

Bilaga 3. Mall för kvalitetsgranskning av observationsstudier

VERSION 2012:2.1

Granskningen av en studie gäller i första hand studiekvalitet, dvs risk för systematiska fel och risk för intressekonflikter (A). I den sammanvägda bedömningen av alla inkluderade studier enligt GRADE inkluderar man också studiernas överensstämmelse (B), överförbarhet (C), precision (D), publikationsbias (E), effektstorlek (F), dos-respons samband (G) och sannolikhet att effekten är underskattad (H).

Författare: _____ År: _____ Artikelnummer: _____

Alternativet ”oklart” används när uppgiften inte går att få fram från texten. Alternativet ”ej tillämpligt” väljs när frågan inte är relevant.

A. Granskning av studiens begränsningar – eventuella systematiska fel (bias)	Ja	Nej	Oklart	Ej tillämpligt
A1. Selektionsbias				
a) Var de observerade grupperna rekryterade på ett tillräckligt likartat sätt?				
b) Var de jämförda gruppernas sammansättning tillräckligt lika vid studiestart?				
c) Har korrigerings av obalanser i baslinjevariabler mellan grupper med olika exponering/behandling gjorts på ett adekvat sätt i den statistiska analysen?				
Kommentarer:				
Bedömning av risk för selektionsbias:				
A2. Behandlingsbias				
a) Var villkoren (utöver den behandling eller exponering som studerades) för grupperna under behandlings-/exponeringstiden tillräckligt likartade?				
b) Var följsamhet gentemot behandling/exponering acceptabel i grupperna?				
Kommentarer:				
Bedömning av risk för behandlingsbias:				

A. Fortsättning	Ja	Nej	Oklart	Ej till- lämpligt
A3. Bedömningsbias (per utfallsmått)				
a) Var utfallsmåttet okänsligt för bedömningsbias?				
b) Var personerna som utvärderade utfallet <i>blindade</i> för studiedeltagarnas exponeringsstatus?				
c) Var personerna som utvärderade utfallet <i>opartiska</i> ?				
d) Var utfallet definierat på ett lämpligt sätt?				
e) Mättes utfallet på ett adekvat sätt med standardiserade/definierade mätmetoder?				
f) Mättes utfallet på ett adekvat sätt med validerade mätmetoder?				
g) Har variationer i exponering över tid tagits med i analysen?				
h) Har utfallet mätts vid optimal(a) tidpunkt(er)?				
i) Var observatörsöverensstämelsen acceptabel?				
j) Har studien tillämpat ett lämpligt statistiskt mått för rapporterad effekt/samband?				
Kommentarer:				
Bedömning av risk för bedömningsbias:				
A4. Bortfallsbias (per utfallsmått)				
a) Var bortfallet tillfredsställande lågt i förhållande till populationens storlek?				
b) Var bortfallet lika stort inom grupperna?				
c) Var relevanta baslinjevariabler lika fördelade mellan bortfallen i interventions- och kontrollgruppen alternativt mellan olika exponeringsgrupper?				
d) Var relevanta baslinjevariabler lika fördelade mellan analys- och bortfallgruppen?				
e) Var den statistiska hanteringen av bortfallet adekvat?				
Kommentarer:				
Bedömning av risk för bortfallsbias:				

A. Fortsättning	Ja	Nej	Oklart	Ej till- lämpligt
A5. Rapporteringsbias				
a) Följde studien ett i förväg fastlagt studieprotokoll?				
b) Var utfallsmåtten relevanta?				
c) Mättes biverkningar/komplikationer på ett systematiskt sätt?				
d) Var tidpunkterna för rapporterad analys relevanta?				
Kommentarer:				
Bedömning av risk för rapporteringsbias:				
A6. Intressekonflikter				
a) Föreligger, baserat på författarnas angivna bindningar och jäv, låg eller obefintlig risk att studiens resultat har påverkats av intressekonflikter?				
b) Föreligger, baserat på uppgifter om studiens finansiering, låg eller obefintlig risk att studien har påverkats av en finansiär med ekonomiskt intresse i resultatet?				
c) Föreligger låg eller obefintlig risk för annan form av intressekonflikt (t ex att författarna har utvecklat interventionen)?				
Kommentarer:				
Bedömning av risk för intressekonflikt:				
Sammanvägning av risk för bias (per utfallsmått)				
	Låg	Medelhög	Hög	
A1. Selektionsbias				
A2. Behandlingsbias				
A3. Bedömningsbias				
A4. Bortfallsbias				
A5. Rapporteringsbias				
A6. Intressekonfliktbias				
Kommentarer:				
Sammanfattande bedömning av risk för systematiska fel (bias):				

Underlag för sammanvägd bedömning enligt GRADE

B. Bristande överensstämmelse mellan studierna

Hanteras endast på syntesnivå

C. Granskning av studiens överförbarhet

	Ja	Nej	Delvis	Ej till- lämpligt
a) Överensstämmer sammanhanget och kontrollgruppens villkor med den situation som SBU/HTA-rapportens slutsatser avser?				
b) Är den inkluderade studiepopulationen tillräckligt lik den population som SBU/HTA-rapportens slutsatser avser?				
c) Är interventionen relevant för de förhållanden som SBU/HTA-rapportens slutsatser avser?				

Kommentar:

Bedömning av brister i överförbarhet:

D. Granskning av precision

	Ja	Nej	Delvis	Ej till- lämpligt
a) Är precisionen acceptabel med hänsyn till antal inkluderade individer och antal händelser (utfall)?				

Kommentar:

E. Granskning av publikationsbias

Hanteras endast på syntesnivå

F. Granskning av effektstorlek

	Ja	Nej	Delvis	Ej till- lämpligt
a) Var effekten stor (t ex RR <0,5 eller >2,0)?				
b) Var effekten mycket stor (t ex RR <0,2 eller >5,0)?				

Kommentar:

G. Granskning av dos-responssamband

	Ja	Nej	Delvis	Ej till- lämpligt
a) Finns stöd för ett dos-responssamband mellan exponering och utfall?				

