

REVIDERAD 2017

6 Kvalitetsgranskning av studier

Detta kapitel beskriver olika upplägg av behandlingsstudier, vilka metodologiska problem som är förknippade med respektive studietyp och hur man granskar studierna.

Randomiserade studier värderas generellt sett högst när undersökningen gäller behandlingseffekter. Anledningen är att de ger större förutsättningar för att kontrollera för faktorer som inte har med själva interventionen att göra. När det gäller att bedöma risker kan dock observationsstudier eller fall-kontrollstudier vara att föredra. Tvärsnittsstudier och fallserier utan kontroll har sämre tillförlitlighet och ingår sällan i SBU-projekt.

Avsnittet avslutas med metodproblem för systematiska översikter och hur översikterna granskas.

Granskningsmallar

Granskningen syftar till att bedöma i vilken utsträckning studiens resultat beror på systematiska fel (bias). SBU har utvecklat separata granskningsmallar (checklistor) för de olika studietyperna (Bilaga 2–3). Mallarna tar i ett antal frågor upp olika kvalitetsaspekter som påverkar studiernas tillförlitlighet. De är konstruerade så att det önskvärda svaret på varje fråga är ”ja”. I praktiken kan det ofta vara svårt att besvara frågorna. Information kan till exempel vara otydlig. Det är därför viktigt att projektgruppen tränar gemensamt på mallarna. Det ger möjlighet att reda ut otydligheter och hur frågorna ska tolkas. Det kan

vara värdefullt att testa överensstämmelsen i bedömning mellan olika personer under projektets lopp för att säkerställa att bedömningarna görs på likartat sätt.

Innan själva granskningen påbörjas bör projektgruppen besluta om några aspekter är speciellt viktiga för kvaliteten medan andra kanske inte ens är relevanta. Det är sällan som studier kommer att uppfylla samtliga kvalitetskriterier, utan bilden kommer att vara blandad. Mallen visar vilka brister en studie har – sedan måste granskarna avgöra i hur stor utsträckning bristerna påverkar tillförlitligheten.

Observera att granskningsmallarna *enbart är ett stöd* för att bedöma studiernas kvalitet. De är inte avsedda att användas så att man sätter krav på ett visst antal ”ja” för att studien ska åsättas en viss studiekvalitet.

Det kan vara klokt att spara den ifyllda granskningsmallen tillsammans med studien och att skriva ner de överväganden som gjorts i samband med granskningen. Detta underlättar vid den kommande skrivningen av rapporten.

För att bedöma risken för systematiska fel i enskilda studier används del A (A1–A6) i granskningsmallen. För att resultaten även ska kunna användas i en sammanvägd bedömning enligt GRADE krävs en del ytterligare information i form av sammanställningar av samtliga ingående studier. Det gäller bristande överensstämmelse mellan studierna (B), studiens överförbarhet (C) samt granskning av studiernas sammanlagda precision (D), publikationsbias (E), effektstorlek (F), dos–respons samband (G) och sannolikhet att effekten är underskattad (H). Sammanvägningen enligt GRADE sker vid ett senare tillfälle än vid granskningen av en enskild studie, men det kan vara lämpligt att vid läsningen av en enskild studie samtidigt kommentera dessa faktorer (B–H).

Randomiserad kontrollerad studie (RCT)

Randomisering innebär att man slumpmässigt fördelar studiedeltagarna mellan olika behandlingsalternativ, insatser eller diagnostiska test. Fördelen med randomisering är att slumpförfarandet ska leda till att alla andra faktorer än behandlingen ska fördelas lika på de båda grupperna. Skillnaden i effekt mellan försöks- och kontrollgrupperna är då sannolikt bara beroende på behandlingen (resonemanget gäller även för studier med mer än två olika grupper).

Det finns en internationell överenskommelse, CONSORT statement, om hur randomiserade studier bör redovisas [1]. CONSORT kan även ses som stöd för vilka aspekter av en studie som är viktiga att bedöma och är ett underlag för SBU:s granskningsmall. Mallen finns i Bilaga 2.

Randomisering

Randomiseringsförfarandet påverkar mycket hur tillförlitlig studien är. En bra randomisering leder till att riskfaktorer hos försökspersonerna i grupperna blir jämnt fördelade. Proceduren för randomiseringen måste därför vara detaljerat beskriven. Speciellt i äldre studier (publicerade före år 1990) är det vanligt att randomiseringen inte beskrivs alls. Här får man avgöra från fall till fall hur den bristande informationen påverkar studiekvaliteten. Gruppernas jämförbarhet bör, trots randomiseringen, alltid värderas (Exempel 6.1). Riskerna för selektionsbias tas upp i granskningsmallen (Bilaga 2, A1 Selektionsbias).

