

Evidensbaserad läkemedelsvärdering

Hans Liedholm, Lund

Inledning

Det här kapitlet innehåller både övergripande och detaljerade uppgifter av betydelse för förståelsen av evidensbaserad medicin (EBM). Exempel på svårigheterna att ta ställning till information inom och utanför den medicinska världen ges liksom till den desinformation som sker med eller utan avsikt. Metoder och pålitliga källor för kritisk läsning exemplifieras.

Lärande

För läkare är arbetet en fråga om kontinuerligt lärande. Efter den första studie- och utbildningstiden finner man snart att man kanske inte läser så mycket originalarbeten som man kanske föresatt sig. Dessa ersätts ofta av snabblästa referat och översikter där man inte informeras om förutsättningarna (vad gäller urval, patientsammansättning, tekniska förutsättningar, tolkning etc) (1).

Efter grundutbildningen finns behov av ökade kunskaper och färdigheter – det är naturligt att vi primärt lär av äldre och mer vana kollegor. Lathundar, klinikföreskrifter, avdelningars PM, SK-kurser och nationella riktlinjer ger stadga i den fortsatta utbildningen. Se även utdrag ur målbeskrivning 2010 för specialistutbildningen i Faktaruta 1. En källa som återopas i den kliniska handläggningen är kvalitetsregistren (2). Dessa utgör grund för bedömningar av hur sjukvården fungerar i olika

Ur målbeskrivning 2010 för specialistutbildningen

Den specialistkompetenta läkaren ska ha förmåga till ett medicinskt vetenskapligt syn- och förhållningssätt, kunskap om forskningsmetodik, inklusive epidemiologiska grundbegrepp, samt om metoder för evidensbaserad medicin (EBM) och granskning av vetenskaplig information.

Den specialistkompetenta läkaren ska ha kunskap om och kompetens i evidensbaserat förbättrings- och kvalitetsarbete. Målet är att kunna initiera, delta i och ansvara för kontinuerligt systematiskt förbättringsarbete med betoning på helhetsperspektiv, patientsäkerhet, patientnytta, mätbarhet och lärandestyrning för att kritiskt kunna granska och utvärdera den egna verksamheten.

landsting, varför rangordning av innehåll ofta genomförs och diskuteras.

Läs mera!

Professor Lars Werkö skrev 1999 en artikel i Läkartidningen under rubriken "Tidningsläsning – fortbildning för aktiva läkare" (3). "För att följa med i vad som händer inom den egna specialiteten och inom den medicinska utvecklingen i stort – både för sin egen skull och för att kunna bemöta frågor från oroliga patienter – är det nödvändigt för varje läkare att inte bara läsa utan också kritiskt granska en ansevärd textmängd varje vecka." Det var en tidsödande verksamhet i ett ansträngt arbetsliv, påpekade han. Hans förslag var: "Lägg tid-

skriftsläsning på schemabunden tid eller bilda tidskriftsklubbar.”

”Att sätta sig in i och smälta all denna information kan endast ske om man är befriad från andra göromål, eller andra aktiviteter, vilket antingen innebär att ägna en fri helg åt detta eller att arrangera andra sätt att få fri tid till att följa litteratur på ett meningsfullt sätt (3).”

Systematisk vidareutbildning med föreläsningar och diskussioner sker ofta i organiserad form på sjukhuskliniker, också systematisk utbildning i mindre eller större grupper. En del av informationen ges av läkemedelsindustrin. En kritisk granskning av denna information är svår att genomföra, men många exempel på avvikelser (selekterad och felaktig information) finns beskrivna (4).

Medicinska litteraturen utökas och förändras
Uppgifter om ny kunskap i den medicinska världen rapporteras snabbt. Det har också blivit allt vanligare att de traditionella tidskrifterna publicerar angelägna artiklar på Internet innan de kommer ut i pappersversion samtidigt som antalet enbart Internetbaserade tidskrifter ökar. Tidskrifter eller artiklar som har epitetet ”open access”, dvs läsbara och nedladdningsbara från Internet utan prenumeration ökar också. Vetenskapliga rapporter måste läsas mycket kritiskt, läsaren får inte ta några resultat för givna. Ändamålet med kritisk läsning är att upptäcka om metoderna kan ge användbar information och om slutsatserna som författarna dragit följer av studieresultaten.

En översikt av vetenskapliga rapporter i medicinska tidskrifter visar att det behövs två saker för att förstå deras innehåll. Det ena är kunskap om de förhållanden som undersöks (sjukdomen, terapierna etc) och det andra är kunskap om de vetenskapliga metoder som använts och den därpå följande tolkningen av resultaten – särskilt kunskap om medicinsk statistik.

Kan systematiska översikter och metaanalyser ge svar på frågan?

För att få reda på aktuell, sammanlagd kunskap om en viss terapi vill man gärna läsa någon form av översikt. En sådan översikts-

Tolkning av en studies resultat och tiden efter publicering

A

Utgångsläget är en randomiserad, kontrollerad studie (RCT) i ett angeläget ämne som publiceras i en ”känd” tidskrift.

- Räcker tidskriftsvalen för att läsaren ska kunna acceptera resultat och slutsatser (5)?
Nej. Det är en god början men inte tillräckligt. Viktigt att uppmärksamma är att ”supplement” även till ”kända” tidskrifter vanligen inte är ”referee”-bedömda.