Kommentar:

H. Sannolikhet att effekten är underskattad pga "confounders"	Ja	Nej	Delvis	Ej till- lämpligt
Vid enstaka tillfällen kan evidensstyrkan höjas om det är mycket sannolikt att effekten är underskattad.				
a) Finns det starkt stöd för att "confounders" som studien inte kunnat ta hänsyn till skulle stärka sambandet?				
Kommentar:				

Mall för kvalitetsgranskning av observationsstudier: förklaringar

Mallen är i första hand tänkt att användas för granskning av studiekvalitet i prospektiva kohortstudier (del A). I den mån retrospektiva kohortstudier med historiska kontroller, retrospektiva fallserier, tvärsnittsstudier eller andra icke-randomiserade studietyper är aktuella att använda kan mallen användas för dessa med vissa tillägg/anpassningar. Granskningsmallen avser att ge ett systematiskt underlag till stöd för att bedöma risken för att en given effekt i en studie systematiskt snedvridits (bias) under forskningsarbetet. Konsekvensen av detta är att effekten antingen underskattas eller överskattas jämfört med en "sann" effekt. Även effektens riktning kan ha missbedömts.

Syftet med mallen är att skapa ett systematiskt och transparent underlag för att diskutera hur stor risken är att skattade effektmått/samband i en enskild studie är systematiskt snedvridna. Någon algoritm för att räkna samman kvalitetspoäng erbjuds alltså inte.

För att resultaten ska kunna användas för evidensgradering enligt GRADE krävs ytterligare information i form av sammanställningar av samtliga ingående studier. Det gäller bristande överensstämmelse mellan studierna (B), studiens överförbarhet (C), precision (D), risk för publikationsbias (E), effektstorlek (F), dos-respons samband (G) och granskning av sannolikhet att effekten är underskattad (H). Denna sammanvägning sker vid ett senare tillfälle, men det kan vara lämpligt att vid läsningen av en enskild studie samtidigt kommentera dessa faktorer.

A. Granskning av studiens begränsningar – eventuella systematiska fel

A1. Risk för selektionsbias ("selection bias")

Med "selektionsbias" avses systematiska fel som är relaterade till hur studien har hanterat urval av försökspersoner (motsvarande) samt indelning i interventions- och kontrollgrupper.

Risk för selektionsbias kan föreligga då interventionsgruppen inte är tillräckligt lik kontrollgruppen vid baslinjen. Kända, såväl som okända, risk- och skyddsfaktorer bör vara tillräckligt lika i de båda grupperna för att inte snedvrider resultatet. Några viktiga förväxlingsfaktorer ("confounders") är ålder, kön, bakomliggande sjukdomshistoria och samsjuklighet. En annan viktig faktor är socioekonomi, som sannolikt är den starkaste riskfaktorn för sjuklighet och för tidig död. Genom att använda statistiska metoder som matchning, stratifiering, multivariat regressionsanalys eller "propensity score"-metodik kan man dock korrigera för kända förväxlingsfaktorer (se A1c).

Risken för selektionsbias är också hög om den åtgärd som studeras är särskilt lämplig att sätta in på vissa försökspersoner som har en särskilt hög eller låg chans för att svara väl på åtgärden.

- A1a. Finns det en klar definition av jämförelsegruppen? Har de jämförda grupperna rekryterats på så pass likvärdiga sätt att resultaten inte har snedvridits? Har jämförelsegruppen hämtats från den allmänna befolkningen eller från ett begränsat urval? Om jämförelsegruppen är en historisk kontrollgrupp finns det anledning att vara särskilt försiktig vid värderingen. En viktig fråga är om samma metodik användes för att rekrytera till interventions- respektive jämförelsegrupp.
- A1b. Uppgifter som kan visa väsentliga skillnader mellan grupperna finns ofta i en inledande tabell eller under bakgrundsdata ("baseline characteristics").
- A1c. Metoder som kan användas i detta sammanhang är matchning/restriktion, stratifierad analys, multivariat modellanalys (t ex regressionsanalys) eller "propensity score"-metodik.

Eftersom observationsstudier (2+) i GRADE-systemet genom sin studiedesign redan från början antas ha större risk för selektionsbias än randomiserade studier så måste riskerna för selektionsbias vara mycket stora för att bedömas som höga i detta avsnitt.

A2. Risk för behandlingsbias ("performance bias")

Med "behandlingsbias" avses systematiska fel som är relaterade till hur studien har behandlat personer som tillhör interventionsgruppen, respektive jämförelsegruppen.

Risk för behandlingsbias föreligger då interventions- eller kontrollgruppen exponeras för något annat än det som jämförelsen syftar till att mäta. Den effekt studien funnit kan då åtminstone delvis bero på dessa skillnader och på så vis snedvrider resultatet. Skillna-

der kan t ex avse felaktig behandling, ofullständig behandling, behandlingsavbrott eller tillägg utanför studieprotokollet. Risken för systematiska fel kan minska om det finns strukturerad kontroll av implementeringen (t ex en checklista eller en manual).

- A2a. Om studien syftar till att skatta effekten av en given behandling/riskfaktor (eventuellt i relation till alternativ behandling) bör kontrollgruppen exponeras för exakt samma sak som behandlingsgruppen bortsett från själva behandlingen. Om annat förekommer kan effekten överskattas eller underskattas. Detta gäller även effektens riktning, dvs risk för behandlingsbias föreligger. Finns det t ex socioekonomiska skillnader mellan behandlings- och kontrollgrupp? Risken är särskilt stor när det gäller preventiva eller symtomlindrande åtgärder som olika grupper av välinformerade individer kan efterfråga vilket kan göra att effekten/risken underskattas. Även om grupperna exponeras för olika faktorer som kan påverka utfallet på ett likvärdigt sätt, kan detta minska studiens känslighet för att upptäcka alternativt utesluta en effekt eller ett risksamband.
- A2b. Kontroll av följsamhet gentemot behandling alternativt av exponeringen är fundamental för trovärdigheten i uppnådda resultaten. Speciellt viktigt är detta i de fall resultatet pekar mot avsaknad av effekt/samband vilket kan bero på avsaknad av exponering för riskfaktorn eller interventionen. Den yttersta formen av låg följsamhet är avbrott av behandling eller exponering. Avbrott innebär att försökspersonen avbryter behandlingen eller avslutar exponeringen (utan att ha uppnått det studerade utfallet), men inte nödvändigtvis avbryter uppföljningen (= bortfall, se A4). Det är viktigt att kontrollera:
- totala andelen avbrott
 - skillnaden i andelen avbrott mellan grupperna
 - skillnader i orsak till avbrott mellan grupperna.

A3. Risk för bedömningsbias ("detection bias")

Med "bedömningsbias" avses systematiska fel som är relaterade till hur man i studien har hanterat mätningar av utfall och analys av resultat.