I en randomiserad studie där man prövade om behandling med alfablockerare alfuzosin ökade chansen för framgångsrik dragning av kateter råkade kontrollgruppen bli i genomsnitt fem år äldre jämfört med behandlingsgruppen [2]. Eftersom ålder visades vara den starkaste prediktorn för misslyckad kateterdragning oavsett behandling blev resultatet svårbedömt.

Exempel 6.1
En äldre kontrollgrupp.

Behandlingsbias och blindning

Den andra faktorn som har stor betydelse för studiens kvalitet är graden av blindning. De berörda personernas förväntningar på resultatet av behandlingen kan påverka utfallet eller mätningen av utfallet. Det gäller särskilt för mjuka variabler som till exempel livskvalitet eller personens skattning av symtombörda. Risken för påverkan på effektmått som död eller fraktur är lägre. Så många aktörer som möjligt i studien bör därför vara blindade, det vill säga inte känna till vilken behandling som ges (experimentell eller kontroll). Det ideala är om alla parter (behandlare/handläggare, patient/klient, den som mäter effekten och den som redovisar resultaten) är blindade, så kallad trippelblind studie.

I praktiken kan det vara svårt att blinda studien. Det gäller till exempel inom kirurgi och socialtjänstinterventioner. Om det är etiskt försvarbart kan man använda så kallad sham-kirurgi ("låtsaskirurgi") för att minimera systematiska fel på grund av patienternas förväntningar. Andra exempel är livsstilsinterventioner, psykoterapi och fysioterapi. Även om det inte går att blinda behandlare/handläggare och patient/klient går det att blinda dem som registrerar och utvärderar resultaten. Denna grad av blindning kan användas som ett minimikrav. Interventioner inom socialtjänstområdet är oftast inte heller möjliga att blinda gentemot deltagarna eller undersökaren medan den som utvärderar kan vara blind för vad som gjorts. Frånvaro av blindning kan också tänkas påverka annan understödande behandling eller andra åtgärder som patienten/klienten får och vilket slutligen skulle kunna snedvrída resultatet – en annan aspekt av behandlingsbias.

Frågor kring behandlingsbias återfinns i granskningsmallen (Bilaga 2, A2 Behandlingsbias).

Bedömning av resultatredovisningen

Inför bedömningen av hur tillförlitliga själva resultaten är (Bilaga 2, A3 Bedömningsbias och A5 Rapporteringsbias) behöver vi tillgång till uppgifter om grundförutsättningarna för studien. Vilket är det primära effektmåttet? Hur stor effekt förväntas behandlingen/interventionen/insatsen ge?

Det är sedan viktigt att de resultat som författarna lyfter fram baseras just på det primära effektmåttet. Om det inte finns någon statistiskt säkerställd skillnad i det primära effektmåttet är det inte ovanligt att resultaten avser ett annat effektmått där författarna observerat en skillnad. Andra sätt att försöka trola bort ett negativt resultat är att utföra post hoc subgruppsanalyser för att hitta statistiskt säkerställda skillnader mellan interventions- och kontrollarm.

Inför denna del av granskningen bör granskarna ha definierat vad som är minsta kliniskt relevanta effekt. Stora studier som påvisar en statistiskt signifikant men icke kliniskt relevant effekt får anses ha ett mindre värde.

Bortfall

Den tredje viktiga aspekten på studiekvalitet är bortfall (Bilaga 2, A4 Bortfallsbias). Tillförlitligheten är beroende av om de som ingår i studien följs upp och ingår i analysen. Ett stort bortfall är bland annat problematiskt om effektmåttet är symtombaserat. Det kan tänkas att personer utan symtomlindring/effekt avbryter studien i större utsträckning än de som förbättrats. Särskilt allvarligt är det om bortfallet skiljer sig mellan experiment- och kontrollgruppen. För läkemedelsstudier har SBU som riktmärke tillämpat följande ungefärliga gränser:

- <10 procent: bortfallet påverkar knappast tillförlitligheten och därmed inte heller studiekvaliteten.
- >30 procent: bortfallet påverkar tillförlitligheten så mycket att studien saknar informationsvärde. Studien exkluderas.