B

Studien finns registrerad i ett internationellt studie-register med föransmält protokoll (se s 1200). Studien kan ha varit föremål för en redovisning, t ex på en medicinsk kongress, och då blivit publicerad i abstraktform (angivet som referens i artikelns litteraturlista). Studien är granskad av tidskriftens redaktion, referenterna har kommenterat, vilket kan ha inneburit att artikeln har fått minska i omfattning och en del information kan ha utgått (tilläggsinformation kan eventuellt placeras som elektroniskt tillägg – tabeller m m). I tilläggsinformation finns författarnas bidrag till artikeln redovisade. (Jfr ”hedersförfattare” med Vancouver-regler (6).) I tilläggsinformation i artikeln har författarna redovisat hur och av vem de varit anställda, vem eller vad som bidragit med ekonomiskt anslag till studien och vilka ekonomiska bindningar författarna haft till sponsorer och andra.

- Räcker detta för läsarens tolkning?
Nej

C

Ofta finns det en ledarartikel (editorial), skriven av en oberoende och i ämnet väl insatt person. Läs ledarartikeln: ämnet kan där ha bedömts och blivit insatt i ett vidare perspektiv. Kanske är ledarskribenten inte lika optimistisk som författarna! I vissa nummer av den ”kända” tidskriften kan ytterligare kommenterande artiklar förekomma, dessa är lika viktiga att läsa. Synsättet på den medicinska frågan kan ha ändrats något.

- Detta har klart ökat möjligheterna, inte minst för ”icke-specialisten”, att värdera informationen men räcker detta för läsarens tolkning?
Nej

D

2 f

I senare nummer av tidskriften kommer så ett urval av insändare som frågar författarna om detaljer i artikeln. Författarna svarar i sin tur, eventuellt först efter det att tidskriften tagit in den sista insändaren. På grund av debatten som uppstått kan författarna sedan ha lämnat kompletterande information eller t o m rättat fel, som inte var tydliga från början. Synsättet kan ha ändrats något.

Notera att insändare och allvarligt menade kommentarer också kan ha kommit in i andra tidskrifter. Information om en artikel som kommenterats i tidskriften eller i andra publikationer finns med vid artikelsökning i PubMed. Artikeln kan ha diskuterats på flera ställen, även i nyhetsmedia.

På ett tidigt stadium har man kanske kunnat läsa smäinlägg om studien på tidskriftens Internet-sida. Utvecklingen av de s k sociala medierna – bloggar, Facebook, Twitter, Youtube m fl – har nu också gjort sitt inträde i medicinska tidskrifter. Några tidskrifter använder sig av bloggar (bl a NEJM). I BMJ skriver huvudredaktören tankvärdheter med en stor mängd länkad information.

- Räcker detta för läsarens tolkning? **Ja, det verkar så.**

Det har nu gått en lång tid från den första publiceringen. Tiden går, vad kan ha hänt? Är artikeln kvar i PubMed?

artikel kan i engelskspråkig litteratur vara det som kallas för "review" eller en "systematic review". Varje dag publiceras 75 kliniska studier, 11 systematiska översikter och mängden ökar hela tiden varpå det kan vara svårt att hinna med bedömningen (7). Se Faktaruta 2, s 1195, för exempel på tolkning av en studies resultat och tiden efter publicering.

Systematiska översikter

En systematisk översikt är en artikel som efter en viss mall refererar och kommenterar alla tillgängliga resultat som belyser ett visst medicinskt problem. Metoden försöker minska sannolikheten för systematiska fel (bias). Eftersom bedömningar görs i flera olika steg i en översikt (exempelvis bestämning av inklusions- och exklusionskriterier och kvalitetsbedömning av ingående stu-

E

2 f

Det kan hända att kritiska granskare hittar andra uppgifter som kastar nytt ljus över publikationen, även efter flera år! Ett exempel på långsamheten i den vetenskapliga processen är studien COOPERATE (om kombination av ACE-hämmare och ARB till patienter med icke-diabetisk njursjukdom) som publicerades 2003 i Lancet (8). Trots flera tidigt ifrågasättande insändare fick frågan om studieupplägg en förklaring först sedan en noggrann genomgång av patientdata utförts på ett japanskt sjukhus. Man kunde då inte verifiera sanningshalten av fakta och en undersökningskommitté ansåg att studien måste dras tillbaka (2009). Så skedde också; i Lancet kan man fortfarande läsa artikeln i sitt originalskick, dock numera med ordet RETRACTED stämplat över sidorna. Det tog minst 3 år innan studien drogs tillbaka! Av detta måste man lära sig att läsa synpunkter som kan följa originalstudien.

- Måste slutsatserna omvärderas om artikeln drags tillbaka? **Ja! Studien "finns inte längre".**

F

Vad händer om en författare visat sig ha "uppfunnit" patienterna? De publicerade uppsatserna får nu "dras tillbaka", stor skada!

- Måste slutsatserna omvärderas om det är en multicenterstudie? **Mycket troligt.**

G

Vad händer om artikeln av studien – trots författaranvisningar av tidskriften om att så inte får förekomma – faktiskt publicerats flera gånger (dubbelpublicering). I en Cochranegenomgång 2005 av antikolinerga läkemedel vid överaktiv blåsa fann man att det mest flagranta exemplet på förekomst av multipel publicering var att en och samma studie förekom 18 gånger (9).

- Måste slutsatserna omvärderas? **Osäkert, men vilken av publikationerna "gäller"?**

H

Vad händer om en eller flera av de namngivna författarna är personer som inte deltagit i studierna men som har ett "gångbart namn" (spökskrivare) och att den vetenskapliga sammanställningen är gjord av ett publiceringsföretag, betalt av den läkemedelsindustri som marknadsför preparatet?

