% girls)  <u>Drop-out rate</u> Not reported	<u>Rater of index test</u> Self rating  <u>Rater training, index test</u> NA  <u>Rater of reference test</u> Master's level staff  <u>Rater training</u> Extensive training  <u>Interobserver reliability</u> Not reported	<u>Cut off scores &gt;16</u> Sensitivity: 48% Specificity: 87%	Moderate  <u>Comments</u> Sampling method Adequate  <u>Blinding</u> Satisfactory  <u>Handling of missing data</u> Not reported
Barrera et al 1988 [52] USA	<u>Design</u> Cross sectional  <u>Setting</u> Child psychiatric clinic  <u>Population</u> n=65 inpatients Mean age: 14,6 years (range 12 to 17 years)  <u>Inclusion criteria</u> Not reported  <u>Exclusion criteria</u> Not reported  <u>Prevalence MDD</u> 42% (27/65)	<u>Index test</u> BDI-I  <u>Reference test</u> CAS  <u>Number of patients</u> n=65 (52% girls)  <u>Drop-out rate</u> Not reported	<u>Rater of index test</u> Self rating  <u>Rater training, index test</u> NA  <u>Rater of reference test</u> Research assistants  <u>Rater training</u> Several weeks of training including supervised inter- views  <u>Interobserver reliability</u> Not reported	<u>Cut off score 11</u> Sensitivity: 82% Specificity: 53%  <u>Cut off score 16</u> Sensitivity: 56% Specificity: 74%	Moderate  <u>Comments</u> Sampling method Not clearly described  <u>Blinding</u> Satisfactory  <u>Handling of missing data</u> Unclear  <u>Other comments</u> Only one interviewer had graduate training in clinical psychology

BDI = Beck depression inventory; CAS = Child assesment scheudule; DICA-IV = Diag-  
nostic interview schedule for children and adolescents, fourth edition; K-SADS = Kiddie-  
schedule for affective disorders and schizophrenia; MD = Doctor of medicine; MDD =  
Major depressive disorder; NA = Not applicable



**Table 3.4.10** Sensitivity and specificity for Children's Depression Inventory (CDI) as screening instrument for depression with semistructured or structured 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
Sørensen et al 2005 [68] Denmark	<p><u>Design</u> Cross sectional, clinical study</p> <p><u>Setting</u> Psychiatry, in and outpatients from a psychiatric Hospital for Children and Adolescents</p> <p><u>Population</u> n=221 children, consecutively referred to psychiatry for the first time because of any mental problem</p> <p><u>Inclusion criteria</u> Age: 8–13 years</p> <p><u>Exclusion criteria</u> Communication problems</p> <p><u>Prevalence MDD</u> 13%</p>	<p><u>Index test</u> CDI</p> <p><u>Reference test</u> K-SADS-PL (children &amp; parents)</p> <p><u>Number of patients</u> n=199 (26% girls; 12 inpatients) Mean age: 10.6 years</p> <p><u>Drop-out rate</u> 25%</p>	<p><u>Rater of index test</u> Children</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater of reference test</u> Clinicians</p> <p><u>Rater training</u> Training course in ICD-10/ DSM-IV and K-SADS-PL</p> <p><u>Interobserver reliability</u> Total agreement on 3 consecutive videos as entry criterion. The rater re-assessed 20 interviews during the study</p> <p>MDD <math>\kappa</math> 0.76</p>	<p><u>MDD at cut off 12</u> <u>K-SADS child</u> Sensitivity: 79% Specificity: 66%</p> <p><u>MDD at cut off 12</u> <u>K-SADS parent</u> Sensitivity: 63% Specificity: 64%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Probably blind. The interviews were performed after the CDI had been filled in</p> <p><u>Handling of missing data</u> Not reported</p> <p><u>Other comments</u> 14 boys and 8 girls were excluded due to communication problems but they did not differ in gender or age</p>

The table continues on the next page

Table 3.4.10 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
Shemesh et al 2005 [67] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Primary care clinics</p> <p><u>Population</u> 125 children, consecutively referred with somatic illnesses considered to have an emotional problem but not under current psychiatric care</p> <p><u>Inclusion criteria</u> Age 8–19 years old English or Spanish speaking, diagnosed with somatic disorder ≥6 months previously</p> <p><u>Exclusion criteria</u> Cognitive barriers that precluded understanding of the study procedure, medical illness that required an inpatient admission during the evaluation, active suicidal risk or current suicidal ideation, current psychiatric treatment</p> <p><u>Prevalence depressive disorders</u> 20% (16/81)</p>	<p><u>Index test</u> CDI</p> <p><u>Reference test</u> Best estimate diagnosis on a consensus meeting based on all information including K-SADS-PL (parent &amp; children)</p> <p><u>Number of patients</u> n=102 (58% girls) Mean age: 13.4 years</p> <p><u>Drop-out rate</u> 21%</p>	<p><u>Rater, index test</u> Children's self rating</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater, reference test</u> Doctorate level psychology interns</p> <p><u>Rater training, reference test</u> Training sessions using taped interviews and at least two supervised patient interviews</p> <p><u>Interobserver reliability</u> κ K-SADS vs best estimate: 0.688</p>	<p><u>CDI, cut off 12.5 for depressive disorders</u> Sensitivity: 60% Specificity: 78%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> Not reported</p>

CDI = Children's depression inventory; DSM-IV = Diagnostic and statistical manual of mental disorder, fourth edition; ICD-10 = International statistical classification of diseases and related health problems, tenth revision; K-SADS = Kiddie-schedule for affective disorders and schizophrenia for school-age children; K-SADS-PL = Kiddie-schedule for affective disorders and schizophrenia for school-age children present and life time; MDD = Major depressive disorder; NA = Not applicable

**Table 3.4.11** Sensitivity and specificity for Centre of Epidemiological Studies Depression Scale for Children (CES-DC) as screening instrument for depression with semistructured or structured 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
Fendrich et al 1990 [71] USA	<p><u>Design</u> Retrospective, cohort</p> <p><u>Setting</u> Research setting</p> <p><u>Population</u> 220 children, high and low risk for major depression by virtue of the presence or absence of major depression in one or more of their parents or not 57% girls, age 6–23 years</p> <p><u>Inclusion criteria</u> Having completed all 20 items on the CES-DC at initial assessment</p> <p><u>Exclusion criteria</u> None</p> <p><u>Prevalence</u> MDD 17%</p>	<p><u>Index test</u> CES-DC</p> <p><u>Reference test</u> K-SADS included in the Best estimate diagnoses</p> <p><u>Number of patients</u> n=166</p> <p><u>Drop-out rate</u> None</p>	<p><u>Rater index test</u> Children's self assessment</p> <p><u>Rater, reference test</u> Not described</p> <p><u>Rater training</u> Not described</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>MDD at cut off 16</u> Sensitivity: 71% Specificity: 62%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Longitudinal sample</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> Participants with incompletely filled in questionnaires were excluded</p> <p><u>Other comments</u> Time between tests unclear</p>

CES-DC = Center for epidemiological studies depression scale for children; K-SADS = Kiddie-schedule for affective disorders and schizophrenia; MDD = Major depressive disorder

**Table 3.4.12** Child Behavior Checklist screening instrument for bipolar syndromes with semistructured or structured 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
Kahana et al 2003 [36] USA	<p><u>Design</u> Cohort</p> <p><u>Setting</u> Clinic for child psychiatry</p> <p><u>Population</u> Referred from two pediatric psychopharmacology research units Outpatients, mean age: 10.9 years SD: 3.3 years (35% girls)</p> <p><u>Inclusion criteria</u></p> <p><u>Exclusion criteria</u> Psychiatric disorder due to a general medical condition, pervasive developmental disorder, mental retardation</p> <p><u>Prevalence</u> 23% (100/439) BPSD I, 15% (66/439) BPSD II, cyclothymia, BPSD NOS</p>	<p><u>Index test</u> CBCL, aggressive subscale</p> <p><u>Reference test</u> K-SADS</p> <p><u>Number of patients</u> n=439</p> <p>Four comparison groups: BPSD=166 DBD=115 MDD=90</p> <p>Any individual without BPSD=250</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater, index test</u> Self rating</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater, reference test</u> Three child and adolescent psychiatrists</p> <p>11 highly trained research assistants</p> <p><u>Rater training</u> Five K-SADS interviews along with an experienced rater. Joint rating sessions every 10th interview</p> <p><u>Interobserver reliability</u> <math>\kappa &gt; 0.85</math></p>	<p><u>ROC-AUC</u> 0.81</p>	<p>High</p> <p><u>Comments</u> <u>Sampling method</u> Probably consecutively included patients</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> Completed rate of CBCL are reported: 157/166 BPSD group 82/90 MDD group 109/115 DBD group</p>

The table continues on the next page

Table 3.4.12 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
Youngstrom et al 2006 [92] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Child-psychiatric outpatient clinic</p> <p><u>Population</u> n=642 (39% girls) visiting for outpatient evaluation, of whom 500 consented and completed all tests</p> <p><u>Inclusion criteria</u> Age: 5–17 years Both parent and youth completing the assessment process</p> <p><u>Exclusion criteria</u> Not fluent in spoken English, pervasive developmental disorder, mental retardation</p> <p><u>Prevalence BPSD</u> 76% (216/500)</p>	<p><u>Index test</u> CBCL</p> <p><u>Reference test</u> K-SADS-PL or K-SADS-E</p> <p><u>Number of patients</u> n=500</p> <p><u>Drop-out rate</u> None</p>	<p><u>Rater of index test</u> Parents</p> <p><u>Training of rater, index test</u> NA</p> <p><u>Rater of reference test</u> Research assistant, all diagnoses were reviewed at a consensus meeting involving a child psychiatrist</p> <p><u>Rater training</u> Highly trained, how is not described</p> <p><u>Interobserver reliability</u> <math>\kappa &gt; 0.85</math> at the item level and complete diagnostic agreement on five cases led by themselves and on five cases led by others</p>	<p><u>ROC-AUC</u> 0.80</p>	<p>High</p> <p><u>Comments</u> <u>Sampling method</u> All eligible patients were included as the “heterogenous” sample</p> <p><u>Blinding</u> Not described</p> <p><u>Handling of missing data</u> 78% had complete data. Listwise deletion was used for missing data</p>

BPSD = Bipolar spectrum disorders; CBCL = Child behaviour check list; DBD = Disruptive behavior disorder; K-SADS = Kiddies-schedule for affective disorders and schizophrenia; K-SADS-E = Kiddies-schedule for affective disorders and schizophrenia, epidemiological edition; K-SADS-PL = Kiddies-schedule for affective disorders and schizophrenia for school-aged children present and life time; MDD = Major depressive disorder; NA = Not applicable; NOS = Not otherwise specified; ROC-AUC = Receiver operating characteristic area – under curve; SD = Standard deviation

**Table 3.4.13** P-GBI as screening instrument for mania 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
Youngstrom et al 2008 [95] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Outpatient psychiatric research clinic</p> <p><u>Population</u> n=637 consecutive outpatients presenting for treatment</p> <p><u>Exclusion criteria</u> Psychiatric disorder due to medical condition, pervasive developmental disorder, mental retardation</p> <p><u>Prevalence BPSD</u> 46% (294/637)</p>	<p><u>Index test</u> P-GBI</p> <p><u>Reference test</u> K-SADS</p> <p><u>Number of patients</u> n=637 (39% girls) Mean age: 11.3 years (SD 3.3)</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of index test</u> Parents</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater of reference test</u> Child and adolescents psychiatrist, 17 research assistants (bachelor or master level)</p> <p><u>Rater training, reference test</u> Highly trained. Every 10th interview was done by 2 raters</p> <p><u>Interobserver reliability</u> Overall <math>\kappa &gt; 0.85</math></p>	<p><u>28 item ROC-AUC</u> 0.83</p> <p><u>10 item ROC-AUC</u> 0.86</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> 72% of all parents answered all the P-GBI items. 98% of the items were complete. Drop-out analyses provided</p>

The table continues on the next page

Table 3.4.13 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
Youngstrom et al 2004 [93] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Research setting, child psychiatric clinic, outpatients</p> <p><u>Population</u> Participants from several pharmacotherapy studies</p> <p>Children: n=318 Mean age: 8.3 years (SD 1.6)</p> <p>Youths: n=324 Mean age: 14.1 years (SD 1.9)</p> <p><u>Inclusion criteria</u> Age: 5–17 years Both the child and the primary caregiver presented for assessment</p> <p><u>Exclusion criteria</u> Not speaking English, pervasive developmental disorder and/or autism or profound mental retardation</p> <p><u>Prevalence BPSD</u> 50% (160/318) age 5–10 41% (132/324) age 11–17</p>	<p><u>Index test</u> P-GBI, A-GBI</p> <p><u>Reference test</u> K-SADS</p> <p><u>Number of patients</u> n=642 (62% girls)</p> <p><u>Drop-out rate</u> A-GBI: 23.5% P-GBI: 4.7%</p>	<p><u>Rater, index test</u> Parents or self rating</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater, reference test</u> Research assistants</p> <p><u>Rater training, reference test</u> Highly trained</p> <p><u>Interobserver reliability</u> Five K-SADS interviews were initially re-rated by an experienced rater and at every 10th interview <math>\kappa &gt; 0.85</math></p>	<p><u>ROC-AUC</u> P-GBI Children 0.81</p> <p>P-GBI Youth 0.84</p> <p>A-GBI Youth 0.67</p>	<p>High</p> <p><u>Comments</u> Sampling method</p> <p>Blinding Satisfactory</p> <p>Handling of missing data Multiple-imputation procedure</p>

The table continues on the next page

Table 3.4.13 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
Findling et al 2002 [96] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Child and adolescent psychiatric outpatient clinic</p> <p><u>Population</u> n=196 (30% girls) consecutively admitted Age: 5–17 years Mean age: 10.3 years (SD 3.1)</p> <p><u>Inclusion criteria</u> Consent</p> <p><u>Exclusion criteria</u> Psychiatric disorder due to medical condition, a pervasive developmental disorder, or mental retardation</p> <p><u>Prevalence BPSD</u> 36% (70/196)</p>	<p><u>Index test</u> GBI</p> <p><u>Reference test</u> K-SADS-E</p> <p><u>Number of patients</u> n=196</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater, index test</u> Parents or self rating</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater, reference test</u> Child and adolescents psychiatrist or 11 research assistants</p> <p><u>Rater training, reference test</u> Research assistants trained to achieve an overall <math>\kappa &gt; 0.85</math></p> <p><u>Interobserver reliability</u> Overall <math>\kappa &gt; 0.85</math></p>	<p><u>P-GBI cut off 17</u> Sensitivity: 91% Specificity: 68%</p> <p><u>ROC-AUC</u> A-GBI 0.82</p> <p><u>P-GBI</u> 0.88</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> Unclear</p>

A-GBI = Adolescent general behavior inventory; BPSD = Bipolar spectrum disorders; GBI = General behavior inventory; K-SADS = Kiddies-schedule for affective disorders and schizophrenia; K-SADS-E = Kiddie-schedule for affective disorders and schizophrenia, epidemiological version; NA = Not applicable; P-GBI = Parent general behavior inventory; ROC-AUC = Receiver operating characteristic – area under curve; SD = Standard deviation



**Table 3.4.14** Child Mania Rating Scale-Parent version as screening instrument for bipolar syndromes 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
Henry et al 2008 [103] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Research child psychiatric clinic</p> <p><u>Population</u> 186 children with BPSD, ADHD or healthy controls</p> <p><u>Inclusion criteria</u> BPSD I and II and NOS, ADHD, or HC</p> <p><u>Exclusion criteria</u> Head injury, epilepsy, or mental retardation, medical illness, medications or substance that could alter their moods</p> <p><u>Prevalence BPSD</u> 33% (50/150)</p>	<p><u>Index test</u> CMRS-P-brief (11 item)</p> <p><u>Reference test</u> WASH-U-KADS</p> <p><u>Number of patients</u> n=150 (42% girls) Mean age: 10.3±2.9 years (range 5 to 17)</p> <p><u>Drop-out rate</u> None</p>	<p><u>Rater</u> Child psychiatrist, one doctoral-level psychiatric nurse, two post-master's degree-level psychology graduates</p> <p><u>Rater training</u> 6 months experience and training, but no description of how. Monthly reliability checks</p> <p><u>Interobserver reliability</u> κ: 0.98 to 1.0</p>	<p><u>CMRS-P brief cut off 10</u> BPSD vs ADHD Sensitivity: 84% Specificity: 83%</p> <p>BPSD vs HC Sensitivity: 92% Specificity: 82%</p>	<p>High</p> <p><u>Comments</u> Sampling method Adequate</p> <p>Blinding Satisfactory</p> <p>Handling of missing data Unclear</p> <p>Other comments The same subjects as in the article by Pavuluri [102], see below</p>

The table continues on the next page

Table 3.4.14 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	Sensitivity Specificity	Study quality Comments
Pavuluri et al 2006 [102] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Research setting child psychiatric clinic</p> <p><u>Population</u> 186 children with BP, ADHD or healthy control</p> <p><u>Inclusion criteria</u> Bipolar disorder I and II and NOS, ADHD, or HC</p> <p><u>Exclusion criteria</u> Head injury, epilepsy, or mental retardation, medical illness, medications or substance that could alter their moods</p> <p><u>Prevalence BPSD</u> 33% (50/150)</p>	<p><u>Index test</u> CMRS-P (21 item)</p> <p><u>Reference test</u> WASH-U-KADS</p> <p><u>Number of patients</u> n=150 (42% girls) Mean age: 10.3±2.9 years (range 5 to 17)</p> <p><u>Drop-out rate</u> None</p>	<p><u>Rater, index test</u> Self rating</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater, reference test</u> Child psychiatrist, one doctoral-level psychiatric nurse, two post-master's degree-level psychology graduates</p> <p><u>Rater training, reference test</u> 6 months experience and training, but no description of how. Monthly reliability checks</p> <p><u>Interobserver reliability</u> κ 0.98 to 1.0</p>	<p><u>Cut off score 20</u> BPSD vs ADHD Sensitivity: 82% Specificity: 94%</p> <p>BPSD vs HC Sensitivity: 84% Specificity: 98%</p>	<p>High</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> Unclear</p> <p><u>Other comments</u> The same subjects as in the article by Henry [103], see above</p>

ADHD = Attention deficit/hyperactivity disorder; BP = Bipolar disorder; BPSD = Bipolar spectrum disorders; CMRS-P= Child mania rating scale – parents; HC = Healthy control; NOS = Not otherwise specified; WASH-U-KADS = Washington University in St-Louis Kiddie schedule for affective disorders and schizophrenia; NA = Not applicable

**Table 3.4.15** YMRS, Young Mania Rating Scale, as screening instrument for bipolar syndromes 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
Youngstrom et al 2005 [109] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Two outpatient facilities; CMHC and AC</p> <p><u>Population</u> Random samples of patients age 5–17 CMHC: n=109 (45% girls) Mean age: 11.1 years (3.1)</p> <p>AC: n=153 (32% girls) Mean age: 10.5 years(3.5)</p> <p><u>Inclusion criteria</u> 5–18 year old, both youth and caregiver had to be English speaking</p> <p><u>Exclusion criteria</u> Pervasive developmental disorder, autism and mental retardation</p> <p><u>Prevalence</u> 34% (89/262) for BP I and II, NOS and Cyclothymia</p>	<p><u>Index test</u> A-YMRS P-YMRS</p> <p><u>Reference test</u> K-SADS</p> <p><u>Number of patients</u> n=262</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater, reference test</u> 4 predoctoral interns, 6 researchers in psychology</p> <p><u>Rater training</u> Trained until they achieved interrater agreement of <math>\kappa &gt; 0.85</math>. Monthly joint rating sessions</p> <p><u>Interobserver reliability</u> BPSD <math>\kappa</math> 0.95 and <math>\kappa</math> 0.91 for all diagnoses</p>	<p><u>ROC-AUC</u> P-YMRS Children 0.66 (95% CI, 0.57; 0.76)</p> <p>P-YMRS Youth 0.70 (95% CI, 0.59; 0.81)</p> <p>A-YMRS Youth 0.50 (95% CI, 0.37; 0.63)</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p>Blinding Satisfactory</p> <p>Handling of missing data Unclear</p> <p>Other comments Number of patients varies between 261 and 265 in different tables</p>

The table continues on the next page

Table 3.4.15 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
Youngstrom et al 2003 [107] USA	<u>Design</u> Cross sectional	<u>Index test</u> YMRS P-YMRS	<u>Rater, index test</u> Parents or self rating	<u>P-YMRS cut off 17</u> Sensitivity: 79% Specificity: 78%	High  <u>Comments</u> Sampling method Unclear
Gracious et al 2002 [106] USA	<u>Setting</u> Outpatient research clinic specializing in mood disorders  <u>Population</u> 117 children and adolescent and their parents participating in research  <u>Inclusion criteria</u> Age: 5–17 years  <u>Exclusion criteria</u> Pervasive developmental disorder, a psychiatric disorder due to a medical condition, mental retardation  <u>Prevalence BPSD</u> 50% (58/117)	<u>Reference test</u> K-SADS  <u>Number of patients</u> n=117 (45.4% girls) Mean age: 12.4±3.1 years  <u>Drop-out rate</u> Not reported	<u>Rater, reference test</u> Research assistants with either bachelor's (n=4) or master's level degrees (n=3)  <u>Rater training</u> Trained to reach an overall $\kappa > 0.85$  <u>Interobserver reliability</u> $\kappa > 0.85$	<u>P-YMRS ROC-AUC</u> 0.85	Blinding Satisfactory  <u>Handling of missing data</u> Unclear

The table continues on the next page

Table 3.4.15 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
Frazier et al 2007 [108] USA	<p><u>Design</u> Cross sectional Two different cohort studies</p> <p><u>Setting</u> Research clinic</p> <p><u>Population</u> 1 014 children, either child psychiatric patients from treatment studies or children with bipolar parents or children recruited by fliers and word of mouth</p> <p>Mean age total sample: 11.11 years (3.51), 62% males, 38% females</p> <p><u>Inclusion criteria</u> Age 4 to 17 years, and both caregivers present for assessment</p> <p><u>Exclusion criteria</u> Oral difficulties, pervasive development disorder, autism, mental retardation or if the child had a current or past medical or neurological history</p> <p><u>Prevalence</u> 53% (476/1 014) bipolar diagnosis</p>	<p><u>Index test</u> YMRS</p> <p><u>Reference test</u> K-SADS</p> <p><u>Number of patients</u> n=1 014</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater, index text</u> Self rating</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater</u> From bachelor's degrees to doctoral level</p> <p><u>Rater training</u> Trained to criterion by completing ratings independently of an experienced rater over the course of three to five interviews</p> <p>New raters led 3 to 5 K-SADS interviews with an experienced rater</p> <p><u>Interobserver reliability</u> Interrater <math>\kappa \geq 0.95</math></p>	<p><u>YMRS ROC-AUC</u> 0.97</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> NA</p> <p><u>Handling of missing data</u> NA</p>

The table continues on the next page

Table 3.4.15 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
Youngstrom et al 2004 [93] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Research setting, child psychiatric clinic, outpatients</p> <p><u>Population</u> Participants from pharmacotherapy studies</p> <p>Children: n=318 Mean age: 8.3 years (SD 1.6)</p> <p>Youths: n=324 Mean age: 14.1 years (SD 1.9)</p> <p><u>Inclusion criteria</u> Age: 5–17 years Both the child and the primary caregiver presented for assessment</p> <p><u>Exclusion criteria</u> Not speaking English, pervasive developmental disorder and/or autism or profound mental retardation</p> <p><u>Prevalence</u> 50% (160/318) age 5–10 41% (132/324) age 11–17</p>	<p><u>Index test</u> YMRS</p> <p><u>Reference test</u> K-SADS</p> <p><u>Number of patients</u> n=642</p> <p><u>Drop-out rate</u> 36%</p>	<p><u>Rater, index test</u> Self rating</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater</u> Research assistants</p> <p><u>Rater training</u> Highly trained</p> <p><u>Interobserver reliability</u> <math>\kappa &gt; 0.85</math></p>	<p><u>ROC-AUC</u> Children 0.83</p> <p>Youth 0.80</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p>Blinding Satisfactory</p>

AC = Academic medical center; A-YMRS = Adolescent young mania rating scale; BPSD = Bipolar spectrum disorder; CI = Confidence interval; CMHC = Community mental health center; K-SADS = Kiddies-schedule for affective disorders and schizophrenia; NOS = Not otherwise specified; P-YMRS = Parent version young mania rating scale; ROC-AUC = Receiver operating characteristics – area under curve; SD = Standard deviation; NA = Not applicable; YMRS = Young mania rating scale

**Table 3.4.16a** Sensitivity and specificity for PHQ-9 as screening instrument for depression 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
Richardson et al 2010 [75] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Health Research Institute health care facility</p> <p><u>Population</u> n=4 000 youths who had seen a health care provider at least once during the last year were mailed the PHQ-2. A subset of 499 youths were, based on their screening, invited to participate</p> <p><u>Inclusion criteria</u> Completed the PHQ-9 screen</p> <p><u>Exclusion criteria</u> –</p> <p><u>Prevalence MDD</u> 4%</p>	<p><u>Index test</u> PHQ-9</p> <p><u>Reference test</u> DISC-IV</p> <p><u>Number of patients</u> n=444 Mean age: 15.3±1.2 years</p> <p><u>Drop-out rate</u> 0.4%</p>	<p><u>Rater, index test</u> Self rating</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater, reference test</u> Not reported</p> <p><u>Rater training, reference test</u> All received 12 hours of classroom and hand-on training and additional project-specific training on the DISC-IV</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>MDD cut off <math>\geq 11</math></u> Sensitivity: 90% Specificity: 78%</p>	<p>Low</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> 17 incompletely filled in forms were excluded</p> <p><u>Other comments</u> The time between index and reference test is unclear. Test reproducibility was not described</p>

DISC-IV = Diagnostic interview schedule for children, fourth edition; MDD = Major depressive disorder; PHQ = Patient health questionnaire; ROC-AUC = Receiver operating characteristic – area under curve

**Table 3.4.16b** continued Sensitivity and specificity for CBQ as screening instrument for bipolar syndromes 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
Papolos et al 2006 [99] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Online research program about juvenile bipolar disorder</p> <p><u>Population</u> 5 120 children and adolescents aged 5–17 years participating in an online investigation</p> <p><u>Inclusion criteria</u> BPSD, ADHD or no psychiatric disorder</p> <p><u>Exclusion criteria</u> Not reported</p> <p><u>Prevalence BPSD</u> 56%</p>	<p><u>Index test</u> CBQ Scoring algorithm for DSM-IV BPSD</p> <p><u>Reference test</u> K-SADS</p> <p><u>Number of patients</u> n=135 self selected, no information about age or gender</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater, index test</u> Self rating</p> <p><u>Rater training, index test</u> NA</p> <p><u>Rater, reference test</u> Four graduate level interviewers</p> <p><u>Rater training, reference test</u> Trained by the project director</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>Cut off according to scoring algorithm for DSM-IV BP</u> Sensitivity: 76% Specificity: 97%</p>	<p>Low</p> <p><u>Comments</u> Sampling method Not adequate</p> <p><u>Blinding</u> Unclear</p> <p><u>Handling of missing data</u> Unclear</p> <p><u>Other comments</u> The time between index and reference test is unclear</p>

ADHD = Attention deficit/hyperactivity disorder; BP = Bipolar disorder; BPSD = Bipolar spectrum disorder; CBQ = Child bipolar questionnaire; DSM-IV = Diagnostic and statistical manual of mental disorder, fourth edition; K-SADS = Kiddie-schedule for affective disorders and schizophrenia for school-age children; NA = Not applicable



**Table 3.4.16c** continued MDQ as screening instrument for bipolar syndromes 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
Wagner et al 2006 [101] USA	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Four outpatient child-psychiatric clinics</p> <p><u>Population</u> 104 adolescents seen for evaluation and treatment</p> <p><u>Inclusion criteria</u> Age: 12–17 years</p> <p><u>Exclusion criteria</u> Not reported</p> <p><u>Prevalence BPSD</u> 39% (41/104)</p>	<p><u>Index test</u> MDQ-A, self-report, parent report, attributional person report</p> <p><u>Reference test</u> K-SADS</p> <p><u>Number of patients</u> n=104 (56% girls) Mean age: 14.5 years (SD 1.6)</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater</u> Not reported</p> <p><u>Rater training</u> Not reported</p> <p><u>Interobserver reliability</u> Not presented</p>	<p><u>ROC-AUC</u> Parent report: 0.80 Self-report: 0.57 Attributional person report: 0.58</p>	<p>Low</p> <p><u>Comments</u> <i>Sampling method</i> Not clearly described</p> <p><i>Blinding</i> Satisfactory</p> <p><i>Handling of missing data</i> Unclear</p> <p><i>Other comments</i> The time between index and reference test is unclear</p>

BPSD = Bipolar spectrum disorders; K-SADS = Kiddie-schedule for affective disorders and schizophrenia for school-age children; MDQ-A = Adolescent mood disorder questionnaire; ROC-AUC = Receiving operating characteristics – area under curve; SD = Standard deviation

## Referenser

1. Chambers WJ, Puig-Antich J, Hirsch M, Paez P, Ambrosini PJ, Tabrizi MA, et al. The assessment of affective disorders in children and adolescents by semistructured interview. Test-retest reliability of the schedule for affective disorders and schizophrenia for school-age children, present episode version. *Arch Gen Psychiatry* 1985;42: 696-702.
2. Carlson GA, Kashani JH, Thomas MD, Vaidya A. Comparison of two structured interviews on a psychiatrically hospitalized population of children. *J Am Acad Child Adolesc Psychiatry* 1987;26:645-648.
3. Herjanic B, Reich W. Development of a structured psychiatric interview for children: agreement between child and parent on individual symptoms. *J Abnorm Child Psychol* 1982;10:307-24.
4. Reich W. Diagnostic interview for children and adolescents (DICA). *J Am Acad Child Adolesc Psychiatry* 2000;39:59-66.
5. Boyle MH, Offord DR, Racine Y, Sanford M, Szatmari P, Fleming JE, et al. Evaluation of the Diagnostic Interview for Children and Adolescents for use in general population samples. *J Abnorm Child Psychol* 1993;21:663-81.
6. Cserel J, Fristad MA. Scaling structured interview data: A comparison of two methods. *J Am Acad Child Adolesc Psychiatry* 2001;40:341-346.
7. Sylvester CE, Hyde TS, Reichler RJ. The Diagnostic Interview for Children and Personality Inventory for Children in studies of children at risk for anxiety disorders or depression. *J Am Acad Child Adolesc Psychiatry* 1987;26:668-675.
8. Matzner F, Silva R, Silvan M, Chowdhury M, Nastasi L. "Preliminary Test-retest Reliability of the KID-SCID", Scientific Proceedings, American Psychiatric Association Meeting; 1997.
9. 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.
10. Sheehan DV, Sheehan KH, Shytle RD, Janavs J, Bannon Y, Rogers JE, et al. Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID). *J Clin Psychiatry* 2010;71:313-326.
11. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000;41:645-55.
12. Cunningham CE, Boyle MH, Hong S, Pettingill P, Bohaychuk D. The Brief Child and Family Phone Interview (BCFPI): 1. Rationale, development, and description of a computerized children's mental health intake and outcome assessment tool. *J Child Psychol Psychiatry* 2009;50:416-23.
13. Boyle MH, Cunningham CE, Georgiades K, Cullen J, Racine Y,

- Pettingill P. The Brief Child and Family Phone Interview (BCFPI): 2. Usefulness in screening for child and adolescent psychopathology. *J Child Psychol Psychiatry* 2009; 50:424-31.
14. Achenbach TM, Edelbrock CS. Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monogr Soc Res Child Dev* 1981;46:1-82.
15. Achenbach & Resorla Manual for ASEBA; 2001.
16. Larsson B, Frisk M. Social competence and emotional/behaviour problems in 6-16 year-old Swedish school children. *Eur Child Adolesc Psychiatry* 1999;8:24-33.
17. Broberg AG, Ekeröth K, Gustafsson PA, Hansson K, Hagglof B, Ivarsson T, et al. Self-reported competencies and problems among Swedish adolescents: a normative study of the YSR. *Youth Self Report. Eur Child Adolesc Psychiatry* 2001;10:186-93.
18. Renouf AG, Kovacs M. Concordance between mothers' reports and children's self-reports of depressive symptoms: a longitudinal study. *J Am Acad Child Adolesc Psychiatry* 1994;33:208-16.
19. Brunshaw JM, Szatmari P. The agreement between behaviour checklists and structured psychiatric interviews for children. *Can J Psychiatry* 1988;33:474-81.
20. Clarke GN, Lewinsohn PM, Hops H, Seeley JR. A self- and parent-report measure of adolescent depression: The Child Behavior Checklist Depression scale (CBCL-D). *Behav Assess* 1992; 14:443-63.
21. Kronenberg Y, Blumensohn R, Apter A. A comparison of different diagnostic tools for childhood depression. *Acta Psychiatrica Scand* 1988;77:194-8.
22. Sprafkin J, Gadow KD, Salisbury H, Schneider J, Loney J. Further evidence of reliability and validity of the Child Symptom Inventory-4: Parent checklist in clinically referred boys. *J Clin Child Adolesc Psychol* 2002;31:513-24.
23. Sterba S, Egger HL, Angold A. Diagnostic specificity and nonspecificity in the dimensions of preschool psychopathology. *J Child Psychol Psychiatry* 2007;48:1005-13.
24. Masi G, Brovedani P, Mucci M, Favilla L. Assessment of anxiety and depression in adolescents with mental retardation. *Child Psychiatry Hum Dev* 2002;32:227-37.
25. Hazell PL, Lewin TJ, Carr VJ. Confirmation that Child Behavior Checklist clinical scales discriminate juvenile mania from attention deficit hyperactivity disorder. *J Paediatr Child Health* 1999;35: 199-203.
26. Najman JM, Heron MA, Hayatbakhsh MR, Dingle K, Jamrozik K, Bor W, et al. Screening in early childhood for risk of later mental health problems: a longitudinal study. *J Psychiatr Res* 2008;42:694-700.
27. Gould MS, Bird H, Jaramillo BS. Correspondence between statistically derived behavior problem syndromes and child psychiatric diagnoses in a community sample. *J Abnorm Child Psychol* 1993;21:287-313.
28. Rey JM, Morris-Yates A. Diagnostic accuracy in adolescents of several depression rating scales extracted from a general

- purpose behavior checklist. *J Affect Disord* 1992;26:7-16.
29. Warnick EM, Bracken MB, Kasl S. Screening efficiency of the child behavior checklist and strengths and difficulties questionnaire: A systematic review. *Child Adolesc Ment Health* 2008;13:140-7.
30. Rey JM, Morris-Yates A. Adolescent depression and the Child Behavior Checklist. *J Am Acad Child Adolesc Psychiatry* 1991;30:423-7.
31. Rishel CW, Greeno C, Marcus SC, Shear MK, Anderson C. Use of the Child Behavior Checklist as a Diagnostic Screening Tool in Community Mental Health. *Res Soc Work Pract* 2005;15:195-203.
32. Biederman J, Faraone S, Mick E, Moore P, Lelon E. Child Behavior Checklist findings further support comorbidity between ADHD and major depression in a referred sample. *J Am Acad Child Adolesc Psychiatry* 1996;35:734-42.
33. Krol NP, De Bruyn EE, Coolen JC, van Aarle EJ. From CBCL to DSM: a comparison of two methods to screen for DSM-IV diagnoses using CBCL data. *J Clin Child Adolesc Psychol* 2006;35:127-35.
34. Mattison RE, Carlson GA, Cantwell DP, Asarnow JR. Teacher and parent ratings of children with depressive disorders. *J Emot Behav Disord* 2007;15:184-92.
35. Ferdinand RF. Validity of the CBCL/YSR DSM-IV scales anxiety problems and affective problems. *J Anxiety Disord* 2008;22:126-34.
36. Kahana SY, Youngstrom EA, Findling RL, Calabrese JR. Employing parent, teacher, and youth self-report checklists in identifying pediatric bipolar spectrum disorders: an examination of diagnostic accuracy and clinical utility. *J Child Adolesc Psychopharmacol* 2003;13:471-88.
37. Ridge NW, Warren JS, Burlingame GM, Wells MG, Tumblin KM. Reliability and validity of the youth outcome questionnaire self-report. *J Clin Psychol* 2009;65:1115-26.
38. Dingle K, Alati R, Williams GM, Najman JM, Bor W, Clavarino A. The ability of YSR DSM-oriented depression scales to predict DSM-IV depression in young adults: a longitudinal study. *J Affect Disord* 2010;121:45-51. Epub 2009 Jun 7.
39. Aebi M, Metzke CW, Steinhausen HC. Prediction of major affective disorders in adolescents by self-report measures. *J Affect Disord* 2009;115:140-9.
40. Goldin S, Hagglof B, Levin L, Persson LA. Mental health of Bosnian refugee children: a comparison of clinician appraisal with parent, child and teacher reports. *Nord J Psychiatry* 2008;62:204-16.
41. Achenbach TM, Dumenci L, Rescorla LA. DSM-oriented and empirically based approaches to constructing scales from the same item pools. *J Clin Child Adolesc Psychol* 2003;32:328-40.
42. Morgan CJ, Cauce AM. Predicting DSM-III-R disorders from the Youth Self-Report: analysis of data from a field

- study. *J Am Acad Child Adolesc Psychiatry* 1999;38:1237-45.
43. Vreugdenhil C, van den Brink W, Ferdinand R, Wouters L, Doreleijers T. The ability of YSR scales to predict DSM/DISC-C psychiatric disorders among incarcerated male adolescents. *Eur Child Adolesc Psychiatry* 2006;15:88-96.
44. Weinstein SR, Noam GG, Grimes K, Stone K, Schwab-Stone M. Convergence of DSM-III diagnoses and self-reported symptoms in child and adolescent inpatients. *J Am Acad Child Adolesc Psychiatry* 1990;29:627-34.
45. Youngstrom EA, Joseph MF, Greene J. Comparing the psychometric properties of multiple teacher report instruments as predictors of bipolar disorder in children and adolescents. *J Clin Psychol* 2008;64:382-401.
46. Geller B, Warner K, Williams M, Zimmerman B. Prepubertal and young adolescent bipolarity versus ADHD: assessment and validity using the WASH-U-KSADS, CBCL and TRF. *J Affect Disord* 1998;51:93-100.
47. Beck JS, Beck AT, Jolly J. *Manual for the Beck Youth Inventories of Emotional and social Impairment*. San Antonio, TX: Psychological Corporation 2001.
48. Thastum M, Ravn K, Sommer S, Trillingsgaard A. Reliability, validity and normative data for the Danish Beck Youth Inventories. *Scand J Psychol* 2009;50:47-54.
49. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561-71.
50. Beck AT, Steer RA, Brown GK. *Manual for the BDI-II*. San Antonio, TX: Psychological Corporation; 1996.
51. Kashani JH, Sherman DD, Parker DR, Reid JC. Utility of the Beck Depression Inventory with clinic-referred adolescents. *J Am Acad Child Adolesc Psychiatry* 1990;29:278-82.
52. Barrera M, Jr., Garrison-Jones CV. Properties of the Beck Depression Inventory as a screening instrument for adolescent depression. *J Abnorm Child Psychol* 1988;16:263-73.
53. Ambrosini PJ, Metz C, Bianchi MD, Rabinovich H, Undie A. Concurrent validity and psychometric properties of the Beck Depression Inventory in outpatient adolescents. *J Am Acad Child Adolesc Psychiatry* 1991;30:51-7.
54. Marton P, Churchard M, Kutcher S, Korenblum M. Diagnostic utility of the Beck Depression Inventory with adolescent psychiatric outpatients and inpatients. *Can J Psychiatry* 1991;36:428-31.
55. Subramaniam G, Harrell P, Huntley E, Tracy M. Beck Depression Inventory for depression screening in substance-abusing adolescents. *J Subst Abuse Treat* 2009;37:25-31.
56. Birlerson P. The validity of depressive disorder in childhood and the development of a self-rating scale: a research report. *J Child Psychol Psychiatry* 1981;22:73-88.
57. Ivarsson T, Lidberg A, Gillberg C. The Birlerson Depression Self-Rating Scale (DSRS): Clinical evaluation in an adolescent inpatient population. *J Affect Disord* 1994;32:115-25.

58. Ivarsson T, Gillberg C. Depressive symptoms in Swedish adolescents: normative data using the Birleson Depression Self-Rating Scale (DSRS). *J Affect Disord* 1997;42:59-68.
59. Fundudis T, Berney TP, Kolvin I, Famuyiwa OO, Barrett L, Bhate S, et al. Reliability and validity of two self-rating scales in the assessment of childhood depression. *Br J Psychiatry Suppl* 1991: 36-40.
60. Kovacs M. The Children's Depression Inventory (CDI). *Psychopharmacol Bull* 1985;21:995-8.
61. Timbremont B, Braet C, Dreesen L. Assessing depression in youth: relation between the Children's Depression Inventory and a structured interview. *J Clin Child Adolesc Psychol* 2004;33:149-57.
62. Nguyen N, Whittlesey S, Scimeca K, DiGiacomo D, Bui B, Parsons O, et al. Parent-child agreement in prepubertal depression: findings with a modified assessment method. *J Am Acad Child Adolesc Psychiatry* 1994;33:1275-83.
63. Craighead WE, Curry JF, Ilardi SS. Relationship of Children's Depression Inventory factors to major depression among adolescents. *Psychol Assess* 1995;7:171-6.
64. Kazdin AE, Colbus D, Rodgers A. Assessment of depression and diagnosis of depressive disorder among psychiatrically disturbed children. *J Abnorm Child Psychol* 1986;14:499-515.
65. Lobovits DA, Handal PJ. Childhood depression: Prevalence using DSM-III criteria and validity of parent and child depression scales. *J Pediatr Psychol* 1985; 10:45-54.
66. Comer JS, Kendall PC. High-end specificity of the children's depression inventory in a sample of anxiety-disordered youth. *Depress Anxiety* 2005; 22:11-9.
67. Shemesh E, Yehuda R, Rockmore L, Shneider BL, Emre S, Bartell AS, et al. Assessment of depression in medically ill children presenting to pediatric specialty clinics. *J Am Acad Child Adolesc Psychiatry* 2005;44:1249-57.
68. Sorensen MJ, Frydenberg M, Thastum M, Thomsen PH. The Children's Depression Inventory and classification of major depressive disorder: validity and reliability of the Danish version. *Eur Child Adolesc Psychiatry* 2005;14:328-34.
69. Weissman MM, Orvaschel H, Padian N. Children's symptom and social functioning self-report scales. Comparison of mothers' and children's reports. *J Nerv Ment Dis* 1980;168:736-40.
70. Olsson G, von Knorring AL. Depression among Swedish adolescents measured by the self-rating scale Center for Epidemiology Studies-Depression Child (CES-DC). *Eur Child Adolesc Psychiatry* 1997;6:81-7.
71. Fendrich M, Weissman MM, Warner V. Screening for depressive disorder in

- children and adolescents: validating the Center for Epidemiologic Studies Depression Scale for Children. *Am J Epidemiol* 1990;131:538-51.
72. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; 16:606-13.
73. Richardson LP, Rockhill C, Russo JE, Grossman DC, Richards J, McCarty C, et al. Evaluation of the PHQ-2 as a brief screen for detecting major depression among adolescents. *Pediatrics* 2010;125: e1097-103.
74. Kumar G, Kim AH, Krefetz D, Steer RA. Screening for major depressive disorders in adolescent psychiatric inpatients with the mood modules from the Primary Care Evaluation of Mental Disorders and the Patient Health Questionnaire. *Psychol Rep* 2001;89:274-8.
75. Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, et al. Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents. *Pediatrics* 2010;126:1117-23. Epub 201 Nov 1.
76. Malmberg M, Rydell AM, Smedje H. Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nord J Psychiatry* 2003;57: 357-63.
77. Smedje H, Broman JE, Hetta J, von Knorring AL. Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". *Eur Child Adolesc Psychiatry* 1999;8:63-70.
78. Indredavik MS, Vik T, Heyerdahl S, Kulseng S, Brubakk A-M. Psychiatric symptoms in low birth weight adolescents, assessed by screening questionnaires. *Eur Child Adolesc Psychiatry* 2005;14:226-36.
79. Vogels AG, Crone MR, Hoekstra F, Reijneveld SA. Comparing three short questionnaires to detect psychosocial dysfunction among primary school children: a randomized method. *BMC Public Health* 2009;9:489.
80. Goodman R, Ford T, Corbin T, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *Eur Child Adolesc Psychiatry*, 2004;13:Suppl 2:II25-31.
81. Blom EH, Larsson JO, Serlachius E, Ingvar M. The differentiation between depressive and anxious adolescent females and controls by behavioural self-rating scales. *J Affect Disord* 2010;122:232-40.
82. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Int Rev Psychiatry* 2003;15:166-72.
83. Biederman J, Petty CR, Monuteaux MC, Evans M, Parcell T, Faraone SV, et al. The Child Behavior Checklist-Pediatric Bipolar Disorder profile predicts a subsequent diagnosis of bipolar disorder and associated impairments in ADHD youth growing up: A longitudinal analysis. *J Clin Psychiatry* 2009;70:732-40.

84. Diler RS, Birmaher B, Axelson D, Goldstein B, Gill M, Strober M, et al. The Child Behavior Checklist (CBCL) and the CBCL-bipolar phenotype are not useful in diagnosing pediatric bipolar disorder. *J Child Adolesc Psychopharmacology* 2009;19:23-30.
85. Faraone SV, Althoff RR, Hudziak JJ, Monuteaux M, Biederman J. The CBCL predicts DSM bipolar disorder in children: A receiver operating characteristic curve analysis. *Bipolar Disord* 2005;7:518-24.
86. Giles LL, DelBello MP, Stanford KE, Strakowski SM. Child Behavior Checklist profiles of children and adolescents with and at high risk for developing bipolar disorder. *Child Psychiatry Hum Dev* 2007;38:47-55.
87. Volk HE, Todd RD. Does the Child Behavior Checklist juvenile bipolar disorder phenotype identify bipolar disorder? *Biol Psychiatry* 2007;62:115-20.
88. Rockhill CM, Russo JE, McCauley E, Katon WJ, Richardson LP, Lozano P. Agreement between parents and children regarding anxiety and depression diagnoses in children with asthma. *J Nerv Ment Dis* 2007;195:897-904.
89. Mick E, Biederman J, Pandina G, Faraone SV. A preliminary meta-analysis of the child behavior checklist in pediatric bipolar disorder. *Biol Psychiatry* 2003;53:1021-7.
90. Biederman J, Wozniak J, Kiely K, Ablon S, Faraone S, Mick E, et al. CBCL clinical scales discriminate pre-pubertal children with structured interview-derived diagnosis of mania from those with ADHD. *J Am Acad Child Adolesc Psychiatry* 1995;34:464-71.
91. Jensen PS, Watanabe HK, Richters JE, Roper M, Hibbs ED, Salzberg AD, et al. Scales, diagnoses, and child psychopathology: II. Comparing the CBCL and the DISC against external validators. *J Abnorm Child Psychol* 1996;24:151-68.
92. Youngstrom E, Meyers O, Youngstrom JK, Calabrese JR, Findling RL. Comparing the Effects of Sampling Designs on the Diagnostic Accuracy of Eight Promising Screening Algorithms for Pediatric Bipolar Disorder. *Biological Psychiatry* 2006;60:1013-9.
93. Youngstrom EA, Findling RL, Calabrese JR, Gracious BL, Demeter C, Bedoya DD, et al. Comparing the diagnostic accuracy of six potential screening instruments for bipolar disorder in youths aged 5 to 17 years. *J Am Acad Child Adolesc Psychiatry* 2004;43:847-58.
94. Klein DN, Dickstein S, Taylor EB, Harding K. Identifying chronic affective disorders in outpatients: validation of the General Behavior Inventory. *J Consult Clin Psychol* 1989;57:106-11.
95. Youngstrom EA, Frazier TW, Demeter C, Calabrese JR, Findling RL. Developing a 10-item mania scale from the Parent General Behavior Inventory for children and adolescents. *J Clin Psychiatry* 2008;69:831-9.
96. Findling RL, Youngstrom EA, Danielson CK, DelPorto-Bedoya D, Papish-David R, Townsend L, et al. Clinical decision-making using the General Behavior Inventory in juvenile bipolarity. *Bipolar Disord* 2002;4:34-42.
97. Reichart CG, van der Ende J, Wals M, Hillegers MH, Ormel J, Nolen WA, et al.