Projektgruppen bör definiera på förhand hur stora bortfall som kan accepteras och i vilken utsträckning de påverkar studiens kvalitet. I vissa fall kan bedömningen påverkas positivt om författarna med hjälp av till exempel en rimlig bortfallsanalys kan argumentera för att bortfallet inte stört resultatet. Man ska dock alltid bedöma bortfallets storlek i förhållande till antal händelser i kontrollgruppen. Ett lågt bortfall kan vara allvarligt om antalet händelser i kontrollgruppen är få.

I studier där man använder kontinuerliga variabler eller skalor använder man sig i vissa fall av beräkningsmetoden *LOCF* (last observation carried forward) för att kompensera för bortfall. Det senast uppmätta resultatet antas då gälla även för de senare tidpunkterna när data saknas. Det finns även andra statistiska metoder för att korrigera för bortfall. Känslighetsanalyser är sannolikt det mest användbara eftersom det ger en fingervisning om effekten kvarstår även under värsta tänkbara betingelser. Ett sådant är till exempel att ingen i bortfallet förbättrades. Kvarstod effekten även i det värsta scenariot?

Intressekonflikter

Studierna måste också granskas utifrån en analys av om eventuella ekonomiska eller andra intressen kan påverka risken för att resultaten inte är tillförlitliga (Bilaga 2, A6 Intressekonflikter). Sådana intressekonflikter gäller kanske framför allt industrisponsrade metoder (läkemedel, medicinteknik) men även när forskare studerar en metod som de själva utvecklat.

Underlag för sammanvägd bedömning enligt GRADE

I en samlad bedömning (Kapitel 10) granskas bristande överensstämmelse mellan studierna (B), studiens överförbarhet (C) samt granskning av studiernas precision (D), publikationsbias (E), effektstorlek (F), dos–respons samband (G) och sannolikhet att effekten är underskattad (H). Några kommentarer kring dessa bedömningar kan vara värda att göra. Granskningen av studiens överförbarhet ska till exempel fokusera på om studierna är tillräckligt lik den population som SBU-/HTA-rapporten behandlar.

Precisionen i data beror i huvudsak på studiens storlek, frekvensen händelser och effektstorleken. Små studier kan vara problematiska av flera skäl. Dels har de svårare att visa statistiskt signifikanta resultat, dels finns det risk för att de är mindre välplanerade än större. Risken för obalanser i kända och okända bakgrundsfaktorer ökar. För små studier finns också en stor risk för typ 2-fel (nollhypotesen accepteras fast den är falsk). Det innebär att författarna inte lyckas påvisa en sann behandlingseffekt eftersom studiepopulationen är för liten för att åstadkomma en statistiskt säkerställd effekt.

Det är viktigt att författarna angett ”tidpunkter” för slutanalys och eventuella interimspanalyser och hur dessa hanterats statistiskt. Annars kan man misstänka att författarna har adderat deltagare successivt till studien tills de har fått en statistisk signifikans.

Själva uträkningen av effekten kan i princip göras på två sätt. I en *per protokollanalys* beräknas resultaten enbart för de personer som har följt hela studieprotokollet (completers). *ITT-analys* (intention to treat) är en mer konservativ beräkningsmetod som syftar till att minska risken för överskattning av behandlingsresultatet. ITT innebär att alla personer som har randomiserats följs upp inom sin behandlingsgrupp oavsett om de har fått rätt behandling eller inte. ITT är den bästa metoden för att mäta effekter av en intervention. Däremot är ofta per protokollanalys att föredra för att mäta biverkningar, eftersom man då bara mäter biverkningarna för dem som har fått behandlingen eller exponerats för en riskfaktor. Utspädningseffekter kan annars göra att man missar eventuella risker. I non-inferiority-studier ska såväl per protokollanalysen som ITT-analysen redovisas.

Effektstorlek (F) och eventuella dos–respons samband (G) påverkar naturligtvis tilltron till resultaten.

Observationsstudier

Observationsstudier kan planeras och utföras på olika sätt och metodologiskt brukar man dela in dem i kohortstudier, fall–kontrollstudier och tvärsnittstudier. Kohortstudier som inkluderas i SBU:s granskningar ska vara kontrollerade, det vill säga ha en jämförelsegrupp.

Riktlinjer och råd för hur man bör värdera kohort- och andra observationsstudier har utvecklats internationellt, till exempel STROBE [3]. Den granskningsmall som SBU använder för observationsstudier finns i Bilaga 3.