- Måste slutsatserna omvärderas? **Osäkert, men beteendet är oärligt.**

Sökord

Innehåll

dier) är systematiska översikter inte befriade från bias. Ett problem som ofta diskuteras är "publication bias", dvs att positiva studier är lättare att publicera än negativa. Upprepade publikationer av samma data och framför allt undanhållande av data från offentliggörande grumlar också bilden och kan medföra att översikter och metaanalyser blir missvisande.

Tyvärr leder ibland systematiska översikter från olika grupper av författare till helt olika resultat. En faktor som påvisats är den bindning författarna har till uppdragsgivaren. Förvirringen hos läsaren blir stor – vem kan man lita på?

Metaanalyser

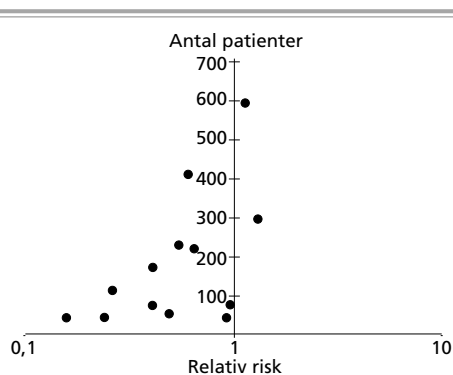
En bättre överblick kan man få om man läser en metaanalys av undersökningar i ämnet. En metaanalys är en metod för att göra en samlad bedömning av ett antal jämförande undersökningar genom att statistiskt sammanföra resultaten. Man gör en omfattande litteratursökning och visar hur och var man sökt. De språkområden och de tidsperioder man undersökt framgår tydligt. Man granskar allt publicerat material och man försöker också få fram opublicerat material. Man ställer upp kriterier för vilka undersökningar som ska ingå i analysen.

Resultaten för varje undersökning redovisas i form av ett jämförande resultatmått (tex en oddskvot med konfidensintervall) som avslutas med en statistisk sammanslagning av resultaten till ett samlat resultatmått (med konfidensintervall). Framställningen ger en överblick över tillgängliga resultat och deras samstämmighet (homogenitet). Resultaten ger en uppfattning om huruvida undersökningarna visat att en metod är bättre än en annan.

Inte heller metaanalyser är "ofelbara". Det metodologiska arbetet är viktigt och de studier som tas med i analysen avgör utfallet. Effekten av en sådan framgång av bilden i Figur 1. Baseras analysen på mindre studier kan slutsatsen i metaanalysen vara missvisande.

Cochrane

Den internationella intresseorganisation, som bl a är specialiserad på att genomföra metaanalyser, är Cochrane Collaboration



Figur 1. Exempel på funnel plot. Varje symbol representerar en studie. Antalet patienter i studierna är avsatt mot den vertikala axeln. Effekten av två behandlingar som jämförts uttrycks som relativ risk – vid värdet "1" är behandlingsresultaten lika. Bilden belyser problemet med att basera slutsatser på flera mindre studier.

(CC), som kan nås via www.cochrane.org. Studierna som presenteras i Cochrane Library är högt rankade och står för kvalitetsvärderad kunskap. Den nackdel man kan se med arbetet i CC är den tid det tar att ta fram materialet för sammanställningarna. Symbolen för CC är den grafiska återgivningen av resultaten i en metaanalys, en sk forest plot. Ett exempel på en grafisk redovisning av en metaanalys visas i Figur 2, s 1199, och Figur 3, s 1200.

Kunskap är en färskvara

Nya data kan lätt ge en ny bild av "sanningen". Därför bör systematiska översikter ersättas av nya. Cochrane Collaboration ser ständigt över sina slutsatser från metaanalyserna och uppdaterar dessa. Det finns således varningstecken utsatta för hur mycket man kan lita på sammanställningarna. I den medicinska litteraturen finns exempel på hur metaanalyser som kommit till vissa slutsatser antingen har avfärdats eller bekräftats sedan enstaka, mycket stora RCT (randomised controlled trials) genomförts. Exempel är intravenös magnesiumterapi (ISIS-4) respektive streptokinasbehandling vid akut hjärtinfarkt (GISSI-1 och ISIS-2) (10–12).

Epidemiologiska studier

Svagheter

Eftersom vi får så mycket information är det viktigt att vara uppmärksam på vad för slags data som presenteras. Det framgår sällan av rubrikerna. En del handlar om vad någon auktoritet skriver om, t ex platsen för en viss terapi, andra om epidemiologiska studier och åter andra om väl genomförda, kontrollerade studier, RCT. De första tillhör oftast den s k åsiktslitteraturen, som har lågt värde på EBM-skalan över trovärdighet. De andra tillhör undersökningar som har viktiga begränsningar.

Epidemiologiska studier – exempel

För att undersöka effekterna av terapi som sänker homocystein (folsyra, vitamin B₁₂ och B₆) på kardiovaskulära händelser har successivt flera placebokontrollerade studier visat att sjukdomsförloppen inte alls påverkas, trots att homocysteinnivån sänkts. Detta har t ex visats i en sekundärpreventiv 2-årsstudie av patienter med TIA och stroke (13), en 3-årsstudie av patienter med färsk hjärtinfarkt (14) samt en 5-årsstudie av patienter med hjärt-kärlsjukdom eller diabetes (15). Det förefaller som om man hela tiden bara påverkat en surrogatvariabel. Epidemiologiska data stämde inte med den kliniska verkligheten!