Risk för bedömningsbias föreligger då det finns skillnader i hur utfallen i interventions- respektive kontrollgruppen bestäms. Den effekt studien funnit kan då åtminstone delvis bero på dessa skillnader och snedvridda resultatet. Bedömningsbias, och därmed studie-kvaliteten som helhet, kan vara olika för olika utfallsmått i en och samma studie. Bedömning under A3 kan därför behöva ske separat för olika utfallsmått i samma studie.

Svaret på vissa av delfrågorna kan göra att andra delfrågor blir mindre eller inte alls relevanta. Exempelvis är relevansen för frågor om blindning (A3b) beroende av hur robust utfallsvariabeln är (A3a).

- A3a. Risken för systematiska fel ökar ju mer subjektiva inslag som finns i bedömningen av utfallet. Medan överlevnad/död är robusta utfallsmått är symtomskalor och livskvalitetsmätningar mycket känsliga för systematiska fel.
- A3b. Om personerna som mäter utfallen (patologen, röntgenologen, psykologen) eller som utvärderar resultatet av mätningen ("forskaren") känner till vilka försökspersoner som fått en viss behandling/exponering kan det öka risken för systematiska fel.
- A3c. Om samma personal som deltar i behandling eller i studiens genomförande också bedömer utfallet ökar risken för systematiska fel. Vid tillfredsställande blindning är dock opartiskhet av liten betydelse.
- A3d. Här handlar det ofta om hur så kallade kompositmått, dvs kombinerade effektmått, är sammansatta eller vilken koppling olika surrogatmått har till klinisk relevans. Vid negativt utfall är det viktigt att det valda utfallsmåttet är tillräckligt känsligt och att konfidensintervallet är tillräckligt smalt för att det ska vara möjligt att utesluta en effekt av klinisk relevant storlek.
- A3e/f. Risken för systematiska fel minskar om mätningen sker med hjälp av en standardiserad eller definierad metod som validerats med avseende på den aktuella populationen.
- A3g. Möjligheten att upptäcka (liksom att utesluta) effekter/samband ökar om exponeringen uppskattats vid upprepade (optimala) tidpunkter under studien.
- A3h. Felaktigt val av tidpunkt för mätning kan göra att utfallet underskattas. Detta är särskilt viktigt vid "non-inferiority" ("inte sämre än")-studier eller då slutsatsen är att effekt saknas.
- A3i. Vid utfallsregistrering kan observatörsvariationen vara en svaghet. Ett exempel är om flera observatörer ska utvärdera röntgenbilder eller cytologiska prov. Då ska observatörsöverensstämmelse mellan alla, eller ett större antal av observatörerna, vara rapporterat. Detta kan ske i form av kappa-överensstämmelse eller "intra-class correlation coefficient" (ICC), beroende på vilken skala som använts.

A3j. De vanligaste mått som används för dikotoma variabler (t ex ja–nej-variabler) är:

- riskkvot ("risk ratio", RR),
- oddskvot ("odds ratio", OR),
- absolut riskreduktion/riskskillnad ("risk difference"), och
- "number needed to treat" (NNT).

"Hazard ratio" (HR) används för att analysera risken över tid.

För kontinuerliga variabler används vanligen absolut skillnad i medelvärde ("difference in means", "mean difference"), standardiserad medelskillnad ("standardized mean difference" \approx "Cohen's d" \approx "Hedges g"), alternativt definieras gränsen för respons och utfallet rapporteras som "responder rate". Vid sådan dikotomisering av kontinuerliga variabler är det viktigt att intervallens gräns(er) motiverats trovärdigt eller är "gängse".

Alla måtten (helst differensen mellan grupperna) ska redovisas med lämpligt precisionsmått, företrädesvis 95 procents konfidensintervall. Bedöm om konfidensintervall eller andra relevanta mått redovisas på ett adekvat sätt eller om det finns en motivering för att sådana uppgifter saknas. Det kan t ex gälla vid totalundersökningar av stora datamaterial.

A4. Bortfallsbias ("attrition")

Med "bortfallsbias" avses systematiska fel som är relaterade till hur studien har hanterat bortfall, dvs personer som har gått med på att delta i en undersökning men som lämnar denna innan deras medverkan/uppföljning är klar. Den engelska termen är "loss to follow-up".

Risk för bortfallsbias föreligger då det finns skillnader i bortfallet mellan interventions- och kontrollgruppen. Den effekt studien funnit kan då åtminstone delvis bero på dessa skillnader och på så vis snedvrider resultatet. Ett generellt stort bortfall, skillnader i bortfallstorlek samt framför allt orsaksskillnader till bortfall ökar risken för systematiska fel. Det bortfall som bedöms här avser bortfall efter inklusion i studien. Man kan aldrig räkna med att bortfall är slumpmässigt.

Stora bortfall ökar generellt sett risken för att resultaten kan vara påverkade av systematiska fel. Bortfallet kan variera mellan olika tidpunkter och olika effektmått. Bortfallsanalysen görs därför separat för de aktuella utfallen. Vid långtidsuppföljning kan man få acceptera något högre bortfall.

Bortfallet kan variera mellan olika tidpunkter i en studie och mellan olika effektmått. Bortfallet är ofta större ju längre tid som har gått. Därmed kan behandlingsresultaten från de sista mättillfällena vara av tveksam validitet, medan resultaten från de första mättillfällena kan vara giltiga.

- A4a. Som ett riktvärde för läkemedelsstudier är risken liten om bortfallet är mindre än 10 procent, medelstor om bortfallet ligger mellan 10 och 19 procent och stor om bortfallet är mellan 20 och 29 procent. Om bortfallet i läkemedelsstudier är 30 procent eller mer är informationsvärdet tveksamt och studien kan eventuellt sorteras bort. Notera att andra värden kan gälla om det inte är läkemedelsstudier. Bortfallet måste också ställas i relation till storleken (och skillnaden) i utfallet. Ju lägre utfall desto större problem även med små bortfall.
- A4c. Skillnader i baslinjevariabler hos bortfall i interventions- och kontrollgrupp alternativt grupper med olika exponering för riskfaktorer är allvarliga eftersom de kan snedvrیدا utfallet särskilt om det rör baslinjefaktorer med direkt koppling till utfallet (t ex sjukdomsstadium vid överlevnadsutfall).
- A4d. Om sammansättningen av personer i bortfallet skiljer från dem som finns kvar i studien, kan det påverka studiens möjlighet att upptäcka relevanta effekter och överförbarhet (t ex att patienter med progredierande sjukdom inte orkar fylla i livskvalitetsfrågeformulär).
- A4e. Vid analys av studier med bortfall används olika så kallade imputeringsmetoder (dvs hur man ersätter missade mätningar, t ex ”last observation carried forward” (LOCF), ”observed cases” (OC) eller interpoleringar). Det är viktigt att utfall med olika imputeringsmetoder redovisas alternativt att man använt den metod som är minst gynnsam för utfallet (konservativ). Detta kan förvisso göra att storleken på effekten underskattas. I så kallade ”non-inferiority”-studier ska man tvärtom använda den imputeringsmetod som gynnar utfallet eftersom man annars kan komma fram till en felaktig slutsats om frånvaro av effekt/skillnad.