När det gäller observationsstudier är kvalitetsproblemen oftast allvarigare för små studier, de med historiska kontroller och studier som inte justerat för viktiga förväxlingsfaktorer, så kallade confounders. Selektionsproblemen kan vara särskilt stora när det gäller preventiva och icke-akuta åtgärder där välinformerade studiedeltagare kan efterfråga specifika insatser.

Kohortstudier

Kontrollerade kohortstudier jämför en grupp som fått behandling eller utsatts för en risk med en grupp som fått alternativ eller ingen behandling respektive inte utsatts för en risk. Kohort betyder grupp och i en kohortstudie följer man individer över tiden framåt (prospektivt) för att se hur det går för dem. När studien startar insamlas oftast en mängd uppgifter om individerna i respektive grupp. Ålder, kön, socioekonomisk situation, levnadsvanor samt sjukdomar är till exempel viktiga och grundläggande uppgifter. Sambandet mellan rökning och flera allvarliga sjukdomar påvisades till exempel redan på 1950-talet genom att en kohortstudie följde hur det gick för brittiska läkare som rökte respektive inte rökte [4]. Med tanke på hur rökvanorna sedan minskat som en följd av denna kunskap är det sannolikt den enskilda studien som räddat flest liv i världen. Våra kunskaper om riskfaktorer för hjärt- och kärlsjukdomar som högt blodtryck, höga kolesterolvärden, rökning och fetma kommer från flera stora kohortstudier som till exempel Framinghamstudien i USA och ”1913 års män i Göteborg”.

En nackdel med kohortstudier är att de kan bli kostsamma och svåra att genomföra vid undersökningar av sällsynta tillstånd eller då det tar lång tid innan man kan mäta utfallet. Det krävs då stora studiepopulationer och långa uppföljningstider. Fall–kontrollstudier kan då vara en mer lämplig och kostnadseffektiv metodik (se nedan). I Sverige och framför allt övriga nordiska länder med väl utvecklade hälsodata- och kvalitetsregister finns dock ofta data redan insamlade för stora studiepopulationer. Det gör att kohortstudier många gånger även kan användas vid studier av sällsynta sjukdomar och krav på långa uppföljningstider (Exempel 6.2).

En dansk registerbaserad kohortstudie analyserade risken för autism bland barn som vaccinerats för mässling, påssjuka och röda hund. Totalt följdes mer än 537 000 barn under åtta års tid. Ingen ökad risk för autism kunde visas för de barn som vaccinerats jämfört med de ovaccinerade [5].

Exempel 6.2
Registerbaserad
kohortstudie.

Fall-kontrollstudier

Fall-kontrollstudier är tillbakablickande (retrospektiva) till sin karaktär. Här letar man först upp fallen, det vill säga individer som drabbats av utfallet (t.ex. sjukdom eller död). Därefter jämförs de med individer som inte drabbats av detta utfall, det vill säga kontrollgruppen. Man kan sedan studera om grupperna skiljer sig åt avseende riskfaktorer eller vilken behandling de fått. Fall och kontroller måste representera samma studiebas och urvalet av kontroller måste ske strikt oberoende av eventuell exponering för behandlingen.

Vanligen måste man samla in data retrospektivt. Datainsamlingen kan göras med hjälp av intervjuer av fall och kontroller eller med hjälp av patient/klientjournaler, eller registerdata (Exempel 6.3). Problemet med intervjuer är att man inte alltid kommer ihåg allt som hänt tidigare i livet och att man tenderar att komma ihåg eller återge minnen olika beroende på om man tillhör experiment- eller kontrollgrupp.

En svensk fall-kontrollstudie avsåg frågan om aspirin och NSAID-preparat kan reducera risken för magcancer [6]. Man intervjuade 567 personer med magcancer och 1 165 personer utan cancer (kontroller) om bland annat deras användning av smärtstillande medel. De som använde aspirin hade en minskad risk för magcancer (OR=0,7, oddskvot) efter kontroll för kön, ålder och socioekonomisk status. Risken minskade med ökad användning av aspirin. Däremot sågs inget samband mellan magcancer och andra smärtstillande medel.

Exempel 6.3
Fall-kontrollstudie.