Låt oss begränsa oss till att säga att den epidemiologiska studiens beräkningar bygger på datainsamling från grupper av patienter, där olika registrerade egenskaper ställs mot varandra med diverse justeringsmetoder för basala olikheter. Här finns ett inbyggt metod- och trovärdighetsproblem. Det man registrerat är det som kan användas och man kan aldrig justera för det man inte registrerat. Epidemiologiska data kan antyda, men aldrig klargöra orsakssamband.

I motsats till epidemiologiska studier utgörs däremot RCT av experimentella data som kan göra orsakssamband bevisbara. Det är därför RCT anses ligga högst på skalan av bevisvärde. Exempel på inskränkningar med RCT är å andra sidan brist på generaliserbarhet, resultaten gäller bara den population som undersöktes (bakgrund, sjukdomsgrad, kön, ålder) och den tidsrymd studien varade. Detta betyder samtidigt att all kunskapsinsamling inte

kan bestå av RCT, både denna metod och den epidemiologiska har sina fördelar och sina begränsningar. Det viktigaste för läsaren är att förstå vad dessa består av.

Vad kan man egentligen lita på?

Det mesta som presenteras i icke referentbedömda artiklar ska betraktas med skepsis.

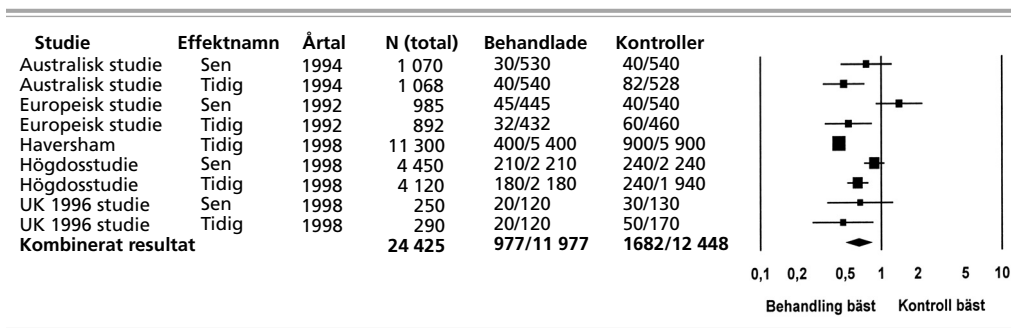
Sådana uppgifter som enbart förekommer i abstraktform (t ex abstrakt till en medicinsk kongress) eller i producenternas inköpta "supplement" till kända tidskrifter, ska betraktas som obevisade, otillräckliga eller opålitliga (16).

Ändring av primärt och sekundärt effektmått *Studiekvalitet gabapentin*

Författarna gick igenom 12 gabapentinstudier för undersökning av vid tidpunkten icke godkänd indikation: profylax mot migrän, behandling av bipolär sjukdom, neuropatisk smärta och nociceptiv smärta (17). Studierna var sponsrade av gabapentintillverkarna. Författarna jämförde interna dokument med publicerade rapporter. Av initialt 20 kliniska prövningar fanns 12 rapporterade. I 8 av de 12 skilde sig de primära resultatmått från dem som beskrivits i protokollen.

Det primära resultatmättet hade ändrats i 5 av 8 publicerade studier med statistiskt signifikanta skillnader som gynnade gabapentin. Av de 21 primära utfallen som var beskrivna i protokollen rapporterades 6 inte alls och 4 rapporterades som sekundära utfall. Av 28 primära utfall som beskrevs i de publicerade rapporterna var 12 nyintroducerade.

Genomgången uppmärksammar risker för allvarlig publikationsbias som är särskilt uttalad bland studier av läkemedel.



Figur 2. Exempel på en metaanalys. Den understa raden visar den kombinerade effekten från olika studier, t ex dödsfall, ny infarkt eller annan händelse. Den horisontella axeln är logaritmisk. Vid värdet "1", som uttrycker relativ risk eller oddskvot, är resultaten i grupperna lika. Symbolernas storlek är proportionella till studiernas storlek. De horisontella strecken utgör omfånget av ett 95% konfidensintervall för värdet.

Jävsförhållanden vid bedömning av ett visst preparat *Rosiglitazon*¹

I en studie ville man undersöka eventuella samband mellan författares jävsförhållanden och deras ställningstaganden till frågan om risker för hjärtinfarkt av rosiglitazon hos patienter med diabetes (18). Det ovanliga i denna undersökning var att man letade mycket brett efter artiklar som kommenterade eller citerade någon av de två huvudstudierna med grunduppgifter kring kontroversen (en metaanalys av små studier och en påföljande stor studie). Alla artiklar, riktlinjer, metaanalyser, översikter, kliniska studier, insändare, kommentarer och ledarartiklar om risker för hjärtinfarkt med rosiglitazon inkluderades.

För varje artikel letade man efter information om ställningstagande och rekommendationer samt information om författarnas uppgivna jävsförhållanden i själva rapporten eller på annan plats de senast föregående två åren.

Man fann 202 artiklar och uppgivna jävsförhållanden förekom i drygt hälften. Författare som hade en mer positiv syn på frågeställningen tenderade generellt att ha en tydligare positiv jävsrelation till tillverkare av antidiabetika i allmänhet och rosiglitazon i synnerhet än de författare som hade en ogynnsam inställning.

