

REVIDERAD 2016

10 Evidensgradering

Det sista steget i utvärderingen är att bedöma hur starkt det samlade vetenskapliga underlaget är. SBU använder det internationellt utarbetade GRADE-systemet [1,2] för att klassificera styrkan på det vetenskapliga underlaget.

GRADE är ett system som utvecklas kontinuerligt via GRADE Working Group där även SBU ingår. Bakgrunden till att GRADE bildades är att det finns en flora av olika system som används parallellt för att gradera evidens och styrka på rekommendationer. Denna mångfald har lett till en viss förvirring, och att många har upplevt att viktiga steg i tidigare utvärderingsrapporter ibland utelämnats eller varit otydliga.

I princip bygger GRADE på erfarenheter från andra graderingssystem men med ett tydligare fokus på risk–nyttaperspektivet. GRADE har redan anammats av internationella aktörer som till exempel WHO, NICE, Cochrane Collaboration och BMJ. I Sverige används GRADE av bland andra SBU och Socialstyrelsen.

GRADE:s evidensgradering bygger på en fyrgradig skala från starkt, måttligt och lågt till mycket lågt vetenskapligt underlag. SBU har valt att ersätta ordet ”lågt” med ”begränsat” och ”mycket lågt” med ”otillräckligt”, men de innebär i princip samma sak. Skälen till att vi ändrat ordvalet är att det överensstämmer bättre med SBU:s tidigare nomenklatur och att vi anser att det ger lite mer vägledning. Evidensstyrkorna blir då med SBU:s terminologi starkt (⊕⊕⊕⊕), måttligt starkt (⊕⊕⊕○), begränsat (⊕⊕○○) och otillräckligt (⊕○○○) vetenskapligt underlag. Begränsat underlag kan vara tillräckligt för att tillämpa metoden i klinisk praxis om andra kriterier är uppfyllda, till exempel

rimlig kostnadseffektivitet. Otillräckligt underlag tydliggör att vi måste ha mer forskning. Förenklat kan man säga att ett starkt vetenskapligt underlag är så stabilt att det är liten risk för att ny forskning kommer att komma fram till nya slutsatser. På samma sätt innebär ett begränsat vetenskapligt underlag att det är högre risk för att nya studier kan förändra slutsatsen.

Detta avsnitt beskriver SBU:s tillvägagångssätt för att använda GRADE. Observera att systemet hittills är utvecklat för interventions-/behandlingsstudier. Det kan dock på likartat sätt användas för evidensgradering av orsakssamband och diagnostisk tillförlitlighet. För studier med kvalitativ metodik testas för närvarande CerQUAL som är inspirerat av GRADE men systemet är fortfarande under utveckling och presenteras inte här. För den intresserade hänvisas till (Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gülmezoglu M, Noyes J, Booth A, Garside R, Rashidian A. PLoS Med. 2015 Oct 27;12(10):

I GRADE-systemet finns också en rekommendationsdel som inte används av SBU.

En sammanfattande resultattabell ger en bra och kortfattad bild av underlaget

Ett lämpligt första steg är att göra en sammanfattande resultattabell. Detta förenklar det fortsatta arbetet. Ett exempel på en sådan resultattabell syns i Tabell 10.1. I tabellen bör den sammanvägda effekten för samtliga viktiga effektmått redovisas separat (t.ex. dödlighet, funktion och livskvalitet). I vissa fall förtecknas även effekter som mer är att betrakta som surrogatmått (t ex HbA_{1c} och blodtryck). Effektmåtten innefattar positiva effekter, men även negativa som biverkningar och komplikationer bör tas med. De olika effektmåtten bör tabelleras hierarkiskt så att de viktigaste står före de mindre viktiga. Som framgår av Tabell 10.1 fyller man i kolumnerna ”Vetenskapligt underlag” och ”Kommentarer” i ett senare skede efter den samlade bedömningen.

Tabell 10.1
Sammanfattande resultattabell (summary of findings). Effekten av antibiotikaproylax jämfört med placebo vid käkkirurgiska ingrepp.

