

Kapitel 6:

Syntes och evidensgradering

När de inkluderade studierna tabellerats ska resultaten sammanställas, vägas ihop och värderas. Det är viktigt att denna syntes görs på ett öppet och reproducerbart sätt och att man eftersträvar att minimera risken för systematiska fel. Det lämpligaste sättet att sammanställa resultaten varierar mellan olika rapporter och ämnen.

SBU har i sitt tillvägagångssätt harmoniserat sig med det internationellt utarbetade GRADE-systemet. GRADE Working Group utvecklar system för att gradera evidens och rekommendationer. Bakgrunden till bildandet av GRADE Working Group är den flora av olika system som idag används av olika organisationer för att gradera evidensstyrka och rekommendationsgrad. Mångfalden har lett till viss förvirring. Dessutom har GRADE-gruppens medlemmar upplevt att viktiga steg i processen ofta utelämnats eller varit otydliga. I princip bygger GRADE på erfarenheter från tidigare graderingssystem med tillägget av ett tydligare risk–nytta perspektiv. Det är egentligen utvecklat för att lättare kunna föras vidare till utveckling av riktlinjer. Delar av GRADE-systemet har anammats av WHO, NICE, Cochrane Collaboration, BMJ och ett flertal andra internationella aktörer. Om många organisationer tillämpar en likartad evidensgradering så underlättas naturligtvis en internationell samverkan inom medicinsk utvärdering.

En sammanfattande resultattabell ger en bra och kortfattad bild av underlaget

För att inledningsvis få en översiktlig bild av den samlade evidensen är metaanalyser samt en sammanfattande resultattabell ofta användbara instrument. Dessa är bra och pedagogiska sätt att presentera data. Ibland kan dock resultaten vara mycket svåra att sammanställa eller vara direkt olämpliga för metaanalys. Då resultaten av de ingående studierna har summerats med hjälp av en metaanalys eller tabellering för varje relevant effektmått sammanställs de i en sammanfattande resultattabell (se Tabell 4). I tabellen redovisas samtliga viktiga effektmått separat (dödlighet, sjuklighet, funktion, livskvalitet samt i vissa fall även effektmått med oklar patientnytta). Effektmåtten innefattar gynnsamma effekter, men även negativa som biverkningar och komplikationer. Effektmåtten bör tabelleras hierarkiskt så att de helt

avgörande står före de mindre viktiga. I vissa fall inkluderas även utfallsmått med mer oklar klinisk relevans, vilka då står sist.

Den sammanfattande resultattabellen ger alltså en bild av förväntade behandlingsvinster, biverkningar samt komplikationer i numeriska termer tillsammans med evidensstyrkan för varje effektmått separat. Kolumnen som avser den sammanlagda evidensstyrkan för effektmåttet fylls i först efter det att man gjort en slutlig sammanvägning av evidensstyrkan.

Tabell 4. Exempel på sammanfattande resultattabell från SBU:s översikt av typ 1-diabetes, intensiv insulinbehandling vs sedvanlig behandling.

| Effektmått | Antal patienter (antal studier) | Medelrisk i standard-grupp (min – max) | Relativ risk (95% KI) | Absolut effekt per 1 000 patienter | Vetenskapligt underlag | Kommentarer |
|--|---------------------------------|--|--------------------------|------------------------------------|------------------------|--|
| Nya fall av allvarlig icke proliferativ näthinnepåverkan eller värre | 842 (3) | 40% (35 till 57 %) | RR 0.43 (0.34 till 0.55) | 226 färre | ⊕⊕⊕⊕ Starkt | Sekundär-preventiva studier |
| Nya fall av betydande äggviteläckage | 866 (3) | 11% (9 till 19 %) | RR 0.47 (0.29 to 0.77) | 57 färre | ⊕⊕⊕⊕ Starkt | Sekundär-preventiva studier |
| Nya fall av betydande nervpåverkan | 1161 (1) | 13% (10 till 17 %) | RR 0.36 (0.24 till 0.55) | 87 färre | ⊕⊕⊕⊕ Starkt | Primär- + sekundär-preventiva studier |
| Makrovaskulära komplikationer | 1630 (4) | 11% (3 till 15 %) | RR 0.38 (0.26 till 0.56) | 70 färre | ⊕⊕○○ Begränsat | I efterhand från forskare inhämtade data Primär- + sekundär-preventiva studier |
| Fall av allvarlig hypoglykemi | 1904 (12) | 30 % (0 till 70 % i intensivgrupp) | RR 1.74 (1.57 till 1.93) | 237 fler | ⊕⊕⊕⊕ Starkt | Alla studier |
| Antal patienter med BMI-ökning mer än 5 kg/m ² | 1246 (1) | 8 % (3 till 15 % i intensivgrupp) | RR 3.73 (2.81 till 4.97) | 220 fler | ⊕⊕⊕⊕ Starkt | Primär- + sekundär-preventiva studier |