Valet mellan fall-kontroll- eller kohortstudie betingas i stor utsträckning av praktiska och ekonomiska aspekter. Kohortstudier kan ofta belysa många olika hypoteser om en adekvat datainsamling skedde vid studiestart. Vid ovanliga utfall är fall-kontrollstudier ofta effektivare än andra studieupplägg, men vid ovanliga exponeringar fungerar de sämre för att påvisa samband.

Tvärsnittsstudier

Tvärsnittsstudier innebär att man mäter förhållanden vid en tidpunkt eller ett tillfälle. De är ett bra sätt för att få en uppfattning av prevalensen av olika tillstånd och om det finns en samvariation mellan olika exponeringar och tillstånd. Statistiska centralbyråns (SCB:s) årliga undersökningar om levnadsförhållanden

är ett bra exempel på en tvärsnittsstudie. Där intervjuas ett slumpmässigt urval av svenska folket årligen om en mängd förhållanden. Data har till exempel använts för att uppskatta prevalensen av tillstånd som inte alltid är lätt att uppskatta utifrån registerdata. Tvärsnittsstudier exkluderas ofta i SBU-projekt eftersom det oftast är svårt att säga något om tids- och orsakssambanden mellan intervention/exponering och utfall.

Kvalitetsvärdering av observationsstudier (se Bilaga 3 för granskningsmall)

Det generellt svåraste metodproblemet att hantera vid observationsstudier jämfört med randomiserade kontrollerade studier är att man måste kontrollera för alla potentiellt viktiga faktorer som kan påverka utfallet. Grupperna i observationsstudier kan skilja sig systematiskt från varandra i viktiga avseenden som kan påverka utfallet (selektionsbias). Om man tillförlitligt mätt dessa skillnader på individnivå går det att med statistiska metoder kontrollera för dem. Om de inte har mätts går det inte att utesluta att eventuella effekter som observerats beror på bakomliggande skillnader mellan grupperna snarare än den behandlingsinsats man vill studera.

Selektionsbias

Såväl den behandlare/handläggare som den behandlade eller brukaren kan påverka vilka som hamnar i den exponerade (försöksgruppen) respektive oexponerade gruppen (kontrollgruppen). Behandlaren kan välja ut deltagare som bedöms ha störst nytta av åtgärden. Dessa kan vara såväl i bättre som sämre utgångsläge än de som inte får behandlingen. Välinformerade studiedeltagare som ofta är friskare eller har en lättare problematik kan ha större kunskap om olika behandlingsalternativ och kan därför kräva specifika insatser som andra inte känner till. Kvalitetsgranskningen av observationsstudier måste därför ha fokus på att bedöma om grupperna är jämförbara, eller om det finns förväxlingsfaktorer (confounders, se nedan) eller andra systematiska fel (Bilaga 3, A1 Selektionsbias). Det är ofta särskilt viktigt att justera analyserna med hänsyn till studiedeltagarnas hälsotillstånd och socioekonomiska förhållanden [7].

Man kan aldrig helt utesluta risken för selektionsbias vid observationsstudier, men när det gäller okända eller oväntade biverkningar är det dock mindre sannolikt [8]. Det kan bland annat bero på att behandlaren i sådana fall inte har kunskap om riskerna och därigenom inte samma möjlighet att selektera vilka som ska få respektive inte få interventionen i fråga sett ur ett riskperspektiv. En illustration av hur resultaten kan påverkas av selektionsbias finns i Exempel 6.4. En systematisk metaanalys kring biverkningar visade att det inte fanns någon skillnad i effektstorlek mellan randomiserade studier och observationsstudier [9]. Författarnas slutsats var att systematiska översikter av biverkningar inte ska begränsas till någon specifik studietyp.

Förväxlingsfaktorer

Förväxlingsfaktorer (confounding factors) kan vara ett stort problem i observationsstudier. Det innebär att man drar slutsatser om samband mellan ett

tillstånd och en bakgrundsfaktor medan tillståndet i verkligheten förorsakas av en annan variabel som samvarierar med den studerade bakgrundsvariabeln.

Orsaksförväxling kan kontrolleras på olika sätt. Vid planeringen av en studie kan man begränsa urvalet av försökspersoner så att förväxlingsproblem inte uppstår, genom att utesluta personer med en av de två samvarierande variablerna. Alternativa sätt att kontrollera för confounders är att stratifiera och matcha materialet eller kontrollera med hjälp av olika typer av multivariata statistiska metoder. Stratifiering av materialet innebär att man delar upp materialet i undergrupper med olika exponering för att minska risken för förväxling. Exempel på sådana variabler kan vara kön, ålder och rökning. Matchning är ett annat sätt att minska risken för förväxling. Matchning innebär att man väljer personer som är lika med avseende på de förväxlingsfaktorer man vill kontrollera. Med hjälp av statistiska metoder, till exempel regressionsmodeller, kan man samtidigt konstanthålla för flera variabler som kan förväxlas. En annan metodik för att kontrollera för confounders är propensity score [10]. Man kan tala om så kallad ”residual confounding”, det vill säga även om man justerar för kända störkfaktorer återstår okända som kan påverka resultatet.