Effektmått	Studie-design Antal patienter (antal studier)	Medelrisk i standard-grupp (min–max)	Relativ risk (95 % KI)	Absolut effekt per 1 000 patienter	Vetenskapligt underlag	Kommentarer
Sårinfektioner vid operation av käkfrakturer	RCT 461 (3)	39 % (20–62 %)	RR 0,25 (0,15 till 0,41)	259 färre		

KI = Konfidensintervall; **RCT** = Randomiserad kontrollerad studie; **RR** = Relativ risk

I exemplet med sårinfektioner vid operationer av käkfrakturer (Tabell 10.1) fanns tre RCT med sammanlagt 461 patienter där det gick att göra en meta-analys och på så sätt få fram en sammanvägd effekt i numeriska termer [3]. I många fall går det dock inte att få fram så preciserade data på medelvärden och riskdifferens. När data är för heterogena är det inte möjligt att väga ihop resultaten i en poolad metaanalys med en absolut eller relativ riskdifferens. Effekten kan då istället redovisas som min–maxvariation för studierna.

Om det finns gott om välgjorda randomiserade kontrollerade studier (RCT) behöver normalt sett inte observationsstudier inkluderas i bedömningen av positiva effekter. Om det inte finns randomiserade studier eller om de randomiserade studierna är bristfälliga eller har för kort uppföljningstid kan observationsstudier ge viktig tilläggsinformation och bidra till den samlade evidensgraderingen i både positiv och negativ riktning. RCT och observationsstudier redovisas dock separat i resultattabellen. Vid bedömning av risker är det ofta viktigt att inkludera observationsstudier eftersom RCT i de flesta fall inte är utformade för att besvara frågor om långsiktiga risker.

Preliminär evidensstyrka

Som nämnts ovan sätts en evidensstyrka för varje utfallsmått som finns med i resultattabellen.

I arbetet med GRADE utgår man från en preliminär evidensstyrka. Den baseras enbart på vilket studieupplägg (studiedesign) som studierna som ingår i det vetenskapliga underlaget består av. Den preliminära evidensstyrkan justeras sedan uppåt eller neråt beroende på ett antal kvalitetsfaktorer som beskrivs närmare nedan. Om underlaget till största delen består av randomiserade studier där risken för systematiska fel är minst bedöms det preliminärt som starkt.

När det gäller diagnostiska tillförlitlighetsstudier (accuracy) har GRADE-gruppen tagit ställning för att den preliminära evidensstyrkan ska utgå från stark även för observationsstudier [4]. Detta kan diskuteras, men SBU accepterar denna utgångspunkt tills vidare. Däremot är det mycket viktigt att man analyserar om förbättrad diagnostik i slutändan ger patientnytta och inom socialtjänst om strukturerad bedömning i slutändan ger någon klientnytta.

Faktaruta 10.1
Preliminär evidensstyrka
baserad på studiedesign
och skäl för ned- eller
uppgredning av
evidensgraderingen.

Preliminär evidensstyrka för interventionsstudier:

Evidensstyrka	Studiedesign
Stark (⊕⊕⊕⊕)	Randomiserade studier
Måttligt stark (⊕⊕⊕○)	
Begränsad (⊕⊕○○)	Observationsstudier; kohort- och fall-kontrollstudier
Otillräcklig (⊕○○○)	Fallstudier

Sedan kan evidensstyrkan sänkas eller höjas enligt nedanstående:

Sänk gradering om	Höj gradering om
<ul style="list-style-type: none">• Brister i studiekvalitet (maximalt -2)• Bristande överensstämmelse mellan studierna (maximalt -2)• Brister i överförbarhet/relevans (maximalt -2)• Bristande precision (maximalt -2)• Hög sannolikhet för publikationsbias (maximalt -1)	<ul style="list-style-type: none">• Stora effekter och inga sannolika confounders (maximalt -2)• Tydligt dos-respons samband (maximalt -1)• Confounders som inte är med i analysen borde leda till bättre behandlingsresultat i kontrollgruppen, dvs. hög sannolikhet att effekten underskattas

Åtta faktorer påverkar den slutliga evidensstyrkan

För att fastställa den slutliga evidensstyrkan bedömer expertgruppen hur tillförlitligt det vetenskapliga underlaget är. Den preliminära evidensstyrkan sänks om underlaget är osäkert med avseende på:

- studiekvalitet
- samstämmighet/överensstämmelse
- överförbarhet/relevans
- precision i data
- risk för publikationsbias.