A5. Rapporteringsbias (”reporting bias”)

Med ”rapporteringsbias” avses systematiska fel som är relaterade till hur studien har hanterat protokoll och rapportering.

Den effekt studien funnit kan åtminstone delvis bero på att vissa resultat rapporteras, medan andra inte rapporteras. Effekten riskerar då att såväl överskattas som underskattas. Även effektens riktning kan ha påverkats.

- A5a. Tillgång till studiens protokoll är av stort värde för att bedöma betydelsen av rapporterade fynd eftersom det inte är ovanligt att studier med *negativa* resultat inkluderar förklarande efteranalyser ("explanatory"- eller post hoc-analyser) för att t ex finna vissa subgrupper inom den studerade patientgruppen som kan ha nytta av behandlingen alternativt hos vilka samband identifieras. Dessa analyser kan fylla en viktig hypotesgenererande funktion, men slutsatserna i en primärt *negativ* studie får aldrig baseras på sådana subgruppsanalyser. När en studie visar ett *positivt* utfall för sitt primära utfallsmått är däremot subgruppsanalyser av stort värde för att bedöma generaliserbarheten av resultatet.
- A5b. Det är viktigt att det går att klargöra vilka utfall som mätts, analyserats respektive rapporterats. Utfall som mätts eller analyserats men inte rapporterats och därför inte tagits hänsyn till i den statistiska analysen gör att betydelsen av interventionen/sambandet kan missbedömas.
- A5c. En undermålig rapportering av risker med en intervention riskerar att överskatta dess ändamålsenlighet (nytta/risk).
- A5d. Olika effekter kan ha mätts vid upprepade tillfällen men det är viktigt att det inte gjorts fler *analyser* av studien än vad som angetts i protokollet (och den statistiska planen medger). Det är också viktigt att det framgår om den redovisade analysen är en slutanalys eller en förplanerad interimanalys. Ad hoc interimanalyser är självklart mycket problematiska speciellt i öppna studier där de kan misstänkas vara datadrivna. Även "lege artis" interimanalyser riskerar att överskatta effekter av en intervention. När det gäller studier som inte påvisar någon effekt av en intervention är det viktigt att tidpunkten för analys är optimalt vald för att kunna påvisa en möjlig effekt.

A6. Intressekonflikter ("other considerations")

Om författare till studien kan vinna något på ett givet resultat, kan detta medföra en överskattning eller underskattning av effekten i den riktning som författaren skulle vinna på. Det kan t ex vara problematiskt om författarna själva har utvecklat den intervention som studerades.

Sammanvägning

För att bedöma den sammanvägda evidensen med hjälp av GRADE, krävs att även nedanstående faktorer vägs samman i en slutgiltig bedömning.

B. Bristande överensstämmelse mellan studierna ("heterogeneity")

Görs t ex om möjligt i form av metaanalyser eller liknande, men utgår i granskning av enskild studie.

C. Bristande överförbarhet ("indirectness of evidence")

Med "överförbarhet" avses möjligheten att tillämpa studiens upplägg, diskussion och resultat i svenska förhållanden.

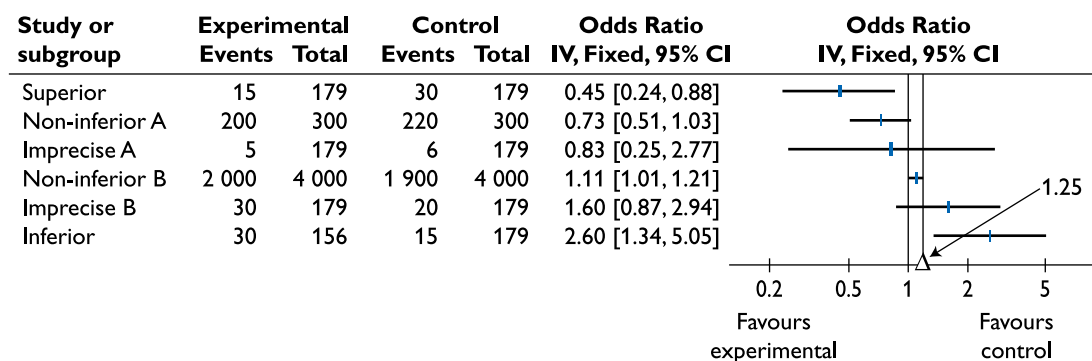
Om population, intervention, kontrollalternativ eller effektmått i en studie avviker från dem som specificerats som adekvat för svenska förhållanden och därmed frågeställningen i översikten, så föreligger överförbarhetsproblem. Den effekt studien funnit kan då åtminstone delvis avvika från den "sanna" effekten med avseende på hur population, intervention, kontrollgruppens villkor eller effektmått som specificerats i översikten. Effekten kan alltså såväl underskattas som överskattas vad gäller svenska förhållanden.

Det är viktigt att populationen i studien motsvarar den population som är aktuell i SBU/HTA-rapporten.

För att kunna bedöma överförbarhet med hjälp av GRADE för ett sammanvägt effektmått krävs att de ingående studierna beaktas som en helhet.

D. Bristande precision ("imprecision")

Här beaktas två aspekter av precision. För det första, om syftet är att testa om interventionen är bättre än kontrollvillkoret så räcker det här med att studera om konfidenstervall täcker linjen för "ingen skillnad" ("1" vid binära utfallsmått samt "0" vid kontinuerliga utfallsmått). Täcks denna linje är precisionen bristande. Resultaten i "Superior", "Non-inferior B" och "Inferior" har god precision i detta avseende (Figur B3.1). För det andra, om syftet är att testa huruvida interventionen inte är sämre än kontrollinterventionen (ofta rörande biverkningar), krävs även en i förväg



Figur B3.1 Illustration av olika tester med skogsdiagram ("forest plot").

kliniskt definierad gräns för hur mycket sämre interventionen får vara utan att det är ett problem ("suggested appreciable harm", kliniskt relevant skillnad). Om konfidensintervallet inte täcker denna gräns är precisionen god och man kan då dra slutsatsen att interventionen inte var sämre än kontrollgruppen. I Figur B3.1 har gränsen satts till 1,25. Tre exempel på resultat som kan illustrera detta är "Superior", "Non-inferior A" samt "Non-inferior B". Exempel på dålig precision utgörs av "Imprecise A" och "Imprecise B". Observera att datakvaliteten är viktig vid bedömning av precisionen i "non-inferiority"-utfall. Exempelvis kan en dålig rapportering av biverkningar göra att resultatet ser ut att vara lika i båda behandlingsarmarna.

