

SBU UTVÄRDERAR • RAPPORT 276/2018

Patient Health Questionnaire-9 (PHQ-9) som stöd för diagnostik och bedömning av svårighetsgrad av depression

En systematisk översikt

Partiell uppdatering av SBU-rapporten

Diagnostik och uppföljning av förstämningssyndrom, nr 212

Rapportserie Denna rapport hör till serien SBU Utvärderar. En allsidig vetenskaplig utvärdering som innefattar systematisk översikt, ekonomiska/hälsoekonomiska analyser samt genomgång av sociala och etiska aspekter. Ämnessakkunniga deltar i arbetet med stöd av medarbetare på SBU, patienter/brukare. Rapporten granskas av oberoende experter. Graden av vetenskaplig tillförlitlighet i de sammanvägda resultaten bedöms på ett systematiskt sätt och rapportens slutsatser godkänns av SBU:s nämnd.

ISSN 1400-1403

Innehållsdeklaration

✓ Utvärdering av ny/etablerad metod	✓ Framtagen i samarbete med sakkunniga patienter/brukare medverkat
✓ Systematisk litteratursökning	Etiska aspekter
✓ Relevansgranskning	Ekonomiska aspekter
✓ Kvalitetsgranskning	Sociala aspekter
✓ Sammanvägning av resultat	✓ Granskad av SBU:s kvalitets- och prioriteringsgrupp
✓ Evidensgradering gjord av SBU	✓ Godkänd av SBU:s nämnd
Evidensgradering gjord externt	
Baseras på en systematisk litteraturoversikt	
Konsensusprocess	

Nyckelord Depression, Diagnostik, Uppföljning, Patientskattning, PHQ-9, 'Patient health Questionnaire 9', Diagnostic Accuracy, Diagnostisk noggrannhet

Utgiven Februari 2018

Giltighetstid 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

Beställ Denna rapport (nr 276) kan beställas från Strömberg distribution. Telefon: 08-779 96 85 • Fax: 08-779 96 10 • E-post: sbu@strd.se

Produktion Grafisk produktion av Anna Edling, SBU. Tryckt av Elanders Sverige AB, Mölnlycke, 2018. Omslagsfoto: Shutterstock

Diarienummer SBU2017/59

Citera denna rapport SBU. Patient Health Questionnaire-9 (PHQ-9) som stöd för diagnostik och bedömning av svårighetsgrad av depression. En systematisk översikt. Stockholm: Statens beredning för medicinsk och social utvärdering (SBU); 2018. SBU-rapport nr 276. ISBN 978-91-88437-18-1.

Innehåll

Sammanfattning och slutsatser	5
1 Introduktion	11
2 Bakgrund	13
3 Metod för den systematiska översikten	17
Frågor	17
Inklusionskriterier	18
Litteratursökning	18
Urval av studier	18
Bedömning av enskilda studier	18
Sammanvägning av resultat	19
Det vetenskapliga underlagets styrka	19
4 Resultat	21
Sensitivitet och specificitet	21
— Primärstudier	21
— Systematiska översikter	26
— Bedömning av vetenskapligt stöd	27
Svårighetsgrad och känslighet för förändring	28
5 Diskussion	29
6 Överväganden för praktik och forskning	31
Användning av PHQ-9	31
Kunskapsluckor	32
7 Projektgrupp, externa granskare, råd och nämnd	33
Projektgrupp	33
— Sakkunniga	33
— SBU	33
Externa granskare	34
Bindningar och jäv	34
SBU:s vetenskapliga råd – Eira	34
SBU:s nämnd	35
8 Ordförklaringar och förkortningar	37
9 Referenser	41

Bilaga 1 Sökstrategier	tillgänglig på www.sbu.se/276
Bilaga 2 Exkluderade studier	45
Diagnostisk tillförlitlighet	45
Svårighetsgrad och monitorering	47
Bilaga 3 Studier med hög risk för systematiska fel (bias)	49
Bilaga 4 Granskningsmallar	tillgänglig på www.sbu.se/276

Sammanfattning och slutsatser

Slutsatser

Strukturerade frågeformulär kan användas som ett stöd i den diagnostiska processen och uppföljningen av egentlig depression (eng. major depression). Ett sådant formulär är Patient Health Questionnaire (PHQ-9) med nio frågor där patienterna skattar sina symtom på depression. PHQ-9 utvärderades i SBU:s rapport Diagnostik och uppföljning av förstämningssyndrom, som publicerades år 2012. Sedan dess har flera nya studier tillkommit. Dessutom finns ett intresse för att använda formuläret i Sverige. Syftet med översikten var att göra en aktuell utvärdering av PHQ-9.

Denna systematiska litteraturoversikt ledde fram till följande slutsatser:

- ▶ I primärvården kan PHQ-9 vara till hjälp i handläggning av misstänkt depression. PHQ-9 har begränsat vetenskapligt stöd för god känslighet och träffsäkerhet när det gäller screening av depression i primärvård. Det finns också ett begränsat vetenskapligt stöd för att PHQ-9 är känsligt för förändring av sjukdomstillståndet. Formuläret kan därför även användas för att följa effekter av behandling. Däremot går det inte att bedöma om PHQ-9 är tillförlitligt för screening av depression inom kroppslig eller psykiatrisk specialistvård eftersom det finns för få studier.

- Ett viktigt användningsområde för självskattningsformulär är som stöd för bedömning av svårighetsgrad. Liksom för övriga självskattningsformulär är det oklart om PHQ-9 kan klassificera svårighetsgrad av depression på ett tillförlitligt sätt eftersom det vetenskapliga underlaget består av några få studier med stora metodbrister. Här behövs därför mera forskning.

Denna utvärdering kommer därmed fram till ett delvis annat resultat än den tidigare SBU-rapporten vad gäller känslighet och träffsäkerhet för PHQ-9.

Bakgrund

Bedömningsformulär kan vara ett stöd i handläggning av patienter med misstänkt depression

Depression är vanligt förekommande och i den allmänna befolkningen uppskattas förekomsten under 12 månader till cirka 7 procent. De flesta patienter handläggs i primärvården där prevalensen är högre, i genomsnitt 12 procent. Det är vanligt med andra samtidiga sjukdomar, både kroppsliga och psykiska. Depression är också förenat med ökad risk för suicid och förtida död.

Diagnosen depression baseras på i vilken utsträckning patienten uppfyller kriterier i klassifikationssystemen Diagnostic and Statistical Manual of Mental Disorders, DSM, eller International Statistical Classification of Diseases and Related Health Problems, ICD. Diagnosen sätts efter samtal med patienten i kombination med utredningar för att utesluta kroppsliga sjukdomar. Diagnosen kan vara svår att ställa eftersom symtomen ofta kan vara gemensamma med andra tillstånd. Systematiska översikter har visat att läkare inom primärvård och öppenvård psykiatri missar diagnosen vid ett första besök för cirka hälften av patienterna. Svårighetsgraden definieras i DSM-systemet som lindrig, måttlig eller svår och grundas på dels antalet symtom och hur mycket de påverkar patientens funktion, dels på hur länge de har pågått.

Det finns bedömningsformulär som kan användas som komplement till samtalet. De kan användas för screening av patienter med symtom som kan tyda på depression, för att bedöma svårighetsgraden av tillståndet och för att följa effekter av en behandling. Patient Health Questionnaire, PHQ-9, är ett sådant formulär och är utvecklat för att användas både för diagnostik och uppföljning. PHQ-9 innehåller nio frågor om symtom. Patienten skattar hur ofta de har haft de olika symtomen under de senaste två veckorna.

Utvärdering av egenskaper hos PHQ-9

Vid utvärdering jämförs bedömningsformuläret (indextest, här PHQ-9) med ett referenstest, som representerar bästa möjliga metod. Kriterier i DSM används vanligen som referenstest. Diagnostisk tillförlitlighet är ett mått på hur väl ett test kan skilja mellan sjuk och frisk, där gränsen ofta kallas för tröskelvärde. Studier om diagnostisk tillförlitlighet genomförs oftast så att studiedeltagarna genomgår båda testerna med så kort tid emellan som möjligt. Indextestets

känslighet (sensitivitet) anger hur stor andel av dem som har tillståndet ifråga som identifieras. Träffsäkerheten (specificitet) anger hur stor andel av dem som inte har tillståndet som identifieras av testet.

Ett formulär som ska användas för att mäta effekter av behandling måste kunna detektera förändringar i svårighetsgrad över tid på ett tillförlitligt sätt (eng. responsiveness). Förändringen ställs i relation till förändringen hos referenstestet.

Metod för den systematiska översikten

Den systematiska litteraturoversikten genomfördes med den metodik som beskrivs i SBU:s Metodbok [23]. Översikten omfattade tre frågor:

- Vilken diagnostisk tillförlitlighet har PHQ-9 jämfört med referenstest?
- Kan PHQ-9 klassificera svårighetsgrad av depression på ett tillförlitligt sätt?
- Är PHQ-9 känsligt för förändring av svårighetsgrad av depression?

Översikten syftade till att uppdatera SBU-rapporten från år 2012. Studiedeltagarna skulle vara mellan 18 och 65 år och vårdsökande. Studierna skulle vara genomförda inom sjukvården i Europa, Nordamerika, Australien eller Nya Zeeland. För diagnostisk tillförlitlighet skulle tiden mellan PHQ-9 och referenstestet, vanligen en strukturerad intervju baserade på DSM-kriterierna, vara högst en vecka. Litteratur söktes i databaserna PubMed, PsycInfo och Embase i december år 2016 kompletterat med litteratur från referenslistor.

Studier granskades med avseende på risken för systematiska fel (bias) och risken att resultaten inte är tillämpliga (eng. applicability) med hjälp av checklistan QUADAS 2. Risken för bias klassificerades som hög eller acceptabel. Endast studier med acceptabel risk för bias användes i analyserna, oavsett bedömningen av tillämplighet. Systematiska översikter granskades med bedömningsformuläret AMSTAR.

Tillförlitligheten för sammanvägda resultat från analyserna, det vill säga känslighet, träffsäkerhet, svårighetsgrad och känslighet för förändring, bedömdes med stöd av GRADE. Det vetenskapliga underlaget analyserades med avseende på övergripande risk för bias, i vilken grad studiernas resultat överensstämde med varandra ("inconsistency"), hur stor osäkerheten i det sammanvägda resultatet var (konfidensområdets storlek, "imprecision"), risk för problem med tillämplighet (överförbarhet, "indirectness") samt risk för snedvridning av resultatet på grund av att studier med negativa resultat inte publicerats ("publicationsbias").