Evidensstyrkan kan dras ner med ett eller två steg för varje faktor beroende på hur stora bristerna är. Om bristerna är mindre allvarliga kan man notera det utan att gradera ner. Om det finns mindre allvarliga brister med avseende på flera av faktorerna kan det leda till att evidensstyrkan totalt sett dras ner ett steg. Man bör dock komma ihåg att observationsstudier redan i utgångsläget graderats ned på grund av studiedesign och därför inte generellt bör graderas ned ytterligare på grund av brist på confounding-kontroll. Vid allvarliga brister i confounding-kontrollen kan det dock vara motiverat att även nedgradera observationsstudier. Däremot kan brister som rör samstämmighet, överförbarhet, precision och risk för publikationsbias motivera nedgradering av observationsstudier.

I vissa fall finns det också skäl till att *öka* evidensstyrkan med ett eller två steg. Detta gäller när det vetenskapliga underlaget består av stora, välgjorda observationsstudier med god kontroll för förväxlingsfaktorer. De tre faktorer som kan höja evidensstyrkan är:

- stora effekter
- dos-responssamband
- hög sannolikhet att effekten i studien är underskattad.

Studiekvalitet

Under granskningsfasen bedömdes varje studies kvalitet individuellt. I detta steg görs en samlad värdering av hela materialet. I första hand vägs de traditionella faktorerna som till exempel randomisering, blindning och bortfall in. Hur välgjorda är studierna med avseende på randomisering totalt sett? Är några mycket välgjorda medan resterande studier har ett oklart randomiseringsförfarande? Även andra faktorer som är belysta i granskningsmallarna kan vara viktiga att beakta, likaså ämnesspecifika metodproblem som experterna har identifierat.

För kohortstudier och andra observationsstudier är frågan om jämförbarhet mellan försöks- och kontrollgrupp central. Det innebär att bedömningen av den sammanvägda studiekvaliteten i hög grad beror på i vad mån studierna har kontrollerat för förväxlingsfaktorer (confounders, se vidare i Kapitel 6).

Observationsstudier visade att östrogenbehandling minskade risken för hjärt- och kärlsjukdom. Eftersom man visste att östrogenbehandling var vanligare bland kvinnor från högre socialgrupper kunde man ha misstänkt att den observerade effekten var överskattad och därigenom nedgraderat tilltron till sambandet.

Exempel 10.1
Risk för att effekten överskattats.

Diagnostiska studier kan graderas ner om rekryteringen inte är konsekutiv, om utvärderarna inte är blindade och om studien har någon form av verifikationsbias.

GRADE-klassificeringen innebär att evidensstyrkan kan justeras ned ett steg om det finns allvarliga kvalitetsbrister i underlaget och två steg om begränsningarna är mycket allvarliga. Observera att GRADE i sig inte kräver att underlaget ska ha minst medelhög kvalitet. För SBU:s vidkommande, där studier med låg kvalitet sorterats bort, är det i praktiken mycket ovanligt att evidensstyrkan dras ner två steg på grund av metodologiska brister.

Samstämmighet/överensstämmelse

Här bedöms i vilken utsträckning studierna kommer fram till samma resultat. Pekar de åt samma håll och är effektstorleken av jämförbar storlek i de

olika studierna? Metaanalyser kan vara en god hjälp för att bedöma graden av samstämmighet.

Samstämmigheten kommer att vara beroende av hur likartade studierna är med avseende på population, exakt hur interventionen genomfördes och vilken kontrollgrupp som användes. En annan viktig faktor är om större delen av studierna har genomförts av samma forskargrupp.