- The use of the GBI in a population of adolescent offspring of parents with a bipolar disorder. *J Affect Disord* 2004;80:263-7.
98. Reichart CG, van der Ende J, Wals M, Hillegers MH, Nolen WA, Ormel J, et al. The use of the GBI as predictor of bipolar disorder in a population of adolescent offspring of parents with a bipolar disorder. *J Affect Disord* 2005;89:147-55.
99. Papolos D, Hennen J, Cockerham MS, Thode HC, Jr., Youngstrom EA. The child bipolar questionnaire: a dimensional approach to screening for pediatric bipolar disorder. *J Affect Disord* 2006;95:149-58.
100. 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.
101. Wagner KD, Hirschfeld RM, Emslie GJ, Findling RL, Gracious BL, Reed ML. Validation of the Mood Disorder Questionnaire for bipolar disorders in adolescents. *J Clin Psychiatry* 2006;67:827-30.
102. Pavuluri MN, Henry DB, Devineni B, Carbray JA, Birmaher B. Child mania rating scale: development, reliability, and validity. *J Am Acad Child Adolesc Psychiatry* 2006;45:550-60.
103. Henry DB, Pavuluri MN, Youngstrom E, Birmaher B. Accuracy of brief and full forms of the Child Mania Rating Scale. *J Clin Psychol* 2008;64:368-81.
104. West AE, Celio CI, Henry DB, Pavuluri MN. Child Mania Rating Scale-Parent Version: A valid measure of symptom change due to pharmacotherapy. *J Affect Disord* 2011;128:112-9.
105. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 1978;133:429-35.
106. Gracious BL, Youngstrom EA, Findling RL, Calabrese JR. Discriminative validity of a parent version of the Young Mania Rating Scale. *J Am Acad Child Adolesc Psychiatry* 2002;41:1350-9.
107. Youngstrom EA, Gracious BL, Danielson CK, Findling RL, Calabrese J. Toward an integration of parent and clinician report on the Young Mania Rating Scale. *J Affect Disord* 2003;77:179-90.
108. Frazier TW, Demeter CA, Youngstrom EA, Calabrese JR, Stansbrey RJ, McNamara NK, et al. Evaluation and comparison of psychometric instruments for pediatric bipolar spectrum disorders in four age groups. *J Child Adolesc Psychopharmacol* 2007;17:853-66.
109. Youngstrom E, Meyers O, Demeter C, Youngstrom J, Morello L, Piiparinen R, et al. Comparing diagnostic checklists for pediatric bipolar disorder in academic and community mental health settings. *Bipolar Disord* 2005;7:507-17.
110. Ambrosini PJ. Historical development and present status of the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS).

- J Am Acad Child Adolesc Psychiatry 2000;39:49-58.
111. Renou S, Hergueta T, Flament M, Mouren-Simeoni MC, Lecrubier Y. [Diagnostic structured interviews in child and adolescent's psychiatry]. *Encephale* 2004; 30:122-34.
112. Hodges K. Structured interviews for assessing children. *J Child Psychol Psychiatry* 1993;34:49-68.
113. Socialstyrelsen. Barn- och ungdomspsykiatrins metoder – En nationell inventering; 2009.
114. Lauth B, Levy SR, Juliusdottir G, Ferrari P, Petursson H. Implementing the semi-structured interview Kiddie-SADS-PL into an in-patient adolescent clinical setting: impact on frequency of diagnoses. *Child Adolesc Psychiatry Ment Health* 2008;2:14.
115. Sørensen MJ, Thomsen PH, Bilenberg N. Parent and child acceptability and staff evaluation of K-SADS-PL: A pilot study. *Eur Child Adolesc Psychiatry* 2007;16:293-7.
116. Youngstrom E. Improving the assessment of juvenile bipolar disorder. 5R01 MH066647. Closing Report Eric Youngstrom.
117. Zimmerman M, Mattia JJ. Psychiatric diagnosis in clinical practice: is comorbidity being missed? *Compr Psychiatry* 1999; 40:182-91.

## 3.5 Bedömningsformulär för vuxna

### Evidensgraderade resultat

#### Intervjuformulär

- I psykiatrisk öppen- och slutenvård har intervjuformuläret SCID-I utan tillgång till journal en sensitivitet på 86 procent (95 % KI, 73 till 94) för att identifiera egentlig depression, med LEAD som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○) och specificitet på 92 procent (95 % KI, 88 till 95) (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- I psykiatrisk öppen- och slutenvård har intervjuformuläret SCID-I utan tillgång till journal en sensitivitet på 79 procent (95 % KI, 66 till 88) för att identifiera bipolärt syndrom, med LEAD som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 94 procent (95 % KI, 90 till 97) (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- I psykiatrisk öppen- och slutenvård har intervjuformuläret SCID-I utan tillgång till journal en sensitivitet på 61 procent (95 % KI, 46 till 74) för att identifiera schizoaffektiva syndrom, med LEAD som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 97 procent (95 % KI, 94 till 99) (starkt vetenskapligt underlag ⊕⊕⊕⊕).
- Det finns för få studier med tillräcklig kvalitet för att bedöma om de semistrukturerade intervjuerna SADS och SCAN har tillräcklig tillförlitlighet för att kunna användas som stöd för att ställa en förstärkningsdiagnos med SCID-I som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att bedöma om de strukturerade intervjuerna CIDI, DIS och SPIFA är tillförlitliga för att ställa diagnosen egentlig depression med strukturerad eller semistrukturerad intervju som referensstandard i kliniska populationer (otillräckligt vetenskapligt underlag ⊕○○○).

- I psykiatrisk öppenvård har MINI en sensitivitet på 95 procent (95 % KI, 93 till 97) och en specificitet på 84 procent för depression (95 % KI, 80 till 87) med strukturerad eller semistrukturerad intervju som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- I psykiatrisk öppenvård har MINI en sensitivitet för pågående mani på 83 procent (95 % KI, 71 till 92) jämfört med strukturerad eller semistrukturerad intervju (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 95 procent (95 % KI, 93 till 97) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- I psykiatrisk öppenvård har MINI en sensitivitet för mani någon gång under livstiden en sensitivitet på 79 procent (95 % KI, 71 till 86) jämfört med strukturerad eller semistrukturerad intervju (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 93 procent (95 % KI, 90 till 95) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- I primärvård, psykiatrisk öppenvård och onkologisk vård har PRIME-MD en sensitivitet på 70 procent (95 % KI, 63 till 70) (begränsat vetenskapligt underlag; ⊕⊕○○) för ”något förstämningssyndrom” och en specificitet på 85 procent (95 % KI, 82 till 88) med strukturerad eller semistrukturerad intervju som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- I primärvård, psykiatrisk öppenvård och onkologisk vård har PRIME-MD en sensitivitet under 70 procent för depression (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 85 procent (95 % KI, 82 till 88) med strukturerad eller semistrukturerad intervju som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

### **Riktad screening av depression**

- För riktad screening av vuxna personer med egentlig depression i primärvård, neuropsykiatrisk öppenvård och onkologisk vård har BDI-II med tröskelvärdet 14 en sensitivitet på 90 procent (95 % KI, 83 till 95) jämfört med strukturerad eller semistrukturerad intervju (begränsat vetenskapligt underlag ⊕⊕○○). Det går inte att bedöma

specificiteten eftersom studierna inte är tillräckligt samstämmiga (otillräckligt vetenskapligt underlag ⊕○○○).

- Det finns för få studier med tillräcklig kvalitet för att bedöma sensitivitet och specificitet för CDSS med tröskelvärdet 7 jämfört med strukturerad eller semistrukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○). Det saknas vetenskapligt underlag som stödjer att tröskelvärdet sju är det optimala för att identifiera egentlig depression hos vuxna patienter med schizofreni.
- Det finns för få studier av tillräcklig kvalitet för att bedöma sensitivitet och specificitet för CES-D med semistrukturerad eller strukturerad intervju som referens med tröskelvärdet 16 (otillräckligt vetenskapligt underlag ⊕○○○).
- Under första året efter förlossningen har EPDS med tröskelvärde 12/13 en sensitivitet på 72 procent (95 % KI, 69 till 76) för depression hos kvinnor med SCID-I som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 88 procent (95 % KI, 86 till 89) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Det går inte att bedöma sensitivitet och specificitet för depression hos gravida kvinnor med EPDS, med strukturerade eller semistrukturerade intervjuer som referensstandard, eftersom det finns för få studier med tillräcklig kvalitet (otillräckligt vetenskapligt underlag ⊕○○○).
- Det går inte att bedöma sensitivitet och specificitet för HADS som formulär för riktad screening för depression hos vuxna med strukturerad eller semistrukturerad intervju som referensstandard eftersom resultaten har för stor spridning (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Det går inte att bedöma sensitiviteten för PHQ-9, baserat på algoritm, som formulär för riktad screening av depression hos vuxna jämfört med strukturerad eller semistrukturerad intervju eftersom studierna har för stor spridning (måttligt starkt underlag ⊕⊕⊕○). Specificiteten

i ett brett antal populationer är 93 procent (95 % KI, 92 till 93) jämfört med strukturerad eller semistrukturerad intervju (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

- Det går inte att bedöma sensitiviteten för PHQ-9, med tröskelvärdet 10, som formulär för riktad screening av depression hos vuxna jämfört med strukturerad eller semistrukturerad intervju eftersom studierna har för stor spridning (måttligt starkt vetenskapligt underlag ⊕⊕⊕○). Specificiteten i ett brett antal populationer är 84 procent (95 % KI, 83 till 86) jämfört med strukturerad eller semistrukturerad intervju (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Det går inte att bedöma sensitivitet och specificitet för MDI som formulär för riktad screening för depression och med strukturerad eller semistrukturerad intervju som referensstandard eftersom det saknas studier.
- Det finns för få studier av tillräcklig kvalitet för att kunna bedöma sensitivitet och specificitet för SHQ och SPAQ som formulär för riktad screening av årstidsbunden depression med semistrukturerad eller strukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **Riktad screening för bipolära syndrom**

- För vuxna i specialiserad vård för affektiva syndrom och i psykiatrisk öppenvård har HCL-32 med tröskelvärdet 14 poäng en sensitivitet på 82 procent (95 % KI, 78 till 85) för bipolärt syndrom, med SCID-I som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○). Det går inte att bedöma specificiteten pga stor spridning i resultaten (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier för att kunna bedöma om HCL-32 med tröskelvärdet 14 poäng är tillförlitligt för bipolärt syndrom typ II gentemot unipolär depression (otillräckligt vetenskapligt underlag ⊕○○○).

- För vuxna i psykiatrisk öppenvård och i specialiserad vård för affektiva syndrom, för nyblivna mödrar med tidigare affektivt syndrom och för patienter med psykiska trauman i primärvård har MDQ med tröskelvärde 7 poäng och kriteriet minst måttlig funktionsnedsättning en sensitivitet på 72 procent (95 % KI, 68 till 76) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○), för bipolärt syndrom, med SCID-I som referensstandard. Specificiteten är 78 procent (95 % KI, 76 till 80) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- För vuxna i psykiatrisk öppenvård och i specialiserad vård för affektiva syndrom har MDQ med tröskelvärde 7 poäng och modifierat kriterium minst mild funktionsnedsättning en sensitivitet på 78 procent (95 % KI, 71 till 84) (begränsat vetenskapligt underlag ⊕⊕○○) och specificitet på 79 procent (95 % KI, 76 till 82) för bipolärt syndrom, med SCID-I som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- För vuxna i psykiatrisk öppenvård och i specialiserad vård för affektiva syndrom har MDQ med tröskelvärde 7 poäng och kriteriet minst måttlig funktionsnedsättning en specificitet på 85 procent (95 % KI, 82 till 88) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○) för bipolärt syndrom typ II, med SCID-I som referensstandard. Det går inte att bedöma sensitiviteten eftersom studierna har för låg precision (otillräckligt vetenskapligt underlag ⊕○○○).
- För vuxna i psykiatrisk öppenvård och i specialiserad vård för affektiva syndrom har MDQ med tröskelvärde 7 poäng och modifierat kriterium minst mild funktionsnedsättning en specificitet på 81 procent (95 % KI, 78 till 84) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○) för bipolärt syndrom typ II, med SCID-I som referensstandard. Det går inte att bedöma sensitiviteten eftersom studierna har för låg precision (otillräckligt vetenskapligt underlag ⊕○○○).

## Bedömning av svårighetsgrad

- Det finns för få relevanta studier för att bedöma om BDI är tillförlitligt för att mäta svårighetsgrad av depression med CGI-S eller DSM-IV-kriterier som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att bedöma om CDSS är tillförlitligt för att mäta svårighetsgrad av depression hos patienter med schizofreni med CGI-S eller DSM-IV-kriterier som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier för att bedöma om HDRS-17 är tillförlitligt för att mäta svårighetsgrad av depression med DSM-IV-kriterier eller CGI-S som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att kunna bedöma om någon av de olika versionerna av IDS är likvärdig med DSM-IV-kriterier eller CGI-S för att uppskatta depressionsdjup vid egentlig depression (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att kunna bedöma om MADRS är tillförlitligt för att mäta svårighetsgrad av depression med SCID-I eller CGI-S som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det går inte att bedöma om olika versioner av Zung (SDS) är tillförlitliga för att mäta svårighetsgrad av depression med SCID-I eller CGI-S som referensstandard eftersom det saknas studier.
- Det finns för få studier med tillräcklig kvalitet för att bedöma om AS-18 är tillförlitligt för att mäta svårighetsgrad av mani, depression och blandtillstånd av bipolärt syndrom med SCID-I eller CGI-BP som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).



- Det går inte att bedöma om BDRS är tillförlitligt för att mäta svårighetsgrad av bipolär depression med CGI-BP eller SCID-I som referensstandard eftersom det saknas studier.
- Det finns för få studier av tillräcklig kvalitet för att bedöma om MAS är tillförlitligt för att mäta svårighetsgrad av mani med CGI-BP eller SCID-I som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att kunna bedöma om YMRS är tillförlitligt för att mäta svårighetsgrad av mani med CGI-BP eller SCID-I som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).
- Det finns för få studier med tillräcklig kvalitet för att kunna bedöma om LCM-p är tillförlitligt för att följa svårighetsgrad av depression och mani med CGI-BP eller SCID-I som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **Uppföljning av behandlingseffekter**

- Det saknas studier som undersöker om formulär för att bedöma svårighetsgrad av affektiva syndrom kan användas för att följa förlopp över tid med DSM-IV-kriterier som referens.

### **Samstämmighet av skattning av depression mellan kliniker och patient**

- Överensstämmelsen mellan patient och kliniker inom psykiatrisk verksamhet för skattning av depressionssymtom med hjälp av formuläret IDS är god (ICC 0,88; 95 % KI, 0,87 till 0,90) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Överensstämmelsen mellan expert- och patientskattad MADRS, mätt fyra veckor efter behandlingstart är tillräcklig (ICC >0,7) (begränsat vetenskapligt underlag ⊕⊕○○).

## Introduktion

Syftet med utvärderingen var att undersöka den diagnostiska tillförlitligheten för intervjuformulär, formulär för riktad screening och formulär för att bedöma svårighetsgrad av affektiva syndrom mätt som sensitivitet och specificitet.

Den optimala diagnostiska proceduren är LEAD (Longitudinal, Experts, All Data procedure), ibland benämnd Best Estimate (Kapitel 1). LEAD kan därmed ses som referensstandard. Ett alternativ kan vara att använda semistrukturerade eller strukturerade intervjuformulär som referensstandard.

Avsnittet inleds med en granskning av studier som jämför den vanligast använda referensstandard SCID-I med LEAD. Därefter jämförs övriga semistrukturerade och strukturerade intervjuer som används i Sverige (enkätsvar, se Kapitel 3.1) med LEAD eller andra strukturerade/semistrukturerade intervjuer.

Sedan följer en utvärdering av formulär för riktad screening för depression respektive bipolärt syndrom. Med riktad screening avser vi ett test som bidrar till att identifiera individer där man behöver gå vidare i den diagnostiska processen för att utesluta eller säkerställa diagnos.

Det följande avsnittet redovisar granskningen av formulär för att bedöma svårighetsgrad av depression respektive bipolära syndrom. Avslutningsvis granskar vi studier som jämför expertskattningar och självskattningar för formulär som finns i båda versionerna.

De formulär som ingår i granskningen sammanfattas i Tabell 3.5.1.

**Tabell 3.5.1** Sammanfattning av resultat av granskningen.

Formulär	Syfte	Referensstandard	Evidensgraderat resultat
SCID-I	Semistrukturerad intervju Egentlig depression	LEAD/Best estimate	Sensitivitet: 86% Specificitet: 92%
SCID-I	Semistrukturerad intervju Bipolärt syndrom	LEAD/Best estimate	Sensitivitet: 79% Specificitet: 94%
SCID-I	Semistrukturerad intervju Schizoaffektivt syndrom	LEAD/Best estimate	Sensitivitet: 61% Specificitet: 97%
SADS	Semistrukturerad intervju Egentlig depression	LEAD/Best estimate	Otillräckligt
SCAN	Semistrukturerad intervju Egentlig depression	Semistrukturerad intervju	Otillräckligt
CIDI	Strukturerad intervju Egentlig depression	Strukturerad/semi-strukturerad intervju	Otillräckligt
DIS	Strukturerad intervju Egentlig depression	Strukturerad/semi-strukturerad intervju	Otillräckligt
MINI	Strukturerad intervju Egentlig depression	Strukturerad/semi-strukturerad intervju	Sensitivitet: 95% Specificitet: 84%
MINI	Strukturerad intervju Pågående mani	Strukturerad/semi-strukturerad intervju	Sensitivitet: 83% Specificitet: 95%
MINI	Strukturerad intervju Mani någon gång under livet	Strukturerad/semi-strukturerad intervju	Sensitivitet: 79% Specificitet: 93%
PRIME-MD	Strukturerad intervju Något förstämningssyndrom	Strukturerad/semi-strukturerad intervju	Sensitivitet: 70% Specificitet: 85%
PRIME-MD	Strukturerad intervju Egentlig depression	Strukturerad/semi-strukturerad intervju	Sensitivitet: <70% Specificitet: 85%
SPIFA	Strukturerad intervju Förstämningssyndrom	Strukturerad/semi-strukturerad intervju	Otillräckligt
BDI-II	Riktad screening (självskattning) Egentlig depression	Strukturerad intervju	Sensitivitet: 90% Specificitet: – Otillräckligt
CDSS	Riktad screening (intervju) Egentlig depression vid schizofreni	Strukturerad/semi-strukturerad intervju	Otillräckligt

Tabellen fortsätter på nästa sida

**Tabell 3.5.1** fortsättning

<b>Formulär</b>	<b>Syfte</b>	<b>Referensstandard</b>	<b>Evidensgraderat resultat</b>
CES-D	Riktad screening (självskattning) Egentlig depression	Strukturerad/semi-strukturerad intervju	Otillräckligt
EPDS	Riktad screening (självskattning) Post partum depression	Semistrukturerad intervju/SCID-I	Sensitivitet: 72% Specificitet: 88%
HADS	Riktad screening (självskattning) Egentlig depression	Strukturerad/semi-strukturerad intervju	Sensitivitet: 33–100% Specificitet: 32–94%
PHQ-9	Riktad screening (självskattning) Algoritm-baserat Egentlig depression	Strukturerad/semi-strukturerad intervju	Sensitivitet: 34–84% Specificitet: 93%
PHQ-9	Riktad screening (självskattning) Tröskelvärde Egentlig depression	Strukturerad/semi-strukturerad intervju	Sensitivitet: 54–100% Specificitet: 84%
MDI	Riktad screening (självskattning) Egentlig depression	Strukturerad/semi-strukturerad intervju	Otillräckligt
SHQ	Riktad screening (självskattning) Årtidsbunden depression	Strukturerad/semi-strukturerad intervju	Otillräckligt
SPAQ	Riktad screening (självskattning) Säsongsbunden depression	Strukturerad/semi-strukturerad intervju	Otillräckligt
HCL-32	Riktad screening (självskattning) Hypomani, någon gång under livet	Strukturerad/semi-strukturerad intervju	Sensitivitet: 82% Specificitet: – Otillräckligt
MDQ	Riktad screening (självskattning) Bipolärt spektrum-syndrom, någon gång under livet	Strukturerad/semi-strukturerad intervju	<i>Normalkriterier</i> Sensitivitet: 72% Specificitet: 78%  <i>Genève-kriterier</i> Sensitivitet: 78% Specificitet: 79%

Tabellen fortsätter på nästa sida

**Tabell 3.5.1** fortsättning

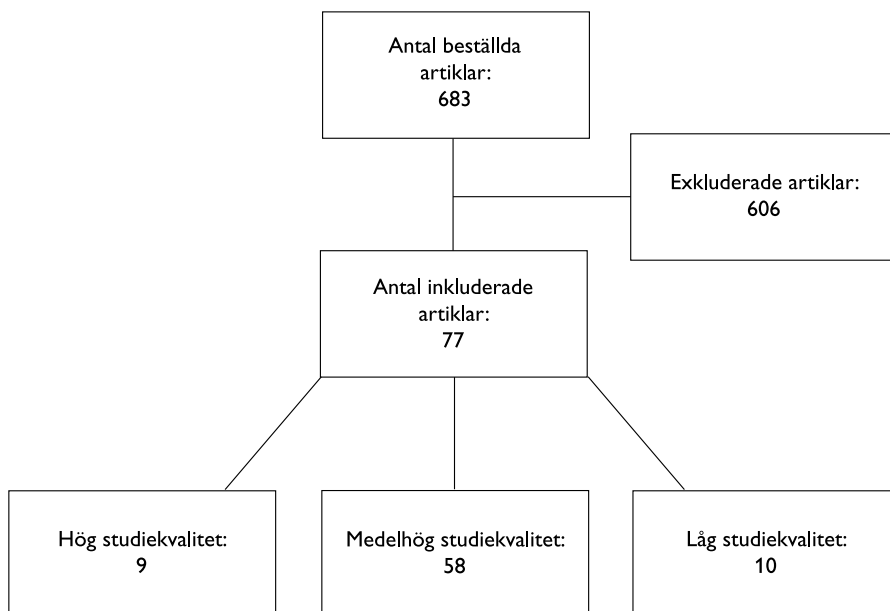
<b>Formulär</b>	<b>Syfte</b>	<b>Referensstandard</b>	<b>Evidensgraderat resultat</b>
BDI-I & II	Svårighetsgrad, depression (självskattning)	DSM-IV-kriterier	Otillräckligt
CDSS	Svårighetsgrad, depression (intervju/självskattning)	CGI-S/DSM-IV-kriterier	Otillräckligt
HDRS-17	Svårighetsgrad, depression (intervju/självskattning)	CGI-S/DSM-IV-kriterier	Otillräckligt
IDS (QIDS)	Svårighetsgrad, depression (självskattning)	CGI-S/DSM-IV-kriterier	Otillräckligt
MADRS	Svårighetsgrad, depression (intervju/självskattning)	CGI-S/DSM-IV-kriterier	Otillräckligt
SDS	Svårighetsgrad, depression (självskattning)	CGI-S/DSM-IV-kriterier	Otillräckligt
AS-18	Svårighetsgrad, bipolärt syndrom (självskattning)	CGI-BP/DSM-IV-kriterier	Otillräckligt
BDRS	Svårighetsgrad bipolärt syndrom (semistrukturerad intervju/självskattning)	CGI-BP/DSM-IV-kriterier	Otillräckligt
MAS	Svårighetsgrad bipolärt syndrom (intervju/självskattning)	CGI-BP/DSM-IV-kriterier	Otillräckligt
YMRS	Svårighetsgrad bipolärt syndrom (intervju/självskattning)	CGI-BP/DSM-IV-kriterier	Otillräckligt
LCM	Svårighetsgrad bipolärt syndrom (självskattning)	CGI-BP	Otillräckligt
IDS (QIDS)	Samstämmighet kliniker/patient Egentlig depression	Ej tillämpligt	ICC: 0,88
MADRS	Samstämmighet kliniker/patient Egentlig depression	Ej tillämpligt	ICC: >0,7

ICC = Intraclass correlation

## Urval av studier

Av 683 artiklar som lästes i fulltext inkluderades 77 för granskning. De vanligaste orsakerna till att studier exkluderades var att de använt referensstandarder som vi bedömt som otillfredsställande, att det framgick klart att tiden mellan index- och referenstestet var för lång eller att studierna genomförts i andra länder än västerländska.

I och med att vi tillämpade strikta inklusions- och exklusionskriterier var det få, närmare bestämt tio studier, som hade så stora metodologiska brister att de bedömdes hålla låg kvalitet. Det vetenskapliga underlaget för att bedöma formulär för vuxna består därmed av 67 studier (Figur 3.5.1).

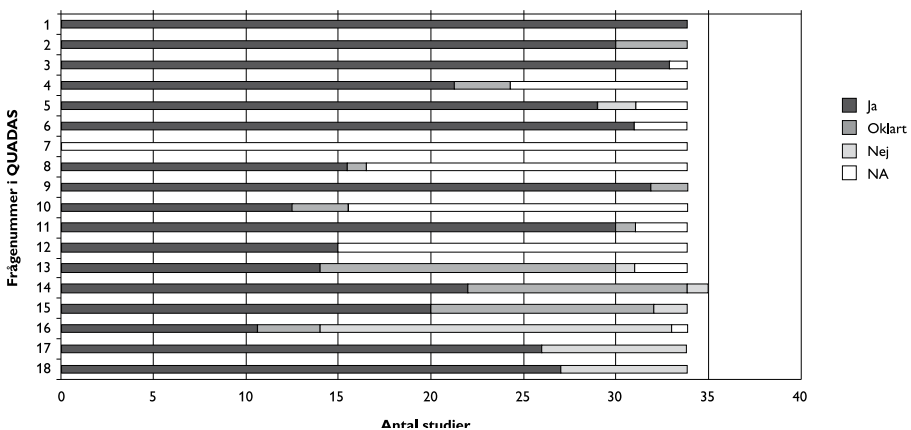


**Figur 3.5.1** Flödesschema för granskning av bedömningsformulär vuxna.

## Studiekvalitet

Av de 67 studierna kunde 35 användas för att dra slutsatser om sensitivitet och specificitet för 15 formulär. Resterande 32 studier undersökte formulär där underlaget visade sig vara otillräckligt.

Figur 3.5.2 visar översiktligt vilka metodproblem som fanns i de 35 studierna. I figuren representeras varje rad av ett påstående (item) i vår QUADAS-mall (se Kapitel 2 och Bilaga 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 framgick att kriteriet inte 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 var interbedömarreliabiliteten (item 16) den svagaste punkten i och med att nära två tredjedelar av studierna inte redovisade några uppgifter.



**Figur 3.5.2** Sammanställning av hur kvaliteten på studier bedömdes med formuläret QUADAS.

## Semistrukturerade diagnostiska intervjuer

SCID-I är en vanlig referensstandard i diagnostiska tillförlitlighetsstudier och valdes som referens även i vårt projekt. Två av våra frågor var: Hur tillförlitlig är SCID-I jämfört med bästa möjliga diagnostik, dvs Best estimate/LEAD, för att ställa diagnoser på förstämningssyndrom? Och vilken sensitivitet och specificitet har övriga diagnostiska intervjuer jämfört med SCID-I eller LEAD? Best estimate/LEAD har beskrivits i Kapitel 1.

### Structured Clinical Interview for DSM-Axis I psychiatric disorders (SCID-I)

SCID-I räknas som den viktigaste referensstandarderna för andra diagnostiska metoder och används regelmässigt inom forskning för att säkerställa diagnosers tillförlitlighet.

SCID-I är en semistrukturerad intervju [1] som används av tränade kliniker för att ställa diagnos enligt kriterierna för Axial I (symtomdiagnos) i DSM-klassifikationen, för närvarande enligt DSM-IV [2]. SCID-I inleds med en översiktsintervju med öppna frågor och därefter genomförs intervjun utifrån moduler för olika diagnostiska grupper. Intervjun inleds vanligen med en sluten fråga och fortsätter med fördjupande uppföljningsfrågor [3,4].

En fundamental princip är att skattningarna baseras på om DSM-kriterier för respektive tillstånd uppfylls, inte nödvändigtvis att svaren är jakande. Intervjuaren behöver ofta be om exempel eller ställa ytterligare frågor för att ha tillräckligt underlag för sin bedömning och bör också använda alla tillgängliga informationskällor. SCID-I för DSM-IV finns i en klinisk version, SCID-I-CV, och i flera mer detaljerade forskningsversioner.

#### *Beskrivning av underlaget*

Tjugosex studier lästes i fulltext varav 24 exkluderades. En studie exkluderades pga att LEAD inte var utförd på sedvanligt sätt [5]. Övriga studier exkluderades pga att de behandlade andra frågeställningar som interbedömarreliabilitet, modifierade metoder, var översikter eller berörde icke-kliniska populationer (se Tabell 3.5.14).



Det vetenskapliga underlaget består därmed av två studier som bedömdes ha medelhög studiekvalitet [6,7]. Båda studierna kommer från USA, bedömarna var välutbildade i att använda SCID-I, och LEAD användes som referensstandard. Studien av Ramirez Basco och medarbetare inkluderade 200 patienter från en psykiatrisk öppenvårdsmottagning och jämförde SCID-I med och utan tillgång till journalhandlingar mot LEAD [7]. Studien av Miller och medarbetare inkluderade 56 patienter från universitetsanknuten psykiatrisk slutenvård [6].

### ***Sammanvägda resultat***

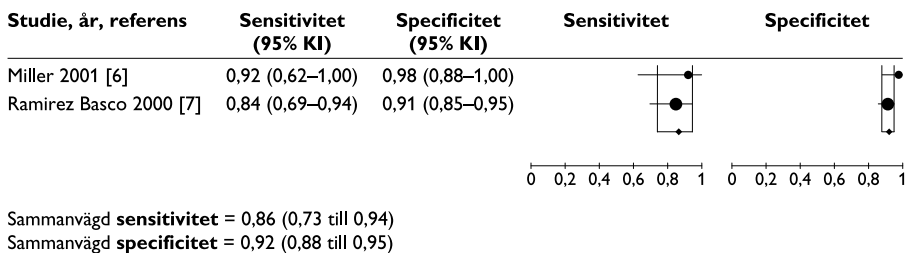
Endast studien av Ramirez Basco och medarbetare rapporterade sensitivitet och specificitet för affektiva syndrom [7], men värdena kunde beräknas med hjälp av givna uppgifter även för studien av Miller och medarbetare [6]. Generell överensstämmelse för alla diagnoser mellan SCID-I och LEAD, mätt med Cohen's  $\kappa$  [8], var 0,82 i den ena studien [6]. I den andra studien var  $\kappa$  0,61 för SCID-I utan journalhandlingar samt 0,76 för SCID-I med tillgång till journalhandlingar [7].

För diagnosen *egentlig depression* var sensitiviteten 84 respektive 92 procent och specificiteten 91 respektive 98 procent i de två studierna (Figur 3.5.3). För *bipolärt syndrom* var sensitiviteten 76 respektive 100 procent och specificiteten 93 respektive 96 procent (Figur 3.5.4). För diagnos av *schizoaffektivt syndrom* var sensitiviteten 53 respektive 76 procent. Båda studierna visade en specificitet på 97 procent (Figur 3.5.5).

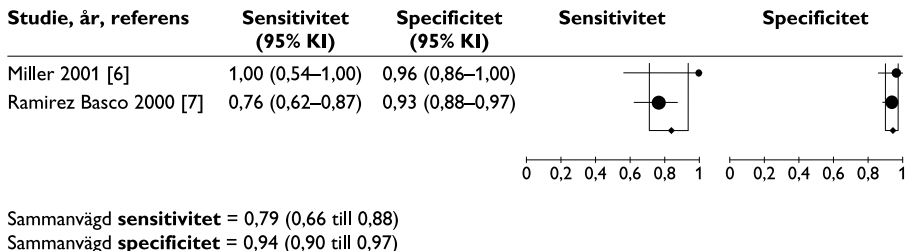
Den sammanvägda sensitiviteten för depression blev 86 procent (95 % KI, 73 till 94). För bipolärt syndrom var det sammanvägda estimatet för sensitivitet 79 procent (95 % KI, 66 till 88). Motsvarande siffror för schizoaffektivt syndrom var 61 procent (95 % KI, 48 till 74).

Den sammanvägda specificiteten för egentlig depression blev 92 procent (95 % KI, 88 till 95), för bipolärt syndrom 94 procent (95 % KI, 90 till 97) samt för schizoaffektivt syndrom 97 procent (96 % KI, 94 till 99).

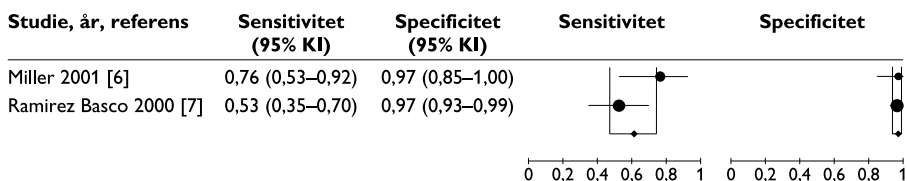
Studien av Ramirez Basco och medarbetare jämförde även SCID-I med och utan tillgång till journal. När bedömaren kunde komplettera SCID-I-intervjun med uppgifter från journalen ökade sensitiviteten från 76 till 90 procent för bipolärt syndrom och från 53 procent till cirka 90 procent för schizoaffektivt syndrom.



**Figur 3.5.3** Forest plot. Spridning av sensitivitet och specificitet för diagnos av egentlig depression med SCID-I utan tillgång till journal.



**Figur 3.5.4** Forest plot. Spridning av sensitivitet och specificitet för diagnos av bipolärt syndrom med SCID-I utan tillgång till journal.



Sammanvägd **sensitivitet** = 0,61 (0,48 till 0,74)

Sammanvägd **specificitet** = 0,97 (0,94 till 0,99)

**Figur 3.5.5** Forest plot. Spridning av sensitivitet och specificitet för diagnos av schizoaffektivt syndrom med SCID-I utan tillgång till journal.

### *Bedömning av evidensstyrka*

Kvaliteten på båda studierna bedömdes som medelhög pga att det saknades uppgifter om interbedömarreliabilitet för SCID-I. Vi gjorde dock inget avdrag på evidensstyrkan eftersom studierna för övrigt bedömdes vara välgjorda och de använde samma referensstandard, dvs LEAD där SCID-I-CV ingick.

Överförbarheten bedömdes som god eftersom materialet innefattade psykiatrisk öppen- och slutenvård och studierna använde samma referensstandard. Precisionen var låg vad gällde sensitivitet för samtliga diagnoser.

Tabell 3.5.2 sammanställer sensitivitet och specificitet för SCID-I avseende egentlig depression, bipolärt syndrom och schizoaffektivt syndrom jämfört med Best estimate.

**Tabell 3.5.2** Sammanvägda resultat avseende diagnoser egentlig depression, bipolärt syndrom och schizoaffektivt syndrom för SCID-I utan tillgång till journal jämfört med LEAD.

Utfallsmått Diagnos	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
<b>Sensitivitet</b>				
Egentlig depression	2 studier n=256	86% (73; 94)	Begränsat ⊕⊕○○	Brister i precision (-2)
Bipolärt syndrom	2 studier n=256	79% (66; 88)	Begränsat ⊕⊕○○	Brister i precision (-2)
Schizoaffektivt syndrom	2 studier n=256	61% (48; 74)	Begränsat ⊕⊕○○	Brister i precision (-2)
<b>Specificitet</b>				
Egentlig depression	2 studier n=256	92% (88; 95)	Starkt ⊕⊕⊕⊕	
Bipolärt syndrom	2 studier n=256	94% (90; 97)	Starkt ⊕⊕⊕⊕	
Schizoaffektivt syndrom	2 studier n=256	97% (94; 99)	Starkt ⊕⊕⊕⊕	

KI = Konfidensintervall

### **Evidensgraderade resultat**

- I psykiatrisk öppen- och slutenvård har SCID-I-intervju utan tillgång till journal en sensitivitet på 86 procent (95 % KI, 73 till 94) (begränsat vetenskapligt underlag ⊕⊕○○) och specificitet på 92 procent (95 % KI, 88 till 95) (starkt vetenskapligt underlag ⊕⊕⊕⊕) för att identifiera egentlig depression, med LEAD som referensstandard.
- I psykiatrisk öppen- och slutenvård har SCID-I-intervju utan tillgång till journal en sensitivitet på 79 procent (95 % KI, 66 till 88) (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 94 procent (95 % KI, 90 till 97) (starkt vetenskapligt underlag ⊕⊕⊕⊕) för att identifiera bipolärt syndrom, med LEAD som referensstandard.

- I psykiatrisk öppenvård och slutenvård har SCID-I-intervju utan tillgång till journal en sensitivitet på 61 procent (95 % KI, 48 till 74) (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 97 procent (95 % KI, 94 till 99) (starkt vetenskapligt underlag ⊕⊕⊕⊕) för att identifiera schizoaffektivt syndrom, med LEAD som referensstandard.

### **Schedule for Affective Disorders and Schizophrenia (SADS)**

SADS är en semistrukturerad intervju som skapades för att ställa psykiatrisk diagnos enligt Research Diagnostic Criteria (RDC-kriterierna). Bedömningen ska göras av tränade kliniker. SADS utvecklades för att kunna användas både i kliniska (somatisk vård och psykiatri) och icke kliniska populationer. Upphovsmännen till SADS var Endicott och kollegor, och originalartikeln publicerades år 1978 [9].

#### ***Beskrivning av underlaget***

Tolv studier lästes i fulltext. Elva av dem exkluderades, främst pga att de inte undersökte relevant frågeställning. Det vetenskapliga underlaget består därmed av en studie som vi bedömde hålla låg kvalitet (Tabell 3.5.15). Studien, av Kosten och medarbetare, rörde vuxna patienter med opiatberoende (n=201) och deras make/maka eller annan anhörig (n=274; n=475) [10]. Patienterna rekryterades utifrån ett bekvämlighetsurval från en universitetsklinik i USA. Referensstandard var LEAD (medicinska journaler, information från intervjuer med anhöriga och information från den strukturerade intervjun SADS). SADS utvärderades för depression. Sensitiviteten var 79 procent och specificiteten var 98 procent.

#### ***Bedömning av evidensstyrka***

Det finns för få relevanta studier och med ett för snävt spektrum av studiedeltagare för att bedöma om SADS har tillräcklig tillförlitlighet för att kunna användas som stöd för att ställa en depressionsdiagnos med LEAD som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## **Schedules for Clinical Assessment in Neuropsychiatry (SCAN)**

SCAN är en semistrukturerad intervju baserad på Present State Examination (PSE-9) som togs fram inom ramen för ett WHO-projekt. SCAN har vidareutvecklats för att kunna ställa diagnos enligt DSM- och ICD-systemen. Bedömningen ska göras av tränade kliniker. SCAN kan användas både kliniskt och i forskning. Upphovsmännen till SCAN var Wing och kollegor, och originalartikeln publicerades år 1990 [11].

### ***Beskrivning av underlaget***

Tre studier lästes i fulltext och samtliga exkluderades. Orsakerna var att de inte undersökte relevant frågeställning, att det var för lång tid mellan index- och referenstest samt att det inte fanns data som medgav beräkning av sensitivitet och specificitet.

### ***Bedömning av evidensstyrka***

Det saknas relevanta studier som undersöker om SCAN är tillförlitligt för att ställa diagnosen egentlig depression med strukturerade eller semistrukturerade intervjuer som referensstandard.

## **Strukturerade diagnostiska intervjuer**

### **Composite International Diagnostic Interview (CIDI)**

CIDI är en strukturerad intervju som kan användas av tränade läkare för att ställa psykiatrisk diagnos, den senaste versionen av CIDI är baserad på ICD-10 och DSM-IV [12]. CIDI skapades primärt för att möjliggöra transkulturella epidemiologiska studier i den allmänna befolkningen. De som utvecklade intervjun utgick från Diagnostic Interview Schedule (DIS) [13] och lade till frågor från Present State Examination (PSE) [14,15].

### ***Beskrivning av underlaget***

Av 26 studier som lästes i fulltext exkluderades 25, främst pga att de inte berörde kliniska populationer. Det vetenskapliga underlaget består därmed av en studie av medelhög kvalitet [16] (se Tabell 3.5.16). Studien omfattade 54 kroppsligt sjuka, manliga patienter på ett veteransjukhus i USA. En modifierad version av CIDI, UM-CIDI [17], jämfördes mot LEAD innefattande bl a SCID-I [16] för pågående och livstidsdiagnos

av egentlig depression. Sensitiviteten för pågående episod av egentlig depression var 67 procent och specificiteten 84 procent.

### ***Bedömning av evidensstyrka***

Det finns för få studier med tillräcklig kvalitet för att bedöma om CIDI är tillförlitligt för att ställa diagnosen egentlig depression i kliniska populationer med strukturerad eller semistrukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **Diagnostic Interview Schedule (DIS)**

DIS utvecklades för att ställa psykiatrisk diagnos enligt DSM-III, Feighner-kriterierna och Research Diagnostic Criteria (RDC). Formuläret togs fram på begäran av National Institute of Mental Health för att användas i en stor amerikansk epidemiologisk studie, Epidemiologic Catchment Area Program (ECA). Upphovsmän var Robins och kollegor och originalartikeln publicerades år 1981 [18]. DIS ska kunna användas av tränade lekmän både i kliniska och icke kliniska populationer.

### ***Beskrivning av underlaget***

Av 19 studier som lästes i fulltext exkluderades 17, beroende på att de inte undersökte relevant frågeställning, att det var för lång tid mellan index- och referenstest eller att de inte undersökte kliniska populationer. Två studier inkluderades. En av dem bedömdes hålla låg studiekvalitet bl a på grund av oklarheter i utfallsmått och inklusionskriterier [19]. Det vetenskapliga underlaget består därmed av en studie som bedömdes ha medelhög studiekvalitet [20] (se Tabell 3.5.17). Studien inkluderade ett randomiserat urval på 129 vuxna slutenvårdspatienter på en rehabiliteringsavdelning för personer med alkohol- och missbruksproblem i USA. Referensmättet för studien var SADS (RDC). DIS utvärderades för egentlig depression och dystymi. För egentlig depression var sensitiviteten 25 procent och specificiteten 90 procent. För dystymi var sensitiviteten 22 procent och specificiteten 92 procent.

### ***Bedömning av evidensstyrka***

Det finns för få studier av tillräcklig kvalitet och med ett för snävt spektrum av studiedeltagare för att bedöma om DIS är tillförlitligt för att ställa diagnosen förstämningssyndrom med strukturerad eller semi-

strukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **Mini International Neuropsychiatric Interview (MINI)**

MINI är en strukturerad intervju som skapades för att ställa psykiatrisk diagnos enligt DSM-IV och ICD-10. Formuläret utvecklades för att kunna användas både kliniskt och i forskning, t ex i epidemiologiska studier, och innehåller 16 diagnoskategorier, däribland egentlig depressionsepisod och manisk episod.

Upphovsmännen till MINI var Sheehan och kollegor, och originalartikeln publicerades år 1998 [21]. MINI ska kunna användas av tränade lekmän.

#### ***Beskrivning av underlaget***

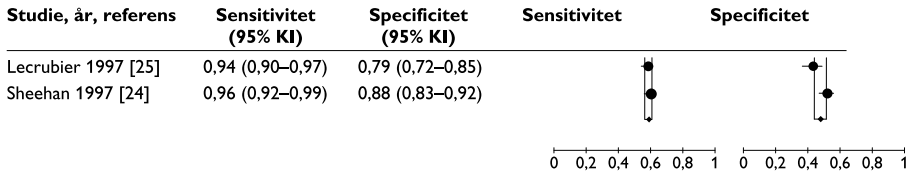
Av nio studier som lästes i fulltext exkluderades fem beroende på att de inte undersökte relevant frågeställning. Av de fyra studierna som inkluderades hade två låg studiekvalitet pga att blindning mellan indexmått och referensmått saknades [22,23]. Det vetenskapliga underlaget består därmed av två studier varav båda bedömdes ha medelhög studiekvalitet [24,25] (se Tabell 3.5.18). Studierna var genomförda på 670 konsekutivt rekryterade vuxna patienter inom psykiatrisk öppenvård i USA och Frankrike.

I studien av Lecrubier och medarbetare användes CIDI som referensstandard [25] medan Sheehan och medarbetare valde SCID-I [24]. MINI utvärderades i båda studierna för egentlig depression, pågående mani och mani någon gång under livstid.

#### ***Sammanvägda resultat***

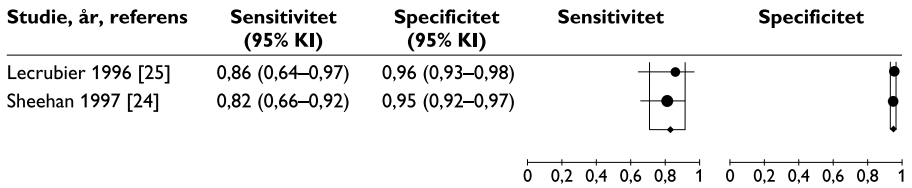
Sensitiviteten vid egentlig depression var 94 respektive 96 procent i de båda studierna och specificiteten 79 respektive 88 procent (Figur 3.5.6). För pågående mani var sensitiviteten 86 respektive 82 procent och specificiteten 96 respektive 95 procent (Figur 3.5.7). När bedömningen avsåg mani någon gång under livet var sensitiviteten 77 respektive 81 procent och specificiteten 92 respektive 94 procent (Figur 3.5.8).





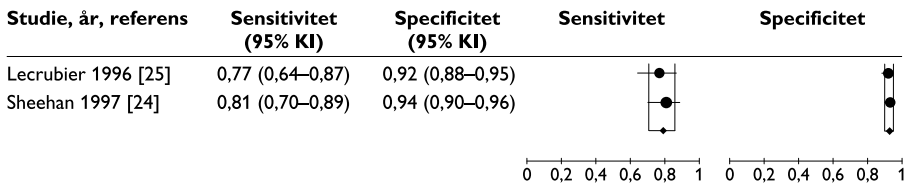
Sammanvägd **sensitivitet** = 0,95 (0,93 till 0,97)  
 Sammanvägd **specificitet** = 0,84 (0,80 till 0,87)

**Figur 3.5.6** Forest plot för bedömning av sammanvägd sensitivitet och specificitet för diagnos av egentlig depression med MINI.