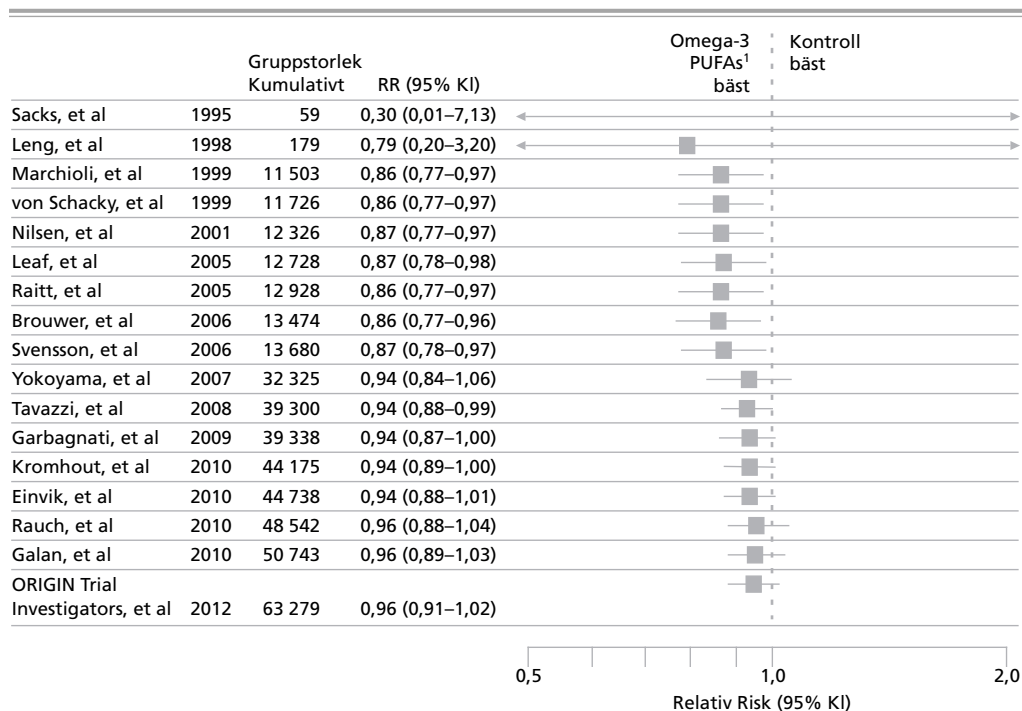
Resultaten understryker betydelsen av relationer mellan forskare och företag för bedömning av läkemedels värde i den vetenskapliga litteraturen och behovet av förändringar i hur uppgiftslämnandet om jäv ska ske för att läsaren bättre ska kunna lita på det vetenskapliga innehållet.

Manipulation av primär effektvariabel är mycket vanligt *Dansk kohort*

Att resultatbeskrivningen i publicerade studier är selektiv har länge misstänkts och direkta bevis har oftast utgjorts av fallrapporter men det finns en dansk undersökning där man studerat en kohort av randomiserade studier (19). Man använde sig av protokoll och publicerade rapporter av randomiserade studier som godkänts av forskningsetikkommittéerna i Köpenhamn och Fredriksberg 1994–1995. Antalet och karaktären av rapporterade och orapporterade studieutfall noterades med utgångspunkt från protokoll och tidskriftsartiklar och kompletterades med en utfrågning av inblandade forskare. Protokoll och publicerade artiklar jämfördes för att identifiera skillnader i primära utfall.

Man identifierade 102 studier med 122 publicerade tidskriftsartiklar och 3 736 utfall, och noterade att 50% av effektutfallen och 65% av biverkningsutfallen var ofullständigt rapporterade. Statistiskt signifi-

1. Sedan 2011 finns inga godkända läkemedel innehållande rosiglitazon i Sverige.



1. PUFAs = polyunsaturated fatty acids, dvs fleromättade fettsyror varav omega-3 är en.

Figur 3. Annat exempel på metaanalys där ingående studier kumulativt ordnats efter antal ingående personer och publikationstid. När den sammanlagda studiepopulationen nått över 30 000 är skillnaden inte längre signifikant (2007).

kanta utfall hade en högre sannolikhet att rapporteras fullt ut jämfört med de icke-signifikanta utfallen, både beträffande effekt och biverkningar. Då publicerade artiklar jämfördes med protokollen hade 62% av studierna åtminstone ett primärt utfall som hade ändrats, nyinförts eller utesluts. Hela 86% av dem som svarat på utfrågningen förnekade existensen av orapporterade utfall trots att det fanns klara bevis på motsatsen.

Studier med positiva och negativa fynd

Studier med positiva fynd publiceras oftare och snabbare än studier med negativa fynd (20). I en undersökning av metodologisk karaktär inom Cochrane undersöktes om kohorter av studier påverkas av påvisad statistisk signifikans, resultatens uppfattade betydelse eller i vilken riktning resultaten pekar. Sökningarna omfattade över 5 000 referenser. Tiden till publicering visade att

översiktliga sammanställningar (systematiska översikter och metaanalyser) med positiva resultat tenderade att publiceras efter 4–5 år, medan sammanställningar med negativa fynd publicerades efter 6–8 år.

I en genomgång av läkemedel som godkänts mellan åren 1998 och 2000 i USA fann man att mer än hälften av de studier som låg till grund för godkännandet hade efter 5 år ännu inte publicerats (!) (21).

Motsvarande slutsatser kom Läkemedelsverket fram till vid en genomgång av studier av antidepressiva läkemedel (22). I den jämfördes 42 placebokontrollerade studier av 5 antidepressiva medel (SSRI) som skickats in till Läkemedelsverket med den faktiska publikationen av samma studier. Av studier som bedömts positiva hittades 19 av 21 studier som separata publikationer. För 21 "negativa studier" återfanns endast 6 som separata publikationer, resten var inlagda i andra publikationer. 21 studier hade publice-

rats minst 2 gånger medan 4 ”negativa” studier förblev opublicerade. Flera studier rapporterade dessutom inte data enligt ITT (intention to treat), utan hade en per protokollanalys, som därmed gav en mer fördelaktig beskrivning av preparaten.