Generellt sett ökar trovärdigheten i det samlade materialet om studierna har gjorts av olika forskargrupper med olika populationer och studierna samstämmigt pekar i samma riktning. Om studierna visar såväl över- som underrisker kan evidensstyrkan minskas med ett steg. Detsamma gäller om effektstorlekarna varierar kraftigt mellan studierna, vilket leder till en ökad osäkerhet.

I vissa fall kan skillnaderna förklaras med olikheter i studierna, till exempel att de undersökt olika populationer. I sådana fall kan det vara mer lämpligt att dela upp materialet och formulera slutsatser för de olika populationerna var för sig.

Överförbarhet/relevans

Med överförbarhet menas i vilken utsträckning det vetenskapliga underlaget är generaliserbart och relevant för svenska förhållanden. Viktiga frågor att överväga är till exempel hur väl populationen överensstämmer med den man ser i daglig svensk praxis, om interventionen utförs på samma sätt som i Sverige, om kontrollgruppen är relevant och om sammanhanget inom vilket det sociala stödet ges är rimligt likartat.

Ett exempel på bristande överförbarhet är att kontrollgruppen får en behandling som inte finns tillgänglig i Sverige. Det går då inte att avgöra hur effektiv interventionen är jämfört med sedvanlig behandling i svensk praxis.

Om överförbarheten och relevansen är bristfällig kan evidensstyrkan justeras ned ett eller två steg. På samma sätt som när det gäller studiekvalitet har sannolikt studier som är mindre relevanta för svenska förhållanden redan sorterats bort (med hjälp av relevansmallen, se Kapitel 5 och Bilaga 1).

Ett specialfall är när det bara finns en studie som mäter effekten med hjälp av ett visst utfallsmått. SBU:s bedömning är att överförbarheten vanligtvis kräver minst två studier. Det innebär att evidensstyrkan i detta fall oftast justeras ner ett steg. Undantag kan till exempel vara när underlaget består av en stor mycket välgjord randomiserad multicenterstudie.

Precision i data

Detta kriterium uppskattar hur osäker den sammanvägda effekten är. Få observationer och breda konfidensintervall i de olika studierna kommer att leda till sämre precision. Precisionen beror av antalet händelser, antal personer i grupperna och den relativa riskminskningen.

Ett hjälpmedel för att bedöma om precisionen är osäker är att göra en power-analys baserad på det totala antalet observationer i de inkluderade studierna. Om antalet observationer i dessa inkluderade studier är mindre än det antal som krävs för att visa statistiskt signifikanta resultat kan det finnas anledning att nedgradera för bristande precision. Om studierna är mycket små bör man även om utfallen är statistiskt signifikanta vara uppmärksam på om skillnaderna i ingångsvärden (baslinjedata) är stora. Skiljer sig baslinjedata mycket åt mellan grupperna kan det finnas anledning att nedgradera för precision i data.

Risk för publikationsbias

Med publikationsbias avses att delar av det vetenskapliga underlaget inte är tillgängligt i publicerade studier. Risken för publikationsbias ökar bland annat om underlaget enbart består av små studier som är utförda av samma forskargrupp och som har stora metodbrister. Utvärderingar av nya metoder faller ofta inom den här kategorin. Analog situation är när underlaget utgörs av företagssponsrade studier av till exempel ett läkemedel. Om samtliga studier av en ny metod har innovatören som huvudförfattare finns det också skäl att överväga publikationsbias.

Effekterna av selektiv publicering framgår väl i en svensk studie från Läkemedelsverket [6]. Systematiska översikter [7] och många andra studier [8,9] pekar entydigt på att studier som sponsrats av industrin eller andra aktörer med egenintressen av resultaten överskattar effekterna av sina produkter. Kostnadseffektanalyser som utförts av läkemedelsföretag visade till exempel mer än två gånger så ofta kostnadseffektkvoter under 20 000 amerikanska dollar per kvalitetsjusterat vunnet levnadsår som icke-industrisponsrade studier [8,9].