Sammanvägd **sensitivitet** = 0,83 (0,71 till 0,92)  
 Sammanvägd **specificitet** = 0,95 (0,93 till 0,97)

**Figur 3.5.7** Forest plot för bedömning av sammanvägd sensitivitet och specificitet för diagnos av pågående mani med MINI.



Sammanvägd **sensitivitet** = 0,79 (0,71 till 0,86)  
 Sammanvägd **specificitet** = 0,93 (0,90 till 0,95)

**Figur 3.5.8** Forest plot för bedömning av sammanvägd sensitivitet och specificitet för diagnos av mani någon gång under livstiden med MINI.

Den sammanvägda sensitiviteten för depression blev 95 procent (95 % KI, 93 till 97) och specificiteten 84 procent (95 % KI, 80 till 87). För pågående mani blev den sammanvägda sensitiviteten 83 procent (95 % KI, 71 till 92) och specificiteten 95 procent (95 % KI, 93 till 97). För bedömning av mani tidigare i livet blev den sammanvägda sensitiviteten 79 procent (95 % KI, 71 till 86) och specificiteten 93 procent (95 % KI, 90 till 95).

En av studierna med låg kvalitet, av Jones och medarbetare, redovisade också sensitivitet och specificitet [23]. Studien fann en sensitivitet på 88 procent och specificitet på 97 procent, vilket stämmer väl med våra resultat.

### ***Bedömning av evidensstyrka***

Vad gäller utfallsmåtten sensitivitet och specificitet rörande egentlig depression fanns det en smärre brist vad gäller studiekvaliteten eftersom interbedömarreliabiliteten på referensmättet SCID-I inte redovisades i en av studierna [24]. Dock ansågs bristen inte vara tillräcklig för att motivera avdrag.

Överförbarheten påverkades av att de båda studierna hade använt olika referensstandarder. Precisionen var bristfällig för sensitivitet vad gäller pågående mani och mani någon gång under livstid (Tabell 3.5.3).

**Tabell 3.5.3** Resultatsammanställning för MINI.

Utfallsmått Diagnos	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
<b>Sensitivitet</b>				
Egentlig depression	2 studier n=720	95% (93; 97)	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (-1)
Pågående mani	2 studier n=720	83% (71; 92)	Begränsat ⊕⊕○○	Bristande överförbarhet (-1) Bristande precision (-1)
Mani någon gång under livstiden	2 studier n=720	79% (71; 86)	Begränsat ⊕⊕○○	Bristande överförbarhet (-1) Bristande precision (-1)
<b>Specificitet</b>				
Egentlig depression	2 studier n=720	84% (80; 87)	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (-1)
Pågående mani	2 studier n=720	95% (93; 97)	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (-1)
Mani någon gång under livstiden	2 studier n=720	93% (90; 95)	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (-1)

KI = Konfidensintervall

### *Evidensgraderade resultat*

- I psykiatrisk öppenvård har MINI en sensitivitet på 95 procent (95 % KI, 93 till 97) och en specificitet på 84 procent för depression (95 % KI, 80 till 87) med strukturerad eller semistrukturerad intervju som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- I psykiatrisk öppenvård har MINI en sensitivitet för pågående mani på 83 procent (95 % KI, 71 till 92) jämfört med strukturerad eller semistrukturerad intervju (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 95 procent (95 % KI, 93 till 97) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

- I psykiatrisk öppenvård med mani någon gång under livstiden har MINI en sensitivitet på 79 procent (95 % KI, 71 till 86) jämfört med strukturerad eller semistrukturerad intervju (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 93 procent (95 % KI, 90 till 95) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

### **The Primary Care Evaluation of Mental Disorders (PRIME-MD)**

PRIME-MD är en strukturerad intervju som skapades för att kunna ställa psykiatrisk diagnos enligt DSM-IV och ICD-10 i primärvård. Den består av moduler för sju diagnostiska kategorier. Upphovsmännen till PRIME-MD var Spitzer och kollegor i samarbete med Pfizer Inc, New York, och originalartikeln publicerades år 1994 [26].

PRIME-MD består av två delar; ett patientformulär (28 frågor som ska besvaras med ja/nej) som fylls i före läkarbesöket och en utvärderingshandledning till läkaren.

#### ***Beskrivning av underlaget***

Av tolv studier som lästes i fulltext exkluderades tio beroende på att de inte undersökte relevant frågeställning eller att det gick för lång tid mellan index- och referenstest. De båda studierna som inkluderades bedömdes ha medelhög studiekvalitet [27,28] (se Tabell 3.5.19).

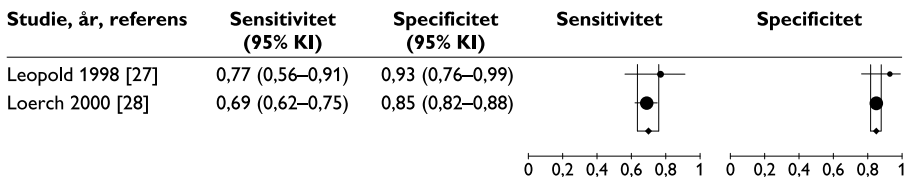
Leopold och medarbetare inkluderade vuxna personer med cancer från USA (n=53) [27]. Referensstandard var SCID-I. I den andra studien, av Loerch och medarbetare, bestod populationen av vuxna personer i primärvård eller psykiatrisk öppenvård (n=704) i Tyskland [28]. Referensmått var Munich CIDI (M-CIDI). Studien föreföll att baseras på ett bekvämlighetsurval.

#### ***Sammanvägda resultat***

Båda studierna redovisade data för ”någon form av förstämningssyndrom”. Sensitiviteten i studierna var 77 respektive 69 procent medan specificiteten var 93 procent respektive 85 procent (Figur 3.5.9a). Den sammanvägda sensitiviteten blev 70 procent (95 % KI, 63 till 76) och specificiteten 85 procent (95 % KI, 82 till 88).

Det gick också att räkna ut sensitivitet och specificitet för egentlig depression. Sensitiviteten var 30 procent respektive 68 procent och specificiteten 95 respektive 84 procent (Figur 3.5.9b). Skillnaden i resultaten för sensitivitet var så stor att det inte var meningsfullt att väga samman värdena. Man kan dock notera att sensitiviteten var låg i båda studierna. Detta sågs även i originalpublikationen av Spitzer och medarbetare som fann en sensitivitet på 57 procent [26].

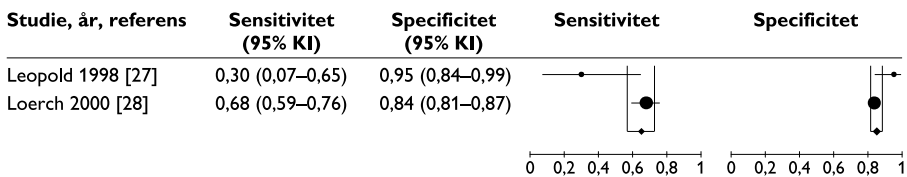
Den sammanvägda specificiteten var 85 procent (82 till 88).



Sammanvägd **sensitivitet** = 0,70 (0,63 till 0,76)

Sammanvägd **specificitet** = 0,85 (0,82 till 0,88)

**Figur 3.5.9a** Forest plot för bedömning av sammanvägd sensitivitet och specificitet för diagnos av förstämningssyndrom med PRIME-MD.



**Sensitivitet** ej sammanvägd

Sammanvägd **specificitet** = 0,85 (0,82 till 0,88)

**Figur 3.5.9b** Forest plot för bedömning av sammanvägd sensitivitet och specificitet för diagnos av egentlig depression med PRIME-MD.

## Bedömning av evidensstyrka

Det fanns en viss risk för bias eftersom interbedömarreliabilitet för referensstandarderna inte redovisades, men bristen var inte tillräckligt allvarlig för att motivera en sänkning av evidensstyrkan. Överförbarheten var bristfällig eftersom studierna använde olika referensstandarder och hade ett begränsat patientspektrum. Det medförde att evidensstyrkan sänktes ett steg.

För diagnosen ”någon form av förstämningssyndrom” hade det sammanvägda estimatet för sensitivitet ett relativt brett konfidensintervall (0,13) vilket medförde en sänkning av evidensstyrkan med ett steg (Tabell 3.5.4).

För diagnosen egentlig depression (Tabell 3.5.4) visade de två studierna stora skillnader i sensitivitet, men sensitiviteten var låg, <70 procent i båda. Spitzers studie med en sensitivitet på 57 procent stödjer vår bedömning. I och med att underlaget enbart bestod av två studier drog vi ner evidensstyrkan ytterligare ett steg. För specificitet gjordes inga ytterligare avdrag i och med att studierna var samstämmiga.

**Tabell 3.5.4** Resultatsammanställning för PRIME-MD.

Utfallsmått	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
<b>Sensitivitet</b>				
Någon form av förstämningssyndrom	2 studier n=757	70% (63; 76)	Begränsat ⊕⊕○○	Bristande överförbarhet (-1) Bristande precision (-1)
Depression	2 studier n=757	<70%	Begränsat ⊕⊕○○	Bristande överförbarhet (-1) Bristande samstämmighet (-1)

Tabellen fortsätter på nästa sida

**Tabell 3.5.4** fortsättning

Utfallsmått	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
<b>Specificitet</b>				
Någon form av förstärknings-syndrom	2 studier n=757	85% (82; 88)	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (-1)
Depression	2 studier n=757	85% (82; 88)	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (-1)

KI = Konfidensintervall

### **Evidensgraderade resultat**

- I primärvård, psykiatrisk öppenvård och onkologisk vård har PRIME-MD en sensitivitet på 70 procent (95 % KI, 63 till 76) (begränsat vetenskapligt underlag ⊕⊕○○) för ”något förstärknings-syndrom” och en specificitet på 85 procent (95 % KI, 82 till 88) med strukturerad eller semistrukturerad intervju som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- I primärvård, psykiatrisk öppenvård och onkologisk vård har PRIME-MD en sensitivitet under 70 procent för depression (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 85 procent (95 % KI, 82 till 88) med strukturerad eller semistrukturerad intervju som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

### **Structured Psychiatric Interview for General Practice (SPIFA)**

SPIFA utvecklades i Norge, Sverige och Finland under slutet av 1990- och början av 2000-talet och är avsett för primärvården [29]. Den norska versionen från 1996 omfattar 19 vanliga psykiatriska diagnoser. I den svenska versionen från 1997 inkluderades även ett antal symptom- och funktionsskattningsskalor. Versionen från 2003 inkluderade ytterligare några psykiatriska diagnoser samt delades upp i en lång och kort version för att vara användbar i olika typer av kliniska sammanhang [29].

### ***Beskrivning av underlaget***

Det finns två publicerade studier om SPIFA. Hedlund och medarbetare utvärderade interbedömarreliabilitet mellan sjuksköterska och psykiater vid diagnostisering med SPIFA, dvs en fråga som inte ingår i denna översikt [30]. Studien exkluderades därför.

Dahl och medarbetare redovisade en norsk klinisk studie, den så kallade SPIFA-336 reliabilitetsstudien, tillsammans med beskrivningar av hur SPIFA använts i Sverige och Finland [29]. I den norska studien jämfördes diagnostik med SPIFA utförd av distriktsläkare mot SCID-I (n=167) och SPIFA (n=164) utförd av psykiatrer. Studien fick låg studie-kvalitet eftersom tidsintervall och blindning mellan testen var oklar.

### ***Evidensgraderat resultat***

Eftersom det saknas studier av god kvalitet går det inte att bedöma om SPIFA är tillförlitligt för att diagnostisera förstämningssyndrom (otillräckligt vetenskapligt underlag ⊕○○○).

## **Riktad screening av depression**

Vår fråga var vilken tillförlitlighet, mätt som sensitivitet och specificitet, formulär för riktad screening har jämfört med en referensstandard. Som framgick i inledningen av kapitlet accepterade vi strukturerade och semi-strukturerade intervjuer som referensstandard.

### **Beck Depression Inventory-II (BDI-II)**

BDI utvecklades av Beck och medarbetare som ett formulär för att mäta svårighetsgrad och originalartikeln publicerades år 1961 [31]. Formuläret har reviderats, senast i mitten på 1990-talet, och anpassades då till DSM-IV-kriterierna. Nuvarande version, BDI-II, kan användas för såväl riktad screening av egentlig depression som för att mäta svårighetsgrad av depressiva symtom [32]. BDI-II är avsedd för att användas från 13 år och uppåt.

Skalan består av 21 frågor enligt en fyragradig skala som går från 0 till 3. För riktad screening av egentlig depression med BDI-II föreslås ett tröskelvärde på 14.



Manualen för BDI-II refererar till fem studier på collegestudenter och personer från psykiatriska mottagningar i USA och Norge som underlag för formulärets validitet och reliabilitet. Studierna är dock inte upplagda så att sensitivitet och specificitet kan bedömas.

### ***Beskrivning av underlaget***

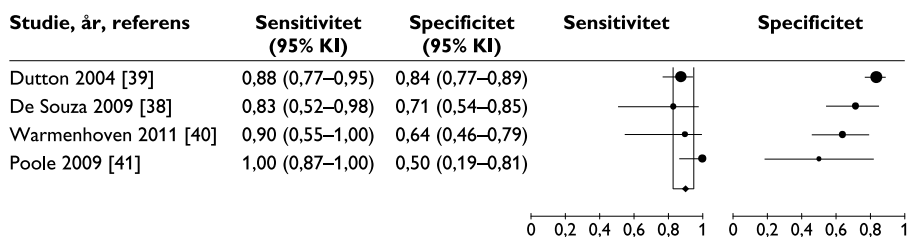
Av 29 studier som lästes i fulltext exkluderades 21 beroende på att de inte undersökte relevant frågeställning, att det var för lång tid mellan index- och referenstest samt att de inte undersökte kliniska populationer. Av de åtta som uppfyllde inklusionskriterierna hade fyra låg studiekvalitet pga att blindning mellan indexmått och referensmått saknades samt oklarheter i utfallsmått [33–37]. Det vetenskapliga underlaget består därmed av fyra studier och samtliga bedömdes ha medelhög studiekvalitet [38–41] (se Tabell 3.5.20). Totalt rekryterades 352 deltagare.

Poole och medarbetare utgick från 36 konsekutivt rekryterade patienter med långvarig smärta i Storbritannien [41]. Referensstandard var SCID-I. De Souza och medarbetare inkluderade 50 vuxna med Huntington's sjukdom inom neuropsykiatrisk öppenvård i Storbritannien och använde SCAN som referensstandard [38]. Två studier hade PRIME-MD som referensstandard. Den ena, av Dutton och medarbetare, rekryterade 220 vuxna personer i primärvård i USA [39]. I den andra som genomfördes av Warmenhoven och medarbetare i Nederländerna ingick 46 vuxna personer med cancer i slutskedet [40].

### ***Sammanvägda resultat***

Sensitiviteten i de fyra studierna låg mellan 83 och 100 procent och specificiteten mellan 50 och 84 procent (Figur 3.5.10). Sammanvägd sensitivitet blev 90 procent (95 % KI, 83 till 95). Vi bedömde att det inte var meningsfullt att väga samman resultaten för specificitet.

Vi gjorde en känslighetsanalys och lade in sensitivitet och specificitet även för de fyra studierna med låg kvalitet. När samtliga studier räknades in förstärktes mönstret av en acceptabel precision på sensitivitet och en stor spridning i specificitet. Det är värt att notera att alla studier utom studien av Dutton undersöker sensitivitet och specificitet för patienter som har kroppsliga sjukdomar förenade med ökad risk för depressionsliknande symtom som t ex energibrist.



Sammanvägd **sensitivitet** = 0,90 (0,83 till 0,95)

**Specificitet** ej sammanvägd

**Figur 3.5.10** Forest plot för BDI-II som formulär för riktad screening för depression.

### Bedömning av evidensstyrka

Det fanns en viss risk för bias i underlaget beroende på att interbedömar-reliabiliteten för referensstandarderna inte redovisades i tre av de fyra studierna. Överförbarheten påverkades i viss mån av att studierna använde olika referensstandarder. Sammantaget ledde det till att evidensstyrkan sänktes ett steg.

Konfidensintervallet för sensitivitet var relativt brett, vilket motiverade avdrag, för bristande precision. (Tabell 3.5.5).

**Tabell 3.5.5** Resultatsammanställning för BDI-II som formulär för riktad screening för depression.

Utfallsmått	Antal studier Antal patienter	Sammanvägt resultat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
<b>Sensitivitet</b>				
Egentlig depression	4 studier n=352	90% (83; 95)	Begränsat ⊕⊕○○	Studiekvalitet och bristande överförbarhet (-1) Precision (-1)

Tabellen fortsätter på nästa sida

**Tabell 3.5.5** fortsättning

Utfallsmått	Antal studier Antal patienter	Samman- vägt resultat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poäng- avdrag
<b>Specificitet</b>				
Egentlig depression	4 studier n=352	Bör ej vägas samman	Otillräckligt ⊕○○○	

### *Evidensgraderat resultat*

För riktad screening av depression hos vuxna personer med olika former av somatisk sjukdom samt inom primärvård har BDI-II med ett tröskelvärde  $\geq 14$  en sensitivitet på 90 procent (95 % KI, 83 till 95) jämfört med strukturerad eller semistrukturerad intervju (begränsat vetenskapligt underlag ⊕⊕○○). Det går inte att bedöma specificiteten eftersom studierna inte är samstämmiga (otillräckligt vetenskapligt underlag ⊕○○○).

### **Calgary Depression Scale for Schizophrenia (CDSS)**

Calgary Depression Scale for Schizophrenia (CDSS) har formen av en strukturerad intervju. Det används av tränade kliniker för att kunna särskilja depression från negativa schizofrena symtom och extrapyramidala symtom pga läkemedelsbehandling, men även för att bedöma svårighetsgrad av depression hos patienter med schizofreni [42]. Intervjun består av åtta strukturerade frågor som avser de två senaste veckorna samt en variabel som baseras på observation under intervjun. Varje fråga (variabel) har kriterier på en femgradig skala (0–4) och med maximalt 36 totalpoäng. Addington och medarbetare har föreslagit att tröskelvärdet 7 är optimalt för riktad screening av egentlig depression [43].

### *Beskrivning av underlaget*

Samtliga 21 studier som lästes i fulltext exkluderades, varav 18 stycken pga att man använt andra formulär som referensstandard eller att de berörde andra frågeställningar. Av de övriga tre hade två använt ett annat tröskelvärde [44,45]. Den tredje studien använde tröskelvärde 7 poäng, men exkluderades eftersom den genomfördes i Brasilien [46].

### ***Bedömning av evidensstyrka***

Det finns för få relevanta studier för att bedöma sensitivitet och specificitet för CDSS med tröskelvärde sju jämfört med strukturerad eller semi-strukturerad som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **Center for Epidemiologic Studies of Depression (CES-D)**

CES-D är ett formulär för självbedömning som består av 20 frågor. Det utvecklades i USA på 1970-talet för att studera förekomsten av depression i olika befolkningsgrupper [192]. Varje fråga representerar symtom eller tecken på depression. Den svarande uppmanas att på en fyragradig svarsskala (0=aldrig eller mindre än en dag; 3=största delen av tiden) markera det alternativ som stämmer bäst med frågans innehåll under de senaste 5–7 dagarna. Resultaten av de tjugo frågorna summeras och poängen kan variera mellan 0 och 60. Ett resultat över 16 poäng definieras som risk för egentlig depression.

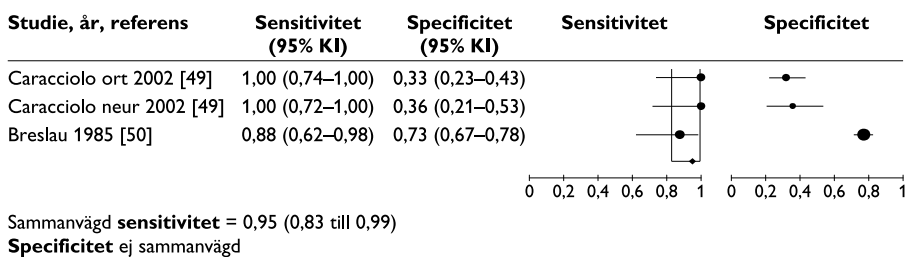
### ***Beskrivning av underlaget***

Av 74 studier som lästes i fulltext exkluderades 70 beroende på att de inte undersökte relevant frågeställning, att det var för lång tid mellan index- och referenstest samt att de undersökta populationerna var irrelevanta. Två studier bedömdes ha låg studiekvalitet [47,48]. Anledningen var att uppgifter om blindning mellan index- och referensstandarder saknades samt att orsakerna till ett omfattande bortfall inte redovisades.

Två studier hade medelhög studiekvalitet [49,50] (se Tabell 3.5.21). Den ena studien utfördes på en italiensk rehabiliteringsklinik med inneliggande patienter [49]. Patienterna delades upp i två grupper; 101 med ortopediska problem och 50 med neurologiska. Referensstandarderna var SCID-I och intervjuerna genomfördes av två certifierade psykologer. Studien redovisade ingen reliabilitet gällande hur samstämmiga de var för att ställa diagnos. Den andra studien baserades på 310 mödrar till funktionshindrade barn, knutna till specialistkliniker i USA [50]. Referensstandarderna var den strukturerade intervjun DIS som utfördes av tränade lekmän och interbedömarreliabiliteten var god.

### Sammanvägd sensitivitet och specificitet

Sensitivitet och specificitet för de tre undersökningsgrupperna redovisas i Figur 3.5.11. Sensitiviteten var genomgående hög, mellan 88 och 100 procent. Den sammanvägda sensitiviteten var 95 procent (95 % KI, 83 till 99). Specificiteten däremot var lägre, mellan 33 och 73 procent. Eftersom spridningen var så stor var det inte meningsfullt att väga samman resultaten.



**Figur 3.5.11** Forest plot med sensitivitet och specificitet för CES-D. Underlaget utgörs av två studier omfattande tre studiegrupper.

### Bedömning av evidensstyrka

Det fanns flera brister i underlaget som riskerade att minska tillförlitligheten i data. Studiekvaliteten drogs ned på grund av att det var oklart om den kliniska bedömningen skett helt blint efter patientens skattning. Överförbarheten påverkades av att det var få studier och med snävt spektrum av deltagare samt att man hade använt olika referensstandarder. Precisionen för sensitivitet var bristfällig. Det mest påtagliga problemet var den bristande samstämmigheten med avseende på specificiteten (Tabell 3.5.6).

**Tabell 3.5.6** Resultatsammanställning för CES-D som formulär för riktad screening.

Utfallsmått	Antal studier Antal deltagare	Sammanvägt resultat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	2 studier n=461	95% 83; 99)	Otillräckligt ⊕○○○	Bristande studiekvalitet (-1) Bristande överförbarhet (-1) Bristande precision (-1)
Specificitet	2 studier n=461	Bör inte vägas samman	Otillräckligt ⊕○○○	

### *Evidensgraderat resultat*

Det finns för få studier av tillräcklig kvalitet för att bedöma sensitivitet och specificitet för CES-D med semistrukturerad eller strukturerad intervju som referens och med tröskelvärde 16 (otillräckligt vetenskapligt underlag ⊕○○○).

### **Edinburgh Postnatal Depression Scale (EPDS)**

EPDS är ett självskattningsformulär med tio frågor som utvecklades för att identifiera depressiva symtom hos nyblivna mödrar [51]. De första månaderna efter förlossningen präglas normalt av aptitförändringar, sömnstörningar och trötthet. Sådana symtom karakteriserar också depression och ingår därför i alla depressionsformulär, vilket kan ge falskt positiva svar. EPDS konstruerades för att kompensera för detta. I ursprungsarbetet föreslogs ett tröskelvärde på 12/13 poäng för dem med trolig depression [51].

EPDS har även testats för screening av fäder till nyfödda samt för screening av kvinnor under graviditeten (då ibland kallad EDS).

Vi har bedömt det vetenskapliga underlaget för såväl post partum-depression som för depression under graviditeten.

### ***Beskrivning av underlaget för postnatal depression***

Av de 52 artiklar som lästes i fulltext uppfyllde 26 inklusionskriterierna. Sex av dem bedömdes ha låg studiekvalitet [52–57]. Två studier bedömdes ha hög studiekvalitet [58,59] och 18 medelhög [51,60–76] (se Tabell 3.5.22).

Materialet var mycket heterogent. Populationerna omfattade bl a rutinundersökta gravida kvinnor, kvinnor efter förlossningen och riskpopulationer. Tid mellan förlossning och undersökning varierade. Följden blev att prevalensen av depression låg mellan 1,5 procent [73] och 41 procent [65]. Dessutom användes olika referensstandarder.

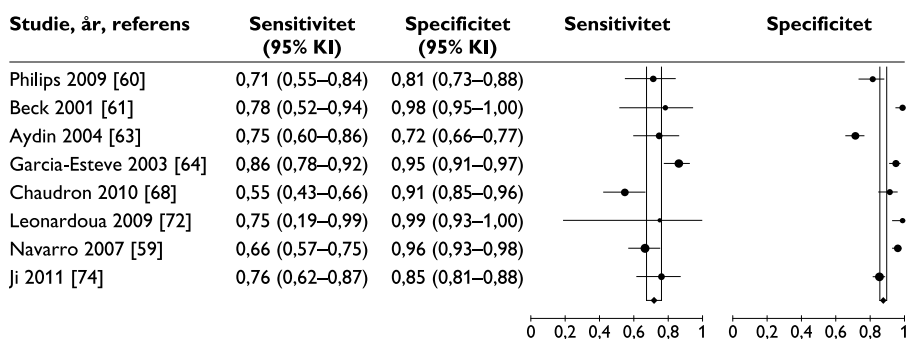
Vi valde att renodla analyserna. Vi inkluderade åtta studier med kvinnor som undersökts med SCID-I som referens [59–61,63,64,68,72,74]. Totalt omfattade studierna 2 208 patienter.

Phillips och medarbetare testade 166 australiensiska riskmödrar som tidigt visat svårigheter med att klara av föräldraskapet [60]. Undersökningen gjordes under barnets första levnadsår, och prevalensen av depression bland mödrarna var 25 procent. Beck och medarbetare undersökte 150 amerikanska mödrar med friska barn [61]. Undersökningen gjordes mellan andra och tolfte veckan efter förlossningen. Prevalensen av depression bland mammorna var 12 procent. I studien av Aydin och medarbetare undersöktes 341 turkiska mödrar i samband med ett primärvårdsbesök som gjordes någon gång under det första året efter förlossningen [63]. Prevalensen av depression bland mammorna var 14 procent. I studien av Navarro och medarbetare undersöktes 405 spanska kvinnor från en större kohort kvinnor som kom för rutinundersökning sex veckor efter förlossningen [59]. Prevalensen av depression var 28 procent. Garcia-Esteve och medarbetares studie omfattade 334 spanska mammor som kommit till ett rutinhälsovårdsbesök sex veckor efter förlossningen [64]. Genom ett tvåstegsförfarande med stratifierade grupper baserade på screening med EPDS valdes kvinnor ut för fördjupad undersökning. Prevalensen av depression bland de intervjuundersökta mammorna var 30 procent. Chaudron och medarbetare rekryterade 198 amerikanska mammor som kom för rutinhälsokontroll under första året efter förlossningen [68]. Prevalensen av depression i hela gruppen var 37 procent.

I studien av Leonardoua och medarbetare ingick 81 grekiska kvinnor som rekryterats efter förlossning [72]. Prevalensen av depression var 5 procent. I studien av Ji och medarbetare undersöktes 534 amerikanska kvinnor som alla hade en känd psykisk sjuklighet [74]. Prevalensen av depression 26 veckor efter förlossningen var 10 procent.

### *Sammanvägda resultat*

Sensitiviteten varierade mellan 55 och 86 procent i de åtta studierna. Det sammanvägda estimatet blev 72 procent (95 % KI, 67 till 76). Specificiteten låg mellan 72 och 99 procent och det sammanvägda estimatet blev 88 procent (95 % KI, 86 till 89) (Figur 3.5.12a).



Sammanvägd **sensitivitet** = 0,72 (0,67 till 0,76)

Sammanvägd **specificitet** = 0,88 (0,86 till 0,89)

**Figur 3.5.12a** Forest plot. Sensitivitet och specificitet för EPDS med SCID-I som referensstandard för post partum-depression hos kvinnor.

### *Bedömning av evidensstyrka*

Det fanns flera brister i det vetenskapliga underlaget som påverkade säkerheten i resultaten (se Tabell 3.5.7). En genomgående svaghet i studierna var att ingen av dem hade redovisat andel frågor i formuläret som hade fyllts i av kvinnorna, vilket betyder att det interna bortfallet är okänt. Dessutom uppgav bara två av studierna interbedömarreliabilitet. Det motiverade en sänkning av evidensstyrkan med ett steg.



Vi bedömde överförbarheten som god eftersom studierna genomförts i flera länder, och samtliga studier använde samma referensstandard (SCID-I).

Samstämmigheten för sensitivitet var relativt god, men två av studierna hade av oförklarliga skäl avvikande resultat med lägre sensitivitet än de övriga vilket drog ner evidensstyrkan ett steg.

**Tabell 3.5.7a** Resultatsammanställning för screening av post partum med EPDS och med SCID-I som referensstandard.

Utfallsmått	Antal studier Antal patienter	Sammanvägt estimat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	8 studier n=2 438	72% (67; 76)	Begränsat ⊕⊕○○	Studiekvalitet (-1) Samstämmighet (-1)
Specificitet	8 studier n=2 438	88% (86; 89)	Måttligt starkt ⊕⊕⊕○	Studiekvalitet (-1)

KI = Konfidensintervall

### *Beskrivning av underlaget för screening av gravida*

Fyra studier uppfyllde inklusionskriterierna [62,66,74,75]. Två av dem, av Bunevicius och medarbetare [62] samt Ji och medarbetare [74], hade SCID-I som referensstandard, medan Adouard och medarbetare använde MINI och Bergink och medarbetare CIDI.

Bunevicius och medarbetare följde 230 gravida kvinnor som kom för rutinundersökning i Litauen [62]. EPDS fylldes i vid ett tillfälle under varje trimester. Prevalensen av depression under första trimestern var 5 procent och under de två övriga trimestrarna 3 procent. Ji och medarbetare undersökte 708 gravida amerikanska kvinnor som tidigare haft någon psykisk sjukdom med 4–6 veckors intervaller [74]. Antalet deltagare varierade mellan olika mättillfällen; prevalensen depression under första trimestern var 15 procent.

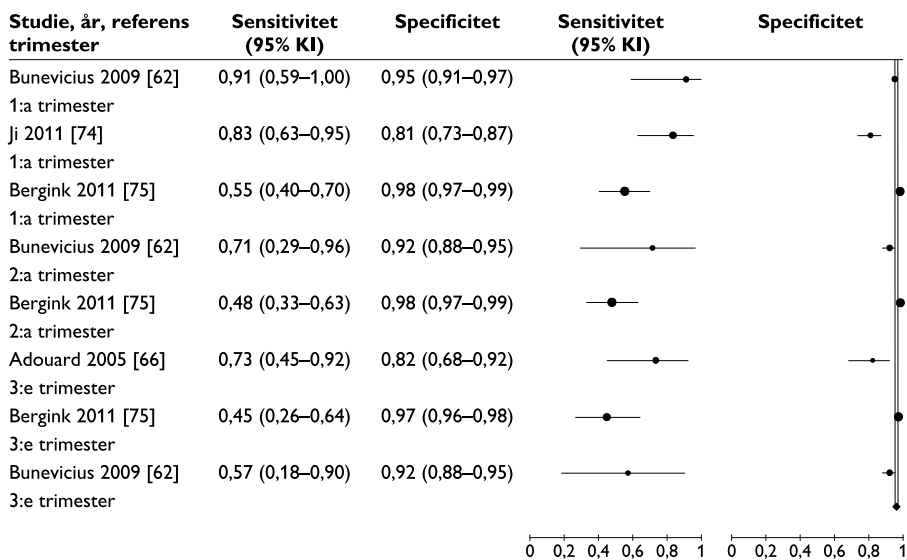
Adouard och medarbetare undersökte 60 franska kvinnor som bedömts ha en obstetiskt riskfylld graviditet under sista trimestern [66]. Alla kvinnorna hade bedömts ha en obstetriskt riskfylld graviditet. Bergink och medarbetare följde 845 gravida kvinnor i Nederländerna, som fyllde i EPDS en gång under varje trimester i samband med rutinundersökning [75].

### ***Sammanvägda resultat***

Analysen försvårades av att studierna mätte med EPDS vid olika graviditetsveckor. Dessutom använde Ji och medarbetare tröskelvärde 12 bara under den första trimestern.

Som framgår av Figur 3.5.12.b gick det inte att bedöma sensitiviteten för EPDS under graviditet. Tre studier redovisade t ex sensitivitet för första trimestern [62,74,75] och den låg mellan 55 och 91 procent. Under tredje trimestern var sensitiviteten i en studie under 50 procent.

Specificiteten varierade i mindre utsträckning under graviditetens gång och låg över 80 procent i samtliga mätningar.



**Sensitivitet** ej sammanvägd  
 Sammanvägd **specificitet** = 0,96 (0,95 till 0,97)

**Figur 3.5.12b** Forest plot. Sensitivitet och specificitet för EPDS med strukturerade och semistrukturerade intervjuer som referensstandard för gravida kvinnor.

### *Bedömning av evidensstyrka*

Det går inte att dra några slutsatser om sensitiviteten. Möjligen antyder data att sensitiviteten sjunker vartefter graviditeten framskrider.

Vad beträffar specificiteten redovisade två studier lägre värden än de övriga [66,74], vilket sänkte evidensstyrkan ett steg. För övrigt låg specificiteten stabilt över graviditeten.

Resultaten sammanfattas i Tabell 3.5.7b nedan.

**Tabell 3.5.7b** Resultatsammanställning för screening av gravida med EPDS med olika strukturerade och semistrukturerade intervjuer som referensstandard.

Utfallsmått	Antal studier Antal patienter	Sammanvägt estimat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	4 studier n=1 752	Bör inte vägas samman	Otillräckligt ⊕○○○	Studiekvalitet (-1) Samstämmighet (-2)
Specificitet	4 studier n=1 752	96% (95; 97)	Måttligt starkt ⊕⊕⊕○	Samstämmighet (-1)

KI = Konfidensintervall

### *Evidensgraderade resultat*

För kvinnor under första året efter förlossningen har EPDS med tröskelvärde 12/13 en sensitivitet på 72 procent (95 % KI, 69 till 76) för depression med SCID-I som referensstandard (begränsat vetenskapligt underlag ⊕⊕○○) och en specificitet på 88 procent (95 % KI, 86 till 89) (begränsat vetenskapligt underlag ⊕⊕○○).

Det går inte att bedöma sensitivitet för depression hos gravida kvinnor med EPDS med strukturerade eller semistrukturerade intervjuer som referensstandard eftersom studierna har låg samstämmighet (otillräckligt vetenskapligt underlag ⊕○○○). Specificiteten är 96 procent (95 % KI, 95 till 97) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

### **Hospital Anxiety and Depression Scale (HADS-D)**

HADS är ett självskattningsformulär som består av 14 frågor. Det utvecklades i Storbritannien under tidigt 1980-tal för att underlätta upptäckten av psykiska problem hos patienter inom medicinsk slutenvård [77].

Syftet med formuläret är att undvika sådana depressionsfrågor som kan förväxlas med somatiska sjukdomssymtom och därigenom minimera falskt positiva resultat. De 14 frågorna är fördelade på två delskalor, ångest (HADS-A, som inte ingår i granskningen) och depression (HADS-D) med sju frågor vardera. Den svarande uppmanas att markera

på en fyrgradig svarsskala (från 0 till 3 där varje grad representerar en beskrivning av frekvens eller intensitet) det alternativ som stämmer bäst med frågans innehåll under den senaste veckan. Resultaten av de sju frågorna summeras och poängen kan variera mellan 0 och 21. Ett resultat över 7 poäng definieras som risk för egentlig depression.

### ***Beskrivning av underlaget***

Av 109 studier som lästes i fulltext exkluderades 97, huvudsakligen beroende på irrelevant frågeställning, population eller språk eller för att det var för lång tid mellan index- och referenstesterna. Av de tolv inkluderade studierna bedömdes tre hålla låg studiekvalitet [38,48,78]. Anledningarna var bristande blindning samt att orsaker till betydande bortfall inte redovisades.

Nio studier utgjorde därmed det vetenskapliga underlaget för bedömning av depressionsmodulen i HADS [79–87] (Tabell 3.5.23). Sammantaget bedömdes sju av studierna nå medelhög studiekvalitet och två av studierna [80,86] hög studiekvalitet. Totalt ingick 1 669 studiedeltagare från öppen- och slutenvård i olika delar av världen.

En kanadensisk studie inkluderade 189 patienter som legat på akutmedicinsk avdelning minst sju dagar [85]. Referensstandard var SCAN. En svaghet med studien var att den inte redovisade vilken personalkategori som gjorde intervjuerna eller vilken träning dessa hade.

Fyra studier var genomförda i sjukhusbaserad öppenvård, tre av medelhög kvalitet [79,81,84] och en av hög kvalitet [86]. Den första omfattade 298 patienter med diabetes på en specialistklinik i Frankrike [79]. MINI användes som referenstest, och psykologer under utbildning gjorde intervjun. Den andra studien undersökte 100 australiensiska patienter som fått hjärnskador efter olycksfall ett halvt till fem och ett halvt år tidigare [81]. SCID-I var referenstest och genomfördes av två läkare under utbildning. De var tränade i SCID-I och hade handledning av en klinisk psykolog. Blindning av undersökarna för indextestens resultat nämndes inte i publikationen, vilket drog ner studiekvaliteten.

I den tredje studien intervjuades 269 patienter med bröstcancer i hemmet [84]. Referensstandard var PSE (Present State Examination). PSE genomfördes av tre intervjuare vars samstämmighet kontrollerats med samskattning av videofilmer för 30 procent av intervjuerna. Den fjärde studien screenade 67 patienter i väntrummet till olika kliniker (obstetrik-gynekologi, invärtesmedicin, psykiatri, öron-näsa-hals-klinik samt smärt-klinik) vid ett sjukhus i Spanien [86]. Femtio procent av patienterna använde psykotropa läkemedel. Två av artikelförfattarna inkluderade patienterna och gav indextesten, tio mentalskötare gjorde intervjuerna (klinisk intervju och PRIME-MD) efter träning i diagnostik av en psykiater.

Två studier utfördes på patienter en tid efter öppen eller slutenvård vid specialistklinik [83,87]. Den första undersökte 82 patienter med tinnitus 24 månader efter första besöket på en öronmottagning på ett universitetssjukhus. En erfaren psykiater tränad i proceduren genomförde alla SCID-I-intervjuerna. Den andra studien undersökte patienter som vårdats pga kranskärlssjukdom vid ett sjukhus i södra Australien. Drygt 500 patienter tillfrågades per brev och telefon om sitt intresse av att delta. Knappt 40 procent av dem inkluderades i studien. Undersökningen ägde rum tre månader efter utskrivningen. HADS skickades ut per post, och en av författarna gjorde alla intervjuerna (MINI) per telefon inom tre dagar.

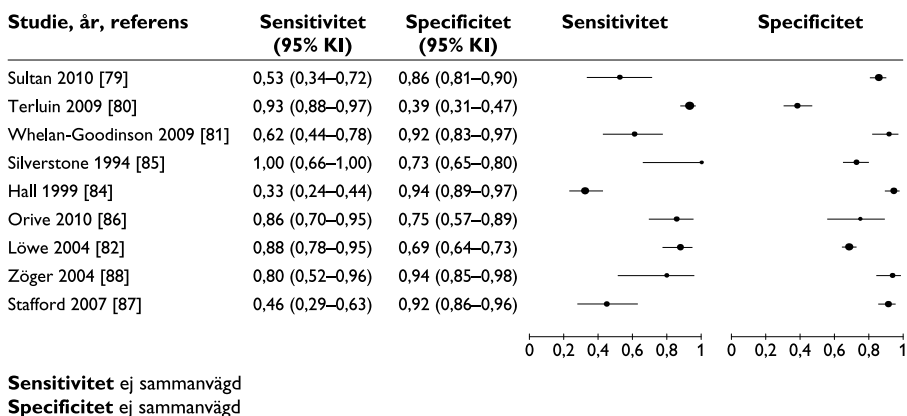
Två studier utfördes i primärvård [80,82]. Den första, från Nederländerna, screenade 307 sjukskrivna patienter eller patienter med psykologiska problem [80]. Referenstesten var CIDI och utfördes av fem lekmän med träning vid WHO CIDI-training Centre. Den andra studien var från Tyskland och studerade patienter på en somatisk poliklinik vid ett sjukhus i Tyskland samt på 12 distriktsläkarmottagningar [82]. Referenstest var SCID-I, och de fyra intervjuerna var tränade och testade med videoinspelningar.

### ***Sammanvägd sensitivitet och specificitet***

Sensitivitet och specificitet från de nio studierna redovisas i Figur 3.5.13. Både sensitiviteten och specificiteten varierade kraftigt mellan studierna.

Sensitiviteten låg mellan 33 och 100 procent och specificiteten mellan 39 och 94 procent (Tabell 3.5.8).

En möjlig orsak till den stora variationen i resultaten för HADS är tillförlitligheten i den referensstandard som använts. Frågan blev då vad som händer om man skärper underlaget till att bara omfatta studier där SCID-I tjänat som referensstandard och utförts av tränad personal. Enligt dessa kriterier återstod tre av de ursprungliga nio studierna. Då varierade både sensitivitet och specificitet något mindre; mellan 62 och 88 procent respektive 69 och 94 procent. Spridningen i resultaten var fortfarande påfallande trots samma referensstandard.



**Figur 3.5.13** Forest plot. Sensitivitet och specificitet för HADS-D. Underlaget utgörs av nio studier.

Vi bedömde att det inte var rimligt att väga samman resultaten för vare sig sensitivitet och specificitet pga den stora spridningen mellan studierna. Sannolikt kommer inte heller ytterligare studier bidra till att spridningen minskar.

### Bedömning av evidensstyrka

Resultaten tyder på att sensitivitet och specificitet för HADS är starkt beroende på sammanhanget, framför allt de individuella patienternas

tolkningar. Studierna var välgjorda och vi bedömde att risken för bias var liten. Överförbarheten begränsades av att studierna använde olika referensstandarder, vilket sänker det vetenskapliga stödet med ett steg.

Vår slutsats är därmed att vi med hög säkerhet kan uttala att sensitivitet och specificitet för HADS inte går att fastställa.

**Tabell 3.5.8** Resultatsammanställning för HADS-D.

Effektmått	Antal deltagare Antal studier	Sammanvägt resultat	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	n=1 669 9 studier	Kan inte fastställas	Måttligt starkt ⊕⊕⊕○	Bristande över- förbarhet (-1)
Specificitet	n=1 669 9 studier	Kan inte fastställas	Måttligt starkt ⊕⊕⊕○	Bristande över- förbarhet (-1)

### *Evidensgraderat resultat*

Det går inte att fastställa sensitivitet och specificitet för HADS som formulär för riktad screening för depression hos vuxna jämfört med strukturerad eller semistrukturerad intervju eftersom samstämmigheten mellan studierna är låg (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

### **Patient Health Questionnaire (PHQ-9)**

PHQ-9 är ett självskattningsformulär med nio frågor som lanserades år 2001. PHQ-9 utvecklades från depressionsmodulen i intervjuformuläret PRIME-MD. Varje fråga representerar ett kriterium för egentlig depression enligt DSM-IV. Den svarande uppmanas på en fyragradig svarsskala markera hur många dagar under de senaste två veckorna personen känt på det sätt som beskrivits i frågan (från 0=inte alls till 3=nästan varje dag). Svaren vägs samman enligt en algoritm. PHQ-9 används även som en kontinuerlig skala där det etablerade tröskelvärdet är 10.

Vi har utvärderat PHQ-9 för algoritm respektive tröskelvärde separat.



### *Beskrivning av underlaget*

Av 74 studier, som lästes i fulltext, exkluderades 61 beroende på irrelevant fråga, population, språk eller att det var för lång tid mellan index och referenstest. Av de tretton som uppfyllde kriterierna för inkludering hade två låg studiekvalitet eftersom det saknades uppgift om blindning samt att inte samtliga deltagare fick referenstesten [53,88].

Elva studier utgjorde det vetenskapliga underlaget [82,86,87,89–96]. Av dessa bedömdes nio studier nå medelhög studiekvalitet [82,87,90–96] och två av studierna hög studiekvalitet [86,89] (Tabell 3.5.24). Tio av dem undersökte den diagnostiska tillförlitligheten för algoritmberad PHQ-9 och sex för PHQ-9 med tröskelvärdet 10.

Studierna omfattade drygt 3 000 patienter och genomfördes inom sjukhusbaserad öppen- och slutenvård eller primärvård. Tre studier ägde rum inom medicinsk slutenvård [90,91,95]. Den första gjordes vid en invärtesmedicinsk enhet på universitetssjukhuset i Genève på cirka 1 000 patienter med invärtesmedicinska sjukdomar, cancer eller infektionssjukdomar [90]. Referenstestet var en psykiatrisk bedömning enligt DSM-IV-kriterier och utfördes av två psykiatrer samt en psykiater under utbildning. Den andra studien omfattade 141 inneliggande patienter på en hudklinik i Rom [91]. SCID-I var referenstest och genomfördes av tränad psykiatrisk personal. Den tredje studien bedrevs vid universitetssjukhus på 1 003 slumpvis utvalda patienter som lades in på invärtesmedicinska och kirurgiska avdelningar [95]. En av författarna, en erfaren klinisk psykolog, genomförde en intervju med alla patienter inom 48 timmar efter att de hade fyllt i PHQ-9.

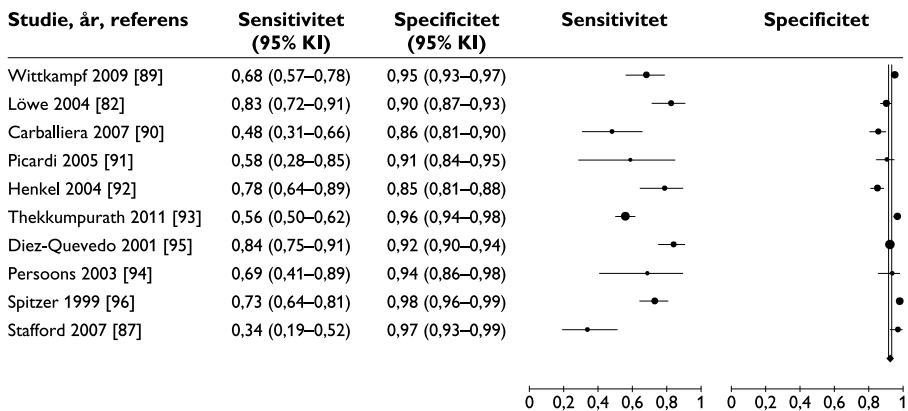
Fyra studier var inriktade på sjukhusbaserad öppenvård [86,87,93,94]. Den första omfattade knappt 800 konsekutiva patienter med cancer som hade minst 15 poäng på HADS [93]. Referenstest var SCID-I utförd per telefon av specialtränade psykologer och sjuksköterskor med veckovis handledning med videotapebedömningar. Den andra studien är beskriven i avsnittet om HADS [86]. Den tredje studien undersökte patienter som vårdats pga kranskärlssjukdom vid ett sjukhus i södra Australien [87]. Patienterna (n=528) tillfrågades per brev och telefon, och undersökningen (n=193) ägde rum tre månader efter utskrivningen. PHQ-9-formu-

läret skickades ut per post, och en av författarna gjorde alla intervjuerna (MINI) per telefon. Den fjärde studien ägde rum vid en öron-näsa-halsklinik på ett universitetssjukhus i Belgien [94]. Konsekutiva öppenvårdspatienter (n=285) med yrsel, av vilka 268 besvarade PHQ-9 och av vilka 97 slumpvis utvalda blev intervjuade med hjälp av MINI, utgjorde materialet. En tränad psykolog gjorde intervjuerna fem dagar efter PHQ-9.