Resultat

Diagnostisk tillförlitlighet

Utvärderingen av diagnostisk tillförlitlighet delades upp på primärvård respektive kroppslig och psykiatrisk specialistvård. Känslighet och träffsäkerhet för PHQ-9 när formuläret används i primärvården har utvärderats i en systematisk översikt av Mitchell och medarbetare [20]. Översikten kvalitetsgranskades inom projektet och bedömdes vara välgjord. Eftersom inga nya studier från primärvården tillkommit kunde översiktens resultat användas. Författarna tillämpade bredare inklusionskriterier än SBU men bedömningen blev att avvikelserna kunde hanteras med hjälp av GRADE. Översikten inkluderade 17 primärvårdsstudier med PHQ-9 vid tröskelvärde 10, med närmare 10 000 deltagare. Studierna hade olika upplägg men många av dem var så kallade väntrumstudier. Metaanalysen gav en känslighet på 81,3 procent (95 % KI, 71,6 till 89,3 %) och träffsäkerhet 85,3 procent (95 % KI, 81,0 till 89,1 %). Det vetenskapliga stödet bedömdes vara begränsat ($\oplus\oplus\circ\circ$). Det berodde för det första på att den utökade tiden mellan PHQ-9 och referenstest, från en till två veckor, leder till en ökad risk för att patienternas tillstånd har hunnit förändras mellan mätpunkterna. Detta ökar osäkerheten i resultatet. För det andra rapporterade författarna att studierna var heterogena.

Det finns också studier som undersökt om PHQ-9 är tillförlitligt för specifika patientgrupper inom kroppslig eller psykiatrisk specialistsjukvård. Det finns dock alltför få studier för att beräkna känslighet och träffsäkerhet för depression när patienten samtidigt har ett annat problem som till exempel hjärtsjukdomar. En sammanvägning av resultat från när formuläret använts inom specialistsjukvård, oavsett om patienterna hade någon samsjuklighet, visar så stor heterogenitet att resultatet inte är relevant.

Svårighetsgrad och känslighet för förändring

Tre studier undersökte om PHQ-9 är tillförlitligt för att bedöma svårighetsgrad av depression. Samtliga exkluderades eftersom de inte hade använt en adekvat referensstandard. Det går därmed inte att bedöma om PHQ-9 kan klassificera lindrig, måttlig och svår depression på ett tillförlitligt sätt (Tabell 1).

Två andra studier undersökte om PHQ-9 är känsligt för förändring. Den ena studien hade hög risk för bias eftersom både PHQ-9 och referenstest genomförts av samma forskare. I den andra, en randomiserad behandlingsstudie, visade PHQ-9 likartad känslighet för förändring som referenstestet efter sex månader. Det vetenskapliga stödet för resultatet är begränsat. Studien mätte känslighet på patienter som kunde ha antingen depression eller kronisk, låggradig depression (dystymi). Dessutom finns risk för intressekonflikter eftersom en av författarna utvecklat PHQ-9.

Utfallsmått	Antal studier Antal deltagare	Estimat	Vetenskapligt stöd
Sensitivitet vid tröskelvärde 10, primärvården	17 (från systematisk översikt) 9 960	81 % (95 % KI, 72 till 89 %)	Begränsat ¹
Specificitet vid tröskelvärde 10, primärvården	17 (från systematisk översikt) 9 960	85 % (95 % KI, 81 till 89 %)	Begränsat ¹
Förmåga att skilja mellan svårighetsgrader av depression, primärvård	0	Underlag saknas	Otillräckligt
Förmåga att mäta förändring (effektstorlek), efter sex månader, primärvård	1 434	Effektstorlek av samma storleksordning som för referensstandard, SCID-I	Begränsat ²

¹ Avdrag för bristande överförbarhet, -1, eftersom tiden mellan testerna kunde vara upp till två veckor samt för bristande samstämmighet, -1.

² Avdrag för bristande överförbarhet, -1, eftersom utfallet mättes för två diagnoser, egentlig depression och kronisk låggradig depression, sammanslaget, samt -1 för annan risk för bias i form av intressekonflikter eftersom en av upphovsmännen till PHQ-9 är delförfattare till denna enda studie.

Tabell 1
Sammanfattning av bedömningarna av vetenskapligt stöd för PHQ-9.

Vad betyder resultaten?

Denna utvärdering av PHQ-9 avser såväl diagnostisk tillförlitlighet vid tröskelvärde 10 som förmåga att klassificera svårighetsgrad och mäta förändring över tid. Den diagnostiska tillförlitligheten beror sannolikt av sammanhanget som formuläret används i. Man kan därför inte förvänta sig att det ska finnas ett medelvärde för sensitivitet respektive specificitet som är giltigt för alla populationer eller vårdnivåer. Det innebär att metaanalyser bör ta hänsyn till sammanhanget för att ge resultat som är kliniskt relevanta.

Användning av PHQ-9

Det vore en fördel om samma formulär kunde användas för såväl stöd i att bekräfta eller avfärda en misstanke om depression, det vill säga screening, som för bedömning av svårighetsgrad och för att följa behandlingseffekter. En tolkning av vår systematiska översikt är att PHQ-9 förmodligen inte uppfyller detta.

Vad gäller screening går det bara att uttala sig om primärvården. Känsligheten är där cirka 81 procent och träffsäkerheten cirka 85 procent. Mera praktiskt användbara mått är positiva (PPV) respektive negativa (NPV) prediktionsvärden. PPV anger sannolikheten för att en individ med ett positivt utslag på testet verkligen har sjukdomen. Omvänt anger NPV sannolikheten för att en individ med ett negativt utslag på testet inte har sjukdomen. Prediktionsvärdena beror av prevalensen och ju högre värdena är, desto mer information

ger de. Med en prevalens på 12 procent i primärvården blir PPV för PHQ-9 42 procent och NPV 97 procent. PHQ-9 fungerar därför bättre för att utesluta depression (högt NPV) än för att stödja en depressionsdiagnos.

Det går inte att bedöma om PHQ-9 kan skilja mellan svårighetsgrader av depression och studier visar att det finns risk för att PHQ-9 överskattar svårighetsgraden av depression. Däremot finns det ett begränsat stöd för att PHQ-9 fungerar tillfredsställande för att mäta förändring och formuläret skulle därmed kunna användas för att följa effekter av behandling.

Det kan noteras att kunskapsläget för andra bedömningsformulär som används i Sverige är otillräckligt, såväl för att stödja diagnostik av depression hos vuxna som att bedöma svårighetsgrad av depression och följa behandlingseffekter.

Kunskapsluckor

Som framgått finns det brister i kunskapen om tillförlitligheten av PHQ-9. Eftersom ett viktigt användningsområde för PHQ-9 är bedömning av svårighetsgrad är det angeläget att utvärdera formuläret med en adekvat referensstandard som till exempel SCID-I.

1 Introduktion

Patient Health Questionnaire-9 (PHQ-9) är ett formulär där patienterna skattar symtom på depression. PHQ-9 används dels som stöd för att bekräfta eller avfärda diagnosen depression och dels för att följa effekter av behandling.

SBU:s rapport Diagnostik och uppföljning av förstämningssyndrom, som publicerades år 2012, granskade det vetenskapliga underlaget för tillförlitligheten för PHQ-9 [1]. SBU fann att formulärets känslighet för depression varierade kraftigt mellan studierna och bedömde att det fanns ett måttligt vetenskapligt stöd för att PHQ-9 inte har någon väldefinierad sensitivitet. SBU fann vidare att det fanns för få studier för att bedöma om PHQ-9 kan användas för att uppskatta svårighetsgrad av depression och följa effekter av behandling.

De senaste åren har ett stort antal ytterligare studier om PHQ-9 publicerats. Nyare systematiska översikter drar slutsatsen att PHQ-9 har god diagnostisk tillförlitlighet [2,3]. Samtidigt rekommenderas PHQ-9 i flera internationella riktlinjer, till exempel de amerikanska [4] och det finns ett intresse i Sverige för att använda formuläret i högre utsträckning.

Syftet med projektet var att uppdatera SBU-rapporten med avseende på det vetenskapliga underlaget för PHQ-9.

2 Bakgrund

Egentlig depression (eng. major depression) är vanligt förekommande och 12-månaders prevalensen uppskattas till runt 7 procent i den allmänna befolkningen [5,6]. De flesta patienter återfinns i primärvården. I en studie från sex europeiska länder var den genomsnittliga förekomsten av depression 12 procent [7]. Det finns inga biologiska markörer för depression utan diagnosen baseras på i vilken utsträckning patienten uppfyller kriterier i klassifikationssystemen Diagnostic and Statistical Manual of Mental Disorders, DSM (f.n. DSM-5) [8] eller International Statistical Classification of Diseases and Related Health Problems - Tenth Revision, ICD-10 [9]. Samsjuklighet är vanlig. Hjärtsjukdomar och diabetes är exempel på kroniska sjukdomar med förhöjd risk för depression [10–12]. Depression förekommer ofta samtidigt med andra psykiatriska tillstånd, till exempel ett eller flera ångestsyndrom [13–15] och är förenat med ökad risk för suicid [16] och förtida död oavsett orsak [16,17].

Inför ställningstagande om lämplig behandling behövs information om dels diagnosen och dels svårighetsgraden av problemen. Diagnosen sätts efter samtal med patienten i kombination med utredningar för att utesluta kroppsliga sjukdomar. Diagnosen kan vara svår att ställa eftersom symtomen ofta kan vara gemensamma med andra tillstånd. SBU:s rapport om Diagnostik och uppföljning av förstämningssyndrom kom fram till att läkare inom primärvård och öppenvård psykiatri missade diagnosen vid ett första besök för cirka hälften av patienterna medan patienter inom slutenvård psykiatri fick korrekt depressionsdiagnos i högre utsträckning [1]. Det kan ta lång tid att ställa diagnosen och tider på upp till ett halvår har rapporterats [18]. Svårighetsgraden definieras i DSM-systemet som lindrig, måttlig eller svår och grundas på dels antalet symtom och hur mycket de påverkar patientens funktion och dels hur länge de har pågått [8].