Sammanvägd evidensstyrka

En sammanvägd evidensstyrka har sin utgångspunkt i projektgruppens analys av det vetenskapliga underlaget (studiekvalitet, samstämmighet, överförbarhet, effektstorlek, risk för publikationsbias, oprecisa data eller andra aspekter som t ex dos-responssamband). I GRADE-systemet har man försökt att i detalj specificera vilka svagheter och styrkor den samlade evidensstyrkan baseras på för varje enskilt effektmått. SBU anser det fortfarande viktigt att värdera kvaliteten i varje inkluderad studie och studier kan exkluderas ur SBU:s analyser pga låg studiekvalitet.

Det viktiga i bedömningen av evidensstyrkan är att projektgrupperna skriftligt kommenterar nedanstående faktorer som motivering till bedömningen av den samlade evidensstyrkan:

- Studiekvalitet
- Samstämmighet/Överensstämmelse
- Överförbarhet/Relevans
- Oprecisa data
- Risk för publikationsbias
- Effektstorlek
- Andra viktiga faktorer, t ex dos-respons

I vad mån man vill tillämpa ett strikt poängsystem för upp- eller nedgradering av evidensstyrkan enligt GRADE-systemet får projektgrupperna själv ta ställning till. Ett sätt kan vara att tillämpa det arbetsblad som utarbetats för att bedöma och fastställa evidensstyrkan (se Bilaga 6). Evidensstyrkan ska fastställas för varje effektmått separat.

- ***Ingående studier***

Ange antal studier, antal patienter samt typ av studier (RCT, observationsstudier med kontrollgrupp etc) som ingår i bedömningen av interventionens påverkan på detta specifika effektmått. Den preliminära evidensstyrkan kan baseras på bedömd studiekvalitet eller studietyp hos inkluderade studier enligt nedan. Dessa kan eventuellt upp- eller nedgraderas när senare kriterier bedömts.

Hög (++++): Underlaget bygger företrädesvis på RCT

Måttlig(+++): Underlaget bygger företrädesvis på kvasirandomiserade studier (tex jämna/udda datum) eller mycket välgjorda och stora kohortstudier med god confounderkontroll.

Begränsad(++): Underlaget bygger företrädesvis på behandlingsstudier med kontrollgrupp eller andra kohortstudier.

Mycket låg, otillräckligt(+): Underlaget bygger företrädesvis på fallserier utan kontrollgrupp, fallrapporter eller dylikt.

- **Studiekvalitet**

I detta steg gör man en samlad värdering av ingående studiers metodologiska kvalitet. Denna värdering har skett individuellt för varje studie med hjälp av SBU:s mallar (se avsnitt för studiekvalitet). I första hand väger man in de traditionella faktorerna såsom randomisering (för RCT), blindning och bortfall, men även andra faktorer belysta i mallarna beaktas liksom eventuella andra ämnesspecifika metodproblem som kan identifieras med hjälp av experterna. För kohortstudier och

andra observationsstudier är frågan om jämförbarhet mellan försöks- och kontrollgrupp central. Det innebär att studiernas kvalitet i hög grad beror på i vad mån man kontrollerat för förväxlingsfaktorer (confounders). Det sammanfattade resultatet av granskningen kan leda till att evidensstyrkan justeras. Bedömningen är arbiträr och baseras inte på en specifik poäng i granskningsmallen.

- **Samstämmighet och överensstämmelse**

För att överhuvud kunna bedöma samstämmigheten krävs flera inkluderade studier.

Trovärdigheten kan öka om studierna gjorts av olika forskargrupper i olika länder med olika patientgrupper och studierna samstämmigt pekar i samma riktning. Om studier pekar i olika riktningar och visar såväl över- som underrisker så kan den samlade evidensstyrkan minskas med -1. I vissa fall kan olikheterna förklaras med olikheter i de inkluderade studierna, tex olika populationer. I dessa fall kan det vara mer lämpligt att undanta vissa populationer från slutsatsen än att sänka graderingen av det vetenskapliga underlaget eller ge olika slutsatser för olika populationer.

Här görs även en bedömning ifall effektstorleken varierar mellan olika studier. Olika effektstorlek medför en ökad osäkerhet beträffande effektstorlek, men behöver inte medföra en sänkning av evidensstyrkan för att en effekt föreligger. Metaanalyser underlättar denna bedömning.

- **Överförbarhet**

Hur väl stämmer studiepopulationen med den patientpopulation man ser i daglig praxis? Kort sagt, är studien rimlig ur ett kliniskt perspektiv? Här bedömer man studiepopulation, interventionens relevans och specificitet, relevansen av jämförelsemetoden, sjukvårdsmiljö, adekvat uppföljningstid med mera.