Fyra studier utfördes i primärvård [82,89,92,96]. En tysk studie har tidigare beskrivits i avsnittet om HADS [82]. Henkel och medarbetare inkluderade 470 konsekutiva patienter från väntrummen för 18 vårdcentraler i Tyskland [92]. Wittkamp och medarbetare studerade 664 patienter bokade för återbesök hos 23 distriktsläkare i Nederländerna [89]. Patienterna skulle vid föregående besök (tre månader innan inkludering i studien) ha uppvisat psykiska problem eller oförklarade somatiska besvär. Referenstestet var SCID-I som gjordes vid en telefonintervju av klinisk personal som stod under handledning. Den fjärde studien genomfördes på 8 primärvårdskliniker där 62 primärvårdsläkare och 21 invärtesmedicinare arbetade [96]. Innan patienterna, med diagnoserna högt blodtryck, ledbesvär, diabetes och lungsjukdomar, besökte läkaren fick de fylla i PHQ-9-formuläret. Referenstesten var SCID-I och utfördes av en klinisk psykolog och tre socionomer.

### ***Sammanvägd sensitivitet och specificitet***

Sensitivitet och specificitet från de tio studierna som undersökte algoritmbaserad PHQ-9 redovisas i Figur 3.5.14a. Spridningen i sensitivitet mellan studierna var stor, mellan 34 och 84 procent. Specificiteten var genomgående hög, mellan 85 och 98 procent.

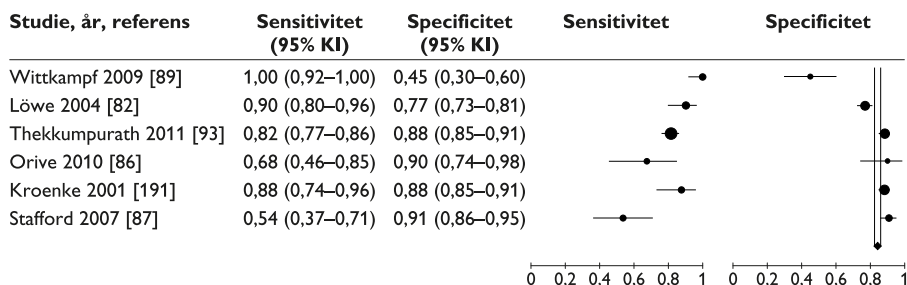


Sensitivitet ej sammanvägd  
 Sammanvägd **specificitet** = 0,93 (0,92 till 0,93)

**Figur 3.5.14a** Forest plot. Sensitivitet och specificitet för algoritmbaserad PHQ-9.

Spridningen i sensitivitet är för stor för att det ska vara meningsfullt att väga samman resultaten från studierna. Däremot är det rimligt att väga samman resultaten för specificitet som då blir 93 procent (95 % KI, 92 till 93).

När PHQ-9 utvärderades med tröskelvärde var fortfarande sensitiviteten osäker och varierade mellan 54 och 100 procent. Det var inte meningsfullt att väga samman resultaten. Specificiteten var likartad i alla studierna och det sammanvägda resultat var 84 procent (Figur 3.5.14b).



Sensitivitet ej sammanvägd  
 Sammanvägd **specificitet** = 0,84 (0,83 till 0,86)

**Figur 3.5.14b** Forest plot. Sensitivitet och specificitet för PHQ-9 med tröskelvärde 10.

### Bedömning av evidensstyrka

Generellt höll studierna hög kvalitet, vilket medförde att risken för systematiska fel var liten. Överförbarheten försämrades av att det var olika yrkeskategorier som genomförde den och att referensstandarderna var olika i studierna. Detta medförde en sänkning av evidensstyrkan med ett steg.

Samstämmigheten för sensitivitet var mycket låg. Vi bedömer att ytterligare studier inte skulle öka den. För specificitet var samstämmighet och precision god (Tabell 3.5.9a och 3.5.9b).

**Tabell 3.5.9a** Resultatsammanställning för algoritmbaserad PHQ-9.

Effektmått	Antal deltagare Antal studier	Sammanvägt resultat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	n=3 225 10 studier	Saknas	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (–1)
Specificitet	n=3 225 10 studier	93% (92; 93)	Måttligt starkt ⊕⊕⊕○	Bristande överförbarhet (–1)

KI = Konfidensintervall

**Tabell 3.5.9b** Resultatsammanställning för PHQ-9 med tröskelvärde 10.

Effektmått	Antal deltagare Antal studier	Samman- vägt resultat (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet	n=2 778 6 studier	Saknas	Måttligt starkt ⊕⊕⊕○	Bristande över- förbarhet (-1)
Specificitet	n=2 778 6 studier	84% (83; 86)	Måttligt starkt ⊕⊕⊕○	Bristande över- förbarhet (-1)

KI = Konfidensintervall

### *Evidensgraderade resultat*

- Det går inte att bedöma vilken sensitivitet algoritmbaserad PHQ-9 har för riktad screening av depression hos vuxna jämfört med strukturerad eller semistrukturerad intervju eftersom samstämmigheten är låg mellan studierna (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Specificiteten för algoritmbaserad PHQ-9 för riktad screening av depression hos vuxna i ett brett antal populationer är 93 procent (95 % KI, 92 till 93) jämfört med strukturerad eller semistrukturerad intervju (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Det går inte att bedöma sensitiviteten för PHQ-9 med tröskelvärde 10 för riktad screening av depression hos vuxna jämfört med strukturerad eller semistrukturerad intervju eftersom samstämmigheten är låg mellan studierna (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- Specificiteten för PHQ-9 med tröskelvärde 10 för riktad screening av depression hos vuxna i ett brett antal populationer är 93 procent (95 % KI, 92 till 93) jämfört med strukturerad eller semistrukturerad intervju (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

### **Major Depression Inventory (MDI)**

MDI är ett självskattningsformulär på tolv frågor som utvecklades i Danmark i början på 2000-talet [193]. Varje fråga representerar kriterier för egentlig depression enligt DSM-IV och ICD-10. Den svarande

uppmanas markera på en sexgradig svarsskala i vilken utsträckning hon under de senaste två veckorna känt på det sätt som beskrivits i frågan (0=ingen del av tiden; 5=nästan hela tiden). Svaren vägs samman enligt en algoritm som beräknar om de självskattade resultaten tyder på egentlig depression. Det finns specifika algoritmer för både DSM-IV och ICD-10.

### ***Beskrivning av underlaget***

Av åtta studier som lästs i fulltext bedömdes ingen uppfylla inklusionskriterierna. Orsakerna var irrelevant frågeställning, referensstandard och population.

### ***Evidensgraderat resultat***

Det går inte att bedöma sensitivitet och specificitet för MDI som formulär för riktad screening för depression med strukturerad eller semistrukturerad intervju som referensstandard eftersom det saknas studier.

## **Seasonal Health Questionnaire (SHQ)**

SHQ avser att täcka tre olika diagnostiska kriterier för säsongbundet affektivt syndrom (Seasonal Affective Disorder; SAD): kriterier enligt DSM-III-R [97], DSM-IV [2] och ICD-10 [98,99]. De så kallade Rosenthalkriterierna för SAD [100] uteslöts, eftersom de snarast uppfattades mäta årstidskänslighet. SHQ omfattar frågor om depressiva symptom, antal episoder samt årstidsmönster för depressiva episoder under de senaste 10 åren.

### ***Beskrivning av underlaget***

Två studier om SHQ identifierades. En av dem exkluderades pga oklarhet kring referensstandard för SAD [99]. Den andra studien inkluderades men bedömdes ha låg kvalitet [101] (Tabell 3.5.25). Det har dessutom visat sig omöjligt att få tillgång till beräkningsmallen för formuläret vilket försvårar tolkningen av resultaten.

Thompson och medarbetare rekryterade cirka 800 patienter från väntrummet hos allmänläkare [101]. De fick diagnosen SAD enligt SHQ om kriterier från ICD-10, DSM-III-R eller DSM-IV var uppfyllda. Ett slumpmässigt urval patienter intervjuades med SCID-I per telefon

(DSM-III-R-kriterier). Sensitiviteten var 59 procent och specificiteten var 97 procent för att identifiera SAD med SHQ. Studien undersökte samtidigt SPAQ, se nedan.

### ***Bedömning av evidensstyrka***

Det finns för få studier av tillräcklig kvalitet för att kunna bedöma om SHQ är tillförlitligt för att identifiera årstidsbunden depression, SAD, i kliniska populationer med semistrukturerad eller strukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **Seasonal Pattern Assessment Questionnaire (SPAQ)**

Seasonal Pattern Assessment Questionnaire (SPAQ) bygger på Rosenthals kriterier [100,102] och avser SAD eller lättare former av SAD.

Patienterna skattar fyra kategorier av frågor:

- Grad av säsongsbunden påverkan av stämningsläge, sömn, socialt intresse, aptit, energi och vikt (totalt 6 frågor med skalstegen 0–4).
- I vilken omfattning detta utgör ett problem (0–5; en skattning med minst 2 = måttligt utgör tröskelvärdet för SAD).
- Kalendariskt mönster av säsongsvariation, t ex vinter- eller sommarmönster.
- Känslighet för olika väderförhållanden.

### ***Beskrivning av underlaget***

Av sju studier som lästes i fulltext exkluderades fem. Anledningarna var att de berörde andra frågor, icke-klinisk population, eller hade oklara kriterier för SAD [103–107]. De två studier som uppfyllde inklusionskriterierna bedömdes ha låg studiekvalitet [101,108]. Båda studierna visade en låg sensitivitet och en hög specificitet (Tabell 3.5.26).

Mersch och medarbetare rekryterade 45 patienter med SAD och 48 med egentlig depression, vilka diagnostiserats med checklista enligt DSM-III-R-kriterier samt 46 psykiatriska öppenvårdspatienter med BDI lägre än 10 poäng och en kontrollgrupp bestående av 25 annonsrekryterade personer som inte haft egentlig depression under det senaste året [108]. I denna studie var sensitiviteten 44 procent och

specificiteten 94 procent för att identifiera SAD med SPAQ och med ovan beskrivna kriterier som referensstandard.

Den tidigare nämnda studien av Thompson och medarbetare [101] undersökte även tillförlitligheten för SPAQ. Sensitiviteten var 38 procent och specificiteten 79 procent för att identifiera SAD med SPAQ och med SCID-I som referensstandard.

### ***Bedömning av evidensstyrka***

Det finns för få studier av god kvalitet för att bedöma sensitivitet och specificitet för SPAQ som formulär för riktad screening för SAD med strukturerad eller semistrukturerad intervju som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## **Riktad screening av bipolärt syndrom**

Den fråga som skulle besvaras var vilken tillförlitlighet, mätt som sensitivitet och specificitet, sådana formulär har jämfört med en referensstandard. Som framgått i inledningen av kapitlet accepterade vi strukturerade och semistrukturerade intervjuer som referensstandard. De formulär som ingick i granskningen var Hypomania Checklist-32 (HCL-32) och Mood Disorder Questionnaire (MDQ).

### **Hypomania Checklist-32 (HCL-32)**

Hypomania Checklist-32 (HCL-32) syftar till att identifiera om patienten haft hypomani någon gång under livet. Formuläret är avsett att underlätta diagnostisering av bipolärt syndrom typ II och andra bipolärspektrum syndrom [109].

HCL-32 innehåller en checklista med 32 frågor om hypomana symtom som ska besvaras med ”ja” eller ”nej”. Antalet ”ja”-svar på de 32 frågorna i checklistan summeras till en totalpoäng. Ursprungsartikeln föreslår att HCL-32-poäng 14 bäst särskiljer bipolära och unipolära syndrom [109]. För HCL-32 undersökte vi även formulärets förmåga att identifiera bipolärt syndrom typ II.



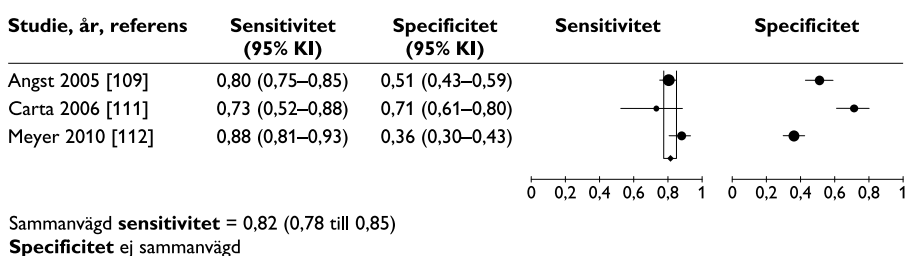
### Beskrivning av underlaget

Av nio artiklar som lästes i fulltext exkluderades fem. Av de fyra som uppfyllde inklusionskriterierna hade en studie låg kvalitet. Det berodde främst på att prövarna inte använde någon strukturerad intervju som referenstest och att blindningen mellan index- och referenstest var oklar [110]. Det vetenskapliga underlaget består därmed av tre studier som alla bedömdes ha medelhög kvalitet [109,111,112] (Tabell 3.5.27).

Totalt innefattade studierna 910 patienter. I studien av Angst och medarbetare rekryterades konsekutivt drygt 400 patienter med bipolärt eller unipolärt syndrom [109]. Cirka hälften var öppenvårdspatienter från en privatklinik för affektiva sjukdomar i Italien medan de övriga fanns i öppen och slutenvård vid specialenheter för affektiva sjukdomar i Umeå och Sunderbyn. Carta och medarbetare rekryterade cirka 125 konsekutiva patienter från en psykiatrisk öppenvårdsmottagning i Italien [111]. Några sökte för psykisk ohälsa och andra för psykiatrisk bedömning inför medicinsk behandling som kunde innebära ökad risk för psykisk ohälsa (t ex interferonbehandling) eller för att få friskhetsintyg. I den tredje studien, av Meyer och medarbetare, ingick cirka 360 konsekutiva patienter med varierande diagnoser vid sju psykiatriska kliniker i södra, östra och norra Tyskland [112].

### Sammanvägda resultat

I de tre studierna låg sensitiviteten mellan 73 och 88 procent och specificiteten mellan 36 och 71 procent (Figur 3.5.15).



**Figur 3.5.15** Forest plot. Spridning av sensitivitet och specificitet för HCL-32.

Den sammanvägda sensitiviteten var 82 procent (95 % KI, 78 till 85). Däremot var studierna så heterogena med avseende på specificitet att det inte var rimligt att väga samman resultaten.

Alla tre studierna undersökte om HCL-32 är tillförlitligt för att identifiera bipolärt syndrom typ II gentemot unipolär depression, men endast studien av Meyer och medarbetare redovisade sensitivitet och specificitet för tröskelvärde 14 poäng [112]. Sensitiviteten var i stort sett lika hög för bipolärt syndrom typ I och typ II medan specificiteten var lägre för bipolärt syndrom typ II.

### ***Bedömning av evidensstyrka***

Säkerheten i det sammanvägda estimatet för sensitivitet bedömdes med GRADE. Det fanns en risk för systematisk bias i underlaget eftersom urval och procedur skilde sig mellan två centra i en studie, det fanns oklarheter med avseende på blindning, internt och externt bortfall i en studie, samt att interbedömarreliabilitet saknades i två studier och var oklar i den tredje studien (Tabell 3.5.10). Det medförde att evidensstyrkan drogs ner ett steg.

Överförbarheten påverkades något av att en av studierna använde modifierade kriterier av SCID-I för att diagnostisera hypomani medan de andra två använde sedvanliga kriterier. Detta innebar ingen sänkning av evidensstyrkan.

**Tabell 3.5.10** Resultatsammanställning för HCL-32.

<b>Utfallsmått</b>	<b>Antal studier Antal patienter</b>	<b>Sammanvägt värde (95% KI)</b>	<b>Bedömning av vetenskapligt underlag</b>	<b>Kommentarer till poängavdrag</b>
Sensitivitet	3 studier n=910	82% (78; 85)	Måttligt starkt ⊕⊕⊕○	Bristande studie-kvalitet (-1)
Specificitet	3 studier n=910	Bör ej vägas samman	Otillräckligt ⊕○○○	

KI = Konfidensintervall

### *Evidensgraderade resultat*

För vuxna patienter i specialiserad vård för affektiv sjukdom och i psykiatrisk öppenvård har HCL-32 med tröskelvärde 14 poäng en sensitivitet på 82 procent (95 % KI, 78 till 85) för att identifiera bipolärt syndrom med SCID-I som referensstandard (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).

Det går inte att bedöma specificiteten för HCL-32 med tröskelvärde 14 poäng med SCID-I som referensstandard pga stor spridning i resultaten (otillräckligt vetenskapligt underlag ⊕○○○).

Det finns för få studier med tillräcklig kvalitet för att kunna bedöma om HCL-32 med tröskelvärde 14 poäng identifierar bipolärt syndrom typ II gentemot unipolärt syndrom (otillräckligt vetenskapligt underlag ⊕○○○).

### **Mood Disorder Questionnaire (MDQ)**

MDQ är avsett att identifiera förekomst av bipolär episod någon gång under livet. MDQ består av tre sektioner: frågor om 13 maniska/hypomana symtom, fråga om några symtom inträffat under samma tidsperiod, samt fråga om grad av funktionsnedsättning på skalan: inga–mindre–måttliga–allvarliga problem. I ursprungsartikeln definierades tröskelvärde för bipolärt syndrom som 7 symtom, jakande svar på fråga om tidsperiod samt minst måttlig funktionsnedsättning [113].

Både under- och överdiagnostik av bipolärt syndrom kan få allvarliga konsekvenser. Förutom frågan om hur tillförlitligt MDQ är jämfört med referensstandard för diagnoserna bipolärt syndrom typ I och typ II sökte vi därför även svar på två ytterligare frågor:

- Kan sensitiviteten för bipolärt syndrom typ II ökas genom att modifiera kriterierna? Hypomani ger mindre uttalade symtom och riskerar lättare att underdiagnostiseras. Flera författare har föreslagit att kriteriet på funktionsnedsättning ska tas bort eftersom detta inte krävs för att ställa diagnosen hypomani. Ett flertal studier har därför undersökt om sensitiviteten för bipolärt syndrom typ II kan ökas genom att modifiera kriterier för tröskelvärde till lägre krav på funktions-

nedsättning (det så kallade Genève-kriteriet). Vi har granskat studier baserade på både original- och Genève-kriterierna.

- Vilka diagnoser är förknippade med risk för hög frekvens falskt positiva med MDQ? I den allmänna befolkningen [114] och i blandade psykiatriska populationer [115] ger MDQ en hög frekvens falskt bipolärt positiva (överdiagnostik). Det är viktigt att klarlägga vilka diagnoser som är associerade med falskt positiva resultat med MDQ.

### ***Beskrivning av underlaget***

Av 30 artiklar som lästes i fulltext inkluderades 15 som utgick från standardalgoritmen för MDQ. En av studierna, som redovisades i två artiklar, hade hög studiekvalitet [115,116]. Sju av studierna hade medelhög kvalitet [112,117–120,122,128], och sex låg kvalitet [113,123–127]. Tre av studierna bedömde även tillförlitligheten hos MDQ när Genève-kriterierna tillämpades [115,117,128] (Tabell 3.5.28 och 3.5.29).

Orsaker till den höga frekvensen falskt positiva med MDQ analyserades i två artiklar [129,130]. Analyserna byggde på data från de patienter som inkluderats i Zimmerman och medarbetares första rapport [115] (Tabell 3.5.29).

Totalt omfattade studierna 2 105 patienter från olika länder och med olika populationer. Gervasoni och medarbetare inkluderade 146 patienter med uni- eller bipolärt syndrom från psykiatrisk öppenvård i Genève [117]. Graves och medarbetare rekryterade 228 traumatiserade patienter, huvudsakligen afroamerikaner, från primärvård i USA [119]. Hardoy och medarbetare inkluderade 154 patienter konsekutivt från psykiatrisk öppenvård i Italien [118]. Kemp och medarbetare undersökte 164 nyligen intagna till fängelse i Ohio, USA [120]. Meyer och medarbetare rekryterade konsekutivt 440 patienter från psykiatriska kliniker i olika delar av Tyskland [112]. Sharma och medarbetare inkluderade 125 nyblivna mammor med anamnes på tidigare egentlig depression, bipolärt syndrom typ I eller II från en förlossningsklinik ansluten till psykiatrisk vård i Kanada [122]. I studien av Weber Rouget och medarbetare inkluderades 96 patienter med antingen bipolärt eller unipolärt syndrom från psykiatrisk öppenvård i Genève [128]. Zimmerman och medarbetare rekryte-

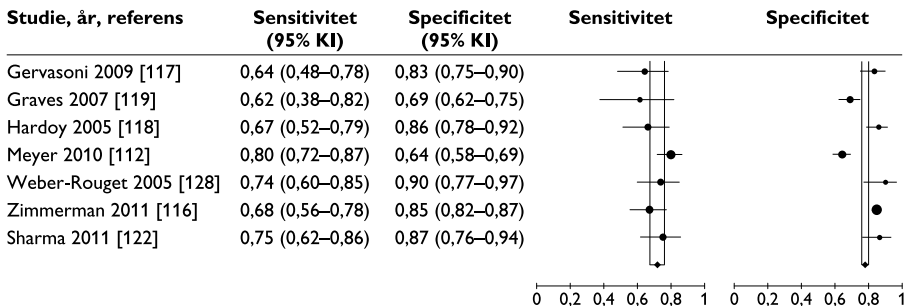
rade 752 patienter från en psykiatrisk öppenvårdsmottagning, associerad till ett akademiskt center i Providence, USA [116].

### *Sammanvägda resultat*

I en preliminär analys av data fann vi att studien av Kemp och medarbetare avvek från övriga [120]. Populationen var speciell (intagna på fängelse) och den strukturerade intervjun MINI var referenstest till skillnad mot de andra studierna som använt SCID-I. För att renodla analysen togs därför studien av Kemp och medarbetare bort från det vetenskapliga underlaget.

Sensitiviteten i de sju kvarvarande studierna låg mellan 62 och 80 procent när standardkriterierna användes. Specificiteten låg mellan 64 och 90 procent (Figur 3.5.16).

Det sammanvägda estimatet för sensitivitet för bipolärt syndrom oavsett typ och med standardkriterier blev 72 procent (95 % KI, 68 till 76). Den sammanvägda specificiteten blev 78 procent (95 % KI, 76 till 80).



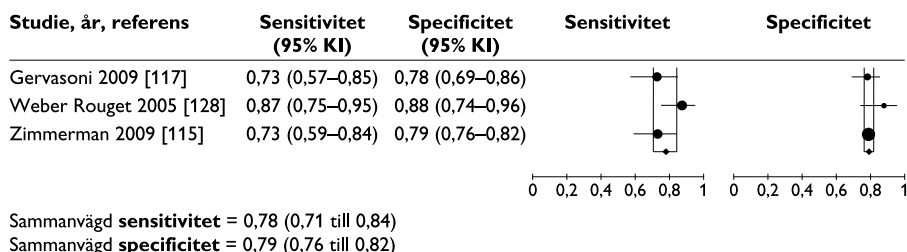
Sammanvägd **sensitivitet** = 0,72 (0,68 till 0,76)

Sammanvägd **specificitet** = 0,78 (0,76 till 0,80)

**Figur 3.5.16** Forest plot. Spridning av sensitivitet och specificitet för MDQ med standardkriterier, alla typer av bipolära syndrom.

När standardalgoritmen ersattes med Genève-algoritmen i tre studier blev sensitiviteten mellan 73 och 87 procent och specificiteten mellan 78 och 88 procent för alla typer av bipolära syndrom (Figur 3.5.17).

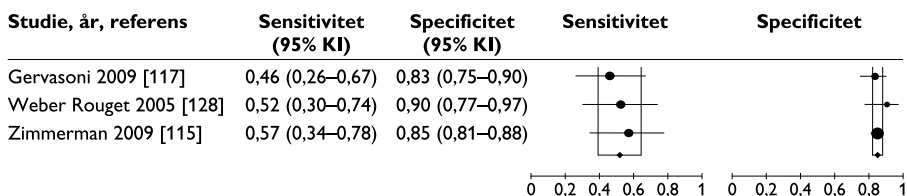
Den sammanvägda sensitiviteten för bipolärt syndrom oavsett typ men baserat på Genève-kriterierna var 78 procent (95 % KI, 71 till 84). Den sammanvägda specificiteten blev 79 procent (95 % KI, 76 till 82).



**Figur 3.5.17** Spridning av sensitivitet och specificitet för MDQ med modifierade kriterier, alla typer av bipolära syndrom.

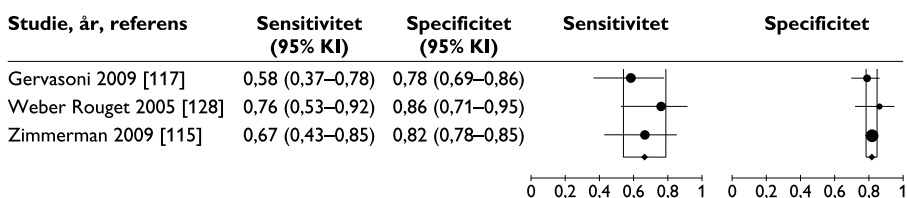
Tre studier undersökte hur väl MDQ kunde identifiera bipolärt syndrom typ II. Som framgår av nedanstående forest plot låg sensitiviteten för bipolärt syndrom typ II mellan 46 procent och 57 procent medan specificiteten låg mellan 83 och 90 procent med standardalgoritmen som grund. Med Genève-algoritmen låg sensitiviteten mellan 58 och 76 procent och specificiteten mellan 78 och 86 procent (Figur 3.5.18).

När det gäller sensitivitet för att särskilja bipolär typ II var den 52 procent (95 % KI, 39 till 64) med standardkriterier och 67 procent (95 % KI, 54 till 78) med Genèvekriterierna. Vad gäller specificitet blev det sammanvägda estimatet 85 procent (95 % KI, 82 till 88) med standardkriterierna respektive 81 procent (95 % KI, 78 till 84) med Genève-kriterierna.



Sammanvägd **sensitivitet** = 0,52 (0,39 till 0,64)  
 Sammanvägd **specificitet** = 0,85 (0,82 till 0,88)

**Figur 3.5.18** Bipolärt syndrom typ II. Spridning av sensitivitet och specificitet för MDQ med standardkriterier.



Sammanvägd **sensitivitet** = 0,67 (0,54 till 0,78)  
 Sammanvägd **specificitet** = 0,81 (0,78 till 0,84)

**Figur 3.5.19** Bipolärt syndrom typ II. Spridning av sensitivitet och specificitet för MDQ med modifierade kriterier.

Diagnoser som är förknippade med förhöjd risk för att ge falskt positivt svar med MDQ redovisas i Tabell 3.5.30 i slutet av avsnittet. I Zimmerman och medarbetares studie var 65 personer falskt positiva och 363 sant bipolärt negativa [129]. Sammanfattningsvis visade studierna att borderline personlighetsstörning, ätstörningar, ADHD, missbruk/beroende, PTSD, specifika fobier, och impuls kontrollstörningar men inte egentlig depression eller dystymi var förenat med en förhöjd risk för falskt positivt svar (uttryckt som oddskvoter).

### **Bedömning av evidensstyrka**

Våra överväganden om evidensstyrkan är sammanfattade i Tabell 3.5.11.

För samtliga fyra jämförelser med MDQ fanns en viss risk för bias pga att bedömarens kompetens var oklar i några studier samt att det saknades interbedömarreliabilitet med avseende på SCID-I i flera studier. Det medförde en sänkning av evidensstyrkan med ett steg.

Eftersom studierna använde samma referensstandard och omfattade breda psykiatriska populationer med sammanlagt nästan 2 000 patienter var överförbarheten god.

Samstämmigheten var tillräckligt god för att inte innebära någon sänkning av evidensstyrkan. Däremot var precisionen i det sammanvägda resultatet mer eller mindre bristfällig vilket påverkade evidensstyrkan (Tabell 3.5.11).

**Tabell 3.5.11** Resultatsammanställning tillförlitlighet av standardalgoritmen av MDQ för alla typer av bipolärt syndrom.

Utfallsmått	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Sensitivitet, alla typer av bipolärt syndrom, standardalgoritmen	7 studier n=1 941	72% (68; 76)	Måttligt starkt ⊕⊕⊕○	Brister i studie-kvalitet (-1)
Alla typer av bipolärt syndrom, Genève-algoritmen	3 studier n=722	78% (71; 84)	Begränsat ⊕⊕○○	Brister i studie-kvalitet (-1) Bristande precision (-1)
Bipolärt syndrom typ II, standardalgoritmen	3 studier n=722	52% (39; 64)	Otillräckligt ⊕○○○	Brister i studie-kvalitet (-1) Bristande precision (-2)
Bipolärt syndrom typ II, Genève-algoritmen	3 studier n=722	67% (54; 78)	Otillräckligt ⊕○○○	Brister i studie-kvalitet (-1) Bristande precision (-2)

Tabellen fortsätter på nästa sida



**Tabell 3.5.11** fortsättning

Utfallsmått	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Specificitet, alla typer av bipolärt syndrom, standardalgoritm	7 studier n=1 941	78% (76; 80)	Måttligt starkt ⊕⊕⊕○	Brister i studie-kvalitet (-1)
Alla typer av bipolärt syndrom, Genève-algoritmen	3 studier n=722	79% (76; 82)	Måttligt starkt ⊕⊕⊕○	Brister i studie-kvalitet (-1)
Bipolärt syndrom typ II, standardalgoritmen	3 studier n=722	85% (82; 88)	Måttligt starkt ⊕⊕⊕○	Brister i studie-kvalitet (-1)
Bipolärt syndrom typ II, Genève-algoritmen	3 studier n=722	81% (78; 84)	Måttligt starkt ⊕⊕⊕○	Brister i studie-kvalitet (-1)

KI = Konfidensintervall

### *Evidensgraderade resultat*

- För vuxna patienter i psykiatrisk öppenvård och i specialiserad vård för affektiva syndrom, för nyblivna mödrar med tidigare affektivt syndrom och för patienter med psykiska trauma i primärvård har MDQ med tröskelvärde 7 poäng och kriteriet minst måttlig funktionsnedsättning en sensitivitet på 72 procent (95 % KI, 68 till 76) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○) för att identifiera bipolärt syndrom, med SCID-I som referensstandard. Specificiteten är 78 procent (95 % KI, 76 till 80) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○).
- För vuxna patienter i psykiatrisk öppenvård och i specialiserad vård för affektiva syndrom har MDQ med tröskelvärde 7 poäng och modifierat kriterium minst mild funktionsnedsättning en sensitivitet

på 78 procent (95 % KI, 71 till 84) (begränsat vetenskapligt underlag ⊕⊕○○) och specificitet på 79 procent (95 % KI, 76 till 82) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○) för att identifiera bipolärt syndrom, med SCID-I som referensstandard.

- För vuxna patienter i psykiatrisk öppenvård och i specialiserad vård för affektiva syndrom har MDQ med tröskelvärde 7 poäng och kriteriet minst måttlig funktionsnedsättning en specificitet på 85 procent (95 % KI, 82 till 88) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○) för att identifiera bipolärt syndrom typ II, med SCID-I som referensstandard. Det går inte att bedöma sensitiviteten eftersom studierna har för låg precision (otillräckligt vetenskapligt underlag ⊕○○○).
- För vuxna patienter i psykiatrisk öppenvård och i specialiserad vård för affektiva syndrom har MDQ med tröskelvärde 7 poäng och modifierat kriterium minst mild funktionsnedsättning en specificitet på 81 procent (95 % KI, 78 till 84) (måttligt starkt vetenskapligt underlag ⊕⊕⊕○) för att identifiera bipolärt syndrom typ II, med SCID-I som referensstandard. Det går inte att bedöma sensitiviteten eftersom studierna har för låg precision (otillräckligt vetenskapligt underlag ⊕○○○).

## Formulär för bedömning av svårighetsgrad av depression

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

### Beck Depression Inventory-I och -II (BDI-I, BDI-II)

För beskrivning av formulären BDI-I och BDI-II, se avsnittet om riktad screening av depression.

#### *Beskrivning av underlaget*

Av elva studier som lästes i fulltext exkluderades sju i huvudsak beroende på att de inte undersökte relevant frågeställning. Ytterligare en studie, av

Hiroe och medarbetare, exkluderades pga att den genomfördes i Japan [131]. Studien höll medelhög kvalitet.

Av de tre som inkluderades hade två låg studiekvalitet pga att blindning mellan indexmätt och referensmätt saknades samt pga oklarhet i tid mellan index- och referenstest [132,133]. En studie av Sprinkle och medarbetare hade medelhög studiekvalitet och utgör därmed det vetenskapliga underlaget [131,134] (Tabell 3.5.31). Studien byggde på ett bekvämlighetsurval av 137 studenter från en hälsovårdsmottagning för universitetsstudier i USA.

### ***Resultat***

I studien av Sprinkle och medarbetare var SCID-I referensstandard [134]. Här var korrelationen 0,83 mellan BDI-II och SCID-I med avseende på svårighetsgrad av depression vilket är ett starkt samband ( $r^2=0,69$ , dvs 69 procent förklarad varians).

### ***Bedömning av evidensstyrka***

Det finns för få studier med tillräcklig kvalitet för att bedöma om BDI-II är tillförlitligt för att bedöma svårighetsgrad av depression med DSM-IV kriterier som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## **Calgary Depression Scale for Schizophrenia (CDSS)**

För beskrivning av Calgary Depression Scale for Schizophrenia (CDSS) se avsnittet om riktad screening av depression.

### ***Beskrivning av underlaget***

Av 20 studier om CDSS, som lästes i fulltext exkluderades 19, i huvudsak beroende på att de använt skattningsskalor (MADRS, HRSD, BDI, PANSS-D) som referensstandard. En studie uppfyllde följaktligen inklusionskriterierna. I studien, av Müller och medarbetare, ingick 119 patienter med stabiliserad, akut schizofreni från en psykiatrisk slutenvårdsklinik i Tyskland [135]. Bedömarna hade adekvat kompetens, träning och interbedömarreliabilitet både för index- och referenstest (CGI-S). Däremot utfördes inte index- och referenstest blindat utan de gjordes av samma bedömare vid samma tillfälle. Det medför att studien bedöms ha låg

kvalitet. Korrelationen mellan de båda metoderna var  $r=0,80$  (Tabell 3.5.32).

### ***Bedömning av evidensstyrka***

Det finns för få studier med tillräcklig kvalitet för att bedöma om CDSS är tillförlitligt för att mäta svårighetsgrad av depression hos patienter med schizofreni med CGI-S eller SCID-I som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **Hamilton Rating Scale for Depression (HDRS-17)**

Hamilton Rating Scale for Depression (HDRS) är ett kliniskt bedömningsformulär som utvecklades på 1960-talet för att bestämma depressionsdjup och följa förlopp över tid hos patienter med belagd depressionsdiagnos [136,137].

Det finns flera versioner av Hamiltonskalan med varierande antal frågor. Vår granskning omfattade endast HDRS-17. Formuläret består av 21 påståenden varav skattningar av 17 påståenden summeras för att avspegla nivåer av depressionsdjup (HDRS-17). Varje påstående består först av ett ledord med efterföljande tre- eller femgradig skala där skalstegen getts förklarande text som beskriver ökande intensitet av symtomen (från 0 till 2 eller 0 till 4). Skattningen görs av en läkare eller psykolog.

### ***Beskrivning av underlaget***

Av 64 studier som lästes i fulltext exkluderades 59. Orsakerna var att frågeställningen inte var relevant, att ingen statistisk prövning presenterades för sambandet mellan HDRS och bedömningen av sjukdomens svårighetsgrad, att population eller språk inte fyllde inklusionskriteriet eller att det var för lång tid mellan index- och referenstest. Av de fem studier som utgjorde det vetenskapliga underlaget bedömdes fyra studier [138–141] vara av låg studiekvalitet pga oklarheter vad gällde blindning och hur data sammanvägts från olika studier.

En studie bedömdes ha medelhög kvalitet [142] (Tabell 3.5.40). Studien utfördes på en geriatrisk klinik i Danmark och omfattade 145 inneliggande äldre deprimerade (98 utan och 47 med demens). I studien analyserades samstämmighet mellan summapoäng för HDRS-17 och CGI-S.

Graden av samstämmighet beräknades enligt en Spearmans rangkorrelation. Korrelationen blev 0,85 mellan HDRS-17 och CGI-S. När patientmaterialet delats in i grupper med respektive utan demens beräknades korrelationen till 0,87 för dem utan och 0,78 för dem med demens. Samstämmigheten mellan bedömarna med avseende på CGI-S var 0,88.

De fyra studierna med låg kvalitet redovisade också korrelationsdata [138–141]. Samstämmigheten mellan HDRS-17 och CGI-S, uttryckt som en korrelation, låg mellan 0,43 och 0,87. Sex av de åtta korrelationerna var på samma nivå som den som redovisades i studien med medelhög kvalitet.

### ***Evidensgraderat resultat***

Det finns för få studier med tillräcklig kvalitet för att bedöma om HDRS-17 mäter svårighetsgrad på ett tillförlitligt sätt och med SCID-I eller CGI-S som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **Inventory of Depressive Symptomatology (IDS och QIDS)**

IDS utvecklades på 1980-talet i USA [143]. Bakgrunden var kritik mot Hamiltonskalan som inte ansågs ta hänsyn till alla kärnsymtom i egentlig depression (enligt DSM-IV). Ett annat syfte med IDS var att utveckla ett formulär som kunde användas både för expert- och självskattningar.

IDS består av 30 påståenden, som inkluderar alla kriterier för egentlig depression samt atypiska symtom. Varje påstående består av ett ledord följt av en fyrgradig skala där skalstegen getts förklarande text som beskriver ökande frekvens av symtomen (från 0 till 3). Resultaten summeras över alla påståendena.

IDS finns i två versioner, benämnda Clinician-Rated (IDS-C) och Self-Rated (IDS-SR). Det finns också en förkortad version, kallad Quick Inventory of Depressive Symptomatology (QIDS). Den består av 16 depressionsfrågor ur IDS och inkluderar frågor som täcker alla kriterier för egentlig depression enligt DSM-IV. Även denna förkortade

version finns som expertskattning (Clinician-Rated, QIDS-C) och som självskattning (Self-Report, QIDS-SR). För att tydligt markera vilken version som avses kallas den långa versionen ibland för IDS-30 och kortversionen för QIDS-16.

### ***Beskrivning av underlaget***

Vi identifierade en studie som uppfyllde inklusionskriterierna men som hade låg studiekvalitet [144] (se Tabell 3.5.33). Det är en amerikansk multicenterstudie med över 1 000 patienter med återkommande depressionsepisoder. Studien undersökte samstämmighet mellan patientskattade depressionssymtom med IDS-30-SR eller QIDS-16-SR och klinisk bedömning av svårighetsgrad. Samstämmigheten mellan patienternas och klinikernas bedömning av svårighetsgrad varierade mellan 0,63 och 0,70 under behandlingsfasen.

### ***Evidensgraderat resultat***

Det finns för få studier för att kunna bedöma om någon av de olika versionerna av IDS mäter depressionsdjup på ett tillförlitligt sätt med SCID-I eller CGI-S som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## **Montgomery-Åsberg Depression Rating Scale (MADRS)**

MADRS är ett kliniskt bedömningsformulär för att bestämma depressionsdjup och följa förlopp över tid [145]. MADRS utvecklades på 1970-talet ur Comprehensive Psychopathological Rating Scale, CPRS [146], och består av de 10 depressionsvariabler i CPRS som visat sig vara känsliga för förändring.

Varje depressionsvariabel består av ett ledord och en beskrivning av vad som efterfrågas. Bedömaren uppmanas att skatta på en sjugradig skala där skalstegen getts förklarande text som beskriver ökande intensitet av symtomen. Skattningen är tänkt att göras vid en psykiatrisk intervju. Både poäng på enskilda variabler och en summerad poäng kan användas för att bedöma djup och förändring.

### ***Beskrivning av underlaget***

Av 16 studier som lästs i fulltext exkluderades 14. De vanligaste orsakerna till att de exkluderades var att frågeställningen inte var relevant och att ingen statistisk prövning presenterades för sambandet mellan MADRS och bedömningen av sjukdomens svårighetsgrad. Båda studierna som inkluderades bedömdes ha låg studiekvalitet [141,147], främst pga att samma bedömare gjort både depressionsskattning och skattning av svårighetsgrad eller att information om procedur vid skattningar (tid mellan skattningar, om skattning varit blindad samt reliabilitetsbedömningar av ingående skattare) saknats i studierna (se Tabell 3.5.34).

Den ena av de två studierna analyserade sammanlagda data från tolv internationella kliniska behandlingsstudier [147]. Fem av dessa studier rörde egentlig depression och omfattade cirka 2 000 patienter. Skattningsdata sammanvägdes från flera tillfällen efter studiestart. Den andra studien omfattade 85 patienter med egentlig depression som vid bedömningstillfället var inneliggande på en psykiatrisk klinik i Tyskland [141]. I de båda studierna var samstämmigheten, beräknad som en korrelation mellan MADRS och sjukdomsdjup (CGI-S), 0,82 respektive 0,87.

### ***Evidensgraderat resultat***

Det finns för få studier med tillräcklig kvalitet för att kunna bedöma om MADRS mäter depressionsdjup på ett tillförlitligt sätt med SCID-I eller CGI-S som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## **Zung Self-Rated Depression Scale (SDS)**

Zung Self-Rated Depression Scale (SDS) är ett självskattningsformulär på 20 frågor som utvecklades i USA i mitten av 1960-talet [148]. Det ursprungliga syftet var att mäta nivå av depressiva symtom hos patienter diagnostiserade med egentlig depression.

Frågorna utgår från diagnostiska kriterier för egentlig depression. Hälften av frågorna är ställda så att förekomsten av centrala depressiva symtom efterfrågas. Den andra hälften av frågorna är reverserade så att avsaknaden av symtom efterfrågas. För varje fråga uppmanas den svarande att skatta förekomsten av symtom under de senaste dagarna på en fyrgradig svarsskala (1=liten del av tiden; 4=mest hela tiden). Svaren vägs

samman över alla frågor (efter omkodning av svaren på de reverserade frågorna) och svarspoängen kan variera mellan 20 och 80.

### ***Beskrivning av underlaget***

Samtliga 15 studier som lästes i fulltext exkluderades eftersom de inte var relevanta.

### ***Evidensgraderat resultat***

Det går inte att bedöma om olika versioner av Zung (SDS) är likvärdiga med DSM-IV-kriterier eller CGI-S för att uppskatta depressionsdjup vid egentlig depression pga att det saknas studier.

## **Formulär för bedömning av svårighetsgrad av bipolärt syndrom**

Vår fråga var vilka formulär som är tillförlitliga för att bedöma svårighetsgrad av bipolära syndrom. Svårighetsgrad enligt Clinical Global Impression Scale for Bipolar Disorder [149] eller enligt DSM-IV:s kriterier utgjorde referensstandard (se Kapitel 3.1).

### **Affective Self Rating Scale-18 (AS-18)**

AS-18 är ett självskattningsformulär som utvecklats i Sverige för att särskilja symtom på mani, depression och blandtillstånd (mixed states) och för att bedöma svårighetsgraden [150]. Formuläret baseras på DSM-kriterier för egentlig depression och mani varvid symtom som finns vid båda tillstånden exkluderats.

AS-18 består av 18 frågor om tillståndet under den senaste veckan. Nio frågor utgör en subskala för depression och nio en subskala för mani. För varje fråga finns fem skalsteg som anger svårighetsgrad med poäng 0–4. Totalpoängen varierar därmed mellan 0–72.

### ***Beskrivning av underlaget***

Vi identifierade bara en studie som undersökte formulärets egenskaper [150] (se Tabell 3.5.35). Studien rapporterade signifikanta korrelationer för subskalorna gentemot CGI-BP hos 61 patienter på en specialiserad klinik för affektiva syndrom i Stockholm. Dessutom korrelerades sub-



skalorna för depression och mani med respektive MADRS och Hypomania Interview Guide – Clinical Version. Spearmans rangkorrelation var 0,68 för depressionsskalan och 0,80 för maniskalan.

Studien bedömdes ha medelhög kvalitet eftersom index- och referenstest gjordes blindat, men antal bedömare, träning eller interbedömarreliabilitet för CGI-BP redovisades inte.

### ***Bedömning av evidensstyrka***

Det finns för få studier av tillräcklig kvalitet för att bedöma om AS-18 är tillförlitligt för att mäta svårighetsgrad av mani, depression och blandtillstånd av bipolärt syndrom med CGI-BP eller SCID-I som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

### **Bipolar Depression Rating Scale (BDRS)**

Bipolar Depression Rating Scale (BDRS) är ett semistrukturerat intervjuformulär som ska användas av tränade kliniker för svårighetsgradering av bipolär depression [151]. BDRS innefattar förutom kärnsymtom för depression också så kallat atypiska symtom och symtom för blandtillstånd (mixed states) vilket kännetecknar bipolär depression till skillnad från unipolär depression. BDRS består av 20 frågor med gradering 0–3. Skattningen avser de senaste dagarna. Eftersom skalan är nyutvecklad har den undersökts i begränsad omfattning.

### ***Beskrivning av underlaget***

Vi identifierade fem studier men samtliga exkluderades. Två av dem undersökte visserligen korrelationer men använde depressions- och maniskalor (MADRS, HDRS, YMRS) som referensstandard [151,152].

### ***Bedömning av evidensstyrka***

Det går inte att bedöma om BDRS är tillförlitligt för att mäta svårighetsgrad av bipolär depression med CGI-BP som referensstandard eftersom det saknas studier.

## **Bech-Rafaelsen Mania Scale (MAS)**

Bech-Rafaelsen Mania Scale (MAS) är ett intervjuformulär som används av tränade kliniker för svårighetsgradering av mani [153]. MAS utvecklades i slutet av 1970-talet i Danmark och finns i en version från 1996 [154]. MAS innehåller 11 påståenden med skalstegen 0–4; maximal poäng är 44. Vanligen avser skattningen de tre senaste dagarna. MAS har främst använts för kliniska läkemedelsprövningar vid akut mani [155].

### ***Beskrivning av underlaget***

Sju studier lästes i fulltext och av dem rörde tre svårighetsgradering med MAS. En av dem exkluderades pga att formuläret hade modifierats [156]. De två andra artiklarna bedömdes ha låg kvalitet.

I en studie av Bech och medarbetare beskrevs enbart medelvärden av MAS vid olika nivåer av CGI, dvs det saknades korrelationer mot CGI-S som utfallsmått [157]. I den andra studien, av Vieta och medarbetare [158], inkluderades konsekutivt 113 patienter med bipolärt syndrom i manisk fas från 18 psykiatriska kliniker i Spanien (Tabell 3.5.36). Studien bedömdes ha låg kvalitet pga att skattningarna inte var blindade utan gjordes av samma bedömare. Utfallsmåttet var inte korrelationer utan en ordinal regression av MAS-poäng med CGI-BP-kategorier som beroende variabel.