Under de senaste decennierna har forskare utvecklat bedömningsformulär som kan användas som komplement till samtalet. Tabell 2.1 visar fyra huvudgrupper av bedömningsformulär. Intervjuer kan vara strukturerade eller semi-strukturerade. De innehåller moduler som täcker flera psykiatriska tillstånd och kan användas som stöd för att bedöma depression. De depressionsspecifika formulärens syftar till att bekräfta eller avfärda diagnosen (eng. case-finding respektive screening) eller till att bedöma svårighetsgrad. Formulär för att mäta depressionsdjup kan även användas regelbundet för att följa effekter av en behandling om de är känsliga för förändring.

Tabell 2.1
Projektets uppdelning
av psykiatriska
bedömningsformulär.

Typ av formulär	Beskrivning	Fylls i av	Exempel
Intervju, strukturerad	Formulär som består av olika moduler som var och en täcker en diagnos i DSM- eller ICD-klassifikationerna. Depression utgör en modul. Patienten kan svara ja och nej. Används som stöd för diagnos	Bedömaren	MINI, PRIME-MD, CDIS, CIS-R
Intervju, semi-strukturerad	Formulär som består av olika moduler som var och en täcker en diagnos i DSM- eller ICD-klassifikationerna. Depression utgör en modul om förstämningssyndrom. Intervjun består av inledande och fördjupande frågor och ställer högre krav på kompetens hos bedömaren än den strukturerade. Används som stöd för diagnos.	Bedömaren	SCID-I
Formulär för skattning av symtom	Specifikt för depression. Frågorna täcker de olika kriterierna i diagnosen. Syftet är att bekräfta eller avfärda en misstanke om depression	Oftast patienten	HADS, PHQ-9,
Formulär för bedömning av svårighetsgrad	Specifikt för depression. Frågorna avser svårighetsgrad och varaktighet av problem.	Oftast patienten	MADRS-S, BDI-II, PHQ-9
Formulär för att följa behandlingseffekter	Samma formulär som för bedömning av svårighetsgrad	Oftast patienten	MADRS-S, PHQ-9

BDI-II = Beck Depression Inventory II; **CDIS** = Computerized Diagnostic Interview Schedule; **CIS-R** = Clinical Interview Schedule-Revised; **HADS** = Hospital Anxiety and Depression Scale; **MADRS-S** = Montgomery Asberg Depression Rating Scale – self rating; **MINI** = Mini International Neuropsychiatric Interview; **PHQ-9** = Patient Health Questionnaire-9; **PRIME-MD** = Primary Care Evaluation of Mental Disorders; **SCID-I** = Structured Clinical Interview for DSM-IV Axis I Disorders

Patient Health Questionnaire, PHQ, har utvecklats som en patientversion av den strukturerade intervjun PRIME-MD [19]. PHQ består av nio frågor (PHQ-9) men finns även som en kortversion, PHQ-2, med endast de två första frågorna i formuläret. Patienten skattar hur ofta de har haft olika symtom under de senaste två veckorna på en fyrgradig Likertskala (0=inte alls; 3=nästan varenda dag). Som stöd för diagnostik kan PHQ-9 användas på två sätt. Resultatet tolkas antingen med hjälp av en algoritm eller genom att räkna samman poäng, där maximalt antal poäng är 27. Utvecklarna av PHQ-9 definierade 10 poäng som tröskelvärde för depression. Många studier av PHQ-9 har syftat till att undersöka om andra tröskelvärden kan fungera bättre men 10 är fortfarande den etablerade gränsen [20,21]. För svårighetsgrad har utvecklarna av PHQ-9

satt gränserna så att de ska vara praktiska att använda. Lindrig depression motsvarar 5–9 poäng, måttlig depression 10–14, måttligt svår depression 15–19 och svår depression 20–27 poäng.

Diagnostisk tillförlitlighet är ett mått på hur väl ett test kan skilja mellan sjuk och frisk. De vanliga måtten för att bedöma diagnostisk tillförlitlighet är känslighet (sensitivitet) och träffsäkerhet (specificitet) där det test som utvärderas, indextestet, jämförs med ett referenstest. Referenstestet ska representera bästa möjliga metod att ställa diagnosen och förutsätts alltid klassificera tillståndet korrekt. Indextestet däremot är inte perfekt utan kan ge såväl falskt positiva som falskt negativa svar. För bedömningsformulären används strukturerade eller semistrukturerade intervjuer baserade på DSM-kriterierna som referenstest. Det bör noteras att det introducerar en viss osäkerhet beroende på subjektiva tolkningar av intervju svaren.

Studier om diagnostisk tillförlitlighet genomförs oftast så att studiedeltagarna genomgår båda testerna med så kort tid emellan som möjligt. Antalet deltagare som anses ha eller inte ha tillståndet ifråga, enligt referenstestet respektive indextestet, läggs in i en så kallad fyrfaltstabell och därifrån beräknas sensitivitet och specificitet (Tabell 2.2).

	Depression enligt referensstandard	Inte depression enligt referensstandard
Depression enligt PHQ-9	TP (sant positiv)	FP (falskt positiv)
Inte depression enligt PHQ-9	FN (falskt negativ)	TN (sant negativ)

Tabell 2.2
Utfallsmått för utvärdering av diagnostisk tillförlitlighet.

$$\text{Sensitivitet} = \frac{TP}{TP+FN} ; \text{Specificitet} = \frac{TN}{FP+TN}$$

Även för bedömning av hur väl formulär kan skatta svårighetsgrad relateras resultatet av testet mot ett referenstest. DSM-systemets kriterier för svårighetsgrad kan användas som referenstest.

Ett formulär som ska användas för att mäta effekter av behandling måste kunna detektera förändringar i svårighetsgrad över tid på ett tillförlitligt sätt (eng. sensitivity to change, eller responsiveness). Ett vanligt mått är standardiserad "response mean", SRM [22]. Den beräknas som skillnaden mellan medelvärdet av poängen vid två tidpunkter, dividerat med poängförändringens standardavvikelse. Förändringen ska ställas i relation till förändringen hos referenstestet. Även här kan DSM användas som referens.

3 Metod för den systematiska översikten

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 granskas med avseende på relevans och risk för systematiska fel (bias). Resultaten från studier som har en acceptabel risk för bias sammanvägs, till exempel i metaanalyser. Slutligen bedöms hur tillförlitligt det sammanvägda resultatet är, uttryckt som styrka på det vetenskapliga stödet ("evidensstyrka"). Hela arbetet bedrivs av läspar med två personer som först oberoende av varandra granskar studierna och därefter kommer överens om sin bedömning.

Detta avsnitt beskriver frågor, urvalskriterier och metodik för rapporten. För mera utförlig information om metodiken hänvisar vi till SBU:s Metodbok [23].

Frågor

- Vilken diagnostisk tillförlitlighet har PHQ-9 med tröskelvärde 10 jämfört med referensstandard?
- Kan PHQ-9 klassificera svårighetsgrad av depression på ett tillförlitligt sätt?
- Är PHQ-9 känsligt för förändring av svårighetsgrad av depression?

Inklusionskriterier

Översikten syftade till att uppdatera SBU-rapporten från år 2012 [1]. Studiedeltagarna skulle vara vuxna, 18–65 år, och vårdsökande. Studierna skulle vara genomförda i Europa, Nordamerika, Australien eller Nya Zeeland. Strukturerade eller semistrukturerade intervjuer som byggde på DSM-IV, DSM-5 eller ICD-10 accepterades som referensstandard för diagnostisk tillförlitlighet. Tiden mellan PHQ-9 och referensstandard skulle vara högst en vecka.

För bedömningar av om PHQ-9 kan klassificera svårighetsgrad respektive är känsligt för förändring accepterades SCID-I och Clinical Global Impression – Severity, CGI-S, som referensstandarder. Tiden mellan bedömningarna fick vara högst 24 timmar för frågan om svårighetsgrad.

Litteratursökning

Sökstrategierna togs fram av en av SBU:s informationsspecialister i samverkan med projektledare och ämnesexpert. Litteratur söktes i databaserna PubMed, PsycInfo och Embase i december år 2016. Sökningarna kompletterades med litteratur från referenslistor för originalstudier och systematiska översikter. Sökstrategierna redovisas i Bilaga 1 (www.sbu.se/276).

Urval av studier

Litteratursökningarna resulterade i artikelsammanfattningar (abstrakts). Projektledaren gallrade bort uppenbart irrelevanta abstrakts med stöd av programmet Rayyan varefter experten granskade de resterande [24]. Potentiellt relevanta artiklar beställdes i fulltext. Projektledaren och ämnesexperten bedömde därefter studiernas relevans. Studier som inte uppfyllde inklusionskriterierna exkluderades.

Bedömning av enskilda studier

Studierna granskades med avseende på risken för systematiska fel (bias) och risken att resultaten inte är tillämpliga (eng. applicability). Checklisten QUADAS 2 användes som stöd för bedömningen av originalstudier [25] (Bilaga 4, www.sbu.se/276). QUADAS 2 är uppbyggd av fyra domäner med underfrågor som täcker risker avseende val av studiedeltagare, det undersökta testet (PHQ-9), referenstestet och den diagnostiska processen (t.ex. tid mellan testerna).

Risken för bias klassificerades i första skedet som hög eller acceptabel. Endast studier med acceptabel risk för bias användes i analyserna, oavsett bedöm-

ningen av tillämplighet. Acceptabel risk i sin tur klassificerades som låg eller medelhög i samband med bedömning av det vetenskapliga underlagets styrka.

Systematiska översikter granskades med AMSTAR [26].

Sammanvägning av resultat

Sammanvägd sensitivitet och specificitet beräknades med bivariat metaanalys i enlighet med rekommendationerna i Cochrane Handbook [27]. För varje studie beräknades antalet sant positiva och negativa, respektive falskt positivt och negativa. Värdena lades in i programvaran RevMan 5.3. Eftersom RevMan 5.3 inte stödjer bivariata metaanalyser, genomfördes delar av analysen med hjälp av ett macro, METADAS, för SAS version 9.3. Metaanalyserna utfördes som random effects bivariat analys med tillhörande 95 procent konfidensområde [27].

Det vetenskapliga underlagets styrka

Styrkan på det vetenskapliga underlaget ("evidensstyrkan") anger hur tillförlitligt det sammanvägda resultatet är. Tillförlitligheten bedömdes med hjälp av GRADE [28]. Det vetenskapliga underlaget analyserades med avseende på övergripande risk för bias, i vilken grad studiernas resultat inte överensstämde med varandra ("inconsistency"), hur stor osäkerheten i det sammanvägda resultatet var (konfidensområdets storlek, "imprecision"), risk för problem med tillämplighet (överförbarhet, "indirectness") samt risk för snedvridning av resultatet på grund av att studier med negativa resultat inte publicerats ("publikationsbias").