Finns det fler studier som är lämpliga att väga samman ska det sammanvägda konfidensintervallet beaktas.

E. Publikationsbias

Sammanvägs på syntesnivå.

F. Effektstorlek

Sammanvägs på syntesnivå. Av praktiska skäl kan det vara bra att notera resultatet för den enskilda studien i granskningsmallen.

G. Dos–respons samband

Slutgiltiga bedömningen hanteras på syntesnivå.

H. Sannolikhet att effekten är underskattad pga "confounders"

Sammanvägs på syntesnivå. Av praktiska skäl kan det vara bra att notera resultatet för den enskilda studien i granskningsmallen. Vid enstaka tillfällen kan evidensstyrkan justeras upp om det är mycket sannolikt att studierna underskattat effekten. Det kan gälla när "confounders" som studien inte kunnat justera för, talar för att effekten är underskattad.

Bilaga 2. Mall för kvalitetsgranskning av randomiserade studier

VERSION 2012:2.1

Granskningen av en studie gäller i första hand studiekvalitet, dvs risk för systematiska fel och risk för intressekonflikter (A). I den sammanvägda bedömningen av alla inkluderade studier enligt GRADE inkluderar man också studiernas överensstämmelse (B), överförbarhet (C), precision (D), publikationsbias (E), effektstorlek (F), dos-respons samband (G) och sannolikhet att effekten är underskattad (H).

Författare: _____ År: _____ Artikelnummer: _____

Alternativet ”oklart” används när uppgiften inte går att få fram från texten. Alternativet ”ej tillämpligt” väljs när frågan inte är relevant. Specificera i kommentarsfältet.

A. Granskning av studiens begränsningar – eventuella systematiska fel (bias)	Ja	Nej	Oklart	Ej tillämpligt
A1. Selektionsbias				
a) Användes en lämplig randomiseringsmetod?				
b) Om studien har använt någon form av begränsning i randomiseringsprocessen (t ex block, strata, minimisering), är skälen till detta adekvata?				
c) Var grupperna sammansatta på ett tillräckligt likartat sätt?				
d) Om man har korrigerat för obalanser i baslinjevariabler, har det skett på ett adekvat sätt?				
Kommentarer:				
Bedömning av risk för selektionsbias:				
A2. Behandlingsbias				
a) Var studiedeltagarna blindade?				
b) Var behandlare/prövare blindade?				
c) Var följsamhet i grupperna acceptabel enligt tillförlitlig dokumentation?				
d) Har deltagarna i övrigt behandlats/exponerats på samma sätt bortsett från interventionen?				
Kommentarer:				
Bedömning av risk för behandlingsbias:				

A. Fortsättning	Ja	Nej	Oklart	Ej till- lämpligt
A3. Bedömningsbias (per utfallsmått)				
a) Var utfallsmåttet okänsligt för bedömningsbias?				
b) Var de personer som utvärderade resultaten blindade för vilken intervention som gavs?				
c) Var personerna som utvärderade utfallet opartiska?				
d) Var utfallet definierat på ett lämpligt sätt?				
e) Var utfallet identifierat/diagnostiserat med validerade mätmetoder?				
f) Har utfallet mätts vid optimala tidpunkter?				
g) Var valet av statistiskt mått för rapporterat utfall lämpligt?				
h) Var den analyserade populationen (ITT eller PP) lämplig för den fråga som är föremål för studien?				
Kommentarer:				
Bedömning av risk för bedömningsbias:				
A4. Bortfallsbias (per utfallsmått)				
a) Var bortfallet tillfredsställande lågt i förhållande till populationens storlek?				
b) Var bortfallet tillfredsställande lågt i förhållande till storleken på utfallet?				
c) Var bortfallets storlek balanserad mellan grupperna?				
d) Var relevanta baslinjevariabler balanserade mellan de som avbryter sitt deltagande och de som fullföljer studien?				
e) Var den statistiska hanteringen av bortfallet adekvat?				
f) Var orsakerna till bortfallet analyserade?				
Kommentarer:				
Bedömning av risk för bortfallsbias:				

A. Fortsättning	Ja	Nej	Oklart	Ej till- lämpligt
A5. Rapporteringsbias				
a) Har studien följt ett i förväg publicerat studieprotokoll?				
b) Angavs vilket/vilka utfallsmått som var primära respektive sekundära?				
c) Redovisades alla i studieprotokollet angivna utfallsmått på ett fullständigt sätt?				
d) Mättes biverkningar/komplikationer på ett systematiskt sätt?				
e) Redovisades enbart utfallsmått som angivits i förväg i studieprotokollet?				
f) Var tidpunkterna för analys angivna i förväg?				
Kommentarer:				
Bedömning av risk för rapporteringsbias:				
A6. Intressekonflikter				
a) Föreligger, baserat på författarnas angivna bindningar och jäv, låg eller obefintlig risk att studiens resultat har påverkats av intressekonflikter?				
b) Föreligger, baserat på uppgifter om studiens finansiering, låg eller obefintlig risk att studien har påverkats av en finansiär med ekonomiskt intresse i resultatet?				
c) Föreligger låg eller obefintlig risk för annan form av intressekonflikt (t ex att författarna har utvecklat interventionen)?				
Kommentarer:				
Bedömning av risk för intressekonflikt:				

Sammanvägning av risk för bias (per utfallsmått)	Låg	Medelhög	Hög
A1. Selektionsbias			
A2. Behandlingsbias			
A3. Bedömningsbias			
A4. Bortfallsbias			
A5. Rapporteringsbias			
A6. Intressekonfliktbias			
Kommentarer:			
Sammanfattande bedömning av risk för systematiska fel (bias):			