### ***Bedömning av evidensstyrka***

Det finns för få studier av tillräcklig kvalitet för att kunna bedöma om MAS är tillförlitligt för att mäta svårighetsgrad av mani med CGI-BP som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## **Young Mania Rating Scale (YMRS)**

Young Mania Rating Scale (YMRS) används av tränade kliniker för svårighetsgradering av mani [159]. Det finns ingen angivelse av tidsrymd för skattningen, men den baseras vanligen på patientens beskrivning av symtom under de två senaste dygnet samt observationer under intervjun. YMRS innehåller 11 påståenden: sju påståenden med poäng 0–4 och fyra påståenden (irritabilitet, talflöde, tankeinhåll och utagerande) som viktas högre (0–8) för att kompensera för dåligt samarbete vid skattningen av svårt sjuka patienter. Poängen varierar mellan 0–60.

YMRS är det internationellt mest använda formuläret för utvärdering av behandling för mani.

### ***Beskrivning av underlaget***

Av åtta artiklar som lästes i fulltext kunde fyra artiklar som berörde svårighetsgradering med YMRS identifieras. Alla hade låg studiekvalitet. En artikel av Berk och medarbetare innehöll poolade data från flera studier som inte gick att utvärdera kvalitetsmässigt och som fokuserade på kriterier för remission [160]. Favre och medarbetare undersökte korrelationer mellan YMRS och MAS [161]. Ursprungsartikeln av Young och medarbetare visade Spearmans rangkorrelationer på 0,88 med den då tillgängliga globala maniskalan, men studien omfattade endast 20 patienter [159]. I den ovan nämnda studien av Vieta och medarbetare undersöktes även svårighetsgradering med YMRS mot CGI-BP [158]. Studien var dock i första hand en valideringsstudie av MAS mot YMRS.

### ***Bedömning av evidensstyrka***

Det finns för få studier med tillräcklig kvalitet för att kunna bedöma om YMRS är tillförlitligt för att mäta svårighetsgrad av mani med CGI-BP som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## **Life Chart Methodology (LCM)**

Life-charting, i fortsättningen benämnt LCM, är utvecklat vid National Institute of Mental Health (NIMH) för att kartlägga långsiktigt förlopp av affektiv sjukdom samt inverkan av behandling och livshändelser [162,163].

LCM är en systematisk metod för att specificera svängningsmönster och resulterar i en grafisk presentation, vilket även kan användas som ett diagnostiskt hjälpmedel för att bestämma undergrupp av affektivt syndrom. Metoden har standardiserats i manualer för beskrivning av retrospektivt förlopp (NIHM-LCM-r) och prospektivt förlopp (NIMH-LCM-p) samt manualer till stöd för patientens egen förloppsbeskrivning (LCM-s/r; LCM-s/p).

### *Beskrivning av underlaget*

Av tio studier som lästes i fulltext inkluderades en studie där tillförlitligheten av LCM-p undersöktes under kliniska förhållanden [164]. Studien bedömdes ha medelhög kvalitet. Två studier exkluderades eftersom de undersökte samband mellan LCM-p och andra referensstandarder (HDRS, YMRS, GAF och IDS-C) än CGI-BP [165,166].

I studien av Meaden och medarbetare jämfördes 40 patienters självskattning med LCM-p med klinikers skattning per telefon med CGI-BP vid tre tillfällen med en månads mellanrum [164]. Spearmans rangkorrelation var 0,32–0,56 under de tre tillfällena för mani och 0,43–0,59 för depression.

### *Bedömning av evidensstyrka*

Det finns för få studier med tillräcklig kvalitet för att kunna bedöma om LCM-p är tillförlitligt för att följa svårighetsgrad av depression och mani med CGI-BP som referensstandard (otillräckligt vetenskapligt underlag ⊕○○○).

## **Uppföljning av behandlingseffekter**

Formulär för att bedöma svårighetsgrad kan användas vid upprepade tillfällen för att följa t ex effekter av behandling. Vi hittade inga studier där formulär jämfördes mot DSM-kriterier.

## **Samstämmighet mellan expertbedömare och patient**

Den fråga som skulle besvaras var om skattning av en expert överensstämmer med patientens (eller närståendes) självskattning när ett formulär finns i båda versioner. Överensstämmelsen kunde studeras med hjälp av olika korrelationsmått. Kravet på studierna var att respektive bedömningar skulle vara oberoende av varandra och att tiden mellan bedömningarna inte fick överstiga ett dygn.

## **Inventory of Depressive Symptomatology (IDS och QIDS)**

IDS har beskrivits i avsnittet Formulär för bedömning av svårighetsgrad av depression hos vuxna.

### ***Beskrivning av underlaget***

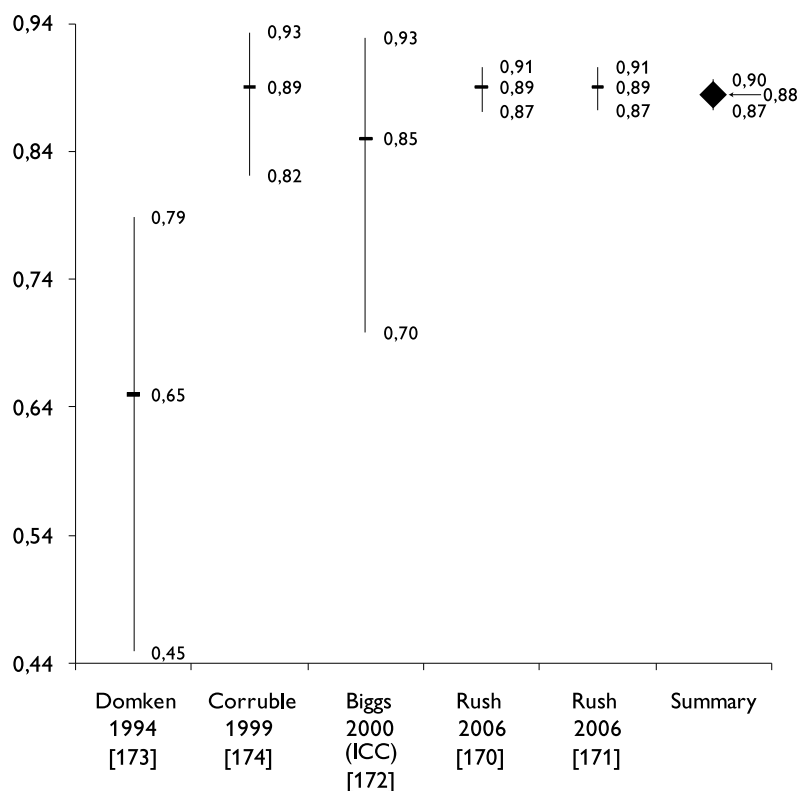
Tjugosex studier granskades i fulltext. Av dem exkluderades 18 pga att populationen inte uppfyllde inklusionskriterierna, att det var för lång tid mellan testerna, att bara skattningen av patienten beskrevs eller att blindning saknades. Tre studier hade låg studiekvalitet [167–169], fem studier medelhög [170–174] (Tabell 3.5.39).

Studierna med medelhög kvalitet omfattade 1 256 patienter. Tre av dem var amerikanska multicenterstudier med över 1 600 deltagare [170,171] varav en [172], utgjorde en pilotstudie som föregått en av de två multicenterstudierna [170]. De två andra var europeiska studier, utförda på psykiatriska akutvårdsavdelningar [173] respektive på psykiatrisk slutenvårdsavdelning [174].

### ***Sammanvägt resultat – samstämmighet mellan expert- och självskattad IDS/QIDS***

Studierna analyserade samstämmighet mellan summapoäng för expert- och patientskattade versioner av IDS och QIDS för patienter med egentlig depression. Sju intraklasskorrelationer (ICC) redovisades och ICC varierade mellan 0,76 och 0,91 (Figur 3.5.21). Den sammanvägda korrelationen blev 0,88 (0,87–0,90).

**Metaanalys  $r_{xy}$  - heterogentitet:  $I^2=0,79$ ,  $\chi^2<0,001$**



**Figur 3.5.21** Metaanalys: Överensstämmelse mellan expert- och patientskattning med IDS.

Studien av Domken och medarbetare avvek från de övriga, något som sannolikt berodde på att patienterna var i en akut fas av sjukdomen och att studien var liten (48 deltagare). Studien påverkade knappast det sammanvägda resultatet.

**Bedömning av evidensstyrka**

Samtliga ingående studier hade medelhög kvalitet. I två av studierna var blindningen oklar (Tabell 3.5.12) men vi bedömde att det inte var

tillräckligt för att sänka evidensstyrkan. Studierna omfattade flera vårdformer och hade genomförts i olika länder, vilket medförde att överförbarheten är god. Samstämmighet och precision påverkade inte heller evidensstyrkan.

**Tabell 3.5.12** Resultatsammanställning för överensstämmelse mellan själv- och klinikerskattning med IDS/QIDS.

Utfallsmått	Antal studier Antal patienter	Sammanvägt värde (95% KI)	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
ICC, samstämmighet mellan kliniker och patienter	5 studier n=1 256	0,88% (0,87; 0,90)	Starkt ⊕⊕⊕⊕	

KI = Konfidensintervall

### *Evidensgraderat resultat*

Överensstämmelsen mellan patient och kliniker inom psykiatrisk verksamhet för skattning av egentligt depressionssymptom med hjälp av formuläret IDS är god (ICC 0,88; 95 % KI, 0,87 till 0,90) (starkt vetenskapligt underlag ⊕⊕⊕⊕).

### **Montgomery-Åsberg Depression Rating Scale (MADRS)**

MADRS är avsett att användas av kliniker för att bestämma depressionsdjup och följa förlopp över tid [145]. Formuläret är beskrivet i avsnittet Formulär för bedömning av svårighetsgrad av depression hos vuxna. Under början av 1990-talet utvecklades en patientversion, MADRS Self-Assessment (MADRS-S) [175].

### *Beskrivning av underlaget*

Av tio studier som lästs i fulltext exkluderades åtta. Orsakerna var ofullständig metodbeskrivning, annan referensstandard eller för lång tid mellan testerna. Det vetenskapliga underlaget utgjordes därmed av två studier med medelhög kvalitet [176,177] (Tabell 3.5.41).

Den ena studien omfattade 63 patienter med förstämningssyndrom (pågående depressionsepisoder enligt ICD-10) vid två psykiatriska öppenvårdsmottagningar i Schweiz [176]. MADRS-skattningarna utfördes av psykiatrer som visat hög samstämmighet i sina bedömningar (intraclass correlation, ICC 0,87). Uppföljande skattningar gjordes fyra veckor efter de initiala skattningarna.

Den andra studien omfattade 400 öppenvårdspatienter med egentlig depression, från ett 30-tal svenska vårdcentraler som deltog i en multicenterstudie av antidepressiva läkemedel [177]. MADRS-skattningarna utfördes av distriktsläkare som via träning uppnått hög samstämmighet i sina skattningar (intraclass correlation, ICC 0,89). Skattningarna genomfördes vid nio tillfällen under 52 veckor.

### ***Samstämmighet mellan expert- och självskattad MADRS***

I den schweiziska studien analyserades samstämmigheten mellan summa-poäng för expert- och patientskattningarna. Pearsons korrelation var 0,81 vid första tillfället och 0,91 fyra veckor senare.

I den svenska studien var ICC mellan expert- och patientskattning 0,47 vid första tillfället och 0,55 vid uppföljning ett år senare (med variation mellan 0,47 och 0,75 över alla nio mättillfällen). Den tredje mätningen gjordes fyra veckor efter inkludering och sammanföll i tid med uppföljningen i den schweiziska studien. ICC var då 0,71 vilket är tillfredsställande högt i det här sammanhanget.

I och med att studierna använder olika sätt att mäta korrelationer är det inte meningsfullt att göra ett sammanvägt estimat av resultaten från de båda studierna. Studierna pekar dock mot att samstämmigheten fyra veckor efter inklusion är tillräcklig.

### ***Bedömning av evidensstyrka***

Båda ingående studierna hade medelhög kvalitet, men i en av studierna var blindningen oklar vilket medförde att evidensstyrkan drogs ned ett steg. Överförbarheten var relativt god. Studierna omfattade flera vårdformer och hade genomförts i olika länder, vilket är en styrka. Överförbarheten minskade dock av att underlaget bara bestod av två



publicerade studier och att studierna hade använt olika mått på korrelation. Vi bedömde att den svenska studiens ICC var ett mera konservativt mått och att korrelationen vid fyra veckor därmed är åtminstone 0,7. I och med att vi valde att uttrycka resultatet som ett intervall var det inte relevant att bedöma underlagets precision.

**Tabell 3.5.13** Resultatsammanställning för överensstämmelse mellan själv- och klinikerskattning med MADRS respektive MADRS-S.

Utfallsmått	Antal studier Antal patienter	Sammanvägt värde	Bedömning av vetenskapligt underlag	Kommentarer till poängavdrag
Korrelation mellan kliniker och patienter efter 4 veckor	2 studier n=463	ICC >0,7	Begränsat ⊕⊕○○	Bristande studie-kvalitet (-1) Bristande överförbarhet (-1)

### *Evidensgraderat resultat*

Överensstämmelsen mellan patient och kliniker inom öppenvården för skattning av svårighetsgrad av depression fyra veckor efter akut fas med hjälp av MADRS är tillräcklig (ICC >0,7) (begränsat vetenskapligt underlag ⊕⊕○○).

## Diskussion

### Intervjuformulär

Få studier uppfyllde våra inklusionskriterier, vilket begränsar våra möjligheter att dra slutsatser om sensitivitet och specificitet för intervjuformulär. Det är viktigt att understryka att frånvaro av studier eller få studier inte i sig betyder att formuläret är dåligt. Översikten har inte inkluderat sedvanliga psykometriska studier av validitet och reliabilitet. Det vi kan uttala oss om är om det finns tillräckligt vetenskapligt underlag för att bedöma intervjuernas tillförlitlighet att diagnostisera förstämningssyndrom i kliniska populationer.

Av åtta intervjuformulär fanns det bara tillräckligt underlag att utvärdera tre: SCID-I, MINI och PRIME-MD.

Vi granskade det vetenskapliga underlaget för SCID-I ur två aspekter: som referensstandard och som intervjuformulär i kliniken. Vi bedömde att SCID-I har tillräcklig sensitivitet och specificitet för depression och mani för att fungera såväl för kliniskt bruk som i forskningen som referensstandard. Sensitiviteten vid bipolära syndrom kan dock vara i lägsta laget för SCID-I som referensstandard om det inte finns tillgång till journalhandlingar.

Vår granskning visade att MINI har god tillförlitlighet för att identifiera och utesluta egentlig depression liksom pågående eller tidigare mani. En svaghet är att MINI enbart är utvärderat inom psykiatrisk vård.

PRIME-MD har däremot inte tillräcklig diagnostisk tillförlitlighet. Studierna visade en acceptabel specificitet för egentlig depression, men sensitiviteten var i en studie så låg som 30 procent. Originalstudien av Spitzer och medarbetare ingår inte i vårt underlag eftersom referensstandarderna var en telefonintervju som endast delvis byggde på en semistrukturerad intervju [26]. Spitzer och medarbetares resultat står dock i samklang med våra med en låg sensitivitet; 57 procent, och en hög specificitet; 98 procent för depression. Sammanfattningsvis är det kliniska värdet av att använda PRIME-MD osäkert, något som kan vara en anledning till att PRIME-MD internationellt börjar ersättas av MINI [178].

### ***Riktad screening för depression***

Vi identifierade ett stort antal studier som undersökte diagnostisk tillförlitlighet hos formulär för riktad screening, men majoriteten av dem jämförde olika formulär med varandra. Vårt underlag baseras på det fåtal studier som använt semistrukturerade eller strukturerade intervjuer som referens. De formulär för depression som vi kan uttala oss om är BDI-II, PHQ-9, HADS och EPDS.

Av dem var det bara BDI-II som visade tillräcklig sensitivitet i studierna och detta oavsett studiepopulation och referensstandard. För PHQ-9 och HADS låg sensitiviteten i studierna från cirka 35 procent och upp till

över 80 procent. Det innebär att det finns en stor risk för att patienter med depression inte upptäcks med PHQ-9 och HADS. Andra systematiska översikter om HADS och PHQ-9 som haft delvis andra inklusionskriterier har kommit fram till likartade resultat [179–181].

En möjlig orsak till resultaten är att formulärens uppbyggnad har betydelse för utfallet av den enskilda patientens skattning. BDI-II mäter depression med flera frågor jämfört med de andra formulärens. Förutsatt att frågorna är valida så innebär flera frågor att formuläret blir mindre känsligt för att slumpfaktorer påverkar resultaten. HADS-D kan vara speciellt känslig, eftersom den endast omfattar sju frågor. Vid få frågor är det större risk att formuläret inte fångar alla aspekter av patientens tillstånd. På samma sätt kan patientens tolkning av de enskilda frågornas innebörd spela större roll. Samma resonemang kan appliceras på PHQ-9 som omfattar nio frågor.

Vi gjorde en fokuserad analys av EPDS vid postnatal depression som tyder på att EPDS har en sensitivitet på cirka 70 procent med SCID-I som referensstandard. Det är lägre än vårt uppsatta riktmärke för formulär för riktad screening på 80 procent, något som diskuteras vidare i Kapitel 9.

Det finns en nyligen publicerad systematisk översikt om EPDS [182]. Sensitiviteten i studierna låg mellan 34 och 100 procent. Författarna konstaterade att det inte gick att dra några slutsatser om den diagnostiska tillförlitligheten, något som sannolikt beror på att de inkluderade studier med olika referensstandarder.

Några svenska studier har undersökt EPDS som formulär för riktad screening för postnatal depression. Två av dem exkluderades och den tredje hade en bristfällig referensstandard [52,183,184]. En ny studie om depression hos gravida exkluderades eftersom tiden mellan indextest och referens var för lång [185]. Med tanke på att vår granskning omfattar ett stort antal länder med olika kultur och organisation av mödravård finns det inte anledning att tro att svenska resultat skulle avvika från övriga världens.

En möjlig begränsning i vårt underlag är att vi avgränsat oss till populationer i Europa, Nordamerika och Australien/Nya Zeeland. Det kan möjligen innebära att vi missat relevanta studier. För PHQ-9, EPDS och HADS består dock underlaget av så många studier så att ytterligare en enstaka studie sannolikt inte skulle innebära någon skillnad i resultat.

### ***Screening för bipolärt syndrom***

När patienter söker för depression finns det starka skäl för att rutinmässigt använda formulär för riktad screening för bipolära syndrom. Om den bipolära diagnosen missas kan läkemedelsbehandling för depression utlösa maniska skov och tätare episoder. Patienter med bipolaritet svarar sämre på behandling med antidepressiva läkemedel vid depression och behöver generellt ett annat behandlingsupplägg än patienter med unipolära depressioner [186]. Det är särskilt svårt att identifiera patienter med bipolärt syndrom typ II eftersom patienterna oftast befinner sig i långdragna depressionsperioder och för att de inte söker hjälp under sina hypomana episoder.

I många kliniska sammanhang är prevalensen av bipolärt syndrom låg. Det innebär att det behövs en hög sensitivitet för att inte missa för många patienter med bipolärt syndrom.

Vår slutsats från granskningen är att MDQ och HCL-32 är värdefulla hjälpmedel men att kliniker bör vara medvetna om deras begränsningar. Våra resultat tyder på att HCL-32 har högre sensitivitet än MDQ men lägre specificitet. Detta styrks av en direkt jämförelse mellan HCL-32 och MDQ av Meyer och medarbetare [112]. MDQ med Genève-kriterier har en acceptabel sensitivitet för bipolära syndrom men inte för bipolärt syndrom typ II enbart.

En hög sensitivitet medför en hög frekvens av falskt positiva resultat, även om testet har hög specificitet. Två större studier från primärvård i USA och Frankrike har t ex funnit en prevalens ”MDQ-positiva” på 8–10 procent. Diagnoserna säkerställdes dock inte med strukturerad diagnostik [187,188]. Zimmerman och medarbetare sökte det tröskelvärdet som skulle ge den enligt honom ”tillräckliga” sensitiviteten 90 procent inom ramen för Rhode Island Methods to Improve Diagnostic

Assessment and Services (MIDAS)-projektet [116]. Zimmerman och medarbetare fann att vid detta tröskelvärde var de flesta som testade positivt falskt positiva.

### ***Bedömning av svårighetsgrad vid ett eller flera tillfällen***

Det fanns för få studier för att kunna avgöra om formulären är samstämmiga med oberoende skattningar av sjukdomens svårighetsgrad. Vår bedömning är att det finns en stor mängd data, där svårighetsgrad vid depression och mani följs över tid, att hämta från effektstudier av t ex läkemedel, men de har sällan analyserats eller presenterats i vetenskapliga artiklar. Möjligen finns data avrapporterade utan att frågeställningen eller analysen tagits upp i publikationens abstrakt. Vi kan inte utesluta att vi i så fall har missat sådana studier.

För tolkningen av resultat finns det ett värde med att ange ungefärliga tröskelvärden mellan normalt stämningsläge, lätt, måttlig eller svår depression respektive mani. I manualer ser man ofta rekommendationer för tolkning av poäng. Det finns en risk att användaren tar rekommenderade tröskelvärden för bokstavliga och inte väger in annan viktig information i sin kliniska bedömning. Vi har funnit att underlaget för rekommenderade tröskelvärden är mycket begränsat. Exempelvis har en arbetsgrupp inom International Society for Bipolar Disorders gett rekommendationer för tolkning av poängen för YMRS och MAS men betonat att dessa kriterier behöver valideras [189].

### ***Överensstämmelse mellan expert- och patient/närstående skattning***

Tre studier med IDS och två med MADRS som inkluderats i granskningen tyder på att man kan uppnå en acceptabel samstämmighet mellan expert- och patientskattat depressionsdjup om experten har gedigen erfarenhet av att använda bedömningsformulär. Ytterligare tre studier av IDS med låg kvalitet kom fram till likartade resultat [143,168,169]. En studie med låg kvalitet som inte följer detta mönster har undersökt MADRS [190]. Studien redovisade en låg korrelation på 0,54 mellan patient, och expertskattad MADRS. Ett problem med studien var att patienterna hade selekterats till studien utifrån expertskattade MADRS-poäng över 30. Detta skulle kunna leda till en för liten variation i materialet för att kunna bedöma överensstämmelsen.

Studier av samstämmighet har fokuserat på om expertbedömare och patienter är överens om depressionsdjup. Vi har inte kunnat identifiera någon systematisk analys av i vilken utsträckning patient eller expert generellt gör en under- eller överskattning. Fyra av de fem inkluderade studierna av IDS/QIDS redovisar dock sådana resultat. Av 17 jämförelser är medelvärdena för patient- respektive expertskattning i det närmaste identiska (inom mindre än en poäng för IDS). En jämförelse redovisar högre medelvärde för expertskattningarna. Resterande nio jämförelser visar att patienterna i genomsnitt rapporterar något högre symtomnivåer. Den svenska studien på MADRS presenterar också resultat som möjliggör en sådan jämförelse. Av nio jämförelser är medelvärdena för expert- respektive patientskattningarna närmast identiska för de (i tiden) två första mätningarna. Vid de sex övriga mättillfällena ger patientskattningarna genomgående högre medelvärden.

Vissa patientgrupper tenderar att underskatta eller överskatta symtomdjup [177]. Vi har dock inte gjort en systematisk bedömning av hur patienters egenskaper som personlighet, kön, ålder och samsjukliga tillstånd inverkar på tillförlitligheten av självskattning. Expertskattning å sin sida kräver utbildning och kontinuerlig samskattning för att vara tillförlitlig, vilket kan vara svårt att åstadkomma i den kliniska vardagen. Översikten indikerar att det i allmänhet är en rätt god överensstämmelse mellan självskattning och expertskattning av IDS och MADRS i studier där bedömarna har god utbildning. För kliniskt bruk kan självskattning och expertskattning ses som kompletterande mått.



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

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	<p><u>Design</u> Cross sectional and longitudinal</p> <p><u>Setting</u> Community mental health centre in Dallas, USA</p> <p><u>Population</u> 210 outpatients recruited, 200 completed both interviews</p> <p><u>Inclusion criteria</u> Not reported</p> <p><u>Exclusion criteria</u> Not reported</p> <p><u>Prevalence</u> BP: 25% (50/200) MDD: 19% (38/200) Schizoaffective: 18% (36/200)</p>	<p><u>Index test</u> I1: SCID-I diagnoses. Including family history of psychiatric illness, general medical history. A life chart was constructed</p> <p>I2: SCID-I-plus-chart diag- noses. Above + medical records</p> <p><u>Reference test</u> LEAD Interview after a week with one of prime investigators (psychiatrist or psycholo- gist) after reviewing all data</p> <p><u>Number of patients</u> n=200 Mean age: 37.6 years (SD 10) 116 (58%) females 134 (67%) Caucasian</p> <p><u>Drop-out rate</u> 5%</p>	<p><u>Rater of index test</u> Three psychiatric nurses</p> <p><u>Training of index test</u> Yes</p> <p><u>Rater of reference test</u> One of three psychiatrists or an psychologist</p> <p><u>Training of reference test</u> NA</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>Bipolar disorder (n=50)</u> SCID-I Sensitivity: 76% Specificity: 93%</p> <p>SCID-I + chart Sensitivity: 88% Specificity: 93%</p> <p><u>Major depressive disorder (n=38)</u> SCID-I Sensitivity: 84% Specificity: 91%</p> <p>SCID-I + chart Sensitivity: 82% Specificity: 96%</p> <p><u>Schizoaffective disorder (n=36)</u> SCID-I Sensitivity: 53% Specificity: 97%</p> <p>SCID-I + chart Sensitivity: 89% Specificity: 95%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Yes, and NA</p> <p><u>Handling of missing data</u> Adequate</p> <p><u>Other comments</u> 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</p>

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

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

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

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

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

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)	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Psychiatric outpatients in France and the USA</p> <p><u>Population</u> Consecutive patients (280 from USA and 40 from France)</p> <p><u>Inclusion criteria</u> Age: &gt;18 years</p> <p><u>Exclusion criteria</u> Dementia, mental retardation or serious medical illnesses</p> <p><u>Prevalence</u> MDD: 45.9% Dysthymia: 1.6% Current mania: 10.3% Lifetime mania: 20.0%</p>	<p><u>Index test</u> MINI</p> <p><u>Reference test</u> SCID-P</p> <p><u>Number of patients</u> n=320 (48% female) 56% met criteria for more than one current SCID-P diagnosis</p> <p>Mean age: 44.8 years (SD 15.1)</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of index test</u> In USA: 10 psychiatrists, 4 research assistants and 2 medical students. In France: 3 psychiatrists and 2 psychologists</p> <p><u>Training of index rater</u> Training packet for the MINI</p> <p><u>Rater of reference test</u> Same as index test rater</p> <p><u>Training of reference test</u> One day training session for the SCID-P</p> <p><u>Interobserver reliability</u> MINI-CR: All of <math>\kappa</math> values were above 0.75, and 16 of 23 were 0.90 or higher</p> <p>SCID-P: Not reported</p>	<p><u>Major depressive disorder</u> Sensitivity: 96% Specificity: 88%</p> <p><u>Dysthymia</u> Sensitivity: 67% Specificity: 99%</p> <p><u>Current mania</u> Sensitivity: 82% Specificity: 95%</p> <p><u>Lifetime mania</u> Sensitivity: 81% Specificity: 94%</p>	<p>Moderate</p> <p><u>Comments</u> <u>Sampling method</u> Adequate, but there was no information about the number that did not consent</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Not reported</p> <p><u>Other comments</u> 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</p>

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

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

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Outpatient palliative care department</p> <p><u>Population</u> Convenience sample n=61 of patients with advanced, non-curable stages of metastatic cancer</p> <p><u>Inclusion criteria</u> Not reported</p> <p><u>Exclusion criteria</u> Unable to read and understand the Dutch language</p> <p><u>Prevalence</u> MDD: 21.7%</p>	<p><u>Index test</u> BDI-II</p> <p><u>Reference test</u> PRIME-MD</p> <p><u>Number of patients</u> n=46</p> <p><u>Drop-out rate</u> 25%</p>	<p><u>Rater of reference test</u> A physician</p> <p><u>Training of reference test</u> Not reported</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>Major depressive disorder</u> Sensitivity: 90% Specificity: 64%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Acceptable</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Not reported</p>

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	<u>Design</u> Cross sectional  <u>Setting</u> Neuropsychiatric clinic in Birmingham  <u>Population</u> Outpatients with Huntington's disease  <u>Inclusion criteria</u> >18 years  <u>Exclusion criteria</u> Not able to give informed consent, not fluent in English  <u>Prevalence</u> MDD: 24% (12/50)	<u>Index test</u> BDI-II  <u>Reference test</u> SCAN  <u>Number of patients</u> n=50 Mean age: 51.2 years (SD 10.35)  <u>Drop-out rate</u> Not reported	<u>Rater of index test</u> NA  <u>Training of index rater</u> NA  <u>Rater of reference test</u> One researcher and one psychologist  <u>Training of reference test</u> Yes, but not specified  <u>Interobserver reliability</u> Not reported	<u>Major depressive disorder</u> Sensitivity: 83% Specificity: 71%	Moderate  <u>Comments</u> Sampling method Adequate  Blinding Unclear  Handling of missing data Not reported
Dutton et al 2004 [39] USA	<u>Design</u> Cross sectional  <u>Setting</u> Waiting rooms of two primary care clinics in southeastern USA  <u>Population</u> n=223 African-American patients fulfilled criteria and consented  <u>Inclusion criteria</u> >18 years  <u>Exclusion criteria</u> Failing to demonstrate at least a 5th grade oral comprehension level  <u>Prevalence</u> MDD: 29.5% (65/220)	<u>Index test</u> BDI-II  <u>Reference test</u> PRIME-MD  <u>Number of patients</u> n=220 Mean age: 49.3 years (SD 10.91)  <u>Drop-out rate</u> 1%	<u>Rater of index test</u> NA  <u>Training of index rater</u> NA  <u>Rater of reference test</u> Six research assistants who all were doctoral students in clinical psychology  <u>Training of reference test</u> Yes but not specified  <u>Interobserver reliability</u> Not reported	<u>Major depressive disorder</u> Sensitivity: 88% (95% CI, 77; 94)  Specificity: 84% (95% CI, 77; 89)	Moderate  <u>Comments</u> 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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Assessment clinic of a specialist pain center in United Kingdom</p> <p><u>Population</u> Convenience sample of patients with chronic pain</p> <p><u>Inclusion criteria</u> Pain episode lasting &gt;12 weeks</p> <p><u>Exclusion criteria</u> Not reported</p> <p><u>Prevalence</u> MDD: 72% (26/36)</p>	<p><u>Index test</u> BDI-II</p> <p><u>Reference test</u> SCID-I</p> <p><u>Number of patients</u> n=36 (64% female) Mean age: 47.8 years (SD 12.85)</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of reference test</u> Two trainee clinical psychologists with experience</p> <p><u>Training of reference test</u> Adequate</p> <p><u>Interobserver reliability</u> <math>\kappa</math> 1.0</p>	<p><u>Major depressive disorder</u> Sensitivity: 100% Specificity: 50%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Not reported</p>

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

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	<u>Design</u> Cross sectional  <u>Setting</u> 4 pediatric specialties clinics in Cleveland, Ohio  <u>Population</u> n=332 mothers to children and adolescents with disabilities (CF, CP, myelodysplasia or multiple handicaps)  <u>Prevalence MDD</u> 5.2%	<u>Index test</u> CES-D, self-administered during interview  <u>Reference test</u> DIS  <u>Number of participants</u> n=319 came to interview complete data from 308 Mean age: 42 years  <u>Drop-out rate</u> 3.4%	<u>Rater</u> Lay female interviewers  <u>Rater training</u> "Trained"  <u>Inter observer reliability</u> $\kappa$ 0.75 (6 months retest, personal vs telephone)	<u>Major depressive disorder</u> Sensitivity: 87.5% Specificity: 73%	Moderate  <u>Comments</u> Sampling method Adequate  Blinding Not reported  Handling of missing data Not reported

CES-D = Center for epidemiologic studies depression scale; CI = Confidence interval;  
DIS = Diagnostic interview schedule; MDD = Major depressive disorder; NP = Neuro-  
logical 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	<p><u>Design</u> Cross sectional, stratified samples for EPDS &lt;10 and ≥10</p> <p><u>Setting</u> Two community-based child health clinics</p> <p><u>Population</u> n=362 women 6 weeks after delivery</p> <p><u>Exclusion criteria</u> Not speaking or reading Norwegian language</p> <p><u>Prevalence MDD</u> 16%</p>	<p><u>Index test</u> EPDS</p> <p><u>Reference test</u> PRIME-MD</p> <p><u>Number of patients</u> n=57 EPDS score ≥10: n=25 EPDS &lt;10: n=31</p> <p><u>Drop-out rate</u> 2%</p>	<p><u>Rater</u> Three experienced general practitioners and one psychiatrist</p> <p><u>Rater training</u> Unclear</p> <p><u>Interobserver reliability</u> 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)</p>	<p><u>For MDD cut off ≥12</u> Sensitivity: 56% Specificity: 81%</p>	<p>High</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Yes</p>

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

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

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Primary health care clinics</p> <p><u>Population</u> 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</p> <p><u>Exclusion criteria</u> Psychiatric treatment history</p> <p><u>Prevalence postnatal MDD</u> 14%</p>	<p><u>Index test</u> EPDS</p> <p><u>Reference test</u> SCID-I</p> <p><u>Number of patients</u> n=341 women Mean age: 26.6±4.8 years</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater</u> A mental health professional</p> <p><u>Rater training</u> Not reported</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>MDD, cut off 12.5</u> Sensitivity: 75.5% (95% CI, 71; 79) Specificity: 71.5% (95% CI, 67; 76)</p>	<p>Moderate</p> <p><u>Comments</u> <u>Sampling method</u> Adequate, consecutively included</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Not reported</p>

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

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

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

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

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

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	<u>Design</u> Cross sectional  <u>Setting</u> Community-based postnatal health service  <u>Population</u> 72 mothers consecutively invited and 12 women invited from a local health clinic Mean age: 26 years  <u>Inclusion criteria</u> Mothers identified by the health visitors at about 6 weeks following delivery as being potentially depressed 12 women without depressive symptoms  <u>Prevalence MDD</u> 25% (21/84)	<u>Index test</u> EPDS  <u>Reference test</u> SPI  <u>Number of patients</u> n=84 women  <u>Drop-out rate</u> Not reported	<u>Rater</u> One psychiatrist and one psychologist  <u>Rater training</u> Trained in SPI  <u>Interobserver reliability</u> Not reported	<u>MDD cut off 12/13</u> Sensitivity: 86% Specificity: 78%	Moderate  <u>Comments</u> Sampling method Unclear  Blinding Yes  Handling of missing data Unclear
Matthey et al 2001 [69] Australia	<u>Design</u> Cross sectional  <u>Setting</u> At home 6–7 weeks post partum  <u>Population</u> 251 couples recruited from preparation for parenthood classes  <u>Prevalence MDD</u> 10.4%	<u>Index test</u> EPDS  <u>Reference test</u> DIS  <u>Number of patients</u> n=238 mothers  <u>Drop-out rate</u> 5%	<u>Rater</u> Research assistant  <u>Rater training</u> Adequate  <u>Interobserver reliability</u> Done but not reported	<u>MDD cut off 12.5</u> Sensitivity: 38% Specificity: 94%  <u>MDD cut off 12.5 for men</u> Sensitivity: 43% Specificity: 98%	Moderate  <u>Comments</u> 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	<u>Design</u> Cross sectional  <u>Setting</u> Residential early parenting centers for mothers with caretaking difficulties  <u>Population</u> 185 consecutive women  <u>Inclusion criteria</u> English speaking women with infants up to one year old  <u>Prevalence MDD</u> 26.8%	<u>Index test</u> EPDS  <u>Reference test</u> CIDI-Auto  <u>Number of patients</u> n=145 Mean age: 33.3±4.3 years  <u>Drop-out rate</u> 5%	<u>Rater</u> Trained interviewers  <u>Rater training</u> Yes, but not specified  <u>Interobserver reliability</u> Not reported	<u>MDD cut off score 12</u> Sensitivity: 76% Specificity: 39%	Moderate  <u>Comments</u> Sampling method Adequate  Blinding Yes  <u>Handling of missing data</u> 5% with missing data was excluded  <u>Other comments</u> Time between reference and index test was prob- ably in 5 days but not clearly stated
Leonardou et al 2009 [72] Greece	<u>Design</u> Cross sectional  <u>Setting</u> Two maternity hospitals  <u>Population</u> 109 women of whom 95 participants were recruited on the second day post partum  <u>Exclusion criteria</u> Language issues  <u>Prevalence MDD</u> 5%	<u>Index test</u> EPDS  <u>Reference test</u> SCID-I-NP  <u>Number of patients</u> n=81 Mean age: 31.5±4.1 years  <u>Drop-out rate</u> 15%	<u>Rater</u> Principal investigator  <u>Rater training</u> Trained but procedure not described  <u>Interobserver reliability</u> NA	<u>MDD cut off 12/13</u> Sensitivity: 80% Specificity: 99%	Moderate  <u>Comments</u> Sampling method Adequate  Blinding Yes  <u>Handling of missing data</u> Unclear  <u>Other comments</u> 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	<u>Design</u> Cross sectional  <u>Setting</u> Research  <u>Population</u> n=202 eligible women from the Mothers' Advisory Clinics and from hospital outpatients psychiatric department  <u>Inclusion criteria</u> Within 6 months post partum  <u>Exclusion criteria</u> Treatment for puerperal psychosis  <u>Prevalence MDD</u> 9%	<u>Index test</u> EPDS  <u>Reference test</u> DIS  <u>Number of patients</u> n=107 women Mean age: 28.4 years (SD 4.18)  <u>Drop-out rate</u> 4%	<u>Rater</u> Psychologist  <u>Rater training</u> Not reported  <u>Interobserver reliability</u> Not reported	<u>MDD cut off 12.5</u> Sensitivity: 100% Specificity: 96%	Moderate  <u>Comments</u> 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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Antenatal clinic attendees' recruitment</p> <p><u>Population</u> 999 consecutive women of whom 454 were eligible</p> <p><u>Inclusion criteria</u> Age between 18 and 40 years. Assessment at 3 months post partum</p> <p><u>Exclusion criteria</u> Language difficulties Single parent &gt;18 weeks pregnant &gt;1 child</p> <p><u>Prevalence MDD</u> 3% (5/199)</p>	<p><u>Index test</u> EPDS</p> <p><u>Reference test</u> PSE</p> <p><u>Number of patients</u> n=199 women</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater</u> Psychiatrists (unknown number)</p> <p><u>Rater training, reference test</u> Adequate</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>MDD cut off 12/13</u> Sensitivity: 70% Specificity: 93%</p>	<p>Moderate</p> <p><u>Comments</u> <u>Sampling method</u> Yes but high external drop-out</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Unclear</p>

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

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Research setting</p> <p><u>Population</u> 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</p> <p><u>Inclusion criteria</u> Caucasian Appropriate knowledge of the Dutch language</p> <p><u>Prevalence MDD</u> 5.6% 12 weeks' gestation 5.4% 24 weeks' gestation 3.4% 36 weeks' gestation</p>	<p><u>Index test</u> EPDS</p> <p><u>Reference test</u> CIDI</p> <p><u>Number of patients</u> n=845</p> <p><u>Drop-out rate</u> 22%</p>	<p><u>Rater</u> One midwife and five experienced psychology students</p> <p><u>Rater training</u> Extensive CIDI training</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>MDD cut off 12 weeks</u> Sensitivity: 55% Specificity: 98%</p> <p>24 weeks Sensitivity: 48% Specificity: 98%</p> <p>36 weeks Sensitivity: 45% Specificity: 97%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p>Blinding Yes</p> <p>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</p>

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

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

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> General practices with 70 physicians</p> <p><u>Population</u> n=370 Patients on sick leave because of psychological problems</p> <p><u>Inclusion criteria</u> Age: 18–60 years Sick leave &lt;3 months</p> <p><u>Exclusion criteria</u> Recognized depressive or anxiety disorder</p> <p><u>Prevalence MDD</u> 49%</p>	<p><u>Index test</u> HADS self rating questionnaire</p> <p><u>Reference test</u> CIDI</p> <p><u>Number of patients</u> n=307 (253 female) Mean age: 39.5±9.2 years</p> <p><u>Drop-out rate</u> 3.9%</p>	<p><u>Rater</u> 5 lay raters</p> <p><u>Rater training</u> Dutch WHO CIDI training and reference centre</p> <p><u>Interobserver reliability</u> Not stated</p>	<p><u>MDD</u> Sensitivity: 93% Specificity: 39%</p>	<p>High</p> <p><u>Comments</u> Blinding Rater blinded</p>

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

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	<u>Design</u> Cross sectional  <u>Setting</u> Accident and emergency departments of medical wards  <u>Population</u> n=189 acutely ill patients consecutively included and still in hospital on the 7th day after injury  <u>Inclusion criteria</u> Age: ≥18 years  <u>Exclusion criteria</u> MMSE <23  <u>Prevalence MDD</u> 5.9%	<u>Index test</u> HADS administered during interview  <u>Reference test</u> SCAN  <u>Number of patients</u> n=153  <u>Drop-out rate</u> 2.5%	<u>Rater</u> Not reported  <u>Rater training</u> Not reported  <u>Inter observer reliability</u> Not reported	<u>MDD</u> Sensitivity: 100% Specificity: 73%	Moderate  <u>Comments</u> Sampling method Adequate  Blinding Yes
Hall et al 1999 [84] United Kingdom	<u>Design</u> Prospective study with a diagnostic test after 3 months  <u>Setting</u> Cancer Research Campaign Interview in patients' homes  <u>Population</u> Women <75 years with early breast cancer (stage I–II)  <u>Prevalence MDD</u> 37.2% (99/266)	<u>Index test</u> HADS  <u>Reference test</u> PSE  <u>Number of patients</u> n=269  <u>Drop-out rate</u> 1%	<u>Rater</u> 3 interviewers  <u>Rater training</u> “Trained”  <u>Interobserver reliability</u> Tested by videotapes in 30% of interviews by independent rater κ performed but not reported	<u>MDD</u> Sensitivity: 33% Specificity: 93%	Moderate  <u>Comments</u> 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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> One hospital</p> <p><u>Population</u> n=167 consecutive patients in waiting rooms to clinics for somatic disease or psychiatry units (50% on psychotropic drugs)</p> <p><u>Inclusion criteria</u> Patients &gt;18 years, attending services for a medical disorder</p> <p><u>Exclusion criteria</u> Severe physical disease, cognitive deterioration, neurological or psychotic disorder compromising the ability to complete the questionnaires and patients with more than 50% incomplete questionnaires</p> <p><u>Prevalence</u> 52% (35/67)</p>	<p><u>Index test</u> HADS</p> <p><u>Reference test</u> Clinical interview, PRIME-MD</p> <p><u>Number of patients</u> n=67 completed HADS (87% women)</p> <p><u>Drop-out rate</u> 1%</p>	<p><u>Rater</u> Two of the authors</p> <p>Interviews by 10 mental health professionals blinded to the result of index test</p> <p><u>Rater training</u> Evaluation of diagnosis of depression of 10 cases from a psychiatrist</p> <p><u>Interobserver reliability</u> κ 0.67–1.0</p>	<p><u>MDD</u> Sensitivity: 0.86% (95% CI, 0.70; 0.95)</p> <p>Specificity: 0.75% (95% CI, 0.57; 0.88)</p>	<p>High</p> <p><u>Comments</u> Sampling method Adequate</p> <p>Blinding Yes</p>

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	<u>Design</u> Cross sectional  <u>Setting</u> Outpatient hospital clinic and 12 family practices  <u>Population</u> Random sample of n=2 050 outpatients at visit ≥18 years with somatic disease  <u>Prevalence MDD</u> 13.2%, 66/501	<u>Index test</u> HADS  <u>Reference test</u> SCID-I  <u>Number of patients</u> n=1 619 consented and filled in questionnaires n=528 completed SCID-I  <u>Drop-out rate</u> 4.9%	<u>Rater</u> 4 trained raters, blind to the result of the questionnaires  <u>Rater training</u> Reliability tested by video-tapes  <u>Inter observer reliability</u> κ (SCID-I) 0.88 (0.47–1.0)	<u>MDD</u> Sensitivity: 88% (95% CI, 78; 95)  Specificity: 69% (95% CI, 64; 73)	Moderate  <u>Comments</u> Sampling method Adequate
Zoger et al 2004 [83] Sweden	<u>Design</u> Cross sectional  <u>Setting</u> Audiological physician at a university hospital  <u>Population</u> n=98 consecutive outpatients with tinnitus (interquartile range 30–71 months). Investigated 24 months after initial visit  <u>Prevalence MDD</u> 18%	<u>Index test</u> HADS  <u>Reference test</u> SCID-I  <u>Number of patients</u> n=82 in study group consented (36% women) Mean age: 50 years  <u>Drop-out rate</u> Not reported	<u>Rater</u> One experienced psychiatrist  <u>Rater training</u> Trained in SCID-I procedure  <u>Inter observer reliability</u> Not reported	<u>MDD</u> Sensitivity: 80% Specificity: 94%	Moderate  <u>Comments</u> 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	<u>Design</u> Cross sectional  <u>Setting</u> One hospital  <u>Population</u> 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  <u>Prevalence MDD</u> 4.7%	<u>Index test</u> HADS at home sent by postal mail  <u>Reference test</u> MINI  <u>Number of patients</u> n=193 fulfilled both tests  <u>Drop-out rate</u> 15.7%	<u>Rater</u> First author did all the interviews  <u>Rater training</u> Not reported  <u>Interobserver reliability</u> Not reported	<u>MDD</u> Sensitivity: 45.7% Specificity: 91.8%	Moderate  <u>Comments</u> Sampling method Adequate

AMI = Acute myocardial infarction; CI = Confidence interval; CID-I = 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
Wittkamp et al 2009 [89] The Netherlands	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> 23 family practitioners (FP)</p> <p><u>Population</u> 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</p> <p><u>Inclusion criteria</u> Symptoms within 3 months</p> <p><u>Exclusion criteria</u> Other psychiatric diseases, somatic problems, language problems</p> <p><u>Prevalence (inverse probability weighting, IPW)</u> 12.3% (9.9–15.2), 82/664</p>	<p><u>Index test</u> PHQ-9</p> <p><u>Reference test</u> SCID-I Telephone interviews</p> <p><u>Number of patients</u> n=664 (443 females) with PHQ-9 score &gt;5 280 with PHQ-9 score &lt;5</p> <p><u>Drop-out rate</u> 27% did not complete both tests</p> <p>5 tests were excluded because of insufficient data</p>	<p><u>Rater</u> Researchers</p> <p><u>Rater training</u> Skilled professional. Ongoing supervision during study</p> <p>Monthly consensus</p> <p><u>Interobserver reliability</u> κ 0.73, 90% agreement</p>	<p><u>Depression (algorithm)</u> (Corrected with IPW) Sensitivity: 68% (95% CI, 57; 78)</p> <p>Specificity: 95% (95% CI, 93; 97)</p> <p><u>Depression (cut off 10)</u> Sensitivity: 100% (95% CI, 92; 100)</p> <p>Specificity: 45% (95% CI, 30; 60)</p>	<p>High</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Yes</p> <p><u>Other comments</u> Time between index and reference test stated as “just before”</p> <p>Not stated that the diagnosis was major depression</p>