Metodologiska skäl talar för att sambanden i observationsstudierna snarast är underskattade. Om felklassificeringen i exposition, till exempel blodtrycksnivåer, grad av missbruk, etcetera, är oberoende av utfallet underskattas alltid sambandets styrka, så kallad regression dilution bias [11]. Ett förhållande som till exempel alltid gäller för kohortstudier där exponeringsinformation insamlas före observationsperioden. Sannolikt klassificeras även data mer noggrant i randomiserade kontrollerade studier än i observationsstudier.

Risken för selektionsbias vid observationsstudier kan illustreras med det ofta citerade exemplet om östrogenbehandling och risk för hjärt- och kärlsjukdom. Flera observationsstudier hade pekat på att östrogenbehandling minskade risken för hjärt- och kärlsjukdom. Detta trots att man i en del studier försökt justera resultaten för en del riskfaktorer. En stor randomiserad kontrollerad studie visade att så inte var fallet. När man i observationsstudierna däremot justerade riskerna med hänsyn till skillnader i socioekonomi försvann riskminskningen [12]. Detta var inte förvånande mot bakgrund av att flera andra studier visat att välutbildade kvinnor i större utsträckning än andra fick östrogenbehandling. Ifall de publicerade observationsstudierna hade haft med socioekonomiska förhållanden i sin analys hade man sannolikt inte dragit felaktiga slutsatser. Förhoppningen om att vitaminer kunde skydda mot hjärt- och kärlsjukdomar och lungcancer kunde inte heller påvisas när man i observationsstudier justerade för socioekonomi [13].

Exempel 6.4 Risk för selektionsbias och behov av kontroll för förväxlingsfaktorer.

Biverkningsstudier

Vid utvärdering av olika åtgärder har observationsstudier tveklöst ett stort värde för att påvisa negativa effekter och risker, det vill säga biverkningar/sido-effekter. Dessa är ofta till sin natur oväntade och en del allvarliga biverkningar/

sidoeffekter är sällsynta, men av sådan dignitet att även mycket låga incidenstal är oacceptabla (Exempel 6.5). Randomiserade kontrollerade studier är sällan dimensionerade, vare sig i termer av urvalsstorlek, uppföljningstid, eller rapporteringsrutiner, för att fånga upp sådana biverkningar/sidoeffekter. Registerbaserade kohort- och fall-kontrollstudier med stora populationer är då ofta ett bra alternativ.

Exempel 6.5
Värdet av
observationsstudier
vid analys av
biverkningar/risker.

Läkemedlet aprotinin har sedan år 1993 använts runt om i världen för att minska blödning vid bland annat by-pass-operationer. Många små randomiserade kontrollerade studier hade inte sett några risker vid användningen av aprotinin. Däremot kunde stora kohort- och fall-kontrollstudier visa på en överrisk för njursvikt och död [14,15]. Det var dock först när en större randomiserad kontrollerad studie senare avbröts på grund av ökad dödlighet inom 30 dagar som medlet drogs in. Sannolikt hade ett antal dödsfall kunnat undvikas om man tagit dessa observationsstudier på större allvar och mer kritiskt granskat metaanalyser av små RCT [16]. Europeiska läkemedelsverket (EMA) har dock tillåtit användningen av aprotinin på särskilda indikationer.

Tabellera studierna

En huvuduppgift i projekten är att extrahera data ur studier med hög och medelhög studiekvalitet och sammanställa dem i tabeller. Syftet är att läsare av rapporten på ett enkelt sätt ska kunna få en överblick över inkluderade studier och hur de har bedömts. Ett annat syfte är att det underlättar det fortsatta arbetet genom att data blir strukturerade.

Om det vetenskapliga underlaget enbart består av studier med låg studiekvalitet ska dessa tabelleras.

Faktaruta 6.1
Läkemedelsverkets
bedömning av
biverkningsrapporter.