I och med att alla kliniska prövningar inom EU från och med maj 2004 måste registreras i en databas (EudraCT, kan nås via <https://eudract.ema.europa.eu/index.html>) har man hoppats att kontrollen av ”publication bias” kan bli enklare och bättre. Ett gemensamt europeiskt register, som utgår från EudraCT, finns också, www.clinicaltrialsregister.eu. Motsvarande webbplats är den amerikanska www.clinicaltrials.gov.

För den som ska studera eller skriva en översiktsartikel är det viktigt med en adekvat tillgång till registrerade studier, både publicerade och opublicerade, oavsett vad de visar.

Sponsrade studier

Det är tidigare känt att slutsatser i RCT är mer positiva gentemot interventionen om undersökningarna sponsrats av företag med ekonomiska intressen av produkterna. I en dansk undersökning av slumpmässigt urval från 370 RCT i metaanalyser i Cochranebiblioteket undersöktes detta förhållande (23). Studieläkemedlet rekommenderades statistiskt signifikant oftare som förstahandsmedel i studier med sponsring från industrin. Slutsatser i sådana studier tycks vara mer positiva i sin tolkning pga systematiska fel (bias) vid tolkningen.

Tidningsredaktörer, granskare och läsare uppmanas därför att noga bedöma om slutsatser i RCT-artiklar stöds av data.

Klinisk forskning som stöds av läkemedelsindustrin påverkar hur doktorer praktiserar medicin. Ett ökat antal kliniska studier under olika faser av en produkts levnads-cykel sponsras av läkemedelsindustrin.

En metaanalys genomfördes för att undersöka om studier som sponsras av läkemedelsindustrin är förknippade med resul-

Rapporteringen av studieutfall är således vanligen ofullständig och överensstämmer inte heller med protokollen. Publicerade artiklar, likaväl som översiktsartiklar som dessa ingår i, är därför opålitliga och övervärderar fördelarna av interventionen.

För att försäkra sig om transparens måste planerade studier registreras och protokollen måste offentliggöras innan studierna avslutas.

tat som är mer gynnsamma för sponsorn och om studiernas metoder skiljer sig från metoder som har andra anslagsgivare (24). Undersökningen baserades på sökning i databaser mellan 1996 och 2002, data från uppgifter i referenserna och personliga filer. Undersökningen gällde många olika diagnoser, studietyper och typer av läkemedel och läkemedelsgrupper.

Studier som hade läkemedelsföretag som sponsorer hade ett utfall som gynnade sponsorn oftare än de som sponsrats av andra. Däremot var de metoder som använts i industrisponsrade prövningar inte sämre.

Studiegranskning av referees – påverkar det kvaliteten?

Granskning av vetenskapliga studier i för-tid med hjälp av sakkunniga (peer review) används över hela världen för att värdera och förbättra kvalitén av insända manuskript till biomedicinska tidskrifter.

Vid en undersökning genomförd av Cochrane 2007 fann man mycket lite evidens för att en sådan förhandsgranskning leder till förbättring av kvalitén av biomedicinsk forskning (25). Metodologiska problem i ämnet är många och komplexa, och frånvaron av evidens om effekt och effektivitet får inte tolkas som att evidens för den saknas. Omfattande studier tycks krävas för att klara ut sakförhållandet.

Anvisningar för studier

Vad presentationer av metaanalyser ska innehålla har arbetats fram i en arbetsgrupp och benämns QUOROM (Quality of Reporting of Meta-analyses, jfr CONSORT).

Motsvarigheten för presentation av observationsstudier benämns MOOSE (Meta-analysis of Observational Studies).

Motsvarigheten för kvalitativa studier benämns RATS (BMC Qualitative Research Review Guidelines – RATS).

Det finns även en standard för studier över diagnostiska tester – STARD (Standards to Improve Reporting of Studies of Diagnostic Tests). Enligt en publicerad översikt från 2006 var ett vanligt fel att studier, i syfte att beskriva nya diagnostiska tester, ofta var kraftigt underdimensionerade (låg power).

Senast i familjen med anvisningar finns nu med förslaget STROBE (Strengthening the Reporting of Observational Studies in Epidemiology), som avser att strama upp beskrivningen av observationsstudier.

Hur påverkas utfallet om studier är registrerade före studiestart?

Den internationella kommittén för redaktörer för medicinska tidskrifter (ICMJE) beslutade år 2005 att författare ska registrera studier före studiestart för att kunna publicera dem i berörda tidskrifter. I en undersökning jämfördes det primära utfallet i registrerade studier med nypublicerade studier i sådana tidskrifter som anses ha "high impact" (26). Man undersökte RCT i 10 tidskrifter inom 3 medicinska områden – kardiologi, reumatologi och gastroenterologi – och ville bestämma om det var vanligt att primärt utfall favoriserades av ett signifikant utfall.

Av 323 inkluderade studier hade ca 45% registrerats i förväg med ett i förväg bestämt primärt utfall. Däremot saknades registrering i 27% och i 14% hade studierna registrerats efter avslut.

Hos de studier som registrerats rätt uppvisade 31% någon avvikelse mellan registrerat och uppvisat utfall. Inverkan av dessa skillnader kunde värderas i ungefär hälften av fallen, varvid ett statistiskt signifikant resultat favoriserades i 82%.