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	<u>Design</u> Cross sectional  <u>Setting</u> Outpatient clinic in a medical hospital and 12 family practices  <u>Population</u> Random sample: n=2 050 outpatients at visit ≥18 years with somatic disease  <u>Prevalence MDD</u> 13.2%	<u>Index test</u> PHQ-9  <u>Reference test</u> SCID-I  <u>Number of patients</u> n=1 619 consented and filled in questionnaires n=528 completed SCID-I  <u>Drop-out rate</u> Missing data on 27 (5%)	<u>Rater</u> 4 trained raters psychologists, blind to the result of the questionnaires  <u>Rater training</u> Reliability tested by video-tapes  <u>Interobserver reliability</u> κ (SCID-I) 0.88 (0.47–1.0)	<u>MDD (algorithm)</u> Sensitivity: 83% (95% CI, 72; 91)  Specificity: 90% (95% CI, 87; 93)  <u>MDD (cut off 11)</u> Sensitivity: 90% (95% CI, 80; 96)  Specificity: 77% (95% CI, 73; 81)	Moderate  <u>Comments</u> Sampling method Adequate
Carballeira et al 2007 [90] Switzerland	<u>Design</u> Cross sectional  <u>Setting</u> Internal medicine units at a university hospital  <u>Population</u> n=1 053 consecutive patients  <u>Inclusion criteria</u> Age 18–64 years, fluent in French  <u>Exclusion criteria</u> Mental and physical disorders  <u>Prevalence MDD</u> 11.3%	<u>Index test</u> PHQ-9  <u>Reference test</u> Psychiatric assessment using DSM-IV criteria  <u>Number of patients</u> n=318 (38% female)  <u>Drop-out rate</u> 8.2%	<u>Rater</u> One licensed psychologist (PHQ-9) 2 psychiatrists and 1 pre-registration psychiatrist (reference test)  <u>Rater training</u> Trained before trial and ongoing evaluation during study every 20th patient  <u>Interobserver reliability</u> κ 0.75 DSM-IV diagnosis	<u>MDD (algorithm)</u> Sensitivity: 50% (95% CI, 30; 70)  Specificity: 86% (95% CI, 81; 91)	Moderate  <u>Comments</u> Sampling methods Adequate  <u>Blinding</u> Yes  <u>Indeterminate results</u> Not reported  <u>Other comments</u> 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	<u>Design</u> Cross sectional  <u>Setting</u> Inpatient ward in a dermatological hospital  <u>Population</u> Patients consecutively admitted to 5 of 8 inpatients' wards at the hospital  <u>Inclusion criteria</u> Patients ≥18 years  <u>Exclusion criteria</u> Patients with cognitive impairment or dementia  <u>Prevalence MDD</u> 8.4%	<u>Index test</u> PHQ-9  <u>Reference test</u> SCID-I  <u>Number of patients</u> 141 (79% of eligible) 56% women  <u>Drop-out rate</u> 9.9%	<u>Rater</u> Trained mental health personnel  <u>Rater training</u> Reliability not reported  <u>Interobserver reliability</u> Not reported	<u>MDD (algorithm)</u> Sensitivity: 55% (95% CI, 46; 63)  Specificity: 91% (95% CI, 86; 96)	Moderate  <u>Comments</u> Sampling method Adequate  <u>Blinding</u> Yes  <u>Other comments</u> Sponsored by Pfizer
Henkel et al 2004 [92] Germany	<u>Design</u> Cross sectional  <u>Setting</u> 18 primary care units  <u>Population</u> Consecutive patients who routinely presented in waiting room  <u>Prevalence</u> 10.2%	<u>Index test</u> PHQ-9  <u>Reference test</u> CIDI  <u>Number of patients</u> n=470 Completed CIDI  <u>Drop-out rate</u> 4.7%	<u>Rater</u> 6 psychologists and 1 psychiatrist  <u>Rater training</u> Yes by a designated training centre  <u>Interobserver reliability</u> "High standard of inter- rater reliability"	<u>MDD (algorithm)</u> Sensitivity: 78% (95% CI, 0.66; 0.87)  Specificity: 85% (95% CI, 0.81; 0.89)	Moderate  <u>Comments</u> Sampling method Adequate  <u>Blinding</u> 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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Clinics of Regional cancer centers</p> <p><u>Population</u> n=4 264 consecutive patients with cancer</p> <p><u>Inclusion criteria</u> ≥15 on HADS</p> <p><u>Exclusion criteria</u> Patients too ill or with significant communication and cognitive difficulties</p> <p><u>Prevalence MDD</u> 34.5% (270/782)</p>	<p><u>Index test</u> PHQ-9</p> <p><u>Reference test</u> SCID-I over telephone</p> <p><u>Number of patients</u> n=782</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater</u> Specially trained psychology graduates and nurses</p> <p><u>Rater training</u> Weekly supervision and review of audiotapes by psychiatrist</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>MDD (algorithm)</u> Sensitivity: 56% (95% CI, 0.55; 0.57)</p> <p>Specificity: 96% (95% CI, 0.95; 0.97)</p> <p><u>MDD cut off 10</u> Sensitivity: 73% (95% CI, 68; 78 )</p> <p>Specificity: 88% (95% CI, 87; 89)</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Not reported</p> <p><u>Other comments</u> Time between tests unclear "several days"</p>

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

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	<u>Design</u> Cross sectional  <u>Setting</u> One hospital  <u>Population</u> n=528 patients admitted for PTCA, AMI or coronary artery bypass graft surgery. Recruitment by postal invita- tion and follow-up phone call 6 weeks after discharge. Assess- ment 3 months post discharge  <u>Prevalence MDD</u> 18.1%	<u>Index test</u> PHQ-9 at home sent by postal mail  <u>Reference test</u> MINI by telephone within 2–3 days  <u>Number of patients</u> n=193 fulfilled criteria, consented and returned the PHQ-9  <u>Drop-out rate</u> 15.7%	<u>Rater</u> First author did all the interviews  <u>Rater training</u> Not stated  <u>Interobserver reliability</u> –	<u>MDD (algorithm)</u> Sensitivity: 34% Specificity: 97%  <u>MDD cut off 10</u> Sensitivity: 54% Specificity: 91%	Moderate
Spitzer et al 1999 [96] USA  Kroenke et al 2001 [191] USA	<u>Design</u> Cross sectional  <u>Setting</u> Waiting room for 8 primary care clinics, 62 primary care and 21 general internal medicine physicians  <u>Population</u> n=3 890 patients with hypertension 25%, arthritis 11%, diabetes 8%, pulmonary disease 7%  <u>Prevalence MDD</u> 7.1%	<u>Index test</u> PHQ-9  <u>Reference test</u> SCID-I by telephone  <u>Number of patients</u> n=3 000 (66% females) Mean age: 46±17.2 years  <u>Drop-out rate</u> 23.3%	<u>Rater</u> One PhD clinical psychologist and three senior psychi- atric social workers  <u>Rater training</u> Not reported  <u>Interobserver reliability</u> Not reported	<u>MDD (algorithm)</u> Sensitivity: 73% (95% CI, 0.59; 0.87)  Specificity: 98% (95% CI, 0.96; 1.00)  <u>MDD cut off 10</u> Sensitivity: 88% (95% CI, 74; 96)  Specificity: 88% (95% CI, 85; 91)	Moderate  <u>Comments</u> Sampling method Adequate  <u>Blinding</u> Yes  <u>Handling of missing data</u> Incomplete question- naires 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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> The otolaryngology department of a tertiary care hospital</p> <p><u>Population</u> Consecutive outpatients with dizziness</p> <p><u>Inclusion criteria</u> Patients 15–75 years</p> <p><u>Exclusion criteria</u> Not reading or writing Dutch, psycho-organic or psychotic disorders. Consulting for medico-legal reasons</p> <p><u>Prevalence MDD</u> 16.5% (16/97)</p>	<p><u>Index test</u> PHQ-9</p> <p><u>Reference test</u> MINI</p> <p><u>Number of patients</u> 310 met inclusion criteria n=268 agreed and completed PHQ-9  n=97 (random sample) were interviewed</p> <p><u>Drop-out rate</u> Unclear</p>	<p><u>Rater</u> A trained psychologist (one of the authors)</p> <p><u>Rater training</u> Not reported</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>MDD (algorithm)</u> Sensitivity: 68.8% Specificity: 94.4%</p>	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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> University hospitals in Spain</p> <p><u>Population</u> n=15 235 medical and surgical inpatients</p> <p>n=9 590 randomly selected</p> <p><u>Inclusion criteria</u> Patients 18–74 years</p> <p><u>Exclusion criteria</u> Intensive care patients. Those admitted to psychiatry and obstetric department or department of substance abuse. Physical impairment or illiteracy or non-Spanish speaking</p> <p><u>Prevalence MDD</u> 8.3%</p>	<p><u>Index test</u> PHQ-9</p> <p><u>Reference test</u> SCID-I</p> <p><u>Number of patients</u> n=1 003 fulfilled criteria and consented</p> <p><u>Drop-out rate</u> No drop-out</p>	<p><u>Rater</u> One of the authors (a clinical psychologist) interviewed all patients face-to-face 48 hours after completion of PHQ-9</p> <p><u>Rater training</u> Not reported “experienced clinical psychologist”</p> <p><u>Interobserver reliability</u> NA</p>	<p><u>MDD (algorithm)</u> Sensitivity: 84% (95% CI, 0.75; 0.94)</p> <p>Specificity: 92% (95% CI, 0.91; 0.92)</p>	<p>Moderate</p> <p><u>Comments</u> <u>Blinding</u> Yes</p>

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

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

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

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

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	<u>Design</u> Cross sectional  <u>Setting</u> One psychiatric clinic in Italy  <u>Population</u> n=123 consecutive patients seeking psychiatric care, or psychiatric evaluation before medical treatment, or applying for certification of mental status (gun, driving license)  <u>Prevalence</u> BP or schizoaffective: 21% (26/123)	<u>Index test</u> HCL-32 Cut off $\geq 14$  <u>Reference test</u> SCID-I  <u>Number of patients</u> n=123 (83% female) Mean age: 37.9 $\pm$ 12.4 years  <u>Drop-out rate</u> Not reported	<u>Rater of reference test</u> Unclear number of physicians in psychiatry  <u>Rater training</u> SCID-I training, but not specified  <u>Interobserver reliability</u> Not reported	<u>BP</u> Sensitivity: 73% Specificity: 69%	Moderate  <u>Comments</u> Sampling method Adequate  Blinding Probable  Handling of missing data Unclear
Meyer et al 2010 [112] Germany	<u>Design</u> Cross sectional  <u>Setting</u> Psychiatric clinics and outpatient clinics across Germany  <u>Population</u> n=488 consecutive patients Mean age: 39.22 years (SD 13.37, range 17–76) (62.7% females)  <u>Prevalence</u> BP: 32% (116/361)	<u>Index test</u> HCL-32 Cut off $\geq 14$  <u>Reference test</u> SCID-I  <u>Number of patients</u> n=361  <u>Drop-out rate</u> 26%	<u>Rater of reference test</u> Unclear number of psychologists and psychiatrists  <u>Rater training of reference test</u> Trained, but not specified  <u>Interobserver reliability</u> Reported only for 1/7 centres; BPI $\kappa$ 0.95 BPII $\kappa$ 0.89 Anxiety disorders $\kappa$ 0.95	<u>BP</u> Sensitivity: 88% Specificity: 36%	Moderate  <u>Comments</u> 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 symptom 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	<p><u>Design</u> Cross sectional and longitudinal</p> <p><u>Setting</u> General outpatient psychiatric clinic</p> <p><u>Population</u> n=183</p> <p><u>Inclusion criteria</u> Symptoms of mood disorder, newly referred for treatment, not previously treated in specialized unit for mood disorders</p> <p><u>Prevalence</u> BP: 30%</p>	<p><u>Index test</u> MDQ Standard criteria</p> <p><u>Reference test</u> SCID-I</p> <p><u>Number of patients</u> 146 (53% females) Median age: 40 years (range 19–64)</p> <p><u>Drop-out rate</u> 20%</p>	<p><u>Rater of reference test</u> Psychiatrist or psychologist</p> <p><u>Rater training</u> Yes but not specified how</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>BP</u> Sensitivity: 63.6% Specificity: 83.3%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Unclear</p> <p><u>Other comments</u> Unclear training and expertise of raters performing the SCID-I</p>

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

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	<u>Design</u> Cross sectional  <u>Setting</u> County jail in  <u>Population</u> n=597 detainees, n=526 completed MDQ at intake, n=146 accepted interview  <u>Prevalence</u> BP: 33.5% (55/164)	<u>Index test</u> MDQ; standard criteria  <u>Reference test</u> MINI  <u>Number of patients</u> n=164 (13% females) Mean age: 33.4 years (SD 10.9)  <u>Drop-out rate</u> Not reported	<u>Rater of reference test</u> Research assistant  <u>Rater training</u> Yes, certified by 100% agreement with expert in 10 interviews and main- tained interrater agreement for at least 80% of individual items  <u>Interobserver reliability</u> Not reported, see above	<u>BP</u> Sensitivity: 47% Specificity: 94%	Moderate  <u>Comments</u> <u>Sampling method</u> Yes, it is a special popula- tion but clinically relevant during intake at jail  <u>Blinding</u> Yes  <u>Handling of missing data</u> Unclear  <u>Other comments</u> Structured interview as reference test. Adequate training but numbers of raters not reported. Reliable analysis of non-participants
Meyer et al 2010 [112] Germany	<u>Design</u> Cross sectional  <u>Setting</u> Seven psychiatric clinics and outpatient clinics across Germany  <u>Population</u> n=488 consecutive patients Mean age: 39.22 years (SD 13.37, range 17–76)  <u>Prevalence</u> BP: 29% (126/440)	<u>Index test</u> MDQ; standard criteria  <u>Reference test</u> SCID-I  <u>Number of patients</u> n=440 (62.7% females)  <u>Drop-out rate</u> 10%	<u>Rater of reference test</u> Unclear number of psycho- logists and psychiatrists  <u>Rater training of reference test</u> Trained, but not described how  <u>Interobserver reliability</u> Reported only for 1/7 centers	<u>BP</u> Sensitivity: 80% Specificity: 64%	Moderate  <u>Comments</u> <u>Sampling method</u> Adequate  <u>Blinding</u> Yes  <u>Handling of missing data</u> 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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Prospective study on the course of MDD and BP during pregnancy and post partum period at a perinatal clinic in a psychiatric hospital</p> <p><u>Population</u> n=224 women were referred 125 completed MDQ 2–4 weeks after delivery</p> <p><u>Inclusion criteria</u> At least 18 years old, history of DSM-IV diagnosis prior to index pregnancy of MDD, BPI or BPII</p> <p><u>Exclusion criteria</u> 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</p> <p><u>Prevalence</u> BP: 46% (57/125) MDD: 54.4% (68/125)</p>	<p><u>Index test</u> MDQ; standard criteria</p> <p><u>Reference test</u> SCID-I</p> <p><u>Number of patients</u> n=125 women Mean age: 28 years (SD 5.17)</p> <p><u>Drop-out rate</u> None</p>	<p><u>Rater of reference test</u> An experienced research coordinator</p> <p><u>Rater training of reference test</u> Extensive training during 8 years</p> <p><u>Interobserver reliability</u> Unclear</p>	<p><u>BP</u> Sensitivity: 75.44% (95% CI, 62.24; 85.87)</p> <p>Specificity: 86.76% (95% CI, 76.36; 93.77)</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Excluded from analysis</p> <p><u>Other comments</u> Information about time period between index and reference test, and the rater's competence was achieved from corresponding author</p>

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	<u>Design</u> Cross sectional and longitudinal  <u>Setting</u> Outpatient psychiatric clinic in Switzerland  <u>Population</u> n=96 Mean age: 45 years (SD 9.8, range 18–63)  <u>Inclusion criteria</u> Patients with mood disorders  <u>Prevalence</u> BP: 56.3% (54/96)	<u>Index test</u> MDQ Standard criteria  <u>Reference test</u> SCID-I  <u>Number of patients</u> 96 (60% females)  <u>Drop-out rate</u> Not reported	<u>Rater of reference test</u> one psychiatrist and one psychologist  <u>Rater training</u> Trained, but not described how  <u>Interobserver reliability</u> Not reported	<u>BP</u> Sensitivity: 74.1% Specificity: 90.5%	Moderate  <u>Comments</u> <u>Sampling method</u> Yes, but from specialized mood disorder clinic and selection criteria not clearly described  <u>Blinding</u> Yes  <u>Handling of missing data</u> Unclear
Zimmerman et al 2011 [116] USA	<u>Design</u> Cross sectional  <u>Setting</u> Community-based outpatient practice affiliated with academic center; The Rhode Island MIDAS project  <u>Population</u> n=773 consecutive outpatients  <u>Prevalence</u> Lifetime BP: 10.6% (80/752)	<u>Index test</u> MDQ Standard criteria  <u>Reference test</u> SCID-I  <u>Number of patients</u> 752 (59.3% females) Mean age: 39.4 years (SD 13.9)  <u>Drop-out rate</u> 3%, 21 with incomplete MDQ	<u>Rater of reference test</u> PhD level psychologists and research assistants  <u>Rater training</u> Highly trained, monitored throughout project  <u>Interobserver reliability</u> Reliability diagnosing BP κ 0.75	<u>BP</u> Sensitivity: 67.5% Specificity: 84.5%	High  <u>Comments</u> <u>Sampling method</u> Adequate  <u>Blinding</u> Yes  <u>Handling of missing data</u> Patients with incomplete data not included in ana- lysis, 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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> General outpatient psychiatric clinic in Geneva, Switzerland</p> <p><u>Population</u> n=183 patients with symptoms of mood disorder, newly referred for treatment</p> <p><u>Exclusion criteria</u> Previously treated in specialized unit for mood disorders</p> <p><u>Prevalence</u> BP: 30% (44/146)</p>	<p><u>Index test</u> MDQ with standard criteria or modified Geneva algorithm</p> <p><u>Reference test</u> SCID-I</p> <p><u>Number of patients</u> n=146 (53% females) Median age: 40 years (range 19–64)</p> <p><u>Drop-out rate</u> 20%</p>	<p><u>Rater of reference test</u> Psychiatrist or psychologist</p> <p><u>Rater training</u> Well-trained, but not described how</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>BP vs UP</u> <u>Standard criteria</u> Sensitivity: 63.6% Specificity: 83.3%</p> <p><u>Geneva algorithm</u> Sensitivity: 72.7% Specificity: 78.4%</p> <p><u>BP I vs UP</u> <u>Standard criteria</u> Sensitivity: 85.0%</p> <p><u>Geneva algorithm</u> Sensitivity: 90.0%</p> <p><u>BP II vs UP</u> <u>Standard criteria</u> Sensitivity: 45.8%</p> <p><u>Geneva algorithm</u> Sensitivity: 58.3%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p>Blinding Yes</p> <p>Handling of missing data Unclear</p>

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

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

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

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Public university, USA</p> <p><u>Population</u> n=137 students visiting the university counselling centre sample</p> <p><u>Prevalence</u> Mood disorder: 48%</p>	<p><u>Index test</u> BDI-II</p> <p><u>Reference test</u> SCID-I</p> <p><u>Number of patients</u> n=137 Mean age: 22.0 years (SD 3.44)</p> <p><u>Drop-out rate</u> No information</p>	<p><u>Rater of reference test</u> 9 counsellors</p> <p><u>Training of reference test</u> Yes, but not described how</p> <p><u>Interobserver reliability</u> Not reported</p> <p><u>Outcome measure</u> Correlation between BDI-II scores and number of MDD symptoms (0–9) and degree of impairment (GAF; mild, moderate and severe)</p>	<p><u>Correlation coefficient</u> r=0.83; r<sup>2</sup>=0.69, ie, 69% explained variance</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Acceptable</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Not reported</p>

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

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	<p><u>Design</u> RCT, double-blind multicenter trial of venlafaxin ER treatment</p> <p><u>Setting</u> 29 sites in the US</p> <p><u>Population</u> 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</p> <p><u>Inclusion criteria</u> Patients &gt;18 years HDRS-17 ≥20 at screening and ≥18 at randomisation one week prior to trial start</p> <p><u>Test situation, period between performance</u> Acute-treatment phase of 10 weeks and continuation phase of 6 months</p> <p>Tests were administrated at 1, 2, 3, 4, 5, 6, 8 and 10 weeks during acute phase and once monthly at continuation phase</p>	<p><u>Index test</u> IDS-sr and QIDS-sr</p> <p><u>Reference tests</u> Clinician rating by CGI-C</p> <p><u>Number of patients</u> n=1 047 acute phase n=715 continuation phase</p> <p>Mean age: 40 years (65% females)</p> <p><u>Drop-out rate</u> -</p>	<p><u>Rater</u> Treating physicians</p> <p><u>Rater training</u> Unclear</p> <p><u>Interobserver reliability</u> Not reported</p> <p><u>Outcome measure</u> Correlations between patients and clinicians rating of MDD severity</p>	<p><u>IDS-sr vs CGI</u> <i>Pearsons correlation</i> Baseline: 0.27 Week 10: 0.67 Month 6: 0.65</p> <p><u>QIDS-sr vs CGI</u> <i>Pearsons correlation</i> Baseline: 0.24 Week 10: 0.63 Month 6: 0.64</p>	<p>Low</p> <p><u>Comments</u> <i>Sampling method</i></p> <p><i>Blinding</i> Unclear</p> <p><i>Other</i> Time between tests unclear</p> <p>Secondary analysis of studies</p>

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	<u>Design</u> Post hoc analysis of pooled data from 5 RCT  <u>Setting</u> All studies in which also CGI was used  <u>Population</u> Outpatients subjected to acute treatment with escitalopram for MDD  <u>Test situation, period between performance</u> 8 weeks	<u>Index test</u> MADRS  <u>Reference tests</u> CGI-S CGI-I  <u>Number of patients</u> n=1 992 (pooled data) MDD  <u>Drop-out rate</u> Not reported	<u>Rater</u> Clinicians/investigators  <u>Rater training</u> Not reported  <u>Interobserver reliability</u> Not reported  <u>Outcome measure</u> Correlation between tests (Thresholds for response and remission)	<u>MADRS vs CGI-S</u> <i>Spearman rank correlation</i> 0.82 across all groups  <u>MADRS vs CGI-I</u> <i>Spearman rank correlation</i> 0.86 across all groups	Low  <u>Comments</u> Study sponsored by Lundbeck A/S, Denmark or Forest Laboratories Inc New York
Müller et al 2003 [141] Germany	<u>Design</u> Longitudinal study of MDD, continuous clinical assessment  <u>Setting</u> Hospitalized patients treated according to clinical routine  <u>Population</u> Inpatients with MDD  <u>Test situation, period between performance</u> Tests administered at the same time of day within one week after admission or before discharge	<u>Index test</u> MADRS  <u>Reference test</u> CGI  <u>Number of patients</u> n=85 (69% female) Mean age: 51.4±14.6 years  <u>Drop-out rate</u> Not reported	<u>Rater</u> One of five treating physicians  <u>Rater training</u> "Trained"  <u>Interobserver reliability</u> ICC >0.85  <u>Outcome measure</u> To assess depression severity using MADRS	<u>MADRS vs CGI</u> <i>Pearsons correlation coefficient</i> 0.87 p<0.0001	Low  <u>Comments</u> <i>Sampling method</i>  <i>Blinding</i> No  <i>Other comments</i> 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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Affective disorder outpatient clinic and inpatient wards at a University Hospital in Sweden</p> <p><u>Population</u> n=61, opportunistically recruited patients (61% females) Mean age: 44 years, (range 17–76)</p> <p><u>Inclusion criteria</u> Not reported</p> <p><u>Exclusion criteria</u> Not reported</p> <p><u>Prevalence</u> BPI: 61% (n=37) BPII: 13% (n=8) BPNOS: 13% (n=8) MDD: 13% (n=8)</p>	<p><u>Index test</u> AS-18</p> <p><u>Reference test</u> CGI-BP</p> <p><u>Number of patients</u> n=61</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of reference test</u> Clinicians</p> <p><u>Training of reference test</u> Not reported</p> <p><u>Interobserver reliability</u> Not reported</p> <p><u>Outcome measure</u> Spearman correlations between depression and mania subscores of the AS-18 with CGI-BP depression and mania scores</p>	<p><u>AS-18-Dep vs CGI-BP-D</u> r=0.68</p> <p><u>AS-18-Man vs CGI-BP-M</u> r=0.80</p> <p><u>AS-18-Dep vs CGI-BP-M</u> r=-0.01</p> <p><u>AS-18-Man vs CGI-BP-D</u> r=0.10</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Yes, but small convenience sample and no structured diagnostic interviews were used for making diagnoses</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Not reported</p>

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> 18 psychiatric centers across Spain</p> <p><u>Population</u> n=215 consecutive outpatients with BPI or BPII on treatment as usual</p> <p><u>Inclusion criteria</u> BPI or BPII, stable patients without restriction in scores, unstable patients with YMRS <math>\geq 18</math> and MADRS <math>\geq 22</math></p> <p><u>Exclusion criteria</u> –</p> <p><u>Prevalence</u> Not applicable</p>	<p><u>Index test</u> MAS</p> <p><u>Reference test</u> CGI-BP</p> <p><u>Number of patients</u> n=113 presenting with manic episode Mean age: 43.1 years</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of index test</u> Clinicians</p> <p><u>Training of index rater</u> Not reported</p> <p><u>Rater of reference test</u> Same clinicians</p> <p><u>Training of reference test</u> Not reported</p> <p><u>Interobserver reliability</u> ICC (95% CI): 0.89 (0.80; 0.97) (n=46)</p> <p><u>Outcome measure</u> One-way ANOVAs of MAS baseline scores with CGI categories as dependent variables</p>	<p><u>MAS vs CGI-BP</u> Adequate</p>	<p>Low</p> <p><u>Comments</u> Sampling method Adequate</p> <p>Blinding No</p> <p>Handling of missing data Not reported</p> <p>Other comments Neither correlation coefficients or Intraclass coefficients reported. No post hoc (Eta<sup>2</sup>) analysis of ANOVA statistics</p>

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> 18 psychiatric centers across Spain</p> <p><u>Population</u> n=215 consecutive outpatients with BPI and BPII on treatment as usual</p> <p><u>Inclusion criteria</u> BPI or BPII, stable patients without restriction in scores, unstable patients with YMRS <math>\geq 18</math> and MADRS <math>\geq 22</math></p> <p><u>Exclusion criteria</u> –</p> <p><u>Prevalence</u> Not applicable</p>	<p><u>Index test</u> YMRS</p> <p><u>Reference test</u> CGI-BP</p> <p><u>Number of patients</u> n=113 presenting with manic episode Mean age: 43.1 years (SD 13.2), 60.7% women</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of index test</u> Clinicians</p> <p><u>Training of index rater</u> Not reported</p> <p><u>Rater of reference test</u> Same clinicians</p> <p><u>Training of reference test</u> Not reported</p> <p><u>Interobserver reliability</u> Not reported</p> <p><u>Outcome measure</u> One-way ANOVAs of baseline YMRS scores with CGI categories as dependent variables</p>	<p><u>YMRS vs CGI-BP</u> Adequate</p>	<p>Low</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> No</p> <p><u>Handling of missing data</u> Not reported</p> <p><u>Other comments</u> Neither correlation coefficients or Intraclass coefficients reported. No post hoc (Eta<sup>2</sup>) analysis of ANOVA statistics</p>

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	<p><u>Design</u> Cross sectional and longitudinal</p> <p><u>Setting</u> Participants were recruited from community through advertisements in newsletters and websites in Illinois</p> <p><u>Population</u> n=40 (70% females) Mean age: 43.5 years (SD 12.2), range 27–72</p> <p><u>Inclusion criteria</u> Current diagnosis of bipolar disorder I or II. At least 18 years old</p> <p><u>Exclusion criteria</u> None</p>	<p><u>Index test</u> LCM-p</p> <p><u>Reference test</u> CGI-BP ratings immediately after telephone ratings of HDRS, YMRS and GAF at three occasions a month apart</p> <p><u>Number of patients</u> Time 1: 35 Time 2: 36 Time 3: 32</p> <p>29 patients had complete data on all occasion</p> <p><u>Drop-out rate</u> See above</p>	<p><u>Rater of reference test</u> One researcher</p> <p><u>Training of reference test</u> Not reported</p> <p><u>Interobserver reliability</u> NA</p> <p><u>Outcome measure</u> 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</p>	<p><u>LCM-HIGH vs CGI-mania</u> Time 1: 0.318 Time 2: 0.383 Time 3: 0.556</p> <p><u>LCM-LOW vs CGI-depression</u> Time 1: 0.430 Time 2: 0.593 Time 3: 0.516</p> <p><u>LCM-LOW vs CGI-mania</u> Time 1: -0.193 Time 2: 0.054 Time 3: 0.144</p> <p><u>LCM-HIGH vs CGI-depression</u> Time 1: -0.072 Time 2: 0.019 Time 3: 0.055</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Yes</p> <p><u>Handling of missing data</u> Excluded from analysis</p> <p><u>Other comments</u> 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</p>

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

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Department of psychiatry at a hospital in Paris</p> <p><u>Population</u> Consecutively admitted inpatients before treatment with antidepressants</p> <p><u>Exclusion criteria</u> Bipolar and schizophrenic disorders, alcohol and drug abuse, neurological disorders, unstable medical disorders</p> <p><u>Test situation, period between performance</u> 4 weeks follow-up; day 10 and 28 Clinician test first</p>	<p><u>Index test</u> IDS-SR</p> <p><u>Reference test</u> IDS-C</p> <p><u>Number of patients</u> n=68 (68.4% females) Mean age: 41.7±12.6 years</p> <p><u>Drop-out rate</u> Day 10: n=4 Day 28: n=5</p>	<p><u>Rater</u> One psychiatrist</p> <p><u>Rater training</u> Not stated</p> <p><u>Outcome measure</u> Pearson correlations, ICC (95% CI) between self vs clinician ratings</p>	<p><u>Day 0</u> Pearson r: 0.79 ICC: 0.76 (95% CI, 0.62; 0.83)</p> <p><u>Day 10</u> Pearson r: 0.93 ICC: 0.91 (95% CI, 0.85; 0.94)</p> <p><u>Day 28</u> Pearson r: 0.89 ICC: 0.85 (95% CI, 0.76; 0.90)</p>	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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> 14 public mental health clinics Texas medication algorithm project</p> <p><u>Population</u> Outpatients with psychotic and non-psychotic depression</p> <p><u>Inclusion criteria</u> ≥18 years DSM-IV criteria for MDD</p> <p><u>Exclusion criteria</u> Patients receiving mental retardation services or assertive community treatment</p> <p><u>Test situation, period between performance</u> Tests at same time (personal communication)</p>	<p><u>Index test</u> IDS-SR QIDS-16-SR</p> <p><u>Reference test</u> IDS-C QIDS-16-C</p> <p>QIDS-16 results were extracted from the IDS tests</p> <p><u>Number of patients</u> n=544 (79% female) Mean age: 42.2±11.1 years</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater</u> Research outcome assessors independent from clinicians without knowledge of the self-rate results</p> <p><u>Rater training</u> Not stated</p> <p><u>Outcome measure</u> Pearson correlations, ICC between self vs clinician ratings</p>	<p><u>IDS-30</u> ICC: 0.88 Pearson r: 0.89 p&lt;0.001</p> <p><u>IDS-30</u> ICC: 0.87 Pearson r: 0.87 p&lt;0.001</p>	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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> 18 primary and 23 speciality care settings participating in the STAR*D clinical trial</p> <p><u>Population</u> n=1 500 consecutive out-patients with non-psychotic major depressive disorder</p> <p><u>Inclusion criteria</u> Age 18–75 years</p> <p><u>Exclusion criteria</u> Patient with psychotic and compulsive disorders, bulimia nervosa, intolerance to protocol treatment and medical conditions such as seizures</p> <p><u>Test situation, period between performance</u> 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</p>	<p><u>Index test</u> QIDS-16-SR</p> <p><u>Reference test</u> QIDS-16-C</p> <p><u>Number of patients</u> Exit: n=582 (63.1% females) Mean age: 42.6±13.1 years</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater</u> Clinical research coordinators independent from clinicians without knowledge of treatment</p> <p><u>Rater training</u> Research outcome assessors trained and certified</p> <p><u>Outcome measure</u> Pearson correlations between self vs clinician ratings</p>	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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> 3 psycho-geriatric outpatient clinics</p> <p><u>Population</u> Patients with dementia Elderly controls without dementia</p> <p><u>Inclusion criteria</u> Controls were relatives to patients or members of local organisations for elderly</p> <p><u>Exclusion criteria</u> Major psychiatric illness or aphasic disorder</p> <p><u>Prevalence</u> Depression: 67% Depression + dementia: 33%</p>	<p><u>Index test</u> HDRS-17</p> <p><u>Reference test</u> CGI</p> <p><u>Number of patients</u> n=145 (102 females, 78.6±6.8 years, 43 male 72.4±5.6 years)</p> <p>Dementia: 47 Non-dementia: 98</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater</u> 7 psycho-geriatricians</p> <p><u>Rater training</u> Co-rating sessions prior to and during the study</p> <p><u>Interobserver reliability</u> ICC: 0.98</p> <p><u>Outcome measure</u> The correlation of HDRS-17 ratings of depression</p>	<p><u>ICC</u> All 0.85</p> <p><u>Non-dementia</u> 0.87</p> <p><u>Dementia</u> 0.78</p>	<p>Moderate</p> <p><u>Comments</u> Blinding Raters blinded</p> <p><u>Handling of missing data</u> Not reported</p>

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	<u>Design</u> Cross sectional  <u>Setting</u> 2 mental health care centres  <u>Population</u> Outpatients with ongoing or starting treatment for mood disorders  <u>Test situation, period between performance</u> Assessment at same instance, self-rate first at inclusion and after 4 weeks	<u>Index test</u> MADRS-S  <u>Reference test</u> MADRS  Face-to-face interview by psychiatrist before MADRS  <u>Number of patients</u> n=63 (50% females) Mean age: 43.9±11.2 years  <u>Drop-out rate</u> Not reported	<u>Rater</u> Psychiatrist and one senior psychiatrist  <u>Rater training</u> Interrater reliability assessed through 23 videotapes ICC: 0.87  <u>Outcome measure</u> Concurrent validity correlations between MADRS and MADRS-S (Pearson's correlation and Cohen's $\kappa$ )	<u>Concurrent validity sum of scores</u> At inclusion (total score) 0.81, p<0.001 After 4 weeks 0.91, p<0.001  <u>Agreement MADRS-S and MADRS</u> 0.69 ( $\kappa$ -statistics)  <u>Correlation between changes of scores on both instrument</u> 0.71, p< 0.001	Moderate
Cunningham et al 2011 [177] Sweden	<u>Design</u> Longitudinal drug trial  <u>Setting</u> General practice  <u>Population</u> Patients with MDD  <u>Inclusion criteria</u> Age: ≥18–70 years DSM-III-R criteria for MDD  <u>Test situation, period between performance</u> Follow-up 6 and 12 months Both tests at same visit MADRS-S in waiting-room before visit	<u>Index test</u> MADRS-S  <u>Reference test</u> MADRS  <u>Number of patients</u> n=400 (71,8% female) Mean age: 47.2±12.3 years  <u>Drop-out rate</u> At baseline n=10 At 52 weeks n=82  Complete data from MADRS-S and MADRS from 249 patients	<u>Rater</u> General practitioners rated their own patients  <u>Rater training</u> Before trial co-rating meetings including videotaped interviews Interrater reliability ICC: 0.89  <u>Outcome measure</u> Agreement of patient (MADRS-S) and general practitioner (MADRS) assessment of response and remission	<u>ICC</u> <u>Baseline</u> 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  <u>Blinding</u> 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

tion 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-IV-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 multi-center 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 Neuropsychiatric 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, Knorrning 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. Carpiello 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. Wittkamp 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 research. 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, Vastenburger 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 checklist (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 Epidemiol 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 Epidemiol 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, Slembariski 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 Epidemiol 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 Bipolar 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, Schlessler 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, Sydsjö 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, Gameraff 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

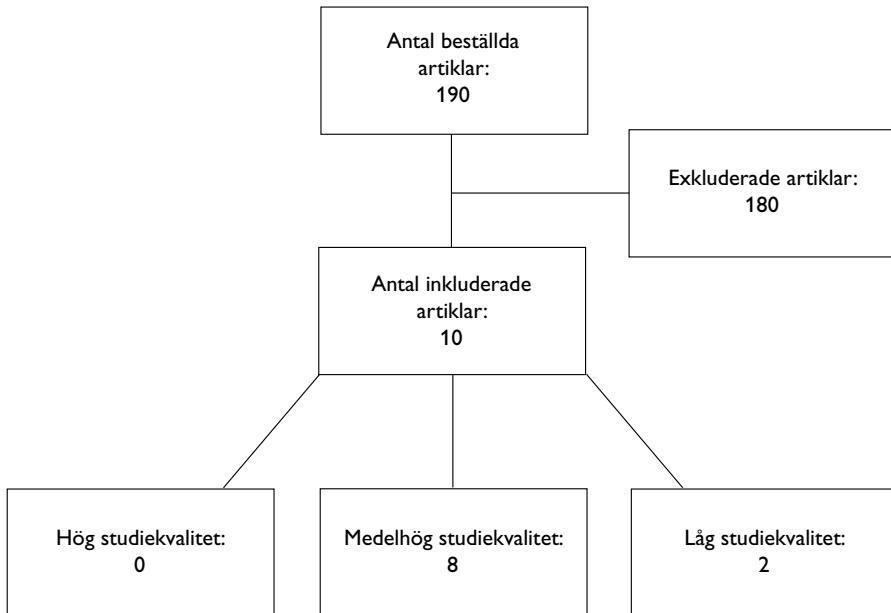
Detta avsnitt granskar det vetenskapliga underlaget för bedömningsformulär som används för äldre. Med äldre avser vi personer som är 65 år och däröver. Mellan 10 och 15 procent av äldre i västvärlden har en depression [1,2]. Depression hos äldre medför en ökad mortalitetsrisk [3–5] även om man tar hänsyn till samsjuklighet. Självmod 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 stämningssläge 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

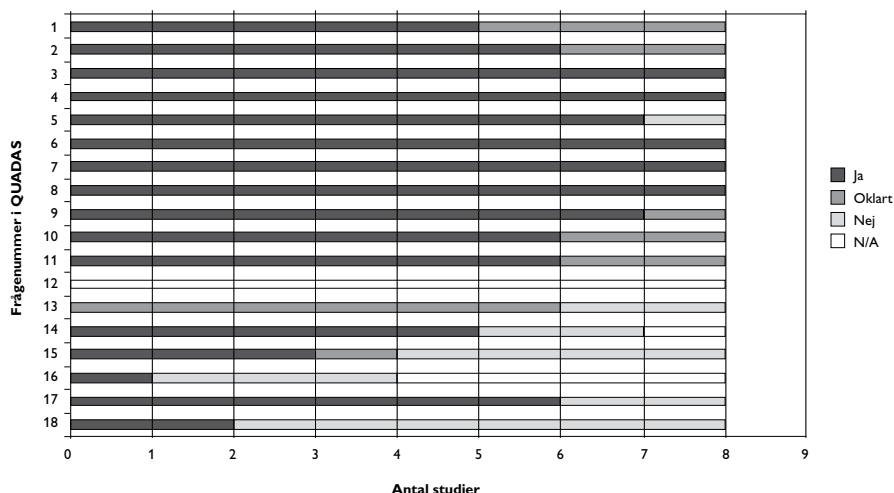


**Figur 3.6.1** Flödesschema äldre.

## 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 referensstandarderna.



**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.

<b>Formulär</b>	<b>Syfte</b>	<b>Referensstandard</b>	<b>Evidensgraderat resultat</b>
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].

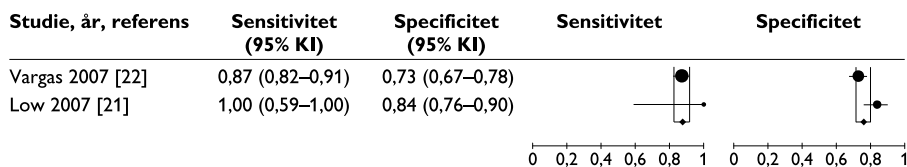
Det vetenskapliga underlaget bestod därmed av två studier som bedömdes ha medelhög kvalitet [21,22]. SCID-I [23] användes som referensstandard i båda studierna. Se Tabell 3.6.6.

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 dubbel-diagnosen 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 tröskelvärde 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. Tröskelvärde 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).



Sammanvägd **sensitivitet** = 0,87 (0,82 till 0,92)

Sammanvägd **specificitet** = 0,76 (0,72 till 0,80)

**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.

Utfalls-mått	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.

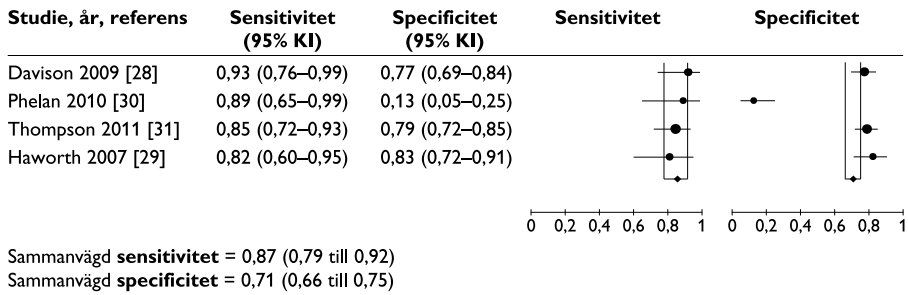
I studien av Thompson och medarbetare ingick 214 amerikanska personer med Parkinsons sjukdom med medelålder 72,5 år [31]. Personerna ingick i ett större forskningsprojekt om Parkinsons sjukdom. Personer med MMSE under 15 exkluderades. Studien utvärderade tillförlitlighet för egentlig depression och minor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



**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 studie-kvalitet (-1) Bristande precision (-1)
Specifitet	4 studier n=539	71% (66; 75)	Begränsat ⊕⊕○○	Bristande studie-kvalitet (-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 specifitet 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 tillägsfrågor 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.

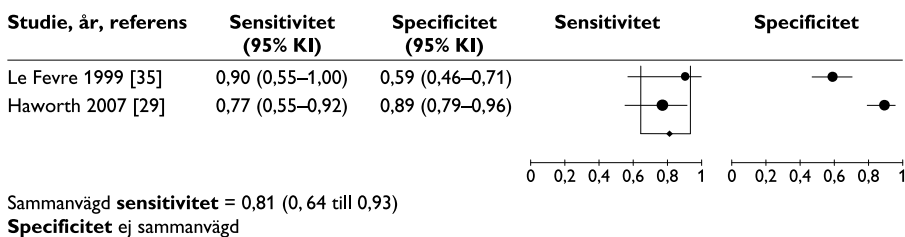
Båda studier hade medelhög studiekvalitet [29,35] (Tabell 3.6.8). Den ena var den ovan nämnda studien av Haworth och medarbetare [29]. Den andra studien, av Le Fevre och medarbetare, omfattade 79 konsekutivt rekryterade personer [35]. De vårdades palliativt vid ett hospice i England, de flesta pga cancer. 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.*

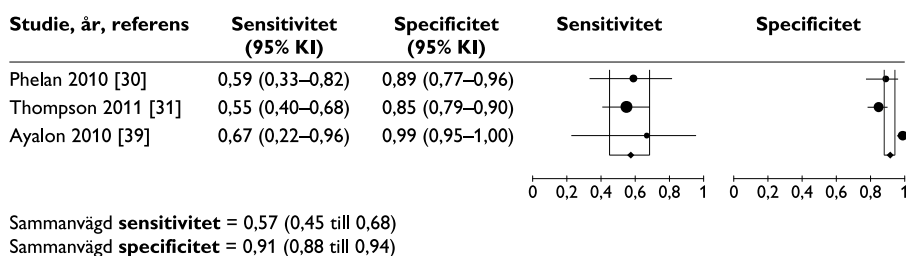
Av sju artiklar som lästes i fulltext exkluderades fyra, främst pga irrelevant referensstandard, att det förflutit för lång tid mellan index- och referenstest samt att tidsintervallet inte redovisades.