Läkemedelsverket gör en bedömning av biverkningsrapporter och graderar det kausala sambandet för de rapporterade biverkningarna som (a) säkert eller sannolikt; (b) möjligt; (c) osannolikt; eller (d) ej bedömbart, baserat på biologisk rimlighet, möjliga verkningsmekanismer, tidssamband och – mer sällan förekommande – resultatet av provokationsstudier där försökspersoner utsätts för en viss exponering eller inte. Vid en sådan bedömning görs inga försök att värdera de enskilda studiernas kvalitet eller den sammantagna evidensgraden.

Tabellerna ska ge information om referens, frågeställning, metod, urval, genomförande, resultat och metodologisk kvalitet. Tabellerna skrivs på engelska. Skälen är att underlätta för andra länder att tillgodogöra sig en del av SBU:s

grundläggande arbete, att artiklarna oftast är på engelska och att man då inte behöver översätta de exakta uttrycken som används i artiklarna samt att tabeller huvudsakligen läses av vetenskapligt skolade personer med god kunskap i att läsa vetenskaplig litteratur. Eventuella tabeller i sammanfattningen ska dock alltid vara på svenska. I Tabell 6.1 ges exempel på hur en tabell kan konstrueras så att den innehåller relevant information.

Tabellen kan användas för att göra en kvalitativ bedömning rörande heterogenitet eller stora variationer mellan studierna. De ger också en samlad bild av kunskapsläget för en specifik frågeställning.

Author Year Reference Country	Study design	Popula- tion charac- teristics	Inter- vention	Follow-up period Drop out rate	Results	Study quality and relevance Comments
	(RCT, CT, cohort, case control etc)	Inclusion/exclusion criteria Setting No at baseline Male/female	Interven- tion (I) (dose, interval, duration) Control (C) (active, placebo, usual care, etc)	(From baseline to follow-up, or from end of interven- tion to follow-up) Drop out (%)	Results (I, C) (Absolute difference, HR, RR, OR, p-value, confidence interval for the difference, sensitivity, specificity, observer reliability, cost-effectiveness, etc)	High, moderate or low study quality if appropriate

Tabell 6.1
Exempel på
tabellstruktur.

C = Control; **CT** = Controlled trial; **HR** = Hazard ratio; **I** = Intervention; **OR** = Odds ratio; **RCT** = Randomised controlled trial; **RR** = Risk ratio

Publicerade systematiska litteraturöversikter och HTA-rapporter

En forskningsfråga i ett projekt kan besvaras med en redan publicerad systematisk litteraturöversikt. Förutsättningarna är att:

- frågan överensstämmer helt med projektets frågeställningar
- litteraturöversikten är tillförlitlig.

I och med att projektgruppen (och SBU) i praktiken ställer sig bakom slutsatserna i en annan publicerad systematisk översikt måste översikten underkastas en noggrann granskning. Ofta har systematiska översikter sådana metodologiska brister att de inte kan användas, annat än som referenslista. I en tvärsnittsstudie av kvaliteten på systematiska översikter som publicerades vintern 2007 saknades till exempel kvalitetsgranskning i cirka 30 procent av översikterna [17].

I syfte att förbättra kvaliteten i systematiska översikter och rapporteringen av metaanalyser har en internationell grupp publicerat en rapport, PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses, www.prisma-statement.org), som är en vidareutveckling av det tidigare så kallade QUOROM statement. I PRISMA och därtill hörande dokument finns rekommendationer för hur tydligheten och transparensen kan förbättras. På sikt kommer därför sannolikt kvaliteten på systematiska översikter att förbättras.

Det första steget i bedömningen är att avgöra om den systematiska översikten är relevant. Överensstämmer översiktens frågor samt inklusions- och exklusionskriterier med projektets fråga? Om översikten har annat syfte eller andra kriterier exkluderas den.

Därefter bedöms översiktens vetenskapliga kvalitet. Den systematiska litteraturöversikten bör bland annat innehålla:

- klart formulerade frågor
- en tydligt beskriven metod för litteratursökning och urval av artiklar
- kvalitetsbedömning av studier som uppfyller översiktens inklusions- och exklusionskriterier
- tabeller som redovisar data för inkluderade studier
- en sammanvägning av studiernas resultat med lämpliga metoder, exempelvis metaanalys
- formuleringar som tyder på att författarna har tagit hänsyn till de ingående studiernas vetenskapliga kvalitet i slutsatserna.