Det är ofta svårt att avgöra om det finns publikationsbias, men det finns några metoder som underlättar bedömningen. Ett sätt är att använda centrala register (t ex www.controlled-trials.com och www.clinicaltrials.gov) över påbörjade kliniska prövningar. En kontroll av vilka studier som finns i registren bör ingå i bedömningen.

Registren har funnits sedan början av 2000-talet och kan ge värdefull information om studier som startats det senaste decenniet. Studier som, enligt registret, förefaller vara avslutade sedan flera år tillbaka kan vara en möjlig källa till publikationsbias. Omvänt kan stora pågående studier leda till en ökad risk för att evidensstyrkan påverkas av deras resultat i framtiden, så även för bedömning av osäkerhet kan registren ge viss vägledning.

Ytterligare ett sätt att påvisa publikationsbias är att titta i databaser för konferensabstrakt. Förekomst av abstrakt för studier som inte publicerats inom ett rimligt antal år är ett starkt stöd för att publikationsbias föreligger. Detta är dock ett arbetskrävande moment.

Om det finns många studier kan risken bedömas med hjälp av en så kallad funnel plot (Bilaga 9). Som en tumregel krävs minst fem studier för att analysen ska vara meningsfull.

Effektstorlek

Detta kriterium kan användas om det finns minst två stora, välgjorda observationsstudier med god kontroll för förväxlingsfaktorer. Om effekten är hög ökar sannolikheten för att det funna sambandet är kausalt.

GRADE anger att den samlade evidensstyrkan kan höjas med ett steg om den sammanvägda effekten definierad som relativ risk (RR) är större än 2 (RR >2,0 alternativt RR <0,5). Om RR >5,0 (alternativt RR <0,2) kan evidensstyrkan höjas med två steg. Om man tidigare bedömt att studiekvaliteten är bristfällig kan man avstå från att höja evidensgraderingen på grund av stor effektstorlek.

Observera att om utfallsmåttet är oddskvoter (OR) kan effekterna vara över-skattade om utfallen är vanliga (>10 procent). Gränserna kan då behöva justeras. Se mer i statistikavsnittet (Bilaga 9) där skillnaden mellan RR och OR förklaras tillsammans med en formel för hur man kan omvandla OR till RR och vice versa.

Exempel 10.2 Stor effekt kan höja evidensstyrkan.

En metaanalys av observationsstudier visade att cykelhjälm reducerade risken för huvudskador med en oddskvot på 0,31 (0,26–0,37) [5]. Detta är en stor effekt och leder till att evidensstyrkan justeras upp ett steg. Om det inte finns någon anledning till att justera ner evidensstyrkan på grund av brister i underlaget skulle det medföra att det vetenskapliga underlaget blir måttligt starkt (⊕⊕⊕○).

Dos-responssamband

Även detta kriterium är begränsat till att gälla stora, välgjorda observationsstudier. Dos-respons kan avse både effekter och risker. Ett dos-responssamband ökar trovärdigheten för att en åtgärd har effekt. Dos-respons kan gälla för både läkemedel och andra insatser. Effekten kan mätas dels inom en studie, dels mellan studier. Generellt är ett dos-responssamband mycket mer trovärdigt när det visats inom en studie än när det har visats genom jämförelser mellan studier.

Ett dos-responssamband kan öka evidensstyrkan med ett steg.

Exempel 10.3 Dos-responssamband kan höja evidensstyrkan.

I SBU-rapporten "Mat vid diabetes" [10] studerades bland annat risken för hjärtinfarkt bland personer med diabetes med varierande alkoholkonsumtion. I tre stora observationsstudier med totalt 10 312 patienter såg man i samtliga studier ett dos-responsförhållande så att gruppen med högre alkoholkonsumtion hade betydligt lägre relativ risk än de med låg eller ingen konsumtion. Expertgruppen uppgraderade därför evidensstyrkan till måttligt starkt vetenskapligt underlag på grund av dos-responsförhållande.