4 Resultat

Litteratur har sökts vid tre tillfällen, dels för den ursprungliga rapporten år 2012 [1], dels för en uppdatering under år 2014 inför publicering i en engelskspråkig tidskrift [29] och dels för denna uppdatering. Vid de två tidigare tillfällena användes samma sökstrategi med breda sökningar på bedömningsformulär. Antalet abstrakts som avsåg PHQ-9 specifikt går därför inte att särredovisa.

Sökningen i databaser för denna uppdatering gjordes i december år 2016. Den resulterade i 1 249 abstrakts för originalstudier. Dessutom identifierades två systematiska litteraturöversikter som bedömdes vara av god kvalitet [20,21].

Sensitivitet och specificitet

Primärstudier

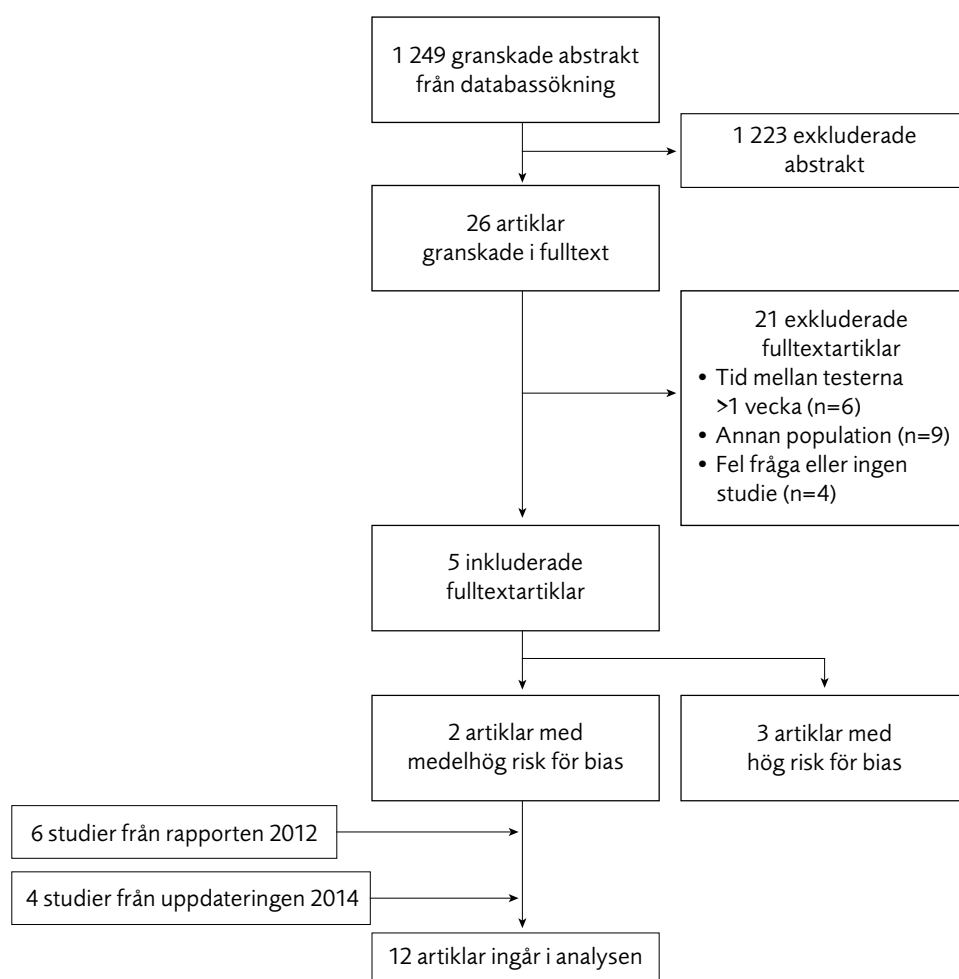
Flödesschema, ingående studier

I den tidigare SBU-rapporten bedömdes sex studier som relevanta och med acceptabel risk för systematiska fel [19,30–34]. Vid den första uppdateringen [29], identifierades ytterligare fyra relevanta studier med acceptabel risk för systematiska fel [35–38].

För frågan om sensitivitet och specificitet lästes 26 abstrakts i fulltext, och fem uppfyllde inklusionskriterierna [39–43]. Två av dem bedömdes ha acceptabel risk för bias [40,42]. Studierna redovisade inte fyrfältsdata men vi fick tillgång till dem efter kontakt med författarna. Studier som exkluderades respektive hade hög risk för bias redovisas i Bilagorna 2 och 3.

Totalt ingick 12 studier med 13 jämförelser i bedömningen. Studierna är sammanfattade i Tabell 4.1

Figur 4.1
Flödesschema för
originalstudier.



Författare År Referens Land	Population	Vårdnivå	Referenstest	Sensitivitet och specificitet
Beard et al 2016 [40] USA	n=869 Inneliggande på klinik för psykoterapi	Specialistvård, psykiatri	MINI	Sensitivitet: 93 % Specificitet: 52 %
Bombardier et al 2012 [35] USA)	n=142 Ryggmärgsskadade	Specialistvård	SCID-I	Sensitivitet: 100 % Specificitet: 80 %
Cassin et al 2013, delstudie 1 [36] Kanada	n=244 Kandidater för fetmakirurgi	Specialistvård	MINI	Sensitivitet: 75 % Specificitet: 48 %
Cassin et al 2013, delstudie 2 [36] Kanada	n=275 Kandidater för fetmakirurgi	Specialistvård	MINI	Sensitivitet: 80 % Specificitet: 46 %
Elderson et al 2011 [37] USA	n=1 024 Hjärtsjuka	Specialistvård	C-DIS	Sensitivitet: 54 % Specificitet: 90 %
Haddad et al 2013 [38] Storbritannien	n=730 Hjärtsjuka	Primärvård	CIS-R	Sensitivitet: 84 % Specificitet: 90 %
Hyphantis et al 2015 [42] Grekland	n=349 Patienter med diabetes, COPD eller reumatoid artrit	Akutintag	MINI	Sensitivitet: 45 % Specificitet: 94 %
Löwe et al 2004 [30] Tyskland	n=501 Oselekerade, väntrum	Primärvård, somatisk specialistvård	SCID-I	Sensitivitet: 90 % Specificitet: 77 %
Orive et al 2010 [31] Spanien	n=53 Oselekerade, väntrum	Sjukhus	PRIME-MD	Sensitivitet: 68 % Specificitet: 89 %
Spitzer 1999/ Kroenke 2001 [19] USA	n=580 Oselekerade, väntrum	Primärvård	SCID-I	Sensitivitet: 88 % Specificitet: 88 %
Stafford et al 2007 [32] Australien	n=193 Hjärtsjuka	Specialistvård	MINI	Sensitivitet: 54 % Specificitet: 91 %
Thekkumpurath et al 2011 [33] Storbritannien	n=782 Regionala cancerkliniker	Specialistvård	SCID-I	Sensitivitet: 82 % Specificitet: 88 %

Tabell 4.1 Beskrivning av ingående studier, sensitivitet och specificitet för PHQ-9 vid tröskelvärde 10 med strukturerad eller semistrukturerad intervju som referenstest.

Tabellen fortsätter på nästa sida

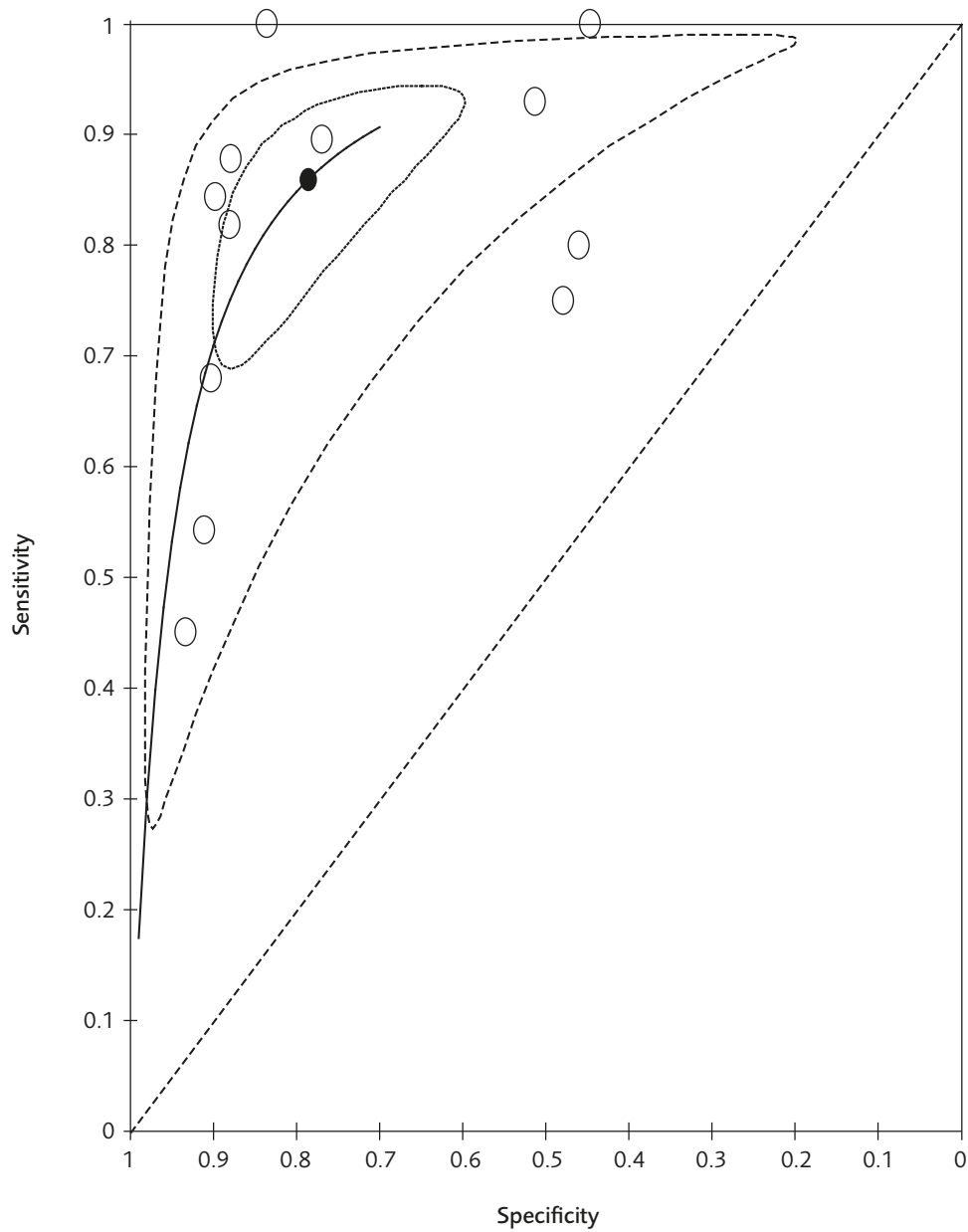
Tabell 4.1
fortsättning

Författare År Referens Land	Population	Vårdnivå	Referenstest	Sensitivitet och specificitet
Wittkampf et al 2009 [34] Nederländerna	n=664 Selekterade, hög risk för depression enligt kriterier	Primärvård	SCID-I	Sensitivitet: 100 % Specificitet: 45 %