Jämförelse av primärt utfall i RCT som var registrerade med den följande publiceringen indikerar att registrering inte hindrar att utfallsbeskrivningen blir selektiv.

Vilken betydelse har det att studier avslutas i förtid?

RCT som avslutas före den planerade studietiden pga tydlig effekt får ofta stor uppmärksamhet och påverkar klinisk praxis (27). Prevalensen av sådana studier, storleken, rimligheten av behandlingseffekterna och hur mycket studierna rapporterar om anledningen till studiestoppet, har tyvärr tidigare varit okänd.

En systematisk genomgång av studier som avslutats i förtid gjordes med sökning av RCT i databaser och i tidskriftsartiklar publicerade i sin helhet för epidemiologi och kvalitet. Tolv granskare fann 143 studier, varav majoriteten hade publicerats i det som kallas för tidskrifter med "high-impact". Studierna var som regel sponsrade av läkemedelsindustrin. Tendensen från åren 2000 till 2004 visade en signifikant ökning av studier med tidigt stopp. I medeltal hade studierna – inom kardiologi, cancer och hiv/aids – bara rekryterat 63% av den planerade studiens storlek efter i median 13 månaders uppföljning. Bara en interimanalys hade genomförts och i median hade 66 patienter uppnått sitt primära effektmått vid studieavbrottet.

RCT som avslutas tidigt för uppnådd fördel – har blivit vanligare – misslyckas ofta med att adekvat berätta om orsaken till studieavbrott och visar orimligt stora behandlingseffekter, särskilt om antalet händelser är litet. Fynden i den systematiska genomgången antyder att läkare bör betrakta resultaten i sådana studier med skepsis.

Tidiga avbrott på basen av interimsanalyser av surrogatvariabler kan omöjliggöra bedömning av effekt på hårda utfallsvariabler som exempelvis överlevnad.

Fusk i forskningen

Oskicklighet och småmanipulationer i forskningsprocessen hittar man ofta exempel på. I Faktaruta 3 finns några godtyckligt valda exempel på forskningsfusk.

VEM kan lastas för fusk?

Tidskriftsredaktionernas ansvar

Kan fler än författarna lastas för fusk? Att stävja fusk i forskningen är en angelägenhet för många, däribland de vetenskapliga tidskrifternas redaktioner, skrev tre medicinska chefredaktörer i Sverige för några år sedan (28). I den svenska diskussionen menade man att rollen nästan helt förbigåtts. I enlighet med vad som föreslagits i USA nyligen bör vi i Sverige få en fristående myndighet till vilken universitetsenheter, vetenskapliga redaktioner, enskilda forskare och allmänheten kan vända sig för utredning eller rådgivning i fall av misstänkt fusk eller annan vetenskaplig ohederlighet.

Tidskrifternas eller de akademiska institutionernas ansvar

Ett slutligt svar på en reaktion över en misstänkliggjord publicerad studie, publicerades i mars 2010 i BMJ (29). På grund av uppdagade felaktigheter i tabeller i den ursprungliga artikeln (år 2006) och en mängd reaktioner på att chefredaktören meddelat att artikeln inte skulle dras tillbaka (2007), begärde författarna själva att artikeln skulle dras tillbaka (2009). En av orsakerna var att författarna av den ursprungliga artikeln "inte kunde återfinna sina originaltabeller" då de tillfrågades om detaljer. Som en reaktion konstruerade tidskriften 2009 en ny funktion i sina författarinstruktioner som innebär att en av författarna ska vara en "arkivförfattare" med ansvar att bevara originaldata efter publiceringen. Tidskriften meddelade att tidskrifter varken kan ansvara för eller har resurser att besvara frågor om hur forskningsarbetet utförts, ett ansvar som i första hand åligger de akademiska institutionerna.

Det intressanta för eftervärlden är en artikel som publicerades 2006, drogs tillbaka först tre år senare (2009) och att chefredaktören slutligen svarade 2010 pga en debatt i en annan tidskrift än den han var chefredaktör för: "Artikeln finns således inte längre!"

Några exempel på forskningsfusk

Exempel 1

År 2000 upptäcktes en känd hematolog och onkolog i Ulm tillsammans med 4 medarbetare ha förfalskat 94 vetenskapliga arbeten mellan 1994 och 1996 – "en av de största förfalskningskandalerna i medicinhistorien" (30). Den första upptäckten gjordes redan 1997 av en misstänksam medarbetare. Resultaten av en undersökningskommission år 2000 ledde till anställningsmässiga och ekonomiska åtgärder. I en ledarartikel i Läkartidningen påpekade en skribent att tillbakadragandet av de "tvivelaktiga publikationerna" får liten effekt, inte minst pga den spridning data redan fått (31). Det allvarligaste problemet med forskningsfusk, dvs att felaktiga eller förfalskade arbeten fortsätter att användas i den vetenskapliga litteraturen, har egentligen fått minst uppmärksamhet och resurser. I fallet Herrmann har 17 av de 19 tillbakadragna artiklarna fortsatt att flitigt citeras. Att hitta fuskarna tycks således vara lättare än att stoppa dem!

Exempel 2

År 2006 blev en korean respektive en norrman mycket uppmärksam på sina forskningsresultat. Senare avslöjades att de helt enkelt hittat på data. Koreanen hade skrivit om stamcells forskning i Science och norrmannen i Lancet 2005 om NSAID och munhåle cancer.