Hög sannolikhet att effekten i studien är underskattad

Vid enstaka tillfällen kan evidensstyrkan justeras upp om det är mycket sannolikt att studierna underskattat effekten. Detta kan gälla om de förväxlingsfaktorer (confounders) som studien inte har kunnat justera för, talar för att effekten är underskattad.

I en systematisk översikt som omfattade 38 miljoner patienter var dödstalen högre på vinstdrivande privata sjukhus än på icke-vinstdrivande privata sjukhus [11]. GRADE Working Group menar att det är sannolikt att patienterna var sjukare på de icke-vinstdrivande sjukhusen och att de vinstdrivande sjukhusen hade större resurser och fler patienter som var väl försäkrade. Evidensen för att dödstalen verkligen är högre på vinstdrivande sjukhus har då stärkts.

Exempel 10.4
Underskattad effekt kan höja evidensstyrkan.

Samlad evidensgradering

Det är viktigt att projektgruppen kommenterar skriftligt hur underlaget påverkats av var och en av ovan diskuterade faktorer, som motivering till bedömningen. För att underlätta arbetet med evidensgradering kan man göra en evidenstabell (Tabell 10.2). I denna noteras den preliminära evidensstyrkan och de justeringar som görs i nästa steg.

När det gäller exemplet antibiotikaproylax vid käckkirurgi med utfallsmåttet sårinfektion kan man i första steget göra en evidenstabell (Tabell 10.2) som underlag för en sammanfattande resultattabell (Tabell 10.3).

I detta fall bedömdes brister i studiekvaliteten motivera en nedgradering liksom de sammanlagda bristerna av att studierna hade få utfall och använt olika antibiotika.

Studier Patienter	Design	Studiekvalitet	Överensstämmelse	Överförbarhet	Opresisa data	Publiceringsbias	Effektstorlek	Dos-respons	Förväxlingsfaktor
3 461	RCT ⊕⊕⊕○	-1*	0	0**	-1***	0	0	0	0

Tabell 10.2
Evidenstabell. Sårinfektioner, antibiotikaproylax jämfört med placebo vid operation av käkfrakturer.

* Ingen blindning, en studie har ingen uppgift om bortfall.

** Olika antibiotika.

*** Få utfall, tillsammans med ** -1.

Den sammanfattande resultattabellen för antibiotikaprofylax redovisas därefter i Tabell 10.3. Den relativa risken var här stor och kunde ha motiverat en uppgradering, men expertgruppen bedömde att bristande studiekvalitet i detta fall inte motiverade en uppgradering.

När det vetenskapliga underlaget för samtliga utfallsmått har bestämts ger resultattabellen en samlad bild av kunskapsläget för en viss fråga.

Tabell 10.3
Sammanfattande resultattabel (summary of findings). Effekten av antibiotikaprofylax jämfört med placebo vid käkkirurgiska ingrepp.

Effektmått	Studie-design Antal patienter (antal studier)	Medelrisk i standard-grupp (min–max)	Relativ risk (95 % KI)	Absolut effekt per 1 000 patienter	Vetenskapligt underlag	Kommentarer
Sårinfektioner vid operation av käkfrakturer	RCT 461 (3)	39 % (20–62 %)	RR 0,25 (0,15 till 0,41)	259 färre	⊕⊕○○ Begränsat	Svagheter i studie-kvalitet (-1) Få utfall (-1)

KI = Konfidensintervall; RCT = Randomiserad kontrollerad studie; RR = Relativ risk

SBU:s bedömning är generellt att enbart en studie av begränsad omfattning är otillräckligt för att utgöra evidens. Det finns i dessa fall en stor osäkerhet om precisionen och ingen annan studie har bekräftat resultaten. Undantag från denna regel måste motiveras, till exempel att det är en stor multicenterstudie eller att det finns andra starka skäl.

Tolkning av evidensstyrkan

Även om utvärderingen inte mynnar ut i en rekommendation så ger den vägledning. Om det vetenskapliga underlaget är otillräckligt indikerar det behov av mer forskning. En begränsad evidensstyrka kan motivera att metoden används under förutsättning att den uppfyller andra krav på acceptabel balans mellan risk och nytta, kostnadseffektivitet och att den är etiskt acceptabel. Vid måttlig eller hög evidensstyrka är det vetenskapliga underlaget gott och motiverar sannolikt att metoden tillämpas under förutsättning att den ekonomiska, etiska och sociala analysen i utvärderingen ger stöd för metoden.