CDIS = Computerized Diagnostic Interview Schedule; **CIS-R** = Clinical Interview Schedule-Revised; **MINI** = Mini International Neuropsychiatric Interview; **PRIME-MD** = Primary Care Evaluation of Mental Disorders; **SCID-I** = Structured Clinical Interview for DSM-IV Axis I Disorders

Studierna omfattade 6 741 deltagare och var genomförda i Nordamerika [19,35–37,40], Europa [30,31–34,38,42] och Australien [32]. En studie ägde rum på en slutenvårdsklinik inom psykiatri [40], fyra i primärvården [19,30,34,38], en på medicinskt akutintag [42] och övriga på öppen- eller slutenvård för olika kroppsliga sjukdomar [30–33,35–37]. Som referenstest användes SCID-I [19,30,33–35], MINI [32,36,40,42], PRIME-MD [31], C-DIS [37] och CIS-R [38].

Fyrfältsdata från studierna lades in i en metaanalys, se Figur 4.2. Som framgår av figuren var spridningen mellan studierna mycket stor. Bedömningen blev att det inte var meningsfullt att väga samman studiernas resultat, men figuren visas som en illustration.



Figur 4.2
 Bivariat metaanalys. Sensitivitet och specificitet för PHQ-9 vid tröskelvärde 10. Den svarta punkten anger medelvärde för sensitivitet och specificitet. De vita punkterna representerar ingående studier. Helledragen linje visar korrelationen mellan sensitivitet och specificitet. 95 % konfidensintervall ligger inom det område som avgränsas med heldragen linje. 95 % prediktionsregion, som avgränsas med streckad linje, ger en uppfattning om var framtida studier kommer att ligga.

Subgruppsanalyser gav inte heller vägledning om hur PHQ-9 fungerar för mer avgränsade grupper eftersom det fanns för få studier och de var heterogena. Underlaget innehöll tre studier med patienter från primärvården [19,30,34]. Sensitiviteten var hög, mellan 88 och 100 procent, men specificiteten varierade från 45 procent till 88 procent. Tre andra inkluderade studier rörde patienter med hjärt- och kärlsjukdom [32,37,38]. Här sågs en hög specificitet, cirka 90 procent, i alla studierna. Sensitiviteten var över 80 procent för hjärtsjuka i en studie från primärvården [38] medan den låg på 54 procent i de två studierna från specialistvård [32,37].

Systematiska översikter

Inga systematiska översikter uppfyllde samtliga inklusionskriterier som användes i SBU-rapporten från år 2012 [1]. Däremot identifierades två nyligen publicerade systematiska översikter som hade så pass likartade kriterier att de kunde vara relevanta [20,21]. Översikten av Moriarty och medarbetare omfattade 36 studier med drygt 21 000 deltagare från hela världen [21]. Översikten undersökte sensitivitet och specificitet för tröskelvärden mellan 7 och 15. I stort sett alla inkluderade studier redovisade sensitivitet och specificitet för tröskelvärde 10. Vid detta tröskelvärde var medelvärdet på sensitiviteten 78 procent (95 % KI, 70 till 84) och specificiteten 87 procent (95 % KI, 84 till 90). Heterogeniteten var hög, $I^2=86\%$, och den enda faktorn som föreföll bidra till heterogeniteten var prevalensen depression. Författarna gjorde därför subgruppsanalyser, uppdelat på primärvård och specialistsjukvård. Resultaten tydde på att sensitiviteten var högre för primärvården (81 %) än för specialistsjukvården (sensitivitet 70 %) medan specificiteten var likartad (87 respektive 88 %). Heterogeniteten var fortfarande i samma storleksordning.

Mitchell och medarbetare fokuserade på primärvården [20]. Översikten representerar det aktuella forskningsläget eftersom inga nya studier tillkommit. De inkluderade 26 studier, varav 17, med närmare 10 000 deltagare, undersökte tröskelvärde 10. Trots avgränsningen till primärvård varierade prevalensen depression i studierna från några få procent till mer än 50 procent, vilket var en trolig orsak till en heterogenitet i samma storleksordning som för översikten av Moriarty och medarbetare. Metaanalysen visade ett medelvärde på sensitivitet 81,3 procent (95 % KI, 71,6 till 89,3) och specificitet 85,3 procent (95 % KI, 81,0 till 89,1).

Bedömning av vetenskapligt stöd

Resultaten från den egna metaanalysen samt metaanalyserna i de två systematiska översikterna som beskrivits ovan ledde till två överväganden. Det ena var att populationen spelar roll för den diagnostiska tillförlitligheten för PHQ-9. Det är därför inte meningsfullt att uppskatta ett medelvärde för sensitivitet och specificitet för ett givet tröskelvärde oavsett population. Det andra var att det kan finnas tillräckligt många studier för att bedöma diagnostisk tillförlitlighet för PHQ-9 med tröskelvärde 10 inom primärvården, förutsatt att några av SBU:s tidigare kriterier inte tillämpades.

Beslutet blev att använda den systematiska litteraturöversikten av Mitchell och medarbetare som underlag [20]. Inga nya studier hade tillkommit sedan översiktens litteratursökningar. Skillnaderna i inklusionskriterier kunde hanteras med GRADE.

Följande överväganden gjordes:

För det första tillämpade översikten ingen geografisk begränsning, vilket SBU-rapporten gjorde. Cirka hälften av studierna i översikten var genomförda i länder som Kina, Malaysia och Östafrika. Att döma av resultaten från studierna spelade det dock inte någon roll i vilka länder studierna var genomförda och något avdrag i GRADE gjordes inte.

För det andra accepterade översikten en längre tid mellan PHQ-9 och referenstest, två veckor istället för en vecka som i SBU-rapporten. Risken för att patientens tillstånd har hunnit förändras mellan mätpunkterna ökar ju längre tiden är mellan testerna, vilket ökar osäkerheten i resultatet. Detta motiverade ett avdrag, -1, för bristande överförbarhet.

För det tredje bedömde översikten visserligen studiernas risk för bias med hjälp av QUADAS 2 men samtliga relevanta studier inkluderades i metaanalyserna oavsett risk. Dock gjordes en känslighetsanalys där enbart studier av hög kvalitet inkluderades. Denna analys visade endast marginella skillnader jämfört med huvudanalysen och därför gjordes inget avdrag med avseende på risk för bias.

Det vetenskapliga stödet påverkas även av att författarna rapporterade att de ingående studierna var heterogena. Utifrån publicerade uppgifter är det svårt att avgöra i vilken utsträckning heterogeniteten ökade osäkerheten. Författarna har kontaktats vid flera tillfällen för klarläggande men de har inte svarat. Ett avdrag, -1, bedömdes vara rimligt.

Sammanfattningsvis finns det därmed ett begränsat vetenskapligt stöd för att PHQ-9 vid tröskelvärde 10 har sensitiviteten 81,3 procent (95 % KI, 71,6 till 89,3) och specificitet 85,3 procent (95 % KI, 81,0 till 89,1) i primärvårdspopulationer.

Det vetenskapliga stödet för att bedöma sensitivitet och specificitet för olika populationer inom specialistsjukvården är otillräckligt då det finns för få studier.

Svårighetsgrad och känslighet för förändring

Tre studier undersökte om PHQ-9 är tillförlitligt för att bedöma svårighetsgrad [34,40,44]. Samtliga exkluderades då de använde andra referensstandarder än SCID-I: MINI [40], Hamilton Depression Rating Scale [34] samt PRIME-MD [44]. Det går därmed inte att bedöma om PHQ-9 kan klassificera lindrig, måttlig och svår depression på ett tillförlitligt sätt (Tabell 4.2).

Två studier undersökte om PHQ-9 är känsligt för förändring [45,46]. En av dem hade hög risk för bias eftersom en forskare genomförde såväl SCID-I som PHQ-9 vid samma tillfälle och bedömningen därmed inte var blindad [45]. Den andra studien inkluderade drygt 400 äldre patienter som deltog i en randomiserad behandlingsstudie och fick aktiv behandling [46]. ”Responsiveness”, det vill säga förmågan att mäta förändring, mättes för två sammanslagna diagnoser: egentlig depression och kronisk, låggradig depression efter sex månader. PHQ-9 hade likartad effektstorlek som SCID-I (-1,3 respektive -1,5). Det vetenskapliga stödet för att PHQ-9 vid tröskelvärde 10 kan mäta förändring på ett tillförlitligt sätt är begränsat (Tabell 4.2).

Tabell 4.2
Sammanfattning
av bedömningarna
av vetenskapligt
stöd för PHQ-9.

Utfallsmått	Antal studier Antal deltagare	Estimat	Vetenskapligt stöd
Sensitivitet vid tröskelvärde 10, primärvården	17 9 960	81 % (95 % KI, 72 till 89)	Begränsat ¹
Specificitet vid tröskelvärde 10, primärvården	17 9 960	85 % (95 % KI, 81 till 89)	Begränsat ¹
Förmåga att skilja mellan svårighetsgrader av depression, primärvård	0	Underlag saknas	Otillräckligt
Förmåga att mäta förändring (effektstorlek), efter sex månader, primärvård	1 434	Effektstorlek av samma storleksordning som för referensstandard, SCID-I	Begränsat ²

KI = Konfidensintervall; **SCID-I** = Structured Clinical Interview for DSM-IV-Axis I disorders

- ¹ Avdrag för bristande överförbarhet, -1, eftersom tiden mellan testerna var högst två veckor samt för bristande samstämmighet, -1
- ² Avdrag för bristande överförbarhet, -1, eftersom utfallet mättes för två diagnoser, egentlig depression och kronisk låggradig depression (dystymi), sammanslaget, samt för risk för intressekonflikter, -1, eftersom en av upphovsmännen till PHQ-9 är delförfattare till denna enda studie.