Underlag för sammanvägd bedömning enligt GRADE

B. Bristande överensstämmelse mellan studierna				
Hanteras endast på syntesnivå				
C. Granskning av studiens överförbarhet	Ja	Nej	Delvis	Ej till- lämpligt
a) Överensstämmer sammanhanget och kontrollgruppens villkor med den situation som SBU/HTA-rapportens slutsatser avser?				
b) Är den inkluderade studiepopulationen tillräckligt lik den population som SBU/HTA-rapportens slutsatser avser?				
c) Är interventionen relevant för de förhållanden som SBU/HTA-rapportens slutsatser avser?				
Kommentar:				
Bedömning av brister i överförbarhet:				
D. Granskning av precision	Ja	Nej	Delvis	Ej till- lämpligt
a) Är precisionen acceptabel med hänsyn till antal inkluderade individer och antal händelser (utfall)?				
Kommentar:				
E. Granskning av publikationsbias				
Hanteras endast på syntesnivå				

F. Granskning av effektstorlek	Ja	Nej	Delvis	Ej till- lämpligt
a) Var effekten stor (t ex RR <0,5 eller >2,0)?				
b) Var effekten mycket stor (t ex RR <0,2 eller >5,0)?				
Kommentar:				

G. Granskning av dos–respons samband	Ja	Nej	Delvis	Ej till- lämpligt
a) Finns stöd för ett dos–respons samband mellan exponering och utfall?				
Kommentar:				

H. Sannolikhet att effekten är underskattad pga ”confounders”
Inte aktuellt på RCT:er

Mall för kvalitetsgranskning av randomiserade studier: förklaringar

Granskningsmallen avser att ge ett systematiskt underlag med vars hjälp man kan bedöma risken för att ett givet utfall man skattat i en studie systematiskt snedvridits (bias) under forskningsarbetet. Konsekvensen av detta är att utfallet antingen underskattas eller överskattas jämfört med ett ”sant” utfall. Även utfallets riktning kan ha missbedömts.

Syftet med mallen är att skapa ett systematiskt och transparent underlag för att diskutera hur stor risken är att skattade utfall i en enskild studie är systematiskt snedvridna. Någon algoritm för att räkna samman kvalitetspoäng erbjuds alltså inte. När det gäller bedömningsbias (A3) och bortfallsbias (A4) behöver granskningen ske per utfallsmått eftersom kvalitetsbrister/bias kan skilja sig för olika utfallsmått.

För att resultaten ska kunna användas för evidensgradering enligt GRADE krävs ytterligare information i form av sammanställningar på syntesnivå, alltså sammanvägning om fler än en studie finns. I vissa fall kan sammanställningar endast ske på syntesnivå, t ex bristande samstämmighet (”inconsistency”), precision och publikationsbias.

A. Granskning av studiens begränsningar – eventuella systematiska fel

A1. Risk för selektionsbias ("selection bias")

Med "selektionsbias" avses systematiska fel som är relaterade till hur studien har hanterat urval av försökspersoner (motsvarande) samt indelning i interventions- och kontrollgrupper.

Risk för selektionsbias föreligger då interventionsgruppen respektive kontrollgruppen inte är tillräckligt lika varandra vid baslinjen avseende kända såväl som okända risk- och skyddsfaktorer. Det utfall man funnit i studien kan då åtminstone delvis bero på dessa skillnader och på så vis snedvrیدا resultatet. Randomiseringen bör ske på ett oförutsägbart sätt och processen bör inte vara möjlig att manipulera. Detta kan förhindras t ex genom att allokeringen sker med hjälp en datorgenererad slumpmässigt serie samt genom att processen är maskerad t ex med slutna kuvert.

Ibland begränsas randomisering för att åstadkomma lika stora grupper (t ex blockrandomisering) eller för att skapa balans mellan grupperna avseende sådana egenskaper hos deltagarna som kan påverka resultaten (t ex stratifierad randomisering). Detta kan öka förutsägbarheten avseende vilken grupp en given individ kommer att hamna i. Detta gäller speciellt om blocken är små respektive om varje stratum innehåller få personer.

A1d. Post hoc-justering av utfallet baserat på skillnader i kända baslinjefaktorer är kontroversiellt. Om det används som ett sätt att testa känsligheten av ett positivt utfall är det okej, men för att ändra ett negativt utfall i grundanalysen till ett positivt krävs mycket hållbara argument.

A2. Risk för behandlingsbias ("performance bias")

Med "behandlingsbias" avses systematiska fel som är relaterade till hur personer som tillhör interventions- respektive kontrollgruppen har behandlats i studien.

Risk för behandlingsbias föreligger då interventions- eller kontrollgruppen exponeras för något annat än det som jämförelsen syftar till att mäta, t ex annan behandling mot aktuell sjukdom än godkänd standardbehandling. Det utfall man funnit i studien kan då åtminstone delvis bero på dessa skillnader och på så vis snedvrیدا resultatet.

Om man vill skatta effekten av en given behandling bör kontrollgruppen (placebo- eller obehandlad kontroll) exponeras för exakt samma sak som behandlingsgruppen bortsett från själva behandlingen. Om annat förekommer kan effekten som redovisas i studien överskatta eller underskatta den sanna effekten, detta gäller även effektens riktning, dvs risk för behandlingsbias föreligger.

Om man vill skatta effekten av en behandling jämfört med en alternativ (aktiv) behandling bör ingen av grupperna exponeras för något annat än det som ingår i de båda behandlingarna. Om annat förekommer kan effekten i studien överskatta eller underskatta den sanna effektskillnaden, detta gäller även effektens riktning, dvs risk för behandlingsbias förekommer.

Skillnader kan avse felaktig behandling, ofullständig behandling, behandlingsavbrott, tillägg utanför studieprotokollet m m. Risken för bias kan minska om behandlare och patienter är ovetande om gruppindelningen (blindad studie) och om det finns strukturerad kontroll av implementeringen (t ex en checklista eller en manual).

A2a/b. Det är önskvärt att både patienter, prövare (och utvärderare, se A3b) är blindade i en studie. Ibland kan det av praktiska skäl vara svårt eller omöjligt att dölja för prövare och/eller patient vilken behandling som ges. Blindningen kan också misslyckats pga karakteristiska effekter eller biverkningar av aktiv behandling, exempelvis muntorrhet vid behandling med neuroleptika och underlivsblödningar vid behandling med östrogen. I vissa fall är det möjligt att ge biverkningsmotverkande medel som tillägg till aktiv behandling för att minska risken för att blindningen äventyras. Andra faktorer som kan försvåra blindningen är bristande likhet mellan tabletter, inhalationspreparat etc avseende utseende eller smak. En stor ”placeboeffekt” i kontrollgruppen kan tala för en lyckad blindning. I vissa studier låter man studiedeltagarna gissa om de fått aktiv behandling eller kontroll.