- **Oprecisa data**

Det här är ett kriterium som kräver bedömningar där man både studerar konfidensintervall och antal observationer. Vid fåtal utfall bör man titta på konfidensintervallet för riskdifferensen (absoluta risken) istället för konfidensintervallet kring den relativa risken.

- **Effektstorlek**

Vid hög effektstorlek så ökar sannolikheten att det funna sambandet är kausalt. GRADE anger att man är beredd att höja den samlade evidensstyrkan med 2 om $RR > 5.0$ (alt $RR < 0.2$).

Om $RR > 2.0$ (alt $RR < 0.5$) så kan den samlade evidensstyrkan höjas med +1. Observera att oddskvoter (OR) i motsats till RR kan överskatta effekterna vid vanliga utfall (> 10 %) och gränserna kan då behöva justeras.

- **Andra viktiga aspekter, t ex dos-respons**

Närvaro av dos-responsförhållande ökar trovärdigheten för åtgärdens effekt där det är relevant.

Dos-respons kan gälla i olika situationer både när det gäller läkemedel och preventiva insatser eller inom en studie och mellan studier. Generellt är ett dos-responssamband mycket mer trovärdigt när det visats i en studie än när det visats genom jämförelser mellan olika studier.

Dos-responssamband kan gälla såväl effekter som risker.

- **Publikationsbias**

Det är ofta svårt att med säkerhet avgöra ifall publikationsbias föreligger, men punkten bör diskuteras i rapporten.. Centrala register (www.controlled-trials.com, www.clinicaltrials.gov) över påbörjade prövningar bör undersökas. En ”funnel plot analys” bör utföras i samband med meta-analys (se kapitel 6 Statistik och metaanalyser). En ”funnel plot” kan vara värdefull, men är osäker och bör tolkas med försiktighet. Vid slutsatser som bygger på små studier av låg kvalitet och från samma forskargrupp föreligger ofta hög risk för publikationsbias. Likaså ökar risken för publikationsbias när det är nya metoder där innovatören av den nya metoden är huvudförfattare eller där enbart företagssponsrade studier av ett läkemedel förekommer. Riskerna för publikationsbias illustreras väl i en svensk studie från Läkemedelsverket (Melander et al, 2003).

Samlad evidensgradering

För varje effektmått görs en samlad värdering av evidensstyrkan enligt de kriterier som angivits ovan. Efter detta graderas evidensstyrkan enligt GRADE i fyra nivåer. Exakt vad denna gradering står för kan diskuteras, men GRADE har formulerat följande definitioner:

- **Hög** Det är osannolikt att framtida forskning kan ändra vår tilltro till den uppskattade effekten (bygger på flera studier med tillförlitlig studieupplägning, utan försvagande faktorer vid syntesen)
- **Måttlig** Det är sannolikt att framtida forskning kan ha en betydelsefull inverkan på vår tilltro till den uppskattade effekten och eventuellt förändra den uppskattade effekten (bygger på studier med tillförlitlig – mindre tillförlitlig studieupplägning, där resultatet delvis påverkas av försvagande respektive förstärkande faktorer vid syntesen)
- **Begränsad/låg** Det är mycket sannolikt att framtida forskning påverkar vår tilltro till den uppskattade effekten och att den sannolikt kommer att förändras (bygger på studier av mindre tillförlitlig studieupplägning där resultatet också kan påverkas av försvagande faktorer vid syntesen)

- **Mycket låg/ Otillräcklig** De uppskattade effekterna är mycket osäkra (bygger på enstaka studier eller studier med låg tillförlitlighet).

I en samlad värdering gäller det också att väga nyttan mot eventuella risker. I GRADE-systemet skiljer man mellan:

- åtgärden har betydligt större positiva effekter än risker
- åtgärden har oftast större nytta än risker, men beslut om åtgärd kan variera mellan enskilda patienter
- det är osäkert om åtgärden gör mer nytta än skada
- åtgärden gör begränsad nytta och riskerna är sannolikt större än nyttan.

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

Utvärderingen ger naturligtvis ändå vägledning för hälso- och sjukvården. Om evidensstyrkan blir mycket låg/otillräcklig så indikerar det behov av mer forskning innan det börjar tillämpas i hälso- och sjukvårdens kliniska vardag. Begränsad/låg evidensstyrka pekar på att det finns visst vetenskapligt underlag som kan motivera att metoden används i hälso- och sjukvården under förutsättning att det uppfyller andra krav på acceptabel nytta-riskbalans, kostnadseffektivitet och att metoden bedöms etiskt rimligt. Vid hög och måttlig 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.