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	Sammanvägt 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 överförbarhet (-1)
Specificitet	3 studier n=436	91% (88; 94%)	Begränsat ⊕⊕○○	Bristande studie-kvalitet (-1) Bristande överförbarhet (-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 träskelvärde. 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 tröskelvärde 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 tröskelvärde 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 formulärens 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 specificiteten 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 tröskelvärde 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 äldrepopulationer 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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> General non-psychiatric outpatient clinics</p> <p><u>Population</u> Non-randomised sample, predominantly women Mean age: 70 years (range 60–96) Caucasians: 78%</p> <p><u>Inclusion criteria</u> Age: ≥60 years</p> <p><u>Exclusion criteria</u> MMSE &gt;13 for illiterate, ≥18 for low to medium-level education, ≥26 for high educational level</p> <p><u>Prevalence MDD</u> 43% (210/484)</p>	<p><u>Index test</u> GDS-30, orally</p> <p><u>Reference test</u> SCID-CV</p> <p><u>Number of patients</u> n=484</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of index test</u> 5 trained interviewers</p> <p><u>Training of index rater</u> Eight hour program, conducted by a pedagogue</p> <p><u>Rater of reference test</u> Psychiatrist</p> <p><u>Training of reference test</u> Trained in the use of SCID-SV</p> <p><u>Interobserver reliability</u> Not specified</p>	<p>Sensitivity: 87% (95% CI, 81.3; 91.0)</p> <p>Specificity: 73% (95% CI, 66.9; 77.8)</p>	<p>Moderate</p> <p><u>Comments</u> Blinding Adequate</p> <p><u>Handling of missing data</u> Not reported</p> <p><u>Other comments</u> Cut off point = 12</p> <p>Index and reference test were completed within one day (personal communication with authors)</p>

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

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

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

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

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

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

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Hospice</p> <p><u>Population</u> n=148 consecutive persons in palliative care were eligible</p> <p><u>Exclusion criteria</u> Seriously ill or close to death</p> <p><u>Prevalence</u> MDD: 18% 13% were classified as depressed using the Endicott criteria</p>	<p><u>Index test</u> HADS-D</p> <p><u>Reference test</u> CIS-R, diagnoses according to ICD-10</p> <p><u>Number of patients</u> n=79 Mean age: 68 years</p> <p><u>Drop-out rate</u> None</p>	<p><u>Index test rater</u> Self rating</p> <p><u>Index test training</u> NA</p> <p><u>Reference test rater</u> One clinician</p> <p><u>Rater training</u> Not reported</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>MDD cut off</u> <math>\geq 8</math> Sensitivity: 91% Specificity: 59%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> Not reported</p>

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

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

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

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

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

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> Two coronary care units</p> <p><u>Population</u> n=224 consecutively</p> <p><u>Inclusion criteria</u> Acute (recent or evolving) MI or unstable angina pectoris. Ability to speak and understand english. Medical stability, freedom of angina pain and of other life threatening medical conditions</p> <p><u>Exclusion criteria</u> Acute delirium, cognitive impairment MMSE: ≤23 Ongoing kidney dialysis</p> <p><u>Prevalence</u> MDD: 5% MDD + dystymic disorder: 5.9% Any depressive disorder: 11.8%</p>	<p><u>Index test</u> BDI-II</p> <p><u>Reference test</u> SCID</p> <p><u>Number of patients</u> 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</p> <p><u>Drop-out rate</u> Not reported</p>	<p><u>Rater of index test</u> Self rating</p> <p><u>Training of index rater</u> Not reported</p> <p><u>Rater of reference test</u> First author</p> <p><u>Training of reference test</u> Not reported</p> <p><u>Interobserver reliability</u> Not reported</p>	<p><u>MDD+ dystymic disorder</u> Sensitivity: 86% Specificity: 89%</p>	<p>Moderate</p> <p><u>Comments</u> Sampling method Adequate</p> <p>Blinding Adequate</p> <p>Handling of missing data Not reported</p> <p>Other comments CI not reported</p> <p>The index and reference tests were given in the same session (personal communication with authors)</p>

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

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	<p><u>Design</u> Cross sectional</p> <p><u>Setting</u> An inpatient based rehabilitation unit</p> <p><u>Population</u> 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</p> <p><u>Exclusion criteria</u> Severe sensory deficits, cognitive impairment, significant aphasia and abuse problems</p> <p><u>Prevalence (SCID)</u> Minor depression: 28.4% MDD: 15%</p>	<p><u>Index test</u> SIDI</p> <p><u>Reference test</u> SCID-CV</p> <p><u>Number of patients</u> n=67 Mean MMSE: 23.8 (SD 4.0)</p> <p><u>Drop-out rate</u> 0%</p>	<p><u>Rater of index test</u> Self rating</p> <p><u>Training of index rater</u> Not reported</p> <p><u>Rater of reference test</u> A trained doctoral-level provider</p> <p><u>Training of reference test</u> Not reported</p> <p><u>Interobserver reliability</u> Not specified</p>	<p><u>Minor and major depression at cut off ≥17</u> Sensitivity: 19% Specificity: 95%</p>	<p><u>Study quality</u> Moderate</p> <p><u>Comments</u> Selection of persons not described</p> <p><u>Blinding</u> Satisfactory</p> <p><u>Handling of missing data</u> Not reported</p> <p><u>Other comments</u> CI not reported</p> <p>All patients did not fulfill DSM-IV criteria for duration of depression</p>

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. Leshner 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 out-patient 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, Lindsay 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	<p><u>Intervention</u> Formal methods to identify postnatal depression</p> <p><u>Control</u> Not using formal methods to identify postnatal depression or usual care</p>	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. Neuro-psychiatric 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
<b>MINI</b> 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
<b>SCID-I</b> 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
<b>BDI-II</b> 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
<b>CBCL</b>	Se Tabell 4.3				
<b>CDI</b> 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
<b>EPDS</b> 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
<b>GDS-15</b> Geriatric Depression Scale	Äldre Självbedömning	Brink 1982 [24]	Ja Carl-Gerhard Gottfries med tillstånd av upphovsmannen Yesavage, men inte återöversatt	15 frågor	Oklart
<b>GDS-30</b> 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 symptom 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 upp- hovsmännen 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

<b>Formulär</b>	<b>Ålder och bedömare</b>	<b>Originalpublikation</b>	<b>Översättning</b>	<b>Omfattning</b>	<b>Kostnad</b>
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 for 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 innehåller 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
<b>GBI</b> General Behavior Inventory	Barn Versioner finns för både föräldra- och självbedömningar	Klein et al 1989 [37]	Ej identifierad	73 frågor	Oklar
<b>HCL-32</b> 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 Stämningssdagböcker	Roy-Byrne et al 1985 [39]	Flera svenska varianter finns	Inte tillämpligt	Oklart
<b>MDQ</b> 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)
<b>YMRS</b> 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 Symptomatology	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

## Referenser

1. Nezu AM, Ronan GF, Meadows EA, McClure KS. Practitioner's Guide to Empirically Based measures of Depression; 2000.
2. Cunningham CE, Boyle MH, Hong S, Pettingill P, Bohaychuk D. The Brief Child and Family Phone Interview (BCFPI): 1. Rationale, development, and description of a computerized children's mental health intake and outcome assessment tool. *J Child Psychol Psychiatry* 2009;50:416-23.
3. 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.
4. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000;41:645-55.
5. Herjanic B, Reich W. Development of a structured psychiatric interview for children: agreement between child and parent on individual symptoms. *J Abnorm Child Psychol* 1982;10:307-24.
6. 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.
7. Matzner F, Silva R, Silvan M, Chowdhury M, Nastasi L. "Preliminary Test-retest Reliability of the KID-SCID", Scientific Proceedings, American Psychiatric Association Meeting; 1997.
8. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 1997;36:980-8.
9. 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.
10. 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.
11. Endicott J, Spitzer RL. A diagnostic interview: The schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry* 1978;35:837-44.
12. 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.

13. 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.
14. Dahl AA, Kruger MB, Dahl NH, Karlson H, Knorrning LV, Stordal E. SPIFA- A presentation of the Structured Psychiatric Interview for General Practice. *Nord J Psychiatry* 2009;1-11.
15. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561-71.
16. Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *J Pers Assess* 1996;67:588-97.
17. Kovacs M. The Children's Depression, Inventory (CDI). *Psychopharmacol Bull* 1985;21:995-8.
18. 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.
19. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1:385-401.
20. Weissman MM, Orvaschel H, Padian N. Children's symptom and social functioning self-report scales. Comparison of mothers' and children's reports. *J Nerv Ment Dis* 1980;168:736-40.
21. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell Scale for Depression in Dementia. *Biol Psychiatry* 1988;23:271-84.
22. Birlleson P. The validity of depressive disorder in childhood and the development of a self-rating scale: a research report. *J Child Psychol Psychiatry* 1981; 22:73-88.
23. 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.
24. Brink TL. Screening tests for geriatric depression. *Clinical Gerontologist: The Journal of Aging and Mental Health* 1982; 1:37-43.
25. 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.
26. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361-70.
27. 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.
28. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report ver-

- sion of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999;282:1737-44.
29. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry* 2000;177:534-9.
30. 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.
31. Rybarczyk B, Winemiller DR, Lazarus LW, Haut A. Validation of a depression screening measure for stroke inpatients. *Am J Geriatr Psychiatry* 1996;4:131-139.
32. Rosenthal MP, Goldfarb NI, Carlson BL, Sagi PC, Balaban DJ. Assessment of depression in a family practice center. *J Fam Pract* 1987;25:143-9.
33. Achenbach TM. The Achenbach system of empirically based assessment (ASEBA): Developmen, findings, theory and applications. 2009. Burlington VT: university of Vermont, Research center for children, youth & families; 2009.
34. Adler M, Liberg B, Andersson S, Isacson 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.
35. Papolos D, Hennen J, Cockerham MS, Thode HC, Jr., Youngstrom EA. The child bipolar questionnaire: a dimensional approach to screening for pediatric bipolar disorder. *J Affect Disord* 2006;95:149-58.
36. Pavuluri MN, Henry DB, Devineni B, Carbray JA, Birmaher B. Child mania rating scale: development, reliability, and validity. *J Am Acad Child Adolesc Psychiatry* 2006;45:550-60.
37. Klein DN, Dickstein S, Taylor EB, Harding K. Identifying chronic affective disorders in outpatients: validation of the General Behavior Inventory. *J Consult Clin Psychol* 1989;57:106-11.
38. 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.
39. 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.
40. 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.
41. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability,

- validity and sensitivity. *Br J Psychiatry* 1978;133:429-35.
42. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56-62.
43. 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.
44. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134:382-9.
45. Svanborg P, Asberg M. A new self-rating scale for depression and anxiety states based on the Comprehensive Psycho-pathological Rating Scale. *Acta Psychiatr Scand* 1994;89:21-8.
46. Zung WW. A Self-Rating Depression Scale. *Arch Gen Psychiatry* 1965;12:63-70.
47. 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.
48. Bech P, Rafaelsen OJ, Kramp P, Bolwig TG. The mania rating scale: scale construction and inter-observer agreement. *Neuropharmacology* 1978;17:430-1.
49. Hesbacher PT, Rickels K, Morris RJ, et al. Psychiatric illness in family practice. *J Clin Psychiatry* 1980;41:6-10.



# 5. Klinisk betydelse av prestanda hos olika tester

---

## Sensitivitet och specificitet – vad betyder de i kliniken?

Ett test karakteriseras av förmågan att upptäcka sjukdom, sensitivitet (positivt test när sjukdom föreligger) respektive att utesluta sjukdom hos en frisk person, specificitet (negativt test vid frånvaro av sjukdom).

För praktiskt bruk kan andra mått vara att föredra. De mått vi kommer att belysa är dels de positiva respektive negativa prediktiva värdena, dels andelen patienter som är falskt positiva eller falskt negativa. Test som ger höga andelar falskt positiva eller falskt negativa resultat får praktiska konsekvenser. Personer som är friska utsätts för onödiga undersökningar och onödig oro. Å andra sidan finns det patienter som är sjuka men inte får den vård de behöver.

Nedan diskuterar vi hur den diagnostiska processen påverkas av sensitivitet och specificitet för några av de vanligare formulären. I våra beräkningar har vi antagit att prevalensen av depression i primärvården är 20 procent och inom den psykiatriska vården 40 procent. För bipolära syndrom har vi använt prevalensen 30 procent för psykiatrisk vård [1,2].

En nyckelfråga är hur hög känslighet och träffsäkerhet som behövs för att formulären ska fungera som ett stöd. Litteraturen är inte entydig. Vi enades om att det krävs minst 80 procent sensitivitet och specificitet för att ett intervjuformulär ska ha god diagnostisk tillförlitlighet (Kapitel 3.1). För riktad screening är det viktigast med en hög sensitivitet för att öka chansen för att deprimerade personer identifieras. Sensitiviteten bör enligt våra bedömningar vara minst 80 procent, men i litteraturen rekommenderas även högre nivåer på sensitivitet (se t ex [3]).



## Diagnos av depression i primärvården

Som framgår av Kapitel 3.3 kom en systematisk översikt med mer än 50 000 patienter fram till att diagnos av läkare i primärvården hade en sensitivitet på 47 procent och en specificitet på 81 procent (Tabell 5.1). Med en prevalens på 20 procent innebär det att läkaren identifierar 10 av 20 som har depression. Läkaren identifierar dessutom 80 procent av de icke-deprimerade korrekt, dvs 65 individer. Tio av tjugo med depression upptäcks inte, samtidigt som en av fem friska får diagnosen depression.

Med hjälp av ett intervjuformulär som MINI ökar andelen deprimerade som identifieras så att nitton av tjugo upptäcks (Tabell 9.2). Andelen friska som får diagnosen depression påverkas inte.

**Tabell 5.1** Klinisk bedömning som enda underlag för diagnos av depression i primärvården jämfört med strukturerade eller semistrukturerade intervjuer. Sensitivitet 47 procent och specificitet 81 procent.

	Sjuk	Frisk	Totalt	Prediktivt värde
Bedömning positiv	100	150	250	PPV: 40%
Bedömning negativ	100	650	750	NPV: 87%
	200	800	1 000	

NPV = Negativt prediktionsvärde; PPV = Positivt prediktionsvärde

**Tabell 5.2** Sammanställning för MINI. Sensitivitet 95 procent och specificitet 84 procent.

	Sjuk	Frisk	Totalt	Prediktivt värde
MINI positiv	190	128	318	PPV: 60%
MINI negativ	10	672	682	NPV: 98%
	200	800	1 000	

NPV = Negativt prediktionsvärde; PPV = Positivt prediktionsvärde

## Diagnostik av depression med PHQ-9

Många vårdprogram rekommenderar att PHQ-9 används vid misstanke om depression. I de studier av formuläret som vi granskade varierade sensitiviteten mycket, från 34 till 84 procent (Kapitel 3.5).

Tabell 5.3 och 5.4 visar andelarna patienter där PHQ-9 ger sant respektive falskt resultat vid dels 34 och dels 84 procent. Vi använde den sammanvägda specificiteten 93 procent för båda värdena på sensitivitet (se Kapitel 3.5). Detta innebär en förenkling eftersom en högre sensitivitet oftast innebär att specificiteten samtidigt sjunker.

**Tabell 5.3** Sammanställning för PHQ-9 med sensitivitet 34 procent och specificitet 93 procent.

	Sjuk	Frisk	Totalt	Prediktivt värde i primärvård
PHQ-9 positiv	7	6	13	PPV: 54%
PHQ-9 negativ	13	74	87	NPV: 85%
	20	80	100	

NPV = Negativt prediktionsvärde; PPV = Positivt prediktionsvärde

**Tabell 5.4** Sammanställning för PHQ-9 med sensitivitet 84 procent och specificitet 93 procent.

	Sjuk	Frisk	Totalt	Prediktivt värde
PHQ-9 positiv	17	6	23	PPV: 74%
PHQ-9 negativ	3	74	77	NPV: 96%
	20	80	100	

NPV = Negativt prediktionsvärde; PPV = Positivt prediktionsvärde

Som framgår av tabellerna kan andelen deprimerade som upptäcks variera mellan en tredjedel och fem sjättedelar. Skattningen blir därmed mycket osäker.

PHQ-9 fungerar bättre för att bekräfta att en person inte är deprimerad. Vid både 34 och 84 procent sensitivitet blir sex av 80 personer falskt positiva, dvs knappt 8 procent. Detta avspeglas i ett högt NPV; 96 procent för primärvård.

PHQ-9 förefaller följaktligen inte vara användbart för riktad screening. Det finns en uppenbar risk för att deprimerade patienter inte upptäcks med testet.

### Riktad screening av bipolära syndrom

MDQ och HCL-32 är två formulär med olika psykometrisk profil. Som framgick i Kapitel 3.5 har HCL-32 en acceptabel sensitivitet på 82 procent medan specificiteten var låg och osäker (mellan 36 och 71 % i tre studier). I vårt räkneexempel har vi använt ett sammanvägt resultat på 48 procent för specificiteten (Tabell 5.5).

MDQ med modifierat kriterium däremot har en sensitivitet på 78 procent och en specificitet på 79 procent (Tabell 5.6).

**Tabell 5.5** Sammanställning för HCL-32 med sensitivitet 82 procent och specificitet 48 procent.

	Sjuk	Frisk	Totalt	Prediktivt värde, psykiatri
HCL-32 positiv	246	364	610	PPV: 40%
HCL-32 negativ	54	336	390	NPV: 86%
	300	700	1 000	

NPV = Negativt prediktionsvärde; PPV = Positivt prediktionsvärde

**Tabell 5.6** Sammanställning för MDQ med modifierade kriterier, med sensitivitet 78 procent och specificitet 79 procent.

	<b>Sjuk</b>	<b>Frisk</b>	<b>Totalt</b>	<b>Prediktivt värde, psykiatrin</b>
MDQ positiv	234	147	381	PPV: 61%
MDQ negativ	66	553	619	NPV: 89%
	300	700	1 000	

NPV = Negativt prediktionsvärde; PPV = Positivt prediktionsvärde

Tabellerna 5.5 och 5.6 visar att vid en prevalens på 30 procent kan bägge formulären relativt väl utesluta bipolärt syndrom. Betydligt fler patienter behöver utredas på misstanke om bipolärt syndrom om man väljer att använda HCL-32 för riktad screening än om man använder MDQ och en stor andel av dessa kommer visa sig vara falskt positiva.

Vid en prevalens på 10 procent, vilket kan vara fallet i primärvård vid riktad screening av patienter som söker för depression, förstärks andelen falskt positiva. Vid positivt testresultat är sannolikheten för bipolärt syndrom endast 15 procent för HCL-32 och 29 procent för MDQ. Värdet av att använda formulären rutinmässigt i primärvård kan därför ifrågasättas pga merarbete, även om ett negativt resultat ger en mycket hög sannolikhet (96–97 %) för att patienten inte har bipolärt syndrom och att ett positivt resultat medför att 8 av 10 sant positiva patienter identifieras. Om det finns möjlighet till en psykiatrisk konsultation kan ett positivt testresultat medföra att patienten får ”rätt” diagnos och behandling.

## Referenser

1. Hantouche EG, Akiskal HS, Lancrenon S, Allilaire JF, Sechter D, Azorin JM, et al. Systematic clinical methodology for validating bipolar-II disorder: data in mid-stream from a French national multi-site study (EPIDEP). *J Affect Disord* 1998; 50:163-73.
2. 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.
3. Gilbody S, Sheldon T, Wessely S. Should we screen for depression? *BMJ* 2006;332:1027-30.

## 6. Etiska och sociala aspekter

---

Detta avsnitt problematiserar kring möjliga etiska eller sociala problem som kan vara förknippade med användning av olika typer av diagnostiska bedömningsformulär för förstämningssyndrom. De frågor som analyseras i kapitlet togs fram vid två möten; ett i projektgruppen och ett inom SBU:s lekmanråd.

Vanligtvis utgår man från de fyra etiska principer som är vägledande inom hälso- och sjukvården: att göra gott, att inte skada, att respektera patientens autonomi och att vara rättvis [1]. För varje princip kan man också bedöma om metoden ger en etisk vinst eller en etisk förlust enligt metodik av Nilstun [2]. Vi har i föregående mån försökt betrakta underlaget ur dessa synvinklar.

### Frågor

- Vilka skulle de etiska konsekvenserna bli om resultaten av den systematiska granskningen tillämpades?
- Vilka är de etiska aspekterna på om bedömningsformulär används på ett sätt som inte är *lege artis*?
- Vilka är de etiska riskerna med riktad screening av förstämningssyndrom?
- Vilka är de etiska problemen med att formulären inte kan användas för alla patienter?
- Finns det risk för ojämlig vård eftersom utbildningen om bedömningsformulär varierar?

## Urval av litteratur

Vi gjorde en systematisk genomgång av publicerad litteratur i databaserna Medline, Embase, PsycInfo och Cinahl. Eftersom vi inte avgränsade studieupplägget inkluderade vi även tvärsnittsstudier och översiktsartiklar.

Sökningarna genererade 1 053 abstrakt varav 24 artiklar beställdes i fulltext. Femton artiklar bedömdes vara relevanta. Ytterligare 20 artiklar tillkom via sökning i referenslistor. Studierna kvalitetsgranskades inte. Bedömningar av den etiska/sociala relevansen av att använda bedömningsformulär grundades på resultaten i Kapitel 3.

## Vilka är de etiska konsekvenserna av resultaten av granskningen?

### Göra gott och inte skada

För att det ska vara etiskt försvarbart att använda en metod måste den vara till nytta för patienten. Vår granskning visar att klinisk bedömning enbart inte är tillräcklig för att ställa diagnoser på förstämningssyndrom. Det är därmed en etisk vinst att komplettera samtalet med ett eller flera bedömningsformulär.

Ett dilemma är att det finns bedömningsformulär där vi inte kan dra några slutsatser om huruvida de ökar den diagnostiska känsligheten och träffsäkerheten. För formulär som används för riktad screening av depression är det viktigt med en hög känslighet för att minska risken för att deprimerade patienter inte upptäcks. Några av de formulär som vi granskade hade så osäkra data på känslighet att det kan innebära en etisk förlust för patienten om de används.

Det finns få studier på barn och äldre. Sannolikt förbättras chansen att korrekt identifiera depression även hos barn och äldre med hjälp av semi-strukturerade och strukturerade intervjuer. Det är därför en etisk förlust att inte använda sådana bedömningsformulär.

## **Vilka är de etiska aspekterna på att använda bedömningsformulär på ett sätt som inte är *lege artis*?**

Det finns flera etiska risker med användning av bedömningsformulär som inte följer *lege artis*. Hit hör över- respektive underanvändning, felaktig användning samt användning av formulär med låg diagnostisk tillförlitlighet, vilket strider mot godhets- och rättvisepinciperna.

Ett exempel på överanvändning är om formulären skulle användas för att ersätta samtalet vid den kliniska bedömningen istället för att komplettera det. Som framgår av Kapitel 7 förefaller det som om bedömningsformulär väl täcker in symtomen på förstämningssyndrom, men för en liten del av patienterna kan viktig information komma att utelämnas. Det innebär då en risk för fel diagnos.

Ett likartat exempel är en överdriven tilltro till poäng på ett bedömningsformulär. Man måste vara medveten om begränsningarna i de formulär som man använder. Om ett formulär för riktad screening identifierar 70 procent av dem som har förstämningssyndrom går fortfarande en tredjedel oupptäckta. Om formuläret utesluter 70 procent av dem som är friska ses en tredjedel som sjuka och får en onödig behandling.

Användning av formulär på annat sätt än vad de är utvecklade för utgör också en risk. Våra enkäter tydde på att de svarande ibland inte var säkra på hur olika formulär är tänkta att användas. MADRS t ex är avsett för bedömning av svårighetsgrad, men användes både som diagnostiskt formulär och för screening. Deltagarna i fokusgrupperna påpekade att aktörer utanför primärvården, som t ex psykiatrin och försäkringskassorna, ofta förväntade sig poäng från ett specifikt formulär, som t ex MADRS. Uppfattningen var att poängen i vissa fall tillmätts större vikt än den kliniska bedömningen. Eftersom samtliga formulär är förknippade med en andel såväl falskt positiva som falskt negativa leder detta till en risk för felaktig bedömning med långtgående konsekvenser.



## Etiska risker med riktad screening

Riktad screening i primärvården ger positiva vinster – att göra gott – om användningen kombineras med ett förbättrat omhändertagande av patienterna, tillgång på utbildad personal och klara remissvägar för fortsatt utredning och behandling [4–6]. Om dessa organisatoriska förhållanden inte är för handen är användning av screening, även riktad sådan, inte bara ett brott mot principen att göra gott utan även mot rättvisepincipen.

Etiska aspekter är speciellt viktiga när det gäller riktad screening av depression hos nyförlösta eftersom moderns hälsa är viktig för såväl mor som barn. Depression hos mamman kan påverka anknytningen till barnet. Det kan i sin tur öka risken för att barnet får störningar i beteende, emotioner och kognition under uppväxten [7,8].

Svenska forskare har analyserat och kritiserat EPDS ur etisk synpunkt. De argumenterar mot rutinanvändning av EPDS för riktad screening av kvinnor efter förlossningen beroende på risk för både falskt positiva och falskt negativa diagnoser [12]. Men EPDS ger ändå sannolikt en etisk vinst i och med att fler kvinnor upptäcks än om bedömningsformuläret inte hade använts. Bågedahl och medarbetare gjorde en journalstudie på barnavårdscentraler som visade att 2 procent av mammorna identifierades som deprimerade under de tre första månaderna efter förlossningen. Med inledande screening med EPDS ökade andelen till 8,4 procent [13].

En viktig aspekt som inte har diskuterats i Sverige i samband med screening av depression efter förlossning är legala och etiska konsekvenser av att barnet är barnhälsovårdens patient men inte kvinnan. Detta framfördes i en artikel av Chaudron och medarbetare som påpekade att information om kvinnan och eventuella resultat av screening inte kan journalföras i barnets journal med hänsyn till den sekretess som normalt används vid psykiska sjukdomar [14].

## **Vilka är de etiska aspekterna på att alla patienter inte kan bedömas med formulär?**

Det finns personer som kan ha svårt att förstå och fylla i bedömningsformulär. Hit hör de som inte helt behärskar språket, som har läs- eller skrivsvårigheter eller nedsatt kognitiv funktion. En annan kategori är patienter som av kulturella skäl uttrycker symtom på förstämningssyndrom på andra sätt än dem som definieras i kriterierna. Ur etisk synvinkel leder det både till risk att skada personer och risk för ojämlik vård.

Användning av formulär som inte uppfattas som relevanta av patienten, t ex på grund av språksvårigheter eller skillnader i syn på psykiska symtom, strider mot autonomiprincipen. Denna säger man bör visa respekt för en patients rätt att ha ett reellt inflytande på handlingar och beslut som berör patienten själv, samt för patientens värderingar, önsknings och åsikter.

### **Konsekvenser av språk- och lässvårigheter**

Bristande kunskaper i svenska språket försvårar diagnostiken. Vid samtalet blir man beroende av tolk och det är svårt att avgöra om tolken förstått frågorna. Användning av bedömningsformulär för patienter som inte helt behärskar språket innebär också risker för missuppfattningar och feltolkningar. Det är dock svårt att avgöra hur stor risken är, vilket visades i en amerikansk studie av Sentell och medarbetare [15]. De undersökte betydelsen av läskunnighet för att förstå specifika svarsalternativ. Individer med olika bakgrund fyllde i BDI. Skillnaden i andelen begrepp som definierats korrekt varierade mellan 0 och 60 procent. Förståelsen av medicinska begrepp var relativt låg, men kunde inte relateras till deltagarnas utbildningsnivå eller läskunnighet.

Ett exempel på risken för diskriminering och ojämlik behandling gäller riktad screening med EPDS. En studie av Magnusson och medarbetare visade att endast 32 procent av svenska barnvårdscentraler screenade alla mödrar oavsett språktillhörighet [16]. De övriga screenade enbart svensktalande kvinnor trots att EPDS finns översatt och validerat till många språk. Detta är ett påtagligt problem ur etisk/social synpunkt. Formuläret EPDS är validerat i många kulturer [9] och i Sverige för

svensktalande kvinnor [10]. Det används vid de flesta barnvårdscentraler i Sverige och rekommenderas av Socialstyrelsen [11].

### **Konsekvenser av kulturella skillnader**

Diagnostik av depression på patienter från kulturer utanför Västeuropa/Nordamerika försvåras av att symtom definieras och beskrivs på andra sätt (se t ex Lesser [17]). Detta påverkar såväl samtalet som användning av formulär som bygger på DSM-kriterier och är ett etiskt och socialt problem i och med att det finns risk för felklassificering.

Skilda kulturella uppfattningar kan vara en orsak till de resultat som erhöles i den så kallade ODIN-studien [18]. Nuevo och medarbetare undersökte validiteten av BDI i fem olika länder – Spanien, Storbritannien, Irland, Finland och Norge – med hjälp av Item Response Theory. För några av frågorna i BDI skilde sig svaren mellan de olika länderna. Orsaken kan bero på felaktig översättning av formuläret men även på skillnader i upplevelse och syn på symtom mellan de olika länderna. Nuevo och medarbetare undersökte dock inte om detta ledde till felklassificeringar.

### **Finns det etiska risker med att utbildningen om bedömningsformulär varierar?**

Bristande eller varierande utbildning om bedömningsformulär och strukturerad diagnostik kan leda till ojämlig vård eftersom diagnostiken inte kan kvalitetssäkras.

De flesta vårdprogram på landstings- eller regionnivå rekommenderar att bedömningsformulär används (Kapitel 7). Samtidigt framkom det i våra praxisundersökningar att en stor andel av läkarna i primärvård och psykiatri inte fått någon utbildning via arbetsgivaren (Kapitel 7). Ur etisk synvinkel är detta betänkligt i och med att huvudmännen är skyldiga att se till att personalen har tillräcklig kompetens [19].

Ett konkret exempel på etiskt problem är att praxis för hur EPDS används varierar i landet. Två studier av Magnusson och medarbetare respektive Massoudi och medarbetare tyder på att personalen inom barnhälsovården inte alltid är utbildad i att informera, tolka och hantera konsekvenser av att använda formuläret [16,20]. Bristande utbildning innebär en risk för att resultaten tolkas på ett felaktigt sätt och strider därmed mot principer för att inte skada någon och för jämlik vård.

## Referenser

1. Beauchamp TL, Childress TE. Principles of Biomedical Ethics, Sixth ed. Oxford University press, New York; 2009.
2. Nilstun T. Medicin och moral. Studentlitteratur. Lund 1991:113-27.
3. Dowrick C, Leydon GM, McBride A, al. E. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ* 2009;338:b663 doi: 10.1136/bmj.b663.
4. O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Ann Intern Med* 2009;151:793-803.
5. Gilbody S, Whitty P, Grimshaw J, Thomas R. Educational and organizational interventions to improve the management of depression in primary care: a systematic review. *JAMA* 2003;289:3145-51.
6. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009;374:609-19.
7. Murray L. The impact of postnatal depression on infant development. *J Child Psychol Psychiatry* 1992;33:543-61.
8. Martins C, Gaffan EA. Effects of early maternal depression on patterns of infant-mother attachment: a meta-analytic investigation. *J Child Psychol Psychiatry* 2000;41:737-46.
9. 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.
10. Wickberg B, Hwang CP. The Edinburgh Postnatal Depression Scale: validation on a Swedish community sample. *Acta Psychiatr Scand* 1996;94:181-4.
11. Socialstyrelsen. Nationella riktlinjer för depression/kvinnor efter förlossning. <http://www.socialstyrelsen.se/nationellariktlinjerfordepressionochangest/sokiriktlinjerna/kvinnorefterforlossning> 2010.
12. Krantz I, Eriksson B, Lundquist-Persson C, Ahlberg BM, Nilstun T. Screening for postpartum depression with the Edinburgh Postnatal Depression Scale (EPDS): an ethical analysis. *Scand J Public Health* 2008;36:211-6.
13. Bagedahl-Strindlund M, Monsen Borjesson K. Postnatal depression: a hidden illness. *Acta Psychiatr Scand* 1998;98:272-5.
14. Chaudron LH, Szilagyi PG, Campbell AT, Mounts KO, McNerny TK. Legal and ethical considerations: risks and benefits of postpartum depression screening at well-child visits. *Pediatrics* 2007;119:123-8.
15. Sentell TL, Ratcliff-Baird B. Literacy and comprehension of Beck Depression Inventory response alternatives. *Community Ment Health J* 2003;39:323-31.
16. Magnusson M, Lindfors A, Tell J. [Big differences in Swedish child health care. Child health care units decide their services – alarming that a national pro-

gram is missing]. *Lakartidningen* 2011; 108:1618-21.

17. Lesser IM. Cultural considerations using the Structured Clinical Interview for DSM-III for mood and anxiety disorders assessment. *J Psychopathol Behav Assess* 1997;19:149-60.

18. Nuevo R, Dunn G, Dowrick C, Vazquez-Barquero JL, Casey P, Dalgard OS, et al. Cross-cultural equivalence of

the Beck Depression Inventory: a five-country analysis from the ODIN study. *J Affect Disord* 2009;114:156-62.

19. Socialstyrelsen. Antidepressiva läkemedel vid psykisk ohälsa. *Studier av praxis i primärvården*. 2006:1-97.

20. Massoudi P, Wickberg B, Hwang P. Screening for postnatal depression in Swedish child health care. *Acta Paediatr* 2007;96:897-901.



## 7. Användning av bedömningsformulär i Sverige

---

Bedömningsformulär för förstämningssyndrom började utvecklas för mer än 50 år sedan, i huvudsak drivet av behovet att mäta effekt av läkemedel i kliniska prövningar eller för att följa förlopp av depression under psykoterapi. Användningen av formulären inom såväl forskning som klinik har varit kantad av tvivel på nyttan och tillförlitligheten samtidigt som nya formulär tillkommit kontinuerligt [1]. Under senare år har bedömningsformulär fått ökad uppmärksamhet, mycket tack vare att de ingår i rekommendationer i riktlinjer som t ex från amerikanska ACHPR [2] och brittiska NICE [3]. I Sverige har Socialstyrelsen inte tagit ställning till hur bedömningsformulär ska användas i de nationella riktlinjerna för depression och ångestsyndrom utan avvaktar vår granskning. Socialstyrelsen rekommenderar dock en väl strukturerad diagnostisk process [4].

Det är osäkert i vilken utsträckning bedömningsformulären används kliniskt i Sverige idag. Vi genomförde därför en kartläggning som syftade till att belysa:

- hur ofta bedömningsformulär används vid förstämningssyndrom
- hur ofta regionala och lokala vårdprogram rekommenderar bedömningsformulär
- hur läkare respektive patienter ser på att använda bedömningsformulär.



## Vetenskapligt underlag

Vårt underlag bestod av publicerade studier, enkäter och en sammanställning av en serie fokusgrupper med läkare från primärvården.

Vi gjorde ingen specifik systematisk sökning på studier om uppfattningar om bedömningsformulär, men ett antal studier föll ut i våra breda sökningar efter studier om diagnostisk tillförlitlighet. Dessa kompletterades med sökning i referenslistorna. Vi granskade inte kvaliteten på de studier som vi bedömde vara relevanta.

Vi använde resultat från en utredning och tre olika enkäter. Socialstyrelsen utredde år 2006 användningen av antidepressiva läkemedel i ett regeringsuppdrag [5]. Utredningen bestod av en granskning av drygt hundra patientjournaler samt en audit via webben (Kapitel 10). Auditen rörde handläggning av patienter med misstänkt depression och gjordes i samarbete med läkemedelskommittéerna och SBU. Drygt 200 läkare deltog och registrerade data för cirka 1 800 patienter. Det fanns en övervikt för läkare som arbetade i Västra Götaland, Gävleborg, Västernorrland och Skåne.

En av enkäterna genomfördes av Utvecklingscentrum för barns psykiska hälsa (UPP-centrum) vid Socialstyrelsen. De riktade en enkät till landets cirka 200 enhetschefer vid 132 BUP-enheter och drygt 85 procent besvarade den [6]. Den andra enkäten, av Magnusson och medarbetare, gick till samtliga 36 barnvårdscentralsenheterna [7]. Samtliga besvarade den.

SBU gjorde en egen enkätundersökning under 2011 med ett slumpmässigt urval av 150 vuxenpsykiatrer och 300 läkare i primärvården. Urvalet var stratifierat för att säkerställa en balans mellan sjukvårdsregionerna. Enkäterna bestod av flervalsfrågor och två öppna frågor och skickades ut per post. Fyra veckor senare skickade vi ut en påminnelse till dem.

Svarsfrekvensen var låg: 56 psykiatrer (36 %) och 123 distriktsläkare (41 %) besvarade enkäten. Majoriteten av dem var över 40 år, och de träffade mellan sex och tjugo patienter per månad med förstämningssyndrom. Nästan alla psykiatrer som deltog hade mer än fem års erfarenhet och över hälften hade mer än 20 års erfarenhet.

Läkarna hade deltagit i utbildningar om bedömningsformulär i varierande omfattning. Ungefär 15 procent av läkarna i primärvården hade utbildats genom arbetsgivaren medan hälften inte fått någon utbildning alls. Av psykiatrerna hade hälften deltagit i läkemedelsföretagens utbildningar och en fjärdedel hade fått utbildning av arbetsgivaren.

SBU fick även tillgång till resultat från fem fokusgrupper som genomfördes inom ett doktorandprojekt på Karolinska Institutet under 2011. Syftet med fokusgrupperna var att kartlägga uppfattningar hos läkare i primärvården. Fokusgrupperna omfattade ett bekvämlighetsurval av 26 läkare från Västra Götaland. Samtliga deltagare var fast anställda på sin vårdcentral. Med få undantag hade de arbetat som specialister i allmänmedicin mellan 5 och 20 år.

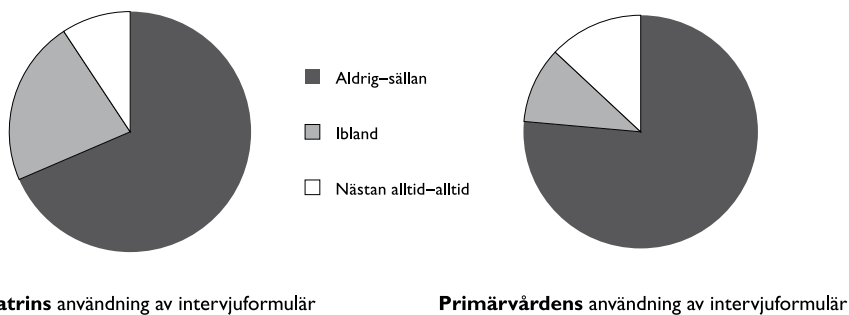
## **Hur vanligt är det att använda bedömningsformulär vid förstämningssyndrom?**

### **Strukturerade och semistrukturerade intervjuer**

Slutsatsen av vår genomgång är att strukturerade eller semistrukturerade intervjuer används regelbundet för en mindre del av patienterna. Socialstyrelsens journalgranskning från utredningen om antidepressiva läkemedel rapporterade t ex att indikationen för behandling av vuxna bara gick att utläsa i 23 av 113 granskade journaler från primärvården. Förskrivningen baserades oftast på en klinisk bedömning utan hänvisning till kriterier [5].

UPP-centrums enkät visade att nära 40 procent av BUP-enheterna använde strukturerade eller semistrukturerade intervjuer någon gång men att endast 12 procent använde dem för mer än tre fjärdedelar av barnen ("ofta"). Den vanligaste intervjun var K-SADS (19 %) följt av DAWBA.

I SBU:s enkät använde drygt var femte läkare från såväl primärvården som vuxenpsykiatri strukturerade eller semistrukturerade intervjuer regelbundet (Figur 7.1). Det vanligaste formuläret i primärvården var PRIME-MD och i psykiatri MINI.



**Figur 7.1** Användning av strukturerade och semistrukturerade intervjuer i vuxenpsykiatri respektive primärvården. (Källa: SBU-enkäter 2011, primärvården 123 svar samt till psykiatri, 56 svar.)

## Riktad screening

Enligt UPP-centrums enkät användes formulär för riktad screening av depression av 78 procent och för mani av cirka 40 procent av de barnpsykiatriska enheterna. De vanligaste alternativen för depression var CDI (46 %) och CBCL (30 %) och för bipolära syndrom CMRS-P (33 %).

I Socialstyrelsens webbaudit från 2006 använde 17 procent av de deltagande allmänläkarna MADRS regelbundet, men det framgick inte hur formuläret användes. Användningen varierade mellan landsting, från 0 till 55 procent av deltagarna.

I vår enkät var användningen större. Var tredje, från primärvården som besvarade enkäten använde "alltid" eller "nästan alltid" formulär för riktad screening. MADRS som expert- eller självskattningsskala valdes oftast (54 personer) följt av HADS (38 personer). En del utnyttjade båda formulären.

Uppgifterna från vuxenpsykiatri är osäkra beroende på att få besvarade enkäten. De som deltog använde i högre utsträckning riktad screening för bipolära syndrom, vanligtvis MDQ, än riktad screening för depression.

Självbedömning av nyförlösta mödrar med EPDS på barnvårdscentrallerna är väletablerad över landet. Magnusson och medarbetares rapport visade att vid 86 procent av enheterna har mer än hälften av sjuksköterskorna utbildats för att instruera och tolka EPDS [7]. Kvinnor med symtom på depression erbjuds oftast samtal med sjuksköterska på barnvårdscentralen eller, i vissa fall, psykolog inom barnhälsovården.

Tidpunkten för att använda EPDS varierar. I t ex Östergötland sker screeningen tre månader efter förlossningen, men i centrala Östergötland används formuläret även i samband med efterkontroll på MVC efter sex veckor. Screening under graviditeten ökar och görs på MVC.

Vi identifierade inte några undersökningar om användning av bedömningsformulär hos äldre. Den tidigare nämnda rapporten från Socialstyrelsen noterade att bedömningsformulär användes i mycket lägre utsträckning för äldre än för vuxna och unga vuxna (7 % jämfört med 25 respektive 22 %) [5].

### **Uppföljning av behandling av depression och mani**

Mer än en tredjedel av de allmänläkare och psykiatrer som besvarade vår enkät följde alltid eller nästan alltid upp insatt behandling mot depression med hjälp av ett formulär, vanligen MADRS eller MADRS-S (70 personer). Ett fåtal nämnde BDI-II och PHQ-9. Det vanligaste bedömningsformuläret för att följa upp behandling mot bipolära syndrom var YMRS och MDQ.

### **Rekommenderas bedömningsformulär för förstämningssyndrom i regionala och lokala vårdprogram?**

Vi har försökt att identifiera lokala och regionala vårdprogram om förstämningssyndrom, dels genom sökning på nätet dels via kontakter med företrädare för psykiatrin från landsting och sjukvårdsregioner.

Vi hittade vårdprogram från Stockholms läns landsting, från olika sjukhus i Västra Götaland, från Skåne, Östergötland, Jönköping, Örebro, Södermanland, Uppsala och Västerbotten. Uppsala, Örebro och Öster-

götland hade vårdprogram för barn- och ungdomspsykiatri. Uppsala hade även ett vårdprogram för äldre med förstämningssyndrom.

Samtliga betonade att diagnostiken ska vara kriteriebaserad, vanligen enligt DSM. Intervjuerna SCID-I och MINI rekommenderades i två vårdprogram för vuxna och K-SADS i två vårdprogram för BUP. Några rekommenderade istället två inledande screeningfrågor för depression (nedstämdhet och intresse) som vägde tungt i den fortsatta diagnostiken. En ytterligare variant var att bygga på formuläret PHQ-9 så att patientens svar på de två första frågorna avgjorde om det var värt att gå vidare med de övriga sju frågorna för diagnos. I ett vårdprogram screenades samtliga nyanmälningar till BUP med BCFPI.

Samtliga vårdprogram rekommenderade användning av ett eller flera formulär för att bedöma svårighetsgrad och för att följa upp insatt behandling. Förutom vårdprogrammet som byggde på PHQ-9 rekommenderade övriga MADRS som expert- eller självskattningsskala för depression. Alternativ till MADRS var någon eller några av skalorna PHQ-9, HDRS, DSRS, HADS och BDI för vuxna och BDI-C, CDI, BUS, DSRS och CBCL för barn och ungdomar.

## **Hur ser läkare och annan vårdpersonal på att använda bedömningsformulär?**

Såväl publicerade studier som enkäterna tyder på att läkare har olika uppfattningar om nyttan av bedömningsformulär.

### **Data från publicerade studier**

Publicerade artiklar redovisar både positiva och negativa uppfattningar om att använda bedömningsformulär. Vi fann nio relevanta publikationer [8–16]. Studier som byggde på intervjuer och fokusgrupper med individer som inte använt formulär exkluderades. Vi exkluderade också studier som jämförde webbaserade formulär med motsvarande i pappersformat.

Två artiklar undersökte hur stor del av psykiatrer och psykologer som använder bedömningsformulär regelbundet med hjälp av enkäter [8,14].

Bruchmuller och medarbetare fann att av nära 2 000 psykiatrer och psykoterapeuter i Schweiz använde bara 15 procent strukturerade intervjuer till sina patienter [14]. Gilbody och medarbetare skickade en enkät per post till 500 slumpvis utvalda psykiatrer i Storbritannien med en svarsfrekvens på 74 procent [8]. Över hälften av psykiatrerna använde aldrig bedömningsformulär. Cirka 10 procent använde regelbundet formulär för riktad screening och uppföljning, vanligen BDI, HADS och HDRS. Drygt 10 procent av läkarna rapporterade att deras arbetsgivare ställde krav på standardiserade bedömningar av psykisk ohälsa. Cirka en tredjedel av dem som besvarade enkäten lämnade frivilliga kommentarer, samtliga negativa. Vanliga invändningar var att formulären inte kan upptäcka subtila förändringar, att tiden inte räckte till och att det krävdes en annan infrastruktur för att hantera dem.

Dowrick och medarbetare genomförde fokusgrupper med brittiska allmänläkare och patienter [13]. Utgångspunkten var att NICE (National Institute for Clinical Excellence) numera kräver att självskattningsskalor används för att läkarna ska få ekonomisk ersättning för handläggning av en patient med depression. Läkarna var skeptiska till skalorna för egen del men ansåg att de kunde vara av värde "för andra". De var osäkra på tillförlitligheten, och om en patientskattning inte överensstämde med läkarens egen uppfattning så vägde skattningen lätt. Läkarna var också osäkra på hur stor risken är för att patienterna manipulerar och inte besvarar frågorna sanningsenligt.

Fem studier testade hur acceptabla olika formulär är för kliniker och fyra rapporterade positiva resultat. Suppiger och medarbetare undersökte en semistrukturerad intervju (ADIS översatt till tyska) [11]. Tio certifierade kliniska psykologer intervjuade knappt 200 personer inom öppen- och slutenvård psykiatrin. Psykologerna var tillfredsställda med intervjuformuläret. De fick fram den relevanta informationen i mer än 90 procent av intervjuerna och uppnådde en positiv relation till patienten i nästan samtliga fall. För 15 procent av patienterna misstänkte psykologerna att patienten inte hade berättat om alla sina problem.

En norsk studie utvärderade MINI som utfördes av annan personal än läkare på en akutpsykiatrisk avdelning [15]. Av 268 konsekutiva patien-

ter gick det att genomföra en intervju med 66 procent. Patienternas och intervjuarens uppfattningar om intervjun skattades med en VAS-skala. Personalen ansåg att intervjun var användbar och trovärdig förutom för patienter som vårdades pga psykos.

Två studier utvärderade PHQ-9. Duffy och medarbetare kom fram till att PHQ-9 var till stor hjälp för diagnos, bedömning av svårighetsgrad och uppföljning inom psykiatri [10]. Studien genomfördes som ett så kallat genombrottsprojekt (systematiskt lokalt förbättringsarbete) med sexton psykiatriska kliniker. När studien startade använde sex kliniker bedömningsformulär för riktad screening och fyra för uppföljning av insatt behandling. Efter tolv månader hade samtliga sexton kliniker infört PHQ-9. Bermejo och medarbetare å andra sidan fann att det var för tidsödande att använda PHQ-9 som väntrumsscreening och att det var svårt att motivera patienterna [12].

Mason och medarbetare intervjuade barnmorskor och sjuksköterskor från en så kallad Primary Care Trust [16]. Samtliga uttryckte belåtenhet med EPDS som fungerade som ett diskussionsunderlag vid samtal med mödrarna. Alla var dock medvetna om att formuläret hade sina begränsningar och endast var ett hjälpmedel.

### **Uppfattningar från svenska fokusgrupper och enkäter**

Många uppfattningar som framkom i enkäter och fokusgrupper överensstämde med dem som redovisats i litteraturen. De flesta av läkarna använde formulären för vissa patienter, t ex sådana som var tystlåtna eller som hade svårt att uttrycka känslor. Generellt sett såg de inte någon större nytta med formulären utan den egna erfarenheten var säkrare, ”Jag är själv det bästa verktyget”. Några påpekade att formulären hade ett pedagogiskt värde.

Flera underströk att formulären är utvecklade för forskning och mätning på gruppnivå och de var osäkra på hur tillförlitliga de var för enskilda patienter. Förutom eventuella begränsningar i själva formulären såg deltagarna problem med patienter som hade otillräckliga kunskaper i svenska. Primärvården skulle behöva ha tillgång till framför allt själv-

skattningsformulär på andra språk. En annan risk var att patienter kunde ha svårt att tolka frågorna i ett självskattningsformulär.

Enkäten visade ett bredare spektrum av uppfattningar än fokusgrupperna. De oftast nämnda fördelarna var att det ledde till en mer objektiv bedömning, att det förenklade uppföljningen och att det förenklade kommunikationen med patienten. Den nackdel som togs upp av flest personer var tidsbrist/att det tog för lång tid.

En central frågeställning för fokusgrupperna var hur konsultationen påverkas av att bedömningsformulär ingår i diagnostik och uppföljning. De negativa synpunkterna övervägde. Många utgick från att formuläret skulle ersätta samtalet. Enstaka läkare noterade att patienten kunde fylla i formuläret vid annat tillfälle.

I enkäterna däremot var det få som uppfattade att konsultationen enbart påverkades negativt (Tabell 7.1). De flesta såg både för- och nackdelar med bedömningsformulär.

**Tabell 7.1** Påverkas konsultationen av att använda bedömningsformulär?  
Svar från SBU:s enkäter till primärvården (123 svar) och psykiatrin (56 svar).

