

Diagnostik och uppföljning av förstämningssyndrom

En systematisk litteraturöversikt

Oktober 2012



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

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 klarlägga 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).

Diagnos	Maniska symtom	Depressiva symtom
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.

Benämning	Används för	Generellt eller specifikt	Bedömare
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.

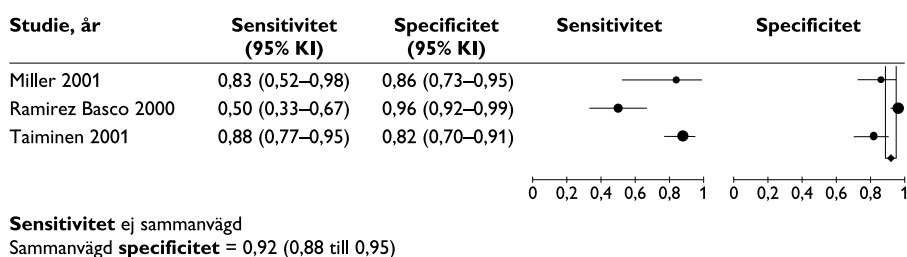
Liksom psykiatrerna kunde distriktsläkarna i högre grad utesluta dem som var friska, och träffsäkerheten uppgick till 81 procent (begränsat vetenskapligt underlag ⊕⊕○○).

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.

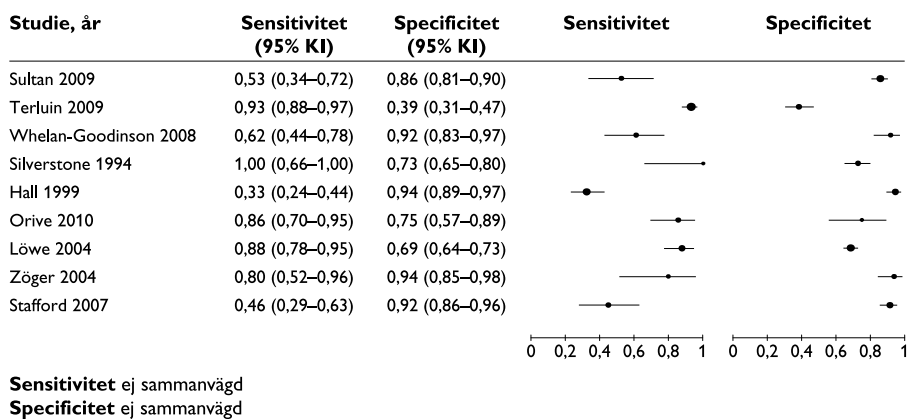
Två versioner av GDS som är avsedda för äldre har också tillräcklig känslighet. Studierna på GDS hade oftast exkluderat deltagare med demens, vilket gör att vi inte kan uttala oss om värdet av formuläret för personer med kognitiv funktionsnedsättning.

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.

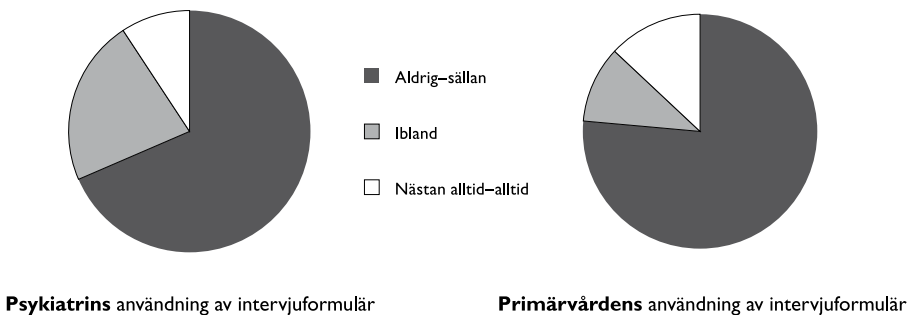
Tabell 6 Översättning och tillgänglighet för formulär med vetenskapligt stöd för tillräcklig sensitivitet och specificitet.

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.

Formulär	Syfte
För barn	
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
För vuxna	
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
För äldre	
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