5 Diskussion

Den tidigare SBU-rapporten kom fram till att PHQ-9 vid tröskelvärde 10 hade specificitet 84 procent men att sensitiviteten inte gick att bedöma eftersom samstämmigheten mellan de sex ingående studierna var för låg. I denna uppdatering kom vi fram till att den diagnostiska tillförlitligheten sannolikt beror av sammanhanget som formuläret används i. Man kan därför inte förvänta sig att det ska finnas ett medelvärde för sensitivitet respektive specificitet som är giltigt för till exempel alla populationer eller vårdnivåer. Det innebär att en metaanalys som inte tar hänsyn till sammanhanget kan ge resultat som inte är rimliga och därmed att resultatet av den metaanalys vi initialt gjorde (Figur 4.2), inte är användbart.

Vi identifierade en annan systematisk översikt [20] som var begränsad till primärvård och beslutade att utgå från den. Vi är medvetna om att vi därmed inte besvarat rapportens fråga, utan gjort en efteranalys. Vid tröskelvärde 10 var sensitiviteten 81,3 procent (95 % KI, 71,6 till 89,3) och specificiteten 85,3 procent (95 % KI, 81,0 till 89,1). Konfidensintervallet för sensitiviteten var brett, vilket tydde på kvarvarande heterogenitet, något som också uppmärksammades av författarna till översikten.

Det finns sannolikt många skäl bakom den kvarstående heterogeniteten. Prevalens av depression är sannolikt en bidragande faktor [20,21]. De ingående studierna redovisade varierande prevalens från några få procent upp till nära 50 procent. Man skulle möjligen få en klarare bild genom att göra subgruppsanalyser och undersöka om det finns några skillnader mellan så kallade vänt-rumsstudier och studier där patienterna sökt för psykiska problem. Formuläret i sig kan principiellt också bidra till heterogenitet. PHQ-9 har dock tidigare

visat sig fungera stabilt oavsett kön och utbildningsnivå, medan patientens ålder kan ha en viss inverkan [47]. Ytterligare en aspekt att ta i beaktande är om översättningar till andra språk än originalspråket engelska orsakar glidningar i betydelse som i sin tur kan påverka hur patienten tolkar frågorna i formuläret.

6 Överväganden för praktik och forskning

Användning av PHQ-9

Det vore en fördel om samma formulär kunde användas för såväl stöd i den diagnostiska processen som för bedömning av svårighetsgrad och för att följa behandlingseffekter. Förespråkarna av PHQ-9 gör också anspråk på att formuläret fungerar väl för samtliga dessa ändamål. Vår systematiska översikt visar dock att PHQ-9 delvis har otillräckligt vetenskapligt stöd.

Vad gäller användning inom diagnostiken är vår slutsats att det för närvarande bara går att uttala sig om primärvården, där PHQ-9 har en sensitivitet på cirka 81 procent och en specificitet på cirka 85 procent. Vilken nytta har man av det? Mera praktiskt användbara mått är positiva (PPV) respektive negativa (NPV) prediktionsvärden. PPV anger sannolikheten för att en individ med ett positivt utslag på testet verkligen har sjukdomen. Omvänt anger NPV sannolikheten för att en individ med ett negativt utslag på testet inte har sjukdomen. Prediktionsvärdena beror av prevalensen och ju högre värdena är, desto mer information ger de. Med en prevalens på 12 procent i primärvården [7] blir PPV för PHQ-9 42 procent och NPV 97 procent. PHQ-9 fungerar således bättre som stöd för att utesluta depression (högt NPV) än för att stödja en depressionsdiagnos.

Det går inte att bedöma om PHQ-9 kan skilja mellan svårighetsgrader av depression. Det är förvånansvärt att internationella riktlinjer som rekommenderat mätning av svårighetsgrad har förordat PHQ-9 med tanke på att bara tre studier har undersökt tillförlitligheten. Ingen av dem använde en adekvat referensstandard. Ett observandum är en debattartikel av Zimmerman [48].

Zimmerman noterade att prevalensen djup depression, mätt med PHQ-9 i flera studier, var orimligt hög jämfört med prevalensen mätt med andra formulär som till exempel Hamilton Depression Rating Scale. Författaren menade därför att PHQ-9 överskattade svårighetsgraden.

Ett sätt att optimera behandlingen av patienter med egentlig depression är att kontinuerligt mäta symtombördan och justera behandlingen om patienten inte förbättras (Measurement-Based Care, MBC) [49,50]. The International Consortium for Health Outcomes Measurement (ICHOM) rekommenderar PHQ-9 för att följa depression (<http://www.ichom.org/medical-conditions/depression-anxiety/>). MBC innebär att mätinstrumentet måste vara känsligt för förändring och här finns ett begränsat stöd för att PHQ-9 fungerar tillfredsställande.

Det kan noteras att kunskapsläget för andra självbedömningsformulär som används i Sverige är otillräckligt, såväl för att stödja diagnostik av depression hos vuxna som att bedöma svårighetsgrad av depression och följa behandlingseffekter [1].

Slutligen, i praktiken kan ett självskattningsformulär som PHQ-9 fylla andra funktioner i patientmötet, som till exempel att underlätta för patienten att beskriva sina symtom och för läkaren att kommunicera en misstanke om depression för patienter som inte vill acceptera en psykiatrisk diagnos [51,52].

Kunskapsluckor

Som framgått finns det stora brister i kunskapen om tillförlitligheten av PHQ-9. Andra systematiska översikter har påtalat att tröskelvärde 10 kanske inte är tillämpligt för andra populationer än primärvårdens men att det finns för litet forskning om andra tröskelvärden [20,21,53]. Eftersom ett viktigt användningsområde för PHQ-9 är bedömning av svårighetsgrad är det angeläget att utvärdera formuläret med en adekvat referensstandard, som till exempel SCID-I.

7 Projektgrupp, externa granskare, råd och nämnd

Projektgrupp

Sakkunniga

LISA EKSELIUS

professor, Institutionen för
neurovetenskap, Uppsala universitet
(var även ordförande för den ursprungliga
rapporten som publicerades 2012)

Extern konsult för metaanalys

LARS BERGLUND

docent, Uppsala universitet

SBU

AGNETA PETTERSSON

projektledare

THERESE KEDEBRING

projektadministratör

SARA FUNDELL

projektadministratör

HANNA OLOFSSON

informationsspecialist

Externa granskare

SBU anlitar externa granskare av sina rapporter. Dessa har kommit med värdefulla kommentarer, som i hög grad bidragit till att förbättra rapporten. I slutversionen av rapporten är det möjligt att SBU inte kunnat tillgodose alla ändrings- eller tilläggsförslag från de externa granskarna, bland annat därför att de inte alltid varit samstämmiga. De externa granskarna står därför inte nödvändigtvis bakom samtliga slutsatser eller andra texter i rapporten.

Externa granskare har varit:

MALIN ANDRÉ
docent, allmänläkare, Uppsala

HANS ÅGREN
professor emeritus,
Göteborgs universitet

Bindningar och jäv

Sakkunniga och granskare har i enlighet med SBU:s krav inlämnat deklARATION rörande bindningar och jäv. Dessa dokument finns tillgängliga på SBU:s kansli. SBU har bedömt att de förhållanden som redovisas där är förenliga med kraven på saklighet och opartiskhet.

SBU:s vetenskapliga råd – Eira

SBU:s vetenskapliga råd har granskat det vetenskapliga underlaget i rapporten.

KJELL ASPLUND
ordförande, professor, Stockholm

INGEMAR ENGSTRÖM
professor, psykiatri, etik,
Örebro universitet

HENRIK ANDERSHED
professor i psykologi, docent i
kriminologi, Örebro universitet

NILS FELTELIUS
docent, Läkemiddelsverket

KRISTINA BENGTTSSON BOSTRÖM
docent, Billingens vårdcentral, Skövde

YLVA NILSAGÅRD
med dr, docent, fysioterapi,
CAMTÖ, Region Örebro Län

CHRISTINA BERGH
professor, Kvinnokliniken,
SU/Sahlgrenska, Göteborg

STEN-ÅKE STENBERG
professor, social forskning,
Stockholms universitet

ANNA EHRENBORG
professor, vårdvetenskap,
Högskolan Dalarna

KATARINA STEEN CARLSSON
fil dr, hälsoekonomi, IHE Lund

SBU:s nämnd

SBU:s nämnd har fattat beslut om slutsatserna i rapporten.