	Ja, positivt	Ja, negativt	Nej	Både ja och nej	Ingen åsikt
Allmänläkare	18	8	29	33	23
Psykiatrer	17	3	6	23	4

Deltagarna i fokusgrupperna underströk att det ofta var andra, utanför primärvården, som förväntade sig en bedömning med ett formulär för preliminär bedömning av risken för depression. Hit hörde såväl psykiatrin som Försäkringskassan. Den vanligaste kommentaren i fokusgrupperna var att deltagarna såg det som oroande att poäng på MADRS i vissa fall kunde tillmätas större värde än den egna kliniska bedömningen.



Ingen av deltagarna i fokusgrupperna hade fått utbildning om användning av bedömningsformulär via arbetsgivaren. Det sågs som en stor nackdel att det huvudsakligen är industrin som utbildar och marknadsför formulären. Formulären förknippades därmed med en viss risk för övermedikalisering.

### **Hur ser patienterna på att bedömningsformulär används?**

Få studier har tagit patientens perspektiv i beaktande trots att patienten är en viktig intressent. Vi identifierade fem artiklar. I samtliga var patienterna positivt inställda till bedömningsformulär och i två av dem noterade författarna att patienterna var mer positivt inställda än läkarna/psykologerna.

Dowrick och medarbetare genomförde även fokusgrupper med patienter i den tidigare beskrivna studien [13]. Patienterna visade sig vara mer positivt inställda till användning av skattningsskalor än läkarna.

Två studier undersökte uppfattningar om MINI. Pinninti och medarbetare intervjuade drygt 100 patienter i psykiatrisk öppenvård [17]. MINI tog i genomsnitt 16 minuter att genomföra (8–31 minuter). De allra flesta ansåg att MINI inte var för tidskrävande, att den täckte alla deras symtom och att de inte såg några problem med den strukturerade proceduren. I den norska MINI-studien av Mordal och medarbetare uppfattade patienterna att intervjun var meningsfull och till stora delar täckte deras symtom [15]. Upplevelserna av intervjun korrelerade negativt med psykossymtom.

Den ovan nämnda studien av Suppiger och medarbetare fann att cirka 80 procent av patienterna kände att intervjun med ADIS var ”helpful” [11]. Nästan samtliga ansåg att relationen till psykologen var positiv och att intervjuaren tog deras problem på allvar.

Bruchmullers och medarbetare enkätstudie omfattade även patienter [14]. De accepterade strukturerade intervjuer i signifikant högre grad än vad terapeuterna själva trodde.

## Diskussion

Svarsfrekvensen på SBU:s enkäter var mycket lägre än förväntat. Detta gör att resultaten måste bedömas med försiktighet. Att resultaten stämmer med tidigare undersökningar och internationell litteratur gör dock att vi tror att de speglar förhållandena i Sverige idag.

Vi fann att strukturerade eller semistrukturerade intervjuer används i liten utsträckning. Formulär för riktad screening är vanligare; cirka hälften av psykiatrer och läkare i primärvården uppger att de använder dem regelbundet. Likaså följer cirka en tredjedel av läkarna upp behandlingen med hjälp av ett bedömningsformulär.

Det finns ett spektrum av åsikter och erfarenheter om bedömningsformulär bland läkarna, något som framkommer i litteraturen och i svenska enkäter. Till fördelarna hör ökad struktur, bättre möjlighet till dokumentation samt stöd för komplicerad diagnostik. Nackdelar som ofta nämns är tidsbrist och att formulären är stelbenta. När det gäller patient-läkarrelationen och påverkan på konsultationen finns inga entydiga svar. Formulären kan upplevas som ett hinder för vissa läkare och vara till nytta för andra.

I studier har patienterna genomgående varit positivt inställda till bedömningsformulär. De flesta har inte sett några problem med att fylla i formulär eller att följa en strukturerad intervju.

## Referenser

1. Demyttenaere K, De Fruyt J. Getting you ask for: on the selectivity of depression rating scales. *Psychother Psychosom* 2003;72:61-70.
2. ACHPR. Agency for Health Care Policy and Research Depression Guideline Panel (1993a). Depression in primary care: Vol. 1. Detection and diagnosis (Clinical Practice Guideline No. 5; ACHPR Publication No. 93 0550). Rockville, MD: U.S. Department of Health and Human Services, Public Health Service; 1993.
3. NICE. Depression: The treatment and management of depression in adults (update). NICE clinical guideline 90 (2009). Available from [www.nice.org.uk/CG90](http://www.nice.org.uk/CG90).
4. Socialstyrelsen. Nationella riktlinjer för vård vid depression och ångestsyndrom 2010 – stöd för styrning och ledning.
5. Socialstyrelsen. Antidepressiva läkemedel vid psykisk ohälsa. Studier av praxis i primärvården. 2006:1-97.
6. Socialstyrelsen. Barn- och ungdomspsykiatrins metoder – En nationell inventering; 2009.
7. Magnusson M, Lindfors A, Tell J. [In Process Citation]. *Lakartidningen* 2011;108:1618-21.
8. Gilbody SM, House AO, Sheldon TA. Psychiatrists in the UK do not use outcomes measures. *National survey. Br J Psychiatry* 2002; 180:101-3.
9. Sørensen MJ, Thomsen PH, Bilenberg N. Parent and child acceptability and staff evaluation of K-SADS-PL: A pilot study. *Eur Child Adolesc Psychiatry* 2007;16: 293-7.
10. Duffy FF, Chung H, Trivedi M, Rae DS, Regier DA, Katzelnick DJ. Systematic use of patient-rated depression severity monitoring: is it helpful and feasible in clinical psychiatry? *Psychiatr Serv* 2008;59:1148-54.
11. Suppiger A, In-Albon T, Hendriksen S, Hermann E, Margraf J, Schneider S. Acceptance of structured diagnostic interviews for mental disorders in clinical practice and research settings. *Behav Ther* 2009;40:272-9.
12. Bermejo I, Niebling W, Berger M, Härter M. Patients' and physicians' evaluation of the PHQ-D for depression screening. *Prim Care Community Psychiatry* 2005;10:125-31.
13. Dowrick C, Leydon GM, McBride A, al. E. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ* 2009;338:b663 doi:10.1136/bmj.b663.
14. Bruchmuller K, Margraf J, Suppiger A, Schneider S. Popular or unpopular? Therapists' use of structured interviews and their estimation

- of patient acceptance. *Behav Ther* 2011; 42:634-43.
15. Mordal J, Gundersen O, Bramness JG. Norwegian version of the Mini-International Neuropsychiatric Interview: feasibility, acceptability and test-retest reliability in an acute psychiatric ward. *Eur Psychiatry* 2010; 25:172-7.
16. Mason L, Poole H. Healthcare professionals' views of screening for post-natal depression. *Community Pract* 2008;81:30-3.
17. Pinninti NR, Madison H, Musser E, Rissmiller D. MINI International Neuropsychiatric Schedule: clinical utility and patient acceptance. *Eur Psychiatry* 2003; 18:361-4.



## 8. Kunskapsluckor

---

Vår granskning visade att kunskapsläget om bedömningsformulärens tillförlitlighet är varierande. Generellt sett är kunskapen om diagnostisk tillförlitlighet sämre när det gäller formulär för barn, ungdomar och äldre.

En stor svaghet är att även om det finns vetenskapligt stöd för enskilda formulär, saknas ofta studier som utvärderar effekter av att använda dem under förhållanden som är relevanta för Sverige. I många fall är också kunskapen otillräcklig om hur tillförlitliga de formulär är som inte översatts till svenska på vedertaget sätt. Studier på användning av diagnostisk stöd vid andra psykiatriska diagnoser och från andra sjukvårdssystem (Kapitel 3.2) tyder dock på att användning av bedömningsformulär förbättrar både vårdprocess och patientens psykiska hälsa.

Vår bedömning är att de mest angelägna forskningsbehoven på övergripande nivå är:

- hur patienterna och klinikerna upplever omhändertagandet med och utan diagnostiska hjälpmedel, hur tillfrisknande, livskvalitet, funktionsförmåga och relation mellan läkare och patient påverkas
- hälsoekonomiska och organisatoriska effekter av en diagnostisk process som kompletteras med bedömningsformulär i primärvård och specialistpsykiatri
- diagnostisk tillförlitlighet för den svenska versionen av formulär som har vetenskapligt stöd i patientgrupper med olika prevalens av förstämningssyndrom.

Det finns också viktiga forskningsbehov som rör enskilda formulär:

- Diagnostisk tillförlitlighet relativt LEAD för intervjuformulären K-SADS, MINI-KID, BCFI och DAWBA för barn och ungdomar.
- Diagnostisk tillförlitlighet jämfört med strukturerade eller semistrukturerade intervjuer för riktad screening av depression med GDS-15 och GDS-30 hos de allra äldsta och hos personer med kognitiv svikt.
- Mätning av svårighetsgrad med Cornellskalan jämfört med en adekvat referensstandard hos personer med depression och samtidig kognitiv svikt.
- Utveckling av formulär med bättre känslighet än EPDS för screening av depression under graviditet och efter förlossning.
- Hur psykiatrisk samsjuklighet påverkar den diagnostiska tillförlitligheten för MDQ och HCL-32. Särskilt viktigt är att klarlägga hur bipolära syndrom kan särskiljas från tillstånd med bristande impuls-kontroll som t ex ADHD och borderline personlighetsstörning.

Metoder för att pröva bedömningsformulärens förmåga att utvärdera svårighetsgrad och sjukdomsförlopp behöver vidareutvecklas.

## 9. Konsekvenser av rapportens resultat

---

Psykiatrisk diagnostik är ett område som tilldrar sig stort intresse, och behovet av kvalitetssäkrad diagnostik poängteras särskilt i Socialstyrelsens nationella riktlinjer för vård vid depression och ångestsyndrom [1] liksom från andra styrande myndigheter internationellt.

En jämförelse mellan vår systematiska översikt och kartläggningar av svensk praktik visar dock att det finns skillnader mellan vetenskap och praxis vad gäller den diagnostiska processen när förstämningssyndrom misstänks. Det gäller såväl hur själva diagnostiken genomförs som eventuella val av bedömningsformulär. En journalstudie från svensk primärvård av Hassler och medarbetare kom t ex fram till att kriterier för depression nämndes i undantagsfall i journalerna [2].

### Önskvärda förändringar

På sikt måste målet vara att primärvård och specialiserad vård använder en kvalitetssäkrad diagnostisk process. Det innebär att tillvägagångssätten skulle närma sig LEAD (*Longitudinal, Experts, All Data procedure*, Kapitel 1).

Diagnosen skulle då baseras på en sammanvägning av information från klinisk intervju, resultat från intervjuformulär och i förekommande fall resultat från laboratorieundersökningar, röntgen, bedömningar från andra yrkesgrupper och information från närstående.

Ett steg på vägen för att optimera diagnostiken är att använda formulär med tillräcklig diagnostisk tillförlitlighet. För verksamheterna kan det innebära såväl att formulär införs som att andra, där evidensen tyder på att tillförlitligheten är osäker, fasas ut (Tabell 9.1 och 9.2 som sammanställer resultaten från vår granskning, Kapitel 3).



**Tabell 9.1** Formulär med vetenskapligt stöd för tillräcklig diagnostisk tillförlitlighet.

Åldersintervall	Intervju	Riktad screening	Svårighetsgrad
Barn och ungdomar		Depression: CBCL, BDI-I Mani: YMRS	
Vuxna	SCID-I, MINI	Depression: BDI-II Post partum-depression: EPDS Mani: HCL-32, MDQ med modifierade kriterier	
Äldre än 65 år		Depression: GDS-15, GDS-30	

**Tabell 9.2** Formulär med vetenskapligt stöd för att den diagnostiska tillförlitligheten inte är tillräcklig.

Åldersintervall	Intervju	Riktad screening	Svårighetsgrad
Barn och ungdomar		CDI	
Vuxna	PRIME-MD	HADS, PHQ-9	
Äldre än 65 år		PHQ-9	

En webbaserad nationell sammanställning över bedömningsformulär skulle underlätta att välja lämpligt formulär. Drygt 60 bedömningsformulär används enligt vår inledande inventering. Huvuddelen av dem har utvecklats i engelskspråkiga länder. En del har genomgått en formell översättningsprocedur och efterföljande utvärdering av den svenska versionen medan andra kan finnas i en eller flera icke officiella översättningar. Sådana översatta formulär är ibland inte alls utvärderade, åtminstone inte i sin svenska version. Det finns också exempel på instrument som cirkulerar utan att frågan om översättningsstatus, validering och rättighet att använda dem är känd.

Idag finns det ingen lättillgänglig sammanställning av bedömningsformulär för sjukvården och dess huvudmän. Socialstyrelsen har påbörjat en sådan metodguide för formulär som används inom socialt arbete [3]. En motsvarande nationell metodguide för psykiatriska bedömningsformulär vore eftersträvsvärd. Utöver information om ändamålet med formuläret och dess prestanda kan sammanställningen innehålla uppgifter om tillgänglighet och eventuella pågående utvärderingsstudier. Dessutom kan utvärderade formulär bli mer spridda och åtkomliga för verksamheterna.

Införandet av en nationell formulärguide kan på så sätt leda till att det blir lättare för verksamheterna att välja formulär med gott vetenskapligt stöd. På samma sätt kan den bidra till att formulär som saknar vetenskapligt stöd fasas ut.

Ytterligare en förändring som är önskvärd är att screening för post partum depression erbjuds samtliga kvinnor, oavsett språktillhörighet. Det är viktigt att det finns en handlingsplan som inbegriper säkerställande av diagnosen och optimal behandling.

## **Konsekvenser av att införa en strukturerad diagnostisk process**

### **För patienterna**

En strukturerad diagnostik förbättrar förutsättningarna för att patienterna får en korrekt diagnos. På så sätt en större andel av de patienter som har förstämningssyndrom bli identifierade och få tillgång till effektiv behandling.

### **För sjukvården**

En uppenbar konsekvens av att införa helt eller delvis strukturerade intervjuer är att tidsåtgången för diagnostiken kan komma att öka. Den tid det tar att gå igenom intervjuformulären varierar från cirka 20 minuter och upp till en timme i okomplicerade fall. Å andra sidan ger intervjuerna möjlighet att tidigt få en uppfattning om samsjuklighet och behov av att remittera patienterna vidare.

Diagnosspecifika bedömningsformulär, oftast så kallade skattningsskalor, behöver ofta inte ta tid från konsultationen eftersom de fylls i av patienten, vilket kan göras i hemmet eller väntrummet (eller motsvarande).

En annan konsekvens är att patientjournalerna skulle behöva anpassas. Om intervjuer och formulär ska kunna införlivas i klinisk rutin måste de kunna anslutas till patientjournalen på ett enkelt sätt.

Slutligen skulle det behövas rutiner för att kvalitetssäkra diagnostiken. Det kan t ex omfatta fortbildning om bedömningsformulär, deras användningsområden, styrkor och svagheter och hur resultaten från dem ska tolkas, samt regelbundna samskattningar.

## Referenser

1. Socialstyrelsen. Nationella riktlinjer för vård vid depression och ångestsyndrom 2010 – stöd för styrning och ledning.
2. Vård av patienter med depression och ångest – en nulägesrapport. Socialstyrelsen; 2009. Artikelnr: 2009-9-9.
3. Socialstyrelsen.  
(<http://www.socialstyrelsen.se/evidensbaseradpraktik/metodguide>).



## 10. Ord- och förkortningslista<sup>1</sup>

---

<b>Algoritm</b>	En beräkningsmodell
<b>Anamnes</b>	Sjukdomshistoria
<b>Atypisk</b>	Något som skiljer sig från det typiska
<b>Best estimate diagnos</b>	Se LEAD
<b>Bias</b>	Ett resultatfel som uppstått genom procedurfel, effektbedömningsfel eller annat mänskligt fel under en undersökning; även fel som görs i bedömningen eller hanteringen av resultaten
<b>Bipolära syndrom</b>	Sjukdomstillstånd som omfattar såväl episoder av förhöjt stämningsläge, ökad energi och aktivitet (mani eller hypomani) som episoder med sänkt stämningsläge, minskad energi och minskad aktivitet (depression)
<b>Bivariat analys</b>	Analysen omfattar två variabler samtidigt, t ex sensitivitet och specificitet
<b>Blindning</b>	Maskering, åtgärder för att hemlighålla vissa centrala omständigheter i en undersökning tills den är avslutad och resultaten ska bearbetas. Viktigaste exemplet: i en blindad klinisk behandlingsprövning är det okänt vilka deltagare som får den ena eller den andra av de prövade behandlingsformerna. En viktig upplysning är vilka av parterna som uppgifterna har hållits hemliga för - deltagare, försöksledare, prövningspersonal och/eller statistiker

---

<sup>1</sup> Förkortningslista – Bedömningsformulär sist i kapitlet.

<b>Depression</b>	Ett sjukdomstillstånd med sänkt stämningsläge, minskad energi och minskad aktivitet
<b>Degenerativ hjärnsjukdom</b>	Sjukdomar i hjärnan som bryter ned hjärnvävnad. Den vanligaste degenerativa hjärnsjukdomen är Alzheimers sjukdom
<b>Differentialdiagnos</b>	Andra tänkbara orsaker till symtom än en viss diagnos
<b>DSM-IV</b>	Diagnostic and Statistical Manual of Mental Disorder, 4th edition, klassifikation av psykiatriska tillstånd
<b>Dystymi</b>	Sjukdomstillstånd med sänkt stämningsläge de flesta dagar under en period av minst två år som inte uppfyller kriterierna för egentlig depression
<b>Egentlig depression</b>	Sjukdomstillstånd med sänkt stämningsläge med sådan svårighetsgrad att den ger upphov till någon form av funktionsnedsättning
<b>Epidemiologi</b>	Vetenskaplig disciplin om sjukdomars utbredning
<b>Estimat</b>	Uppskattat värde
<b>Evidensbaserad</b>	Något som bedöms tyda på att ett visst förhållande gäller (av latinets evidentia ”tydlighet”). I termen evidensbaserad sjukvård är evidens det sammanvägda resultatet av systematiskt insamlade och kvalitetsgranskade vetenskapliga observationer, vilka ska uppfylla bestämda krav på tillförlitlighet så att de sammantaget kan anses utgöra ”bästa tillgängliga bevis” i en viss fråga

<b>Exklusionskriterium</b>	Omständigheter som förhindrar att en person föreslås delta i en undersökning, eller att en viss studie inte tas med i en systematisk granskning. Dessa kriterier ska vara angivna i undersökningens protokoll. Exempel: personen har, förutom den sjukdom som ska studeras, också en annan sjukdom som kan väntas störa bedömningen. Graviditet är ett exklusionskriterium i nästan alla läkemedelsprövningar
<b>Extra-pyramidala symptom</b>	Rörelsebesvär, Parkinsonliknande symptom eller ofrivilliga muskelsammandragningar. Förekommer som biverkan vid vissa läkemedelsbehandlingar
<b>Forest plot</b>	Grafisk presentation av resultatet av en metaanalys
<b>Förstämningsyndrom</b>	Sjukdomstillstånd med förhöjd eller sänkt sinnesstämning
<b>Generaliserbarhet</b>	Hur resultaten i en studie kan appliceras på t ex människor i andra världsdelar
<b>Gold standard</b>	En referensstandard under ideala förhållanden
<b>Heterogenicitet</b>	Skillnader mellan studier avseende t ex patienturval
<b>Hypomani</b>	Sjukdomstillstånd med förhöjt stämningsläge, som inte ger upphov till större funktionsnedsättning
<b>ICD-10</b>	International Statistical Classification of Diseases and related Health Problems, 10th revision. Klassifikation av sjukdomar enligt Världshälsoorganisationen



<b>Inklusionskriterium</b>	De betingelser som ska föreligga för att en person ska kunna erbjudas att delta i en undersökning. Kriterierna kan gälla sjukdomsdiagnos, åldersgrupp m m. Att informerat samtycke givits är ett obligatoriskt inklusionskriterium. Deltagande i undersökningen kräver dessutom att inget av exklusionskriterierna föreligger
<b>Interbedömarreliabilitet</b>	Grad av överensstämmelse mellan bedömare
<b>Ischemisk hjärtsjukdom</b>	Hjärtsjukdom som orsakas av att organet drabbats av syrebrist
<b>Kognitiv svikt</b>	Störningar i kognitiv funktion, t ex nedsatt minne
<b>Konfidensintervall, KI</b>	Ett talintervall som med viss angiven sannolikhet innefattar det sanna värdet av t ex ett medeltal eller en oddskvot. Konfidensintervallet innehåller alla tänkbara värden som inte kan förkastas på grundval av föreliggande data. Vanligen anges övre och nedre gränsen för ett konfidensintervall som har 95 procents sannolikhet
<b>LEAD</b>	Longitudinal, Experts, All Data procedure. LEAD anses vara den optimala referensstandarderna inom psykiatrin. LEAD innebär en samlad värdering från erfarna bedömare av information från samtal med patienten, diagnostiska intervjuer och frågeformulär, information från närstående, observationer av patienten samt erforderliga laboratorieundersökningar
<b>Likertskala</b>	En skala där respondenten får ange i hur hög grad han/hon instämmer i ett påstående (t ex grad av smärta)
<b>Mani</b>	Sjukdomstillstånd med förhöjt stämningsläge som ger upphov till funktionsnedsättning

<b>Metaanalys</b>	Metod att göra en samlad bedömning av ett antal jämförande undersökningar genom att statistiskt sammanföra deras resultat. Genom omfattande litteratursökning anskaffas allt publicerat material inom det valda området. Ibland försöker man också skaffa fram opublicerade data. Hela materialet granskas, och uppställda inklusions- och exklusionskriterier avgör vilka undersökningar som ska accepteras. Metaanalysen redovisar samtliga resultat i form av ett jämförande resultatmått (t ex oddskvot) med tillhörande konfidensintervall samt, genom en statistisk sammanslagning av resultaten, ett samlat resultatmått med tillhörande konfidensintervall. Proceduren ger en överblick över tillgängliga resultat och deras samstämmighet (homogenitet). De summerade jämförelsemåtten ger en sammanfattande uppfattning om huruvida publikationerna visat att en metod är bättre än en annan metod (eller bättre än ingen åtgärd alls)
<b>Minor depression</b>	Sjukdomstillstånd med sänkt stämningsläge som inte uppfyller kriterierna för egentlig depression
<b>Multicenterstudie</b>	Klinisk studie som utförts på flera medicinska centra
<b>Odds kvot, OR</b>	Odds ratio, OR, kvoten mellan två odds. Exempel: odds exponerade/icke exponerade bland personer med viss sjukdom dividerat med motsvarande odds bland friska. Odds kvoten ger en uppfattning om hur starkt sambandet är mellan exponeringen och sjukdomen
<b>Onkologi</b>	Medicinsk disciplin som behandlar tumörer och dess behandling
<b>Palliativ</b>	Vård i slutskedet av livet som syftar till att ge patienten psykiskt välbefinnande

<b>Pilotstudie</b>	Studie som testar i mindre skala innan studien genomförs i full skala
<b>Positivt/negativt prediktionsvärde</b>	Positivt prediktionsvärde anger sannolikheten för att en person som får positivt utfall (dvs resultatet talar för sjukdom) verkligen har sjukdomen. Siffran är kvoten mellan sant positiva och sant + falskt positiva. Negativt prediktionsvärde anger sannolikheten för att en person som får negativt utfall (dvs resultatet talar emot sjukdom) verkligen inte har sjukdomen. Siffran är kvoten mellan sant negativa och sant + falskt negativa
<b>Post partum-depression</b>	Sjukdomstillstånd med sänkt stämningsläge som debuterar kort tid efter förlossning
<b>Precision</b>	Noggrannheten hos en klinisk mätmetod (t ex blodanalys): låg precision innebär stora slumpmässiga avvikelser från sanna värden
<b>Prevalens</b>	Sjukdomsförekomst: antalet personer som vid en viss tidpunkt har en viss sjukdom. Uttrycks t ex som antal sjuka per 1 000 personer eller som andel av hela landets befolkning
<b>Psykometri</b>	Ett forskningsområde som utformar och kvalitetsgranskar frågeformulär
<b>Publikationsbias</b>	Snedvridning av publicerade resultat av studier orsakad av att undersökare, ibland också tidskriftsredaktörer, föredrar att publicera undersökningar som givit positivt resultat, t ex visat att en behandling har effekt. Studier som inte visat någon effekt blir till stor del okända, och bilden av behandlingens värde blir omotiverat gynnsam

<b>Randomiserad</b>	Slumpmässig fördelning av deltagarna mellan grupperna i en undersökning. Randomiseringen är förutsättningen för att man med statistiska metoder ska kunna bedöma sannolikheten för att undersökningens resultat uppkommit genom slumpens verkan. Randomiseringen har dessutom förutsättningar att fördela okända stör-faktorer (confounders) lika mellan grupperna samt göra grupperna önskvärt jämförbara i sin sammansättning; slumpen kan dock åstadkomma vissa skillnader. Randomisering utförs t ex genom att en dator genererar en slumpmässig sekvens (randomiseringskod) som avgör till vilken av undersökningens olika grupper varje ny deltagare ska föras. Randomisering ger tillsammans med blindning skydd mot selektionsbias
<b>Reliabilitet</b>	Exakthet i mätningar, t ex om testet ger samma resultat vid upprepade testomgångar och oberoende av vem som utför testet.
<b>Retrospektiv</b>	Undersökning som är ”tillbakablickande”, dvs utnyttjar data som är anskaffade, respektive händelser som inträffat, innan undersökningen startades. Ett typiskt exempel är fall–kontrollundersökningar
<b>ROC-AUC</b>	Receiver Operating Characteristic – Area Under Curve, är en kurva som visar sambandet mellan sensitivitet och specificitet vid olika tröskelvärden. Ytan under kurvan, ROC-AUC, är ett mått på hur väl testet kan skilja på sjuk eller frisk
<b>Schizoaffektivt syndrom</b>	Sjukdomstillstånd med betydande varaktighet av egentlig depression, maniska symtom eller blandsymtom samtidigt som symtom på schizofreni samt en period på minst två veckor utan symtom på förstämningssyndrom

<b>Screening</b>	Engelska för sållning. Undersökning av en population för upptäckt av vissa icke diagnostiserade sjukdomar eller riskfaktorer för sjukdom. Exempel: allmän mammografiundersökning av kvinnor i vissa åldrar
<b>Semi-strukturerad intervju</b>	Intervju där färdigformulerade frågor kan kompletteras med följdfrågor
<b>Sensitivitet</b>	Egenskap hos diagnosmetod: andelen av sjuka som metoden identifierar korrekt
<b>Sensoriska störningar</b>	Störningar i hjärnans förmåga att hantera sinnesintryck.
<b>Signifikant</b>	Ett statistiskt mått på om en observation beror på slump eller inte
<b>SMD</b>	Standardized Mean Differences, mått som ofta används i metaanalyser
<b>Somatisk</b>	Kroppslig
<b>Specificitet</b>	Egenskap hos diagnosmetod: andelen av friska som metoden identifierar korrekt
<b>Stratifierad</b>	Populationen delas in i grupper efter olika egenskaper t ex ålder eller utbildning
<b>Strukturerad intervju</b>	Intervjun består endast av färdigformulerade frågor

<b>Systematisk översikt</b>	En översikt som avser en tydligt formulerad fråga och som använder systematiska och explicita metoder för att identifiera, välja ut och kritiskt bedöma relevanta studier samt för att samla in och analysera uppgifter från dessa. Analysen görs ofta med hjälp av metaanalyser
<b>Trimester</b>	Sista tredjedelen av graviditeten
<b>Tröskelvärde</b>	Ett gränsvärde som t ex skiljer mellan sjuk och frisk i ett test (eng. cut off)
<b>Tvärsnittsstudie</b>	(Cross sectional study), undersökning av ett antal personer som utförs vid ett enda tillfälle. Exempel: personerna intervjuas om sina kost- och motionsvanor, och man mäter deras kroppsvikt, blodsocker och blodkolesterol
<b>Utfallsmått</b>	Det mått på effekt eller tillförlitlighet som en klinisk studie avser att utvärdera
<b>Validitet</b>	I vilken grad testet mäter vad det är avsett att mäta. I ett frågeformulär ska frågorna vara relevanta för tillståndet, antalet frågor ska vara tillräckligt för att täcka in olika aspekter och frågorna ska tolkas på ett likartat sätt av dem som fyller i formuläret
<b>Årstidsrelaterad depression</b>	Återkommande sjukdomstillstånd med ett sänkt stämningsläge under en del av året, vanligen vintern

## Förkortningslista – Bedömningsformulär

<b>A-YMRS</b>	Adolescent Young Mania Rating Scale
<b>AS-18</b>	Affective Self Rating Scale-18
<b>AS-18-Dep</b>	Affective Self Rating Scale – 18 Depression
<b>AS-18-Man</b>	Affective Self Rating Scale – 18 Mania
<b>ASEBA</b>	Achenbach System of Empirically Based Assessment
<b>BCDRS</b>	Brief Carroll Depression Rating Scale
<b>BCFPI</b>	Brief Child and Family Phone Interview
<b>BDI-I</b>	Beck Depression Inventory-I
<b>BDI-II</b>	Beck Depression Inventory-II
<b>BDRS</b>	Bipolar Depression Rating Scale
<b>BUS</b>	Becks ungdomsskalor
<b>BYI</b>	Beck Youth Inventories
<b>CBCL</b>	Child Behaviour Checklist
<b>CBQ</b>	Child Bipolar Questionnaire
<b>CDSS</b>	Calgary Depression Scale for Schizophrenia
<b>CES-D</b>	Center for Epidemiologic Studies Depression Scale
<b>CES-DC</b>	Center of Epidemiological Studies Depression Scale for Children

<b>CGI</b>	Clinical Global Impressions scale
<b>CGI-BP</b>	Clinical Global Impressions Scale Bipolar Version
<b>CGI-BP-D/ CGI-BP-M</b>	Clinical Global Impressions scale Bipolar version, modified version
<b>CGI-C</b>	Clinical Global Impressions of change
<b>CGI-I</b>	Clinical Global Impressions of Improvement
<b>CGI-S</b>	Clinical Global Impressions, Severity Scale
<b>CIDI</b>	Composite International Diagnostic Interview
<b>CIDI-Auto</b>	Composite International Diagnostic Interview, Computer assisted version
<b>CIDI-SF</b>	Composite International Diagnostic Interview, Short Form
<b>UM-CIDI</b>	Composite Internatinal Diagnostic Interveiw, University of Michigan
<b>CMRS</b>	Child Mania Rating Scale
<b>CMRS-P</b>	Child Mania Rating Scale – Parents
<b>CPRS</b>	Comprehensive Psychopathological Rating Scale
<b>CSDD</b>	Cornell Scale for Depression in Dementia
<b>DAWBA</b>	Development and Well-Being Assessment
<b>DICA</b>	Diagnostic Interview for Children and Adolescents



<b>DICA-IV</b>	Diagnostic Interview Schedule for Children and Adolescents, Fourth edition
<b>DIS</b>	Diagnostic Interview Schedule
<b>DISC-IV</b>	Diagnostic Interview Schedule for Children, Version 4
<b>DSRS</b>	Birleson Depression Self Rating Scale
<b>EPDS</b>	Edinburgh Postnatal Depression Scale
<b>GDS</b>	Geriatric Depression Scale
<b>GPI</b>	General Behavior Inventory
<b>HADS</b>	Hospital Anxiety and Depression Scale
<b>HADS-D</b>	Hospital Anxiety and Depression Scale, Depression subscale Hospital Anxiety and Depression Scale, (depressionsdelen)
<b>HCL-32</b>	Hypomania/Mania Symptom Checklist
<b>HDRS</b>	Hamilton Depression Rating Scale
<b>HSCL-25</b>	Hopkins symptom checklist
<b>IDS</b>	Inventory of Depressive Symptomatology
<b>IDS-c</b>	Inventory of Depressive Symptomatology, Clinician-Rated
<b>IDS-sr</b>	Inventory of Depressive Symptomatology, Self Report

<b>KID-SCID</b>	Structured Clinical Interview for DSM-IV, Childhood Diagnoses
<b>K-SADS</b>	Kiddie-Schedule for Affective Disorders and Schizophrenia for School-Age Children
<b>K-SADS-E</b>	Kiddie-Schedule for Affective Disorders and Schizophrenia for School-Age Children, Epidemiological version
<b>K-SADS-PL</b>	Kiddies-Schedule for Affective Disorders and Schizophrenia for School-Age children present and life time
<b>LCM</b>	Life Chart Methodology
<b>LCM-p</b>	Life Chart Methodology (p= prospektiv)
<b>LCM-s/p</b>	Life Chart Methodology Patient Self Rated Prospective Form
<b>LCM-s/r</b>	Life Chart Methodology Patient Self Rated Retrospective Form
<b>LEAD</b>	Longitudinal, Experts, All Data procedure
<b>MADRS</b>	Montgomery-Åsberg Depression Rating Scale
<b>MADRS-S</b>	MADRS Self-Assessment
<b>MAS</b>	Bech-Rafaelsen Mania Scale
<b>MDI</b>	Major Depression Inventory
<b>MDQ</b>	Mood Disorder Questionnaire

<b>MDQ-A</b>	Mood Disorder Questionnaire – Adolescent version
<b>MINI</b>	Mini International Neuropsychiatric Interview
<b>P-GBI</b>	Parent General Behavior Inventory
<b>PHQ-2</b>	Patient Health Questionnaire-2
<b>PHQ-9</b>	Patient Health Questionnaire
<b>PRIME-MD</b>	Primary Care Evaluation of Mental Disorders
<b>P-YMRS</b>	Parent Version Young Mania Rating Scale
<b>QIDS</b>	Quick Inventory of Depressive Symptomatology
<b>QIDS-16-c</b>	QIDS Clinician Rated
<b>QIDS-16-sr</b>	QIDS Self Report
<b>SADS</b>	Schedule for Affective Disorders and Schizophrenia
<b>SADS-L</b>	Schedule for Affective Disorders and Schizophrenia Lifetime Version
<b>SCAN</b>	Schedules for Clinical Assessment in Neuropsychiatry
<b>SCID-CV</b>	Structured Clinical Interview for DSM-Disorders, Clinical version
<b>SCID-I</b>	Structured Clinical Interview for DSM-IV Axis I disorders
<b>SCID-I-CV</b>	Structured Clinical Interview for DSM-IV, Clinical version

<b>SCID-I-NP</b>	Structured Clinical Interview for DSM-IV, Non patient version
<b>SCID-I-P</b>	Structured Clinical Interview for DSM-IV-Disorders, Patient version
<b>SCL</b>	Symptom Checklist
<b>SDQ</b>	Strength and Difficulties Questionnaire
<b>SDS</b>	Zung Self-Rating Depression Scale
<b>SHQ</b>	Seasonal Health Questionnaire
<b>SIDI</b>	Stroke Inpatient Depression Inventory
<b>SIDP-IV</b>	Structured Interview for DSM-IV Personality
<b>SPAQ</b>	Seasonal Pattern Assessment Questionnaire
<b>SPI</b>	Standardized Psychiatric Interview
<b>SPIFA</b>	Structured Psychiatric Interview for General Practice
<b>TRF</b>	Teacher Report Form
<b>UM-CIDI</b>	Composite International Diagnostic Interview, University of Michigan version
<b>WASH-U-KADS</b>	Washington University in St Louis Kiddie Schedule etc
<b>YMRS</b>	Young Ziegler Mania Rating Scale



# 11. Projektgrupp, externa granskare, bindningar och jäv

---

Medlemmarna i gruppen representerar olika infallsvinklar på kunskapsområdet och gruppen har bestått av följande personer:

## Projektgrupp

*Derya Akcan (informationspecialist)*  
SBU, Stockholm

*Per Allard*  
Docent, Institutionen för klinisk vetenskap/enheten för psykiatri, Umeå universitet

*Kristina Bengtsson Boström*  
Docent, Billingens vårdcentral, Skövde

*Thomas Davidson (hälsoekonom)*  
SBU, Stockholm

*Lisa Ekselius*  
Professor, överläkare, Institutionen för neurovetenskap, Uppsala universitet och Psykiatridivisionen Akademiska sjukhuset, Uppsala

*Petter Gustavsson*  
Professor, Karolinska Institutet, Sektionen för psykologi, Stockholm

*Kerstin Hagenfeldt*  
Professor emeritus, Stockholm

*Susanna Kjellander (biträdande projektledare)*

SBU, Stockholm

*Ingvar Krakau*

Universitetslektor, Institutionen för medicin,  
Karolinska Universitetssjukhuset, Solna

*Anders Norlund (hälsoekonom)*

SBU, Stockholm

*Agneta Pettersson (projektledare)*

SBU, Stockholm

*Mia Ramklint*

Universitetslektor, Specialist i barn- och ungdomspsykiatri,  
Akademiska sjukhuset, Uppsala

*Charlotta Sunnqvist*

Universitetslektor, Institutionen för vårdvetenskaper, Malmö högskola

*Cecilia Svanborg*

Med dr, Specialist i psykiatri, Leg psykoterapeut,  
Karolinska Universitetssjukhuset Huddinge

*Anneth Syversson (projektassistent)*

SBU, Stockholm

*Maria Tillfors*

Professor, Institutionen för juridik, psykologi och socialt arbete,  
Örebro universitet

## Externa vetenskapliga granskare

*Malin André*

Docent, primärvårdens FOU-enheten, Uppsala

*Ata Ghaderi*

Professor, Institutionen för psykologi, Uppsala universitet

*Mikael Landén*

Professor, Institutionen för neurovetenskap och fysiologi,  
Göteborgs universitet

*Ann-Charlotte Smedler*

Professor, Psykologiska institutionen, Stockholms universitet

*Ingmar Skoog*

Professor, Sektionen för psykiatri och neurokemi,  
Göteborgs universitet

*Hans Ågren*

Professor emeritus, Göteborg

## Bindningar och jäv

SBU kräver att alla som deltar i projektgrupper lämnar skriftliga deklara-  
tioner avseende potentiella bindningar eller jäv. Sådana intressekonflikter  
kan föreligga om medlem i gruppen får ekonomisk ersättning från part  
som kan ha intresse i de frågor gruppen studerar. Gruppens ordförande  
och SBU tar därefter ställning till om det finns några omständigheter  
som skulle kunna försvåra en objektiv värdering av kunskapsunderlaget  
och ger vid behov förslag till åtgärder

Sakkunniga och granskare har i enlighet med SBU:s krav inlämnat  
deklaration rörande bindningar och jäv. Dessa dokument finns till-  
gängliga på SBU:s kansli. SBU har bedömt att jäv inte föreligger.



# Rapporter publicerade av SBU

## Gula rapporter (2005–2012)

- 
- Diagnostik och uppföljning av förstämningssyndrom (2012), nr 212
- 
- Implementeringsstöd för psykiatrisk evidens i primärvården (2012), nr 211
- 
- Arbetets betydelse för uppkomst av besvär och sjukdomar – nacken och övre rörelseapparaten (2012), nr 210
- 
- Godartad prostataförstoring med avflödes hinder (2011), nr 209
- 
- Medicinska och psykologiska metoder för att förebygga sexuella övergrepp mot barn (2011), nr 207
- 
- Blödande magsår (2011), nr 206
- 
- Tandförluster (2010), nr 204
- 
- Rotfyllning (2010), nr 203
- 
- Program för att förebygga psykisk ohälsa hos barn (2010), nr 202
- 
- Mat vid diabetes (2010), nr 201
- 
- Antibiotikaprofylax vid kirurgiska ingrepp (2010), nr 200
- 
- Behandling av sömnbesvär hos vuxna (2010), nr 199
- 
- Rehabilitering vid långvarig smärta (2010), nr 198
- 
- Triage och flödesprocesser på akutmottagningen (2010), nr 197
- 
- Intensiv glukossänkande behandling vid diabetes (2009), nr 196
- 
- Patientutbildning vid diabetes (2009), nr 195
- 
- Egna mätningar av blodglukos vid diabetes utan insulinbehandling (2009), nr 194
- 
- Äldres läkemedelsanvändning – hur kan den förbättras? (2009), nr 193
- 
- Transkraniell magnetstimulering (Uppdatering av Kapitel 8 i SBU-rapport 166/2 från 2004) (2007), nr 192. *Publiceras endast i elektronisk version på [www.sbu.se](http://www.sbu.se)*
- 
- Vacciner till barn – skyddseffekt och biverkningar (2009), nr 191
- 
- Öppenvinkelglaukom (grön starr) – diagnostik, uppföljning och behandling (2008), nr 190
- 
- Rörbehandling vid inflammation i mellanörat (2008), nr 189
- 
- Karies – diagnostik, riskbedömning och icke-invasiv behandling (2007), nr 188
- 
- Benartärsjukdom – diagnostik och behandling (2007), nr 187
- 
- Ljusterapi vid depression samt övrig behandling av årstidsbunden depression (Uppdatering av Kapitel 9 i SBU-rapport 166/2 från 2004) (2007), nr 186. *Publiceras endast i elektronisk version på [www.sbu.se](http://www.sbu.se)*
- 
- Dyspepsi och reflux (2007), nr 185
- 
- Nyttan av att berika mjöl med folsyra i syfte att minska risken för neuralrörsdefekter (2007), nr 183
- 
- Metoder för att främja fysisk aktivitet (2006), nr 181
- 
- Måttligt förhöjt blodtryck (Uppdatering av SBU-rapport 170/1 från 2004) (2007), nr 170/1U
-

---

Metoder för tidig fosterdiagnostik (2006), nr 182

---

Hjärnskakning – övervakning på sjukhus eller datortomografi och hemgång? (Uppdatering av rapport från 2000) (2006), nr 180

---

Metoder för behandling av långvarig smärta (2006), nr 177/1+2

---

Riskbedömningar inom psykiatri – kan våld i samhället förutsägas? (2005), nr 175

---

Bettavikelser och tandreglering i ett hälsoperspektiv (2005), nr 176

---

Behandling av ångestsyndrom (2005), nr 171/1+2

---

### **Vita rapporter (1999–2012)**

---

Volym och resultat (2011), nr 205

---

Behandling med vitamin D och kalcium (2006), nr 178

---

Volym och kvalitet (2005), nr 179

---

ADHD hos flickor (2005), nr 174

---

Evidensbaserad äldrevård (2003), nr 163

---

Rökning och ohälsa i munnen (2002), nr 157

---

Placebo (2000), Ges ut av Liber, nr 154

---

Behov av utvärdering i tandvården (2000), nr 152

---

Sveriges ekonomi och sjukvårdens III, Konferensrapport (2000), nr 149

---

Alert – Nya medicinska metoder (2000), nr 148

---

Barn födda efter konstgjord befruktning (IVF) (2000), nr 147

---

Patient-läkarrelationen (1999), Ges ut av Natur och Kultur, nr 144

---

Evidensbaserad omvårdnad: Behandling av patienter med schizofreni (1999), nr 4

---

Evidensbaserad omvårdnad: Patienter med depressionssjukdomar (1999), nr 3

---

Evidensbaserad omvårdnad: Patienter med måttligt förhöjt blodtryck (1998), nr 2

---

Evidensbaserad omvårdnad: Strålbehandling av patienter med cancer (1998), nr 1

---

Evidensbaserad sjukgymnastik: Patienter med ländryggsbesvär (1999), nr 102

---

Evidensbaserad sjukgymnastik: Patienter med nackbesvär (1999), nr 101

---

### **SBU Alert-rapporter (2006–2012)**

---

Vakuumassisterad sårbehandling, nr 2011-09

---

Perifert inlagd central venkateter (PICC), nr 2011-08

---

Analys av foster-DNA i kvinnans blod: icke-invasiv fosterdiagnostik för blodgrupps- eller könsbestämning, nr 2011-07

---

Molekylärdiagnostiska test för män med ökad sannolikhet för prostatacancer, nr 2011-06

---

Datorassisterad granskning inom mammografiscreening (CAD), nr 2011-05

---

Dabigatran för att förebygga stroke vid förmaksflimmer, nr 2011-04

---

Datortomografi för misstänkt kranskärlssjukdom, nr 2011-03

---

Perkutan vertebroplastik och ballongkyfoplastik vid ryggsmärta pga kotkompression som orsakats av osteoporos, nr 2011-02

---

Lasermedierad värmebehandling av levermetastaser, nr 2011-01

---

Kateterburen ablationsbehandling vid förmaksflimmer, nr 2010-06
Urinprov vid diagnostik av klamydia hos kvinnor, nr 2010-05
Hem blodtrycksmätning, nr 2010-04
Tidig och riktad ultraljudsundersökning efter fysiskt trauma, nr 2010-03
Silverförband vid behandling av kroniska sår, nr 2010-02
Cilostazol vid behandling av fönstertittarsjuka (claudicatio intermittens), nr 2010-01
Datorstödd träning för barn med ADHD, nr 2009-05
Dopaminerga medel vid restless legs syndrome, nr 2009-04
Laser vid avlägsnande av karies, nr 2009-03
Leukocytaferes vid inflammatorisk tarmsjukdom, främst ulcerös kolit, nr 2009-02
Kylbehandling av nyfödda barn som drabbats av allvarlig syrebrist under förlossningen, nr 2009-01
Mätning av kväveoxid i utandningsluft vid astma, nr 2008-05
Screening för bukaortaaneurysm, nr 2008-04
Ranibizumab för behandling av åldersförändringar i näthinnans gula fläck, nr 2008-03
EEG-baserad anestesidjupsmonitorering, nr 2008-02
Allmän barnvaccination mot HPV 16 och 18 i syfte att förebygga livmoderhalscancer, nr 2008-01
Självtestning och egenvård vid användning av blodproppsförebyggande läkemedel, nr 2007-05
Operation vid brytningsfel i ögat, nr 2007-04
Datorbaserad kognitiv beteendeterapi vid ångestsyndrom eller depression, nr 2007-03
Perkutan vertebroplastik vid svår ryggsmärta pga kotkompression, nr 2007-02
Pacemaker för synkronisering av hjärtkamrarnas rytm (CRT) vid kronisk hjärtsvikt, nr 2007-01
Nya immunmodulerande läkemedel vid måttlig till svår psoriasis, nr 2006-07
Implanterbar defibrillator, nr 2006-06
Natriuretiska peptider som hjälp vid diagnostik av hjärtsvikt, nr 2006-05
ST-analys i kombination med CTG (STAN) för fosterövervakning under förlossning, nr 2006-04
Individanpassad vård av underburna barn – NIDCAP, nr 2006-03
Sänkning av kroppstemperaturen efter hävt hjärtstopp, nr 2006-02
Bilaterala cochleaimplantat (CI) hos barn, nr 2006-01

## **Rapporter på engelska (2001–2012)**

---

Treatment of Hemophilia A and B and von Willebrand Disease (2011), no 208E
Medical and Psychological Methods for Preventing Sexual Offences Against Children (2011), no 207E
Dementia (2008), three volumes, no 172E
Obstructive Sleep Apnoea Syndrome (2007), no 184E
Interventions to Prevent Obesity (2005), no 173E
Moderately Elevated Blood Pressure (2004), Volume 2, no 170/2
Sickness Absence – Causes, Consequences, and Physicians' Sickness Certification Practice, Scandinavian Journal of Public Health, Suppl 63 (2004), no 167/suppl
Radiotherapy for Cancer (2003), Volume 2, no 162/2
Treating and Preventing Obesity (2003), no 160E
Treating Alcohol and Drug Abuse (2003), no 156E
Evidence Based Nursing: Caring for Persons with Schizophrenia (1999/2001), no 4E
Chemotherapy for Cancer (2001), Volume 2, no 155/2

---