En systematisk litteraturoversikt väger samman resultat från olika studier. SBU använder det internationellt utarbetade GRADE-systemet (<http://www.gradeworkinggroup.org>) för att göra en strukturerad bedömning av tillförlitligheten (evidensstyrkan) hos varje sammanvägt delresultat (utfall) i översikten. Den sakliga grunden för värderingen ska redovisas tydligt så att det är möjligt för andra att granska och göra sin egen bedömning.

Bedömningen av tillförlitlighet innefattar, för varje sammanvägt delresultat:

- hur stor risken är för systematiska fel i studierna (engelska: *bias*, snedvridning),
- hur mycket studierna motsäger varandra (engelska: *inconsistency*, bristande samstämmighet),
- i vilken grad som de studerade förhållandena skiljer sig från den aktuella frågan (engelska: *indirectness*, bristande överförbarhet),
- hur stor den statistiska osäkerheten är (engelska: *imprecision*, bristande precision) samt
- hur stor risken är för snedvriden publicering av studier och resultat (engelska: *publication bias*).

Hänsyn tas också till storleken på delresultatet, eventuellt samband mellan dos och respons samt i vilken riktning som tänkbara snedvridande faktorer kan förväntas verka.

För mer detaljerad beskrivning se metodboken [ref].

Tillförlitligheten graderas i fyra nivåer:

- Det sammanvägda resultatet har **hög tillförlitlighet** (⊕⊕⊕⊕)
(Bedömningen är att resultatet stämmer)
- Det sammanvägda resultatet har **måttlig tillförlitlighet** (⊕⊕⊕○)
(Bedömningen är att det är troligt att resultatet stämmer)
- Det sammanvägda resultatet har **låg tillförlitlighet** (⊕⊕○○)
(Bedömningen är att det är möjligt att resultatet stämmer)
- Det sammanvägda resultatet har **mycket låg tillförlitlighet** (⊕○○○)
(Det går inte att bedöma om resultatet stämmer)

När det helt saknas studier som uppfyller inklusionskriterierna anges "studier saknas", utan gradering av tillförlitligheten.

Referenser

1. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
2. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines 1. Introduction – GRADE evidence profiles and summary of finding tables. *J Clin Epidemiol* 2011;64:383-94.
3. SBU. Antibiotikaprofylax vid kirurgiska ingrepp. En systematisk litteraturöversikt. Stockholm: Statens beredning för medicinsk utvärdering (SBU); 2010. SBU-rapport nr 200. ISBN 978-91-85413-36-2.
4. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;336:1106-10.
5. Thompson DC, Rivara F, Thompson R. Helmets for preventing head and facial injuries in bicyclists. *Cochrane Database of Systematic Reviews* 1999, Issue 4. Art. No.: CD001855. DOI: 10.1002/14651858.CD001855.
6. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3.
7. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.
8. Jørgensen AW, Maric KL, Tendal B, Faurschou A, Gøtzsche PC. Industry-supported meta-analysis compared with meta-analysis with non-profit or no support: differences in methodological quality and conclusions. *BMC Med Res Methodol* 2008;8:60.
9. Bell CM, Urbach DR, Ray JG, Bayoumi A, Rosen AB, Greenberg D, Neumann PJ. Bias in published cost effectiveness studies: systematic review. *BMJ* 2006; 332:699-703.
10. SBU. Mat vid diabetes. En systematisk litteraturöversikt. Stockholm: Statens beredning för medicinsk utvärdering (SBU); 2010. SBU-rapport nr 201. ISBN 978-91-85413-37-9.
11. Devereaux PJ, Choi PT, Lacchetti C, Weaver B, Schünemann HJ, Haines T, et al. A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. *CMAJ* 2002;166:1399-406.