KERSTIN NILSSON
ordförande, professor,
Örebro universitet

PETER ALLEBECK
huvudsekreterare, Forte

SUSANNA AXELSSON
generaldirektör, SBU

HEIKI ERKERS
förbundsordförande,
Akademikerförbundet SSR

EVA FRANZÉN
forsknings- och utvecklingschef,
Statens Institutionsstyrelse

VESNA JOVIC
verkställande direktör,
Sveriges Kommuner och Landsting

JAN-INGVAR JÖNSSON
huvudsekreterare för ämnesrådet
för medicin, Vetenskapsrådet

LARS-TORSTEN LARSSON
avdelningschef, Socialstyrelsen

LARS OHLY
ordförande, Funktionsrätt Sverige

LARS OSCARSSON
professor, Ersta Sköndal högskola

JENNY REHNMAN
avdelningschef, Socialstyrelsen

SINEVA RIBEIRO
förbundsordförande, Vårdförbundet

HEIDI STENSMYREN
ordförande, Sveriges läkarförbund

ANDERS SYLVAN
landstingsdirektör,
Västerbottens Läns Landsting

MAGNUS WALLINDER
förvaltningschef, Ljungby kommun

8 Ordförklaringar och förkortningar

Algoritm	En beräkningsmodell
BDI-II	Beck Depression Inventory-II
Biologisk markör	Företeelse eller storhet som när den kan påvisas respektive överstiger ett visst mätvärde utgör en indikator för att ett visst biologiskt förhållande föreligger, eller kan förväntas uppkomma.
Bivariat analys	Analysen omfattar två variabler samtidigt, t.ex. sensitivitet och specificitet
Blindning	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
CDIS	Computerized Diagnostic Interview Schedule
CGI-S	Clinical Global Impressions, Severity Scale
CIDI	Composite International Diagnostic Interview
CIS-R	Revised Clinical Interview Schedule
Depression	Ett sjukdomstillstånd med sänkt stämningsläge, minskad energi och minskad aktivitet
DSM	Diagnostic and Statistical Manual of Mental Disorders
DSM-5	Diagnostic and Statistical Manual of Mental Disorder, 5 th edition, klassifikation av psykiatriska tillstånd
Dystymi	Sjukdomstillstånd med sänkt stämningsläge de flesta dagar under en period av minst två år som inte uppfyller kriterierna för egentlig depression

Egentlig depression	Sjukdomstillstånd med sänkt stämningsläge med sådan svårighetsgrad att den ger upphov till någon form av funktionsnedsättning
Estimat	Uppskattat värde
Evidensstyrka	Evidensstyrkan är en bedömning av hur tillförlitligt det sammanvägda resultatet är (t.ex. estimat från en metaanalys). SBU tillämpar det internationellt utarbetade evidensgraderingssystemet GRADE. Resultaten försvagas om det finns allvarliga risker för systematiska fel, bristande samstämmighet, bristande överförbarhet till svenska förhållanden, bristande precision i resultatet eller risk för publikationsbias.
HADS	Hospital Anxiety and Depression Scale
Heterogenicitet	Skillnader mellan studier avseende t.ex. patienturval
ICD-10	International Statistical Classification of Diseases and related Health Problems, 10 th revision. Klassifikation av sjukdomar enligt Världshälsoorganisationen
Inklusionskriterier	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 gjvits är ett obligatoriskt inklusionskriterium. Deltagande i undersökningen kräver dessutom att inget av exklusionskriterierna föreligger
Konfidensintervall, KI	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 procent sannolikhet
Likertskala	En skala där respondenten får ange i hur hög grad han/hon instämmer i ett påstående (t.ex. grad av smärta)
MADRS	Montgomery-Åsberg Depression Rating Scale
MADRS-S	MADRS Self-Assessment
Metaanalys	Metod att göra en samlad bedömning av ett antal jämförande undersökningar genom att statistiskt sammanföra deras resultat.
MINI	Mini International Neuropsychiatric Interview
PHQ-2	Patient Health Questionnaire-2
PHQ-9	Patient Health Questionnaire-9
Positivt/negativt prediktionsvärde	Positivt prediktionsvärde anger sannolikheten för att en person som får positivt utfall (d.v.s. 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 (d.v.s. resultatet talar emot sjukdom) verkligen inte har sjukdomen. Siffran är kvoten mellan sant negativa och sant + falskt negativa
Prevalens	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
PRIME-MD	Primary Care Evaluation of Mental Disorders
Publikationsbias	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

Randomiserad	Sluppmässig 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 störfaktorer (confounders) lika mellan grupperna samt göra grupperna önskvärt jämförbara i sin sammansättning
SCID-I	Structured Clinical Interview for DSM-IV Axis I disorders
Screening	Engelska för sållning. Undersökning av en population för upptäckt av vissa icke diagnostiserade sjukdomar eller riskfaktorer för sjukdom. I denna rapport innebär screening en undersökning av individer med risk för depression (egentligen riktad screening)
Semistrukturerad intervju	Intervju där färdigformulerade frågor kan kompletteras med följdfrågor
Sensitivitet	Känslighet d.v.s. andelen av dem som är sjuka enligt referenstestet som klassificeras som sjuka av indextestet
Somatisk	Kroppslig
Specificitet	Träffsäkerhet d.v.s., andelen av dem som är friska enligt referenstestet som klassificeras som friska av indextestet
Strukturerad intervju	Intervjun består endast av färdigformulerade frågor
Systematisk översikt	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
Systematiskt fel (bias)	Ett resultatfel som uppstått genom procedurfel, effektbedömningsfel eller annat mänskligt fel under en undersökning; även fel som görs i bedömningen eller hanteringen av resultaten
Tröskelvärde	Ett gränsvärde som t.ex. skiljer mellan sjuk och frisk i ett test (eng. cut off)
Utfallsmått	Det mått på effekt eller tillförlitlighet som en klinisk studie avser att utvärdera

9 Referenser

1. SBU. Diagnostik och uppföljning av förstämningssyndrom. En systematisk litteraturöversikt. Stockholm: Statens beredning för medicinsk utvärdering (SBU); 2012. SBU-rapport nr 212. ISBN 978-91-85413-52-2.
2. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ* 2012;184:E191-6.
3. Meader N, Mitchell AJ, Chew-Graham C, Goldberg D, Rizzo M, Bird V, et al. Case identification of depression in patients with chronic physical health problems: a diagnostic accuracy meta-analysis of 113 studies. *Br J Gen Pract* 2011;61:e808-20.
4. O'Connor E, Rossom RC, Henninger M, Groom HC, Burda BU, Henderson JT, et al. Screening for Depression in Adults: An Updated Systematic Evidence Review for the U.S. Preventive Services Task Force. Evidence Synthesis No. 128. AHRQ Publication No. 14-05208-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; 2016.
5. Kringlen E, Torgersen S, Cramer V. A Norwegian psychiatric epidemiological study. *Am J Psychiatry* 2001;158:1091-8.
6. 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.
7. King M, Nazareth I, Levy G, Walker C, Morris R, Weich S, et al. Prevalence of common mental disorders in general practice attendees across Europe. *Br J Psychiatry* 2008;192:362-7.
8. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5. Arlington, VA, American Psychiatric Publishing; 2013.
9. World Health Organization (WHO). ICD-10 International Statistical Classification of Diseases and Related Health Problems, 10th Revision. Geneva, World Health Organization; 2010.
10. Gili M, Comas A, Garcia-Garcia M, Monzón S, Antoni SB, Roca M. Comorbidity between common mental disorders and chronic somatic diseases

- in primary care patients. *Gen Hosp Psychiatry* 2010;32:240-5.
11. Bădescu SV, Tătaru C, Kobylinska L, Georgescu EL, Zahiu DM, Zăgrean AM, et al. The association between Diabetes mellitus and Depression. *J Med Life* 2016;9:120-5.
 12. Chang WH, Lee IH, Chen WT, Chen PS, Yang YK, Chen KC. Coexisting geriatric anxiety and depressive disorders may increase the risk of ischemic heart disease mortality-a nationwide longitudinal cohort study. *Int J Geriatr Psychiatry* 2017;32:e25-33.
 13. Anseau M, Dierick M, Buntinkx F, Cnockaert P, De Smedt J, Van Den Haute M, et al. High prevalence of mental disorders in primary care. *J Affect Disord* 2004;78:49-55.
 14. Mergl R, Seidscheck I, Allgaier AK, Möller HJ, Hegerl U, Henkel V. Depressive, anxiety, and somatoform disorders in primary care: prevalence and recognition. *Depress Anxiety* 2007;24:185-95.
 15. Roca M, Gili M, Garcia-Garcia M, Salva J, Vives M, Garcia Campayo J, et al. Prevalence and comorbidity of common mental disorders in primary care. *J Affect Disord* 2009;119:52-8.
 16. Chesney E, Goodwin GM, Fazel S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. *World Psychiatry* 2014;13:153-60.
 17. Lumme S, Pirkola S, Manderbacka K, Keskimäki I. Excess Mortality in Patients with Severe Mental Disorders in 1996-2010 in Finland. *PLoS One* 2016;11:e0152223.
 18. Magruder-Habib K, Zung WW, Feussner JR. Improving physicians' recognition and treatment of depression in general medical care. Results from a randomized clinical trial. *Med Care* 1990;28:239-50.
 19. 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.
 20. Mitchell AJ, Yadegarfar M, Gill J, Stubbs B. Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: a diagnostic meta-analysis of 40 studies. *BJPsych Open* 2016;2:127-38.
 21. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Gen Hosp Psychiatry* 2015;37:567-76.
 22. Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *Int J Integr Care* 2002;2:e15.
 23. SBU. Utvärdering av metoder i hälso- och sjukvården: En handbok. 2 uppl. Stockholm; Statens beredning för medicinsk utvärdering (SBU); 2014.
 24. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210.
 25. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.
 26. 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.
 27. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10 Analysing and Presenting Results. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Cochrane Collaboration; 2010. [Webblänk: http://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/](http://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/)

28. Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *ACP J Club* 2008;149:2.
29. Pettersson A, Boström KB, Gustavsson P, Ekselius L. Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nord J Psychiatry* 2015;69:497-508.
30. Löwe B, Spitzer RL, Gräfe 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. Bombardier CH, Smiley J. Measurement characteristics and clinical utility of the Patient Health Questionnaire-9 among individuals with spinal cord injury. *Rehabil Psychol* 2015;60:211-2.
36. Cassin S, Sockalingam S, Hawa R, Wnuk S, Royal S, Taube-Schiff M, et al. Psychometric properties of the Patient Health Questionnaire (PHQ-9) as a depression screening tool for bariatric surgery candidates. *Psychosomatics* 2013;54:352-8.
37. Elderon L, Smolderen KG, Na B, Whooley MA. Accuracy and prognostic value of American Heart Association: recommended depression screening in patients with coronary heart disease: data from the Heart and Soul Study. *Circ Cardiovasc Qual Outcomes* 2011;4:533-40.
38. Haddad M, Walters P, Phillips R, Tsakok J, Williams P, Mann A, et al. Detecting depression in patients with coronary heart disease: a diagnostic evaluation of the PHQ-9 and HADS-D in primary care, findings from the UPBEAT-UK study. *PLoS ONE* 2013;8:e78493.
39. Amtmann D, Bamer AM, Johnson KL, Ehde DM, Beier ML, Elzea JL, et al. A comparison of multiple patient reported outcome measures in identifying major depressive disorder in people with multiple sclerosis. *J Psychosom Res* 2015;79:550-7.
40. Beard C, Hsu KJ, Rifkin LS, Busch AB, Bjorgvinsson T. Validation of the PHQ-9 in a psychiatric sample. *J Affect Disord* 2016;193:267-73.
41. Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, et al. How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. *J Affect Disord* 2011;134:266-71.
42. Hyphantis T, Kotsis K, Kroenke K, Paika V, Constantopoulos S, Drosos AA, et al. Lower PHQ-9 cutpoint accurately diagnosed depression in people with long-term conditions attending the Accident and Emergency Department. *J Affect Disord* 2015;176:155-63.
43. Rathore JS, Jehi LE, Fan Y, Patel SI, Foldvary-Schaefer N, Ramirez MJ, et al. Validation of the Patient Health Questionnaire-9 (PHQ-9) for depression screening in adults with epilepsy. *Epilepsy Behav* 2014;37:215-20.
44. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606-13.

45. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J Affect Disord* 2004;81:61-6.
46. Löwe B, Unutzer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med Care* 2004;42:1194-201.
47. Cameron IM, Crawford JR, Lawton K, Reid IC. Differential item functioning of the HADS and PHQ-9: an investigation of age, gender and educational background in a clinical UK primary care sample. *J Affect Disord* 2013;147:262-8.
48. Zimmerman M. Symptom severity and guideline-based treatment recommendations for depressed patients: implications of DSM-5's potential recommendation of the PHQ-9 as the measure of choice for depression severity. *Psychother Psychosom* 2012;81:329-32.
49. Harding KJ, Rush AJ, Arbuckle M, Trivedi MH, Pincus HA. Measurement-based care in psychiatric practice: a policy framework for implementation. *J Clin Psychiatry* 2011;72:1136-43.
50. Fortney JC, Unutzer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A Tipping Point for Measurement-Based Care. *Psychiatr Serv* 2017;68:179-88.
51. Baik SY, Gonzales JJ, Bowers BJ, Anthony JS, Tidjani B, Susman JL. Reinvention of depression instruments by primary care clinicians. *Ann Fam Med* 2010;8:224-30.
52. Pettersson A, Björkelund C, Petersson EL. To score or not to score: a qualitative study on GPs views on the use of instruments for depression. *Fam Pract* 2014;31:215-21.
53. Levis B, Benedetti A, Levis AW, Ioannidis JPA, Shrier I, Cuijpers P, et al. Selective Cutoff Reporting in Studies of Diagnostic Test Accuracy: A Comparison of Conventional and Individual-Patient-Data Meta-Analyses of the Patient Health Questionnaire-9 Depression Screening Tool. *Am J Epidemiol* 2017;185:954-64.

Bilaga 2

Exkluderade studier

Diagnostisk tillförlitlighet

Referens	Orsak till exkludering
Allgaier AK, Pietsch K, Frühe B, Sigl-Glöckner J, Schulte-Körne G. Screening for depression in adolescents: validity of the patient health questionnaire in pediatric care. <i>Depress Anxiety</i> 2012;29:906-13.	Fel population, studiedeltagarna var ungdomar
Asih S, Mayer TG, Bradford EM, Neblett R, Williams MJ, Hartzell MM, et al. The Potential Utility of the Patient Health Questionnaire as a Screener for Psychiatric Comorbidity in a Chronic Disabling Occupational Musculoskeletal Disorder Population. <i>Pain Pract</i> 2016;16:168-74.	Fel indextest, PHQ-9 med algoritm
Bhana A, Rathod SD, Selohilwe O, Kathree T, Petersen I. The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. <i>BMC Psychiatry</i> 2015;15:118.	Genomförd i Sydafrika
Bombardier CH, Smiley J. Measurement characteristics and clinical utility of the Patient Health Questionnaire-9 among individuals with spinal cord injury. <i>Rehabil Psychol</i> 2015;60:211-2.	Inte någon studie
Chagas MHN, Tumas V, Rodrigues GR, Machado-De-Sousa JP, Filho AS, Hallak JEC, et al. Validation and internal consistency of Patient Health Questionnaire-9 for major depression in Parkinson's disease. <i>Age Ageing</i> . 2013;42:645-9.	Genomförd i Brasilien

Tabellen fortsätter på nästa sida

Referens	Orsak till exkludering
Costa MV, Diniz MF, Nascimento KK, Pereira KS, Dias NS, Malloy-Diniz LF, et al. Accuracy of three depression screening scales to diagnose major depressive episodes in older adults without neurocognitive disorders. <i>Rev Bras Psiquiatr</i> 2016;38:154-6.	Fel population, äldre
Davis K, Pearlstein T, Stuart S, O'Hara M, Zlotnick C. Analysis of brief screening tools for the detection of postpartum depression: comparisons of the PRAMS 6-item instrument, PHQ-9, and structured interviews. <i>Arch Womens Ment Health</i> 2013;16:271-7.	Fel population, kvinnor med postpartum depression
de Man-van Ginkel JM, Gooskens F, Schepers VPM, Schuurmans MJ, Lindeman E, Hafsteinsdóttir TB. Screening for poststroke depression using the patient health questionnaire. <i>Nurs Res</i> 2012;61:333-41.	Fel population, äldre
de Man-van Ginkel JM, Hafsteinsdóttir T, Lindeman E, Burger H, Grobbee D, Schuurmans M. An efficient way to detect poststroke depression by subsequent administration of a 9-item and a 2-item Patient Health Questionnaire. <i>Stroke</i> 2012;43:854-6.	Fel population och fel referenstest (GDS-15)
Fiest KM, Patten SB, Wiebe S, Bulloch AG, Maxwell CJ, Jette N. Validating screening tools for depression in epilepsy. <i>Epilepsia</i> 2014;55:1642-50.	För lång tid mellan undersökningarna (två veckor)
Gothwal VK, Bagga DK, Bharani S, Sumalini R, Reddy SP. The patient health questionnaire-9: validation among patients with glaucoma. <i>PloS One</i> 2014;9:e101295.	Genomförd i Indien
Inoue T, Tanaka T, Nakagawa S, Nakato Y, Kameyama R, Boku S, et al. Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care. <i>BMC Psychiatry</i> 2012;12:73.	Genomförd i Japan
Janssen EP, Köhler S, Stehouwer CD, Schaper NC, Dagnelie PC, Sep SJ, et al. The Patient Health Questionnaire-9 as a Screening Tool for Depression in Individuals with Type 2 Diabetes Mellitus: The Maastricht Study. <i>J Am Geriatr Soc</i> 2016;64:e201-6.	Fel population (inte vårdsökande)
Patten SB, Burton JM, Fiest KM, Wiebe S, Bulloch AG, Koch M, et al. Validity of four screening scales for major depression in MS. <i>Mult Scler</i> 2015;21:1064-71.	Tid mellan undersökningarna (två veckor)
Prisnie JC, Fiest KM, Coutts SB, Patten SB, Atta CA, Blaikie L, et al. Validating screening tools for depression in stroke and transient ischemic attack patients. <i>Int J Psychiatry Med</i> 2016;51:262-77.	Fel population (äldre) samt tid mellan undersökningarna (två veckor)
Rooney AG, McNamara S, MacKinnon M, Fraser M, Rampling R, Carson A, et al. Screening for major depressive disorder in adults with cerebral glioma: an initial validation of 3 self-report instruments. <i>Neuro Oncol</i> 2013;15:122-9.	Fel fråga (samstämmighet med bedömning av närstående)
Seo JG, Park SP. Validation of the Patient Health Questionnaire-9 (PHQ-9) and PHQ-2 in patients with migraine. <i>J Headache Pain</i> 2015;16:65.	Genomförd i Korea
Sidebottom AC, Harrison PA, Godecker A, Kim H. Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. <i>Arch Womens Ment Health</i> 2012;15:367-74.	Tid mellan undersökningarna (fyra veckor)
Thapar A, Hammerton G, Collishaw S, Potter R, Rice F, Harold G, et al. Detecting recurrent major depressive disorder within primary care rapidly and reliably using short questionnaire measures. <i>Br J Gen Pract</i> 2014;64:e31-7.	Tid mellan undersökningarna (tre veckor)

Tabellen fortsätter på nästa sida

Referens	Orsak till exkludering
van der Zwaan GL, van Dijk SEM, Adriaanse MC, van Marwijk HWJ, van Tulder MW, Pols AD, et al. Diagnostic accuracy of the Patient Health Questionnaire-9 for assessment of depression in type II diabetes mellitus and/or coronary heart disease in primary care. <i>J Affect Disord</i> 2016;190:68-74.	Tid mellan undersökningarna (två veckor)
Volker D, Zijlstra-Vlasveld MC, Brouwers EP, Homans WA, Emons WH, van der Feltz-Cornelis CM. Validation of the Patient Health Questionnaire-9 for Major Depressive Disorder in the Occupational Health Setting. <i>J Occup Rehabil</i> 2016;26:237-44.	Fall-kontrollstudie med lång tid mellan undersökningarna (upp till fyra veckor).

Svårighetsgrad och monitorering

Referens	Orsak till exkludering
Beard C, Hsu KJ, Rifkin LS, Busch AB, Björgvinsson T. Validation of the PHQ-9 in a psychiatric sample. <i>J Affect Disord</i> 2016;193:267-73.	MINI referensstandard
Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. <i>J Gen Intern Med</i> 2001;16:606-13.	PRIME-MD referensstandard
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. <i>Gen Hosp Psychiatry</i> 2009;31:451-9.	Jämförde med Hamilton Depression Rating Scale

Bilaga 3

Studier med hög risk för systematiska fel (bias)

Referens	Orsak
Amtmann D, Bamer AM, Johnson KL, Ehde DM, Beier ML, Elzea JL, et al. A comparison of multiple patient reported outcome measures in identifying major depressive disorder in people with multiple sclerosis. <i>J Psychosom Res</i> 2015;79:550-7.	Högreselektad grupp där PHQ-9 lästes upp per telefon tillsammans med flera andra instrument
Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, et al. How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. <i>J Affect Disord</i> 2011;134:266-71.	Oklar rekryteringsprocess
Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). <i>J Affect Disord</i> 2004;81:61-6.	Ej blindad vid uppföljning; en person utförde intervju och PHQ-9 per telefon vid samma tillfälle.
Rathore JS, Jehi LE, Fan Y, Patel SI, Foldvary-Schaefer N, Ramirez MJ, et al. Validation of the Patient Health Questionnaire-9 (PHQ-9) for depression screening in adults with epilepsy. <i>Epilepsy Behav</i> 2014;37:215-20.	Svagheter i beskrivning av urval, hur PHQ-9 och MINI genomfördes samt av processen

SBU – Statens beredning för medicinsk och social utvärdering

webbplats: www.sbu.se • twitter: [@SBU_se](https://twitter.com/SBU_se) • telefon: 08-412 32 00