Även systematiska litteraturöversikter och HTA-rapporter granskas och värderas med hjälp av en granskningsmall (Bilaga 6). Granskningsmallen baseras på en internationellt utvecklad granskningsmall, AMSTAR [18,19]. AMSTAR består liksom övriga checklistor av en rad frågor som kan besvaras med ”ja”, ”nej”, ”kan inte svara” och ”ej tillämpligt”. Projektgruppen bör i förväg besluta vilka krav som måste besvaras med ”ja” (eller ”ej tillämpligt”) för att kvaliteten på översikten ska anses vara godkänd.

En viktig fråga att ta ställning till är om översikten fångat in alla relevanta artiklar. Om någon studie saknas kan det antingen bero på att sökstrategin inte är tillräckligt bra eller på publikationsbias. Intressekonflikter kan påverka vilka studier som inkluderats, något som kan vara svårt att kontrollera.

För ytterligare kvalitetssäkring rekommenderas att projektgruppen kontrollerar fakta och tolkningar genom att även läsa några av de studier som inkluderats i översikten. Anledningen är att det förekommer att studier har feltolkats vilket påverkar hela analysen. Studier som kodats som randomiserade kan vid närmare betraktande vara observationsstudier.

Tabell 6.2 sammanfattar vad som krävs för att SBU ska acceptera en systematisk översikt för att besvara en fråga.

Alternativ	Kvalitetsgranskning (enligt Bilaga 6)	Slutsats och evidensgradering	Åtgärd
Frågor och inklusions- och exklusionskriterier överensstämmer med projektets	Godkänd	Slutsatserna kan accepteras om inte senare tillkommen litteratur motsäger detta	Översikten inkluderas och kompletteras med eventuellt senare publicerade artiklar. Översikten evidensgraderas med GRADE
	Inte godkänd	Slutsatserna accepteras inte	Översikten exkluderas. Referenslista och annan information utnyttjas i det egna arbetet
Frågorna överensstämmer, men inte inklusions- och exklusionskriterier		Slutsatserna accepteras inte	Översikten exkluderas. Referenslista och annan information utnyttjas i det egna arbetet

Tabell 6.2
Bedömning av användbarhet av systematiska litteraturoversikter och HTA-rapporter.

Referenser

- Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010;8:18. <http://www.biomedcentral.com/1741-7015/8/18>
- McNeill SA, Daruwala PD, Mitchell ID, Shearer MG, Hargreave TB. Sustained-release alfuzosin and trial without catheter after acute urinary retention: a prospective placebo-controlled. *BJU Int* 1999;84:622-7.
- von Elm E, Egger M, Altman DG, Pocock SJ, Vandebroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806-8.
- Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. *BMJ* 1954;1:1451-5.
- Madsen KM, Hviid A, Vestergaard M, Schendel D, Wohlfahrt J, Thorsen P, et al. A population-based study of measles, mumps, and rubella vaccination and autism. *N Engl J Med* 2002;347:1477-82.
- Akre K, Ekström AM, Signorello LB, Hansson L-E, Nyrén O. Aspirin and risk for gastric cancer: a population-based case-control study in Sweden. *Br J Cancer* 2001;84:965-8.
- Rosén M, Axelsson S, Lindblom J. Släng inte ut observationsstudier med badvattnet. Bedöm deras kvalitet istället. *Läkartidningen* 2008;105:3191-4.
- Vandebroucke JP. When are observational studies as credible as randomised trials? *Lancet* 2004;363:1728-31.
- Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomized controlled trials as compared to observational studies: methodological overview. *PLoS Medicine* 2011;8:1-13.
- D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265-81.
- MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, et al. Blood pressure, stroke and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990;335:765-74.
- Humphrey LL, Chan BK, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med* 2002;137:273-84.
- Lawlor DA, Davey Smith G, Bruckdorfer KR, Kundu D, Ebrahim S. Those confounded vitamins: what can we learn

- from the differences between observational versus randomised trial evidence? *Lancet* 2004;363:1724-7.
14. Mangano DT, Tudor JC, Dietzel C. The risk associated with aprotinin in cardiac surgery. *N Engl J Med* 2006;354:353-65.
 15. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med* 2008;358:771-83.
 16. Rosén M. The aprotinin saga and the risks of conducting meta-analysis on small randomised controlled trials – a critique of a Cochrane review. *BMC Health Serv Res* 2009;9:34.
 17. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4:e78.
 18. 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.
 19. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;62:1013-20.