A2c. Kontroll av följsamheten är särskilt viktig då det saknas en signifikant effektskillnad i utfall mellan grupperna. En bristande följsamhet kan minska såväl interventionens effekter som bieffekter. Detta är alltså extra viktigt vid så kallade ”non-inferiority” (”inte sämre än”)-studier men om interventionen har en signifikant effekt är kontroll av följsamheten ofta av mindre betydelse. Undantag är om det var sämre följsamhet i gruppen som fick referensbehandling. Det senare är tänkbart i en placebokontrollerad studie om blindningen varit otillräcklig, alternativt om en referensbehandling har mycket högre frekvens av biverkningar.

A3. Risk för bedömningsbias (”detection bias”)

Med ”bedömningsbias” avses systematiska fel som är relaterade till hur studien har hanterat genomförande av mätningar och analys av resultat.

Risk för bedömningsbias föreligger då det finns skillnader i hur utfallen i interventions- respektive kontrollgruppen bestäms. Det utfall man funnit i studien kan då åtminstone delvis bero på dessa skillnader och på så vis snedvrider resultatet. Bedömningsbias, och

därmed studiekvaliteten som helhet, kan vara olika för olika utfallsmått i en och samma studie. Bedömning under A3 kan därför behöva göras separat för olika utfallsmått i samma studie.

- A3a. Risken för bias ökar ju mer subjektiva inslag som finns i bedömningen av utfallet. Medan överlevnad/död är robusta utfallsmått är symtomskalor och livskvalitetsmätningar mycket känsliga för bias och i princip oanvändbara i oblindade studier.
- A3b. Förutom att den som utvärderar studien är blindad är det också viktigt att det framgår av beskrivningen att all resultatbearbetning utfördes innan prövningskoden bröts.
- A3c. I randomiserade studier är ju ofta prövare och utvärderare samma personer, men i större högkvalitativa studier finns ibland oberoende kommittéer (DSMB) som tar ställning till och utvärderar utfallet.
- A3d. Här handlar det ofta om hur så kallade kompositmått, dvs kombinerade utfallsmått, är sammansatta eller vilken koppling till klinisk relevans som finns för olika surrogatmått.
- A3e. Om mätningen sker med hjälp av en standardiserad metod som validerats med avseende på den aktuella populationen minskar risken för bias.
- A3f. Val av mättidpunkt för att optimera möjligheten att upptäcka en skillnad i utfall är särskilt viktigt i så kallade ”non-inferiority”-studier.
- A3g. De vanligaste mått som används för dikotoma variabler, exempel ja–nej-variabler, är riskkvot (”risk ratio”, RR), oddskvot (”odds ratio”, OR), absolut riskreduktion/riskskillnad (”risk difference”) och ”number needed to treat” (NNT). ”Hazard ratio” (HR) används för att analysera risken över tid. För kontinuerliga variabler används vanligen absolut skillnad i medelvärde (”difference in means”, ”mean difference”) alternativt definieras gränsen för respons och utfallet rapporteras som ”responder rate”. Alla måtten (helst differensen mellan grupperna) ska redovisas med lämpligt precisionssmått, företrädesvis 95 procents konfidensintervall.
- A3h. Resultaten kan analyseras enligt ”intention to treat” (ITT) och/eller per protokoll (PP). En ITT-analys innebär att alla personer som randomiserats följs upp inom sin behandlingsarm oavsett om de fått tilldelad behandling eller inte och är oftast den metod som bör användas. Om resultaten är beräknade på annat sätt än med

ITT finns det risk för att behandlingseffekten blir överskattad. ITT-analysen kan kompletteras med en känslighetsanalys enligt ”worst case scenario” där sämsta tänkbara utfall tillskrivs saknade patienter i den grupp som uppvisar bäst effekt och bästa tänkbara utfall tillskrivs saknade patienter i den grupp som uppvisar sämst effekt. Ibland (speciellt ”non-inferiority”-studier) är det viktigt att även en PP-analys redovisas, vilket innebär att bara de som följt hela studieprotokollet ingår i analysen.

A4. Bortfallsbias (”attrition”)

Med ”bortfallsbias” avses systematiska fel som är relaterade till hur studien har hanterat bortfall, dvs personer som har gått med på att delta i en undersökning men som lämnar denna innan den fullbordas.

Risk för bortfallsbias föreligger då det finns skillnader i bortfallet mellan interventions- och kontrollgruppen. Det utfall man funnit i studien kan då åtminstone delvis bero på dessa skillnader och på så vis snedvrیدا resultatet. Ett generellt stort bortfall, skillnader i bortfallstorlek samt framför allt orsaksskillnader till bortfall ökar risken för bias. Det bortfall som bedöms här avser bortfall efter randomisering. Man kan aldrig räkna med att bortfall är slumpmässigt. Om sammansättningen av personer i bortfallet inte skiljer sig från dem som finns kvar i studien, är dock en bättre situation än om det finns signifikanta skillnader. Nedanstående exempel kan tjäna som *grova* riktvärden:

- litet (<10 %)
- måttligt (10–19 %)
- stort (20–29 %)
- mycket stort (≥30 %). Undersökningen bedöms ofta sakna informationsvärde vilket kan innebära att studien bör exkluderas.

Bortfallet måste också ställas i relation till storleken (och skillnaden) i utfallet. Ju lägre utfall desto större problem även med små bortfall.

Bortfallet kan variera mellan olika tidpunkter i en studie och mellan olika utfallsmått. Bortfallet är ofta större ju längre tid som har gått. Därmed kan behandlingsresultaten från de sista besöken vara av tveksam validitet, medan resultaten från de första besöken kan vara giltiga.

A4e. Vid analys av studier med bortfall används olika så kallade imputeringsmetoder (dvs hur man ersätter missade mätningar, t ex ”last observation carried forward” (LOCF), ”observed cases” (OC) eller interpoleringar). Det är viktigt att utfall med olika imputeringsmetoder redovisas alternativt att man använt den metod

som är minst gynnsam för utfallet (konservativ). Detta kan förvisso göra att storleken på effekten underskattas. I så kallade "non-inferiority"-studier ska man tvärtom använda den imputeringsmetod som gynnar utfallet eftersom man annars kan komma fram till en felaktig slutsats om frånvaro av effekt/skillnad.

A5. Rapporteringsbias ("reporting bias")

Med "rapporteringsbias" avses systematiska fel som är relaterade till hur studien har hanterat rapportering i relation till sitt protokoll.

Det utfall man funnit i studien kan åtminstone delvis bero på att endast vissa resultat rapporteras, medan andra inte rapporteras. Utfallet riskerar då att såväl överskattas som underskattas. Även utfallets riktning kan ha påverkats.

A5a. Det är inte ovanligt att studier med *negativa* resultat inkluderar förklarande efteranalyser ("explanatory"- eller post hoc-analyser) för att t ex finna vissa subgrupper inom den studerade patientgruppen som kan ha nytta av behandlingen. Dessa analyser kan fylla en viktig hypotesgenererande funktion, men slutsatserna i en *negativ* studie får aldrig baseras på sådana analyser. När en studie visar ett *positivt* utfall för sitt primära utfallsmått är däremot subgruppsanalyser av stort värde för att bedöma generaliserbarheten av resultatet.

A5c/d. Även om redovisade utfallsmått är rimliga, definierade i förväg och adekvat rapporterade kan det finnas andra viktiga utfallsmått som utelämnats. Oftast gäller det utfallsmått för att bedöma biverkningar/risker.

A5f. Det är viktigt att inte fler analyser av studien än vad som angetts i protokollet (och den statistiska planen medger) gjorts. Det är också viktigt att det framgår om den redovisade analysen är en slutanalys eller en förplanerad interimanalys. Ad hoc interimanalyser är självklart mycket problematiska speciellt i öppna studier där de kan misstänkas vara datadrivna.

A6. Intressekonfliktbias ("other considerations")

Om författare till studien kan vinna något på ett givet resultat, så kan detta medföra en överskattning eller underskattning av effekten i den riktning som författaren skulle vinna på.

Sammanvägning

För att dimensionen studiebegränsningar ska kunna beaktas när ett betyg sätts för ett sammanvägt utfallsmått med hjälp av GRADE, krävs att alla ovanstående former av risk för bias vägs samman. Detta sker med fördel i diskussion i expertgrupp.

B. Bristande överensstämmelse mellan studierna ("heterogeneity")

Hanteras på syntesnivå.

C. Bristande överförbarhet ("indirectness of evidence")

Med "överförbarhet" avses möjligheten att tillämpa studiens upplägg, diskussion och resultat på de förhållanden som SBU/HTA-rapporten avser.

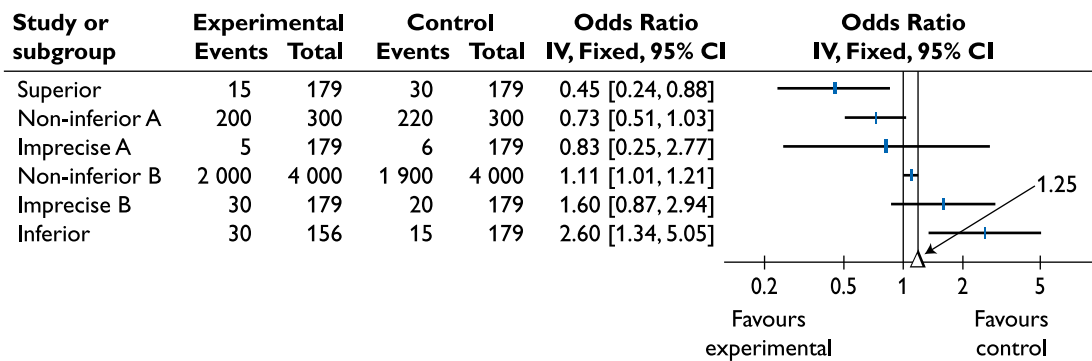
Om population, intervention, kontrollalternativ eller utfallsmått avviker från dem som specificerats i SBU/HTA-översikten föreligger överförbarhetsproblem. Det utfall man funnit i studien kan då åtminstone delvis avvika från det "sanna" utfallet med avseende på hur population, intervention, kontrollalternativ eller utfallsmått har specificerats i översikten. Utfallet kan alltså underskattas såväl som överskattas, vilket även gäller utfallets riktning.

Det är betydligt viktigare att studiepopulationen motsvarar den population man vill dra slutsatser om i SBU/HTA-rapporten, än om studiepopulationen inte är representativ med avseende på syftet i den enskilda studien (t ex beroende på bortfall före randomiseringen).

För att dimensionen överförbarhet ska kunna beaktas när ett betyg sätts med hjälp av GRADE för ett sammanvägt utfallsmått, krävs att ingående studier beaktas som en helhet.

D. Bristande precision ("imprecision")

Här beaktas två aspekter av precision. För det första, om syftet är att testa om interventionen är bättre än kontrollvillkoret räcker det här med att studera om konfidensintervallet täcker linjen för "ingen skillnad" ("1" vid binära utfallsmått samt "0" vid kontinuerliga utfallsmått). Täcks denna linje är precisionen bristande. Resultaten i "Superior", "Non-inferior B" och "Inferior" har god precision i detta avseende (Figur B2.1). För det andra, om syftet är att testa huruvida interventionen inte är sämre än kontrollinterventionen (ofta rörande biverkningar), krävs även en i förväg kliniskt definierad gräns för hur mycket sämre interventionen får vara utan att det är ett problem ("suggested appreciable harm", kliniskt relevant skillnad). Om konfidensintervallet inte täcker denna gräns är precisionen god och man kan då dra slutsatsen att interventionen inte var sämre än kontrollinterventionen. I Figur B2.1 har gränsen satts till 1,25. Tre exempel på resultat som kan illustrera detta är "Superior", "Non-inferior A" samt "Non-inferior B". Exempel på dålig precision utgörs av "Imprecise A" och "Imprecise B". Observera att datakvaliteten är viktig vid bedömning av precisionen i "non-inferiority"-utfall. Exempelvis kan en dålig rapportering av biverkningar göra att resultatet ser ut att vara lika i båda behandlingsarmarna.



Figur B2.1 Illustration av olika tester med skogsdiagram ("forest plot").

Finns det fler studier som är lämpliga att väga samman ska det sammanvägda konfidensintervallet beaktas.

E. Publikationsbias

Hanteras endast på syntesnivå.

F. Effektstorlek

Hanteras i första hand på syntesnivå. Om ingående studiers kvalitet har föranlett nedgradering kan uppgradering för effektstorlek komma ifråga endast efter noggrann övervägning.

G. Dos–respons samband

Sammanvägs på syntesnivå. Av praktiska skäl kan det vara bra att notera resultatet för den enskilda studien i granskningsmallen.

H. Sannolikhet att effekten är underskattad pga "confounders"

Inte aktuellt på RCT:er.