Exempel 3

Ett annat omfattande fusk uppdagades 2009. Chefredaktören för Anesthesia & Analgesia meddelade då att man funnit att minst 21 publicerade data i studier av en amerikansk anestesilog vid den akuta smärtenheten vid Baystate Medical Center (Massachusetts), var påhittade och att man skulle dra tillbaka artiklarna (32). Studierna handlade om postoperativ smärtlindring där bl a Celebra, Bextra, Vioxx, Lyrica och Efexor ingått. Åtminstone en av medförfattarna var påhittad.

Exempel 4

Ytterligare ett exempel gäller en liten studie av vaccination mot mässling, påssjuka och röda hund som 1998 publicerades i Lancet. Författarna hävdade att de bl a funnit ett samband med autism – studien väckte stort obehag. Omfattande genomgång och idogt arbete av en journalist ledde till att den förfalskade studien drogs tillbaka vid årsskiftet 2010/2011 – 12 år senare!

Kan statistiska metoder avslöja fusk?

Med hjälp av statistiska metoder är det möjligt att påvisa fusk i forskning, hävdade

statistikern Adam Taube (33). Ett framgångsrikt exempel hade publicerats i BMJ, där tidskriftens referenter fattade misstankar om ett insänt manus och därför begärde in originaldata som de satte under luppen. Med stor trovärdighet kunde de visa att data var manipulerade (påhittade?) och att observationsmaterialet inte kunde härröra från en korrekt randomiserad, enkelblind studie. Det åligger författarna att visa upp vad som verkligen är gjort i en studie, bevisbördan är deras.

De som fått som uppgift att granska artikeln (referenter, "referees") har en lika viktig uppgift, därtill kommer tidskriftsredaktörerna. Riktlinjer för hur medicinska artiklar ska utformas finns för flertalet studietyper (jfr anvisning till CONSORT). Se även s 1207.

En känd engelsk statistiker ger i en bok ett tankvärt råd vid bedömning av en artikel: "Har författarna inte skrivit det – utgå då från att de inte har gjort det heller!"

Har Evidence Based Medicine blivit ersatt av Market Based Medicine?

Mycket entusiasm har skapats kring bruket av EBM. Läkemedelsindustrin och de kommunikationsföretag som samarbetar med dem använder däremot ofta publicering i den medicinska litteraturen för att marknadsföra sina produkter.

Flera metoder som förnekar och manipulerar negativa data har utvecklats för att hjälpa till att publicera medicinska artiklar – en metod är spökskrivande, med eller utan hjälp av professionella publiceringsföretag. I en publikation (34) ges exempel på en lång rad avslöjanden, från början företagsinterna dokument som blivit tillgängliga med legalt tvång.

Författarna av artikeln föreslår att medan EBM är ett ädelt ideal så är marknadsbaserad medicin det som är den nuvarande realiteten: "Från EBM till MBM."

Hur ska de medicinska tidskrifterna klara det här problemet? Hur ska data spridas? Bättre tillgång till rådata? Möjlighet med registrering av kliniska provningar har inte löst problemet sedan det avslöjats att rapportering av data sker selekterat. Alla som granskar studier i medicinska tidskrifter måste kontrollera uppgifter i prov-

ningsregister. Möjlighet att kontrollera uppgifter hos de regulatoriska myndigheterna borde också vara en utväg, men även där kan man se skillnader mot vad som senare publiceras.

"Innan vi får en förbättrad tillgång till tillräckligt säkra data ska all stor entusiasm för MBM betraktas med skepsis" (34).

Vetenskapens plikter

Nedanstående är till stor del ett citat från en tidningsartikel som skrevs efter mötet med FN:s klimatpanel i Köpenhamn 2009 (35).

Vetenskapsmän ska tänka fritt och lansera nya djärva hypoteser. Sedan ska dessa prövas mot den empiriska verkligheten med yttersta samvetsgrannhet. En vetenskaplig maktkoncentration förutsätter öppenhet och kritisk granskning. I verkligheten är det mycket krångligare. Det kan vara svårt att avgöra vilka fakta som är relevanta för att testa en hypotes. En av vetenskapens grundvalar är att resultat ska vara reproducerbara. Det kan ta lång tid eftersom det viktiga inte är att hitta belegg för hypotesen utan att förvissa sig om att den inte går att motbevisa.

Men vetenskaplighet handlar inte om att forskarna är ofelbara utan om att deras metoder är genomskinliga. Än mer besvärligt är det faktum att vetenskapsmän är människor av kött och blod. De kan förtränga invändningar mot den egna teorin. En anledning kan vara personlig prestige och karriärskäl.

Men det kan lika gärna inträffa därför att en forskare är övertygad om att han eller hon har upptäckt något som har stor betydelse. Om man anser att man har ett fungerande vaccin mot aids som har testats tillräckligt kan man tycka att det är oansvarigt att kräva ytterligare studier som innebär att sjuka människor inte får hjälp.

Lösningen på detta problem är öppenhet och pluralism. Forskning måste vara genomskinlig. Alla ska kunna granska såväl resultat som de metoder som har använts.

En enkel metod för att bedöma en klinisk läkemedelsstudie – en RCT

Tre viktiga punkter ska kunna identifieras i en artikel.

1. Mål, primär effektvariabel
2. Statistik
3. Resultat

Mål, primär effektvariabel

Vilket var det primära målet med provningen? Målet ska vara tydligt och väl preciserat. Det anges nästan alltid i slutet av inledningen av en artikel eller i avsnittet där de statistiska metoderna beskrivs. Effektvariabeln bör helst vara enkel och lättförståelig – och av kliniskt intresse – för att räknas. Fortfarande publiceras många artiklar där man inte hittar någon preciserad primär frågeställning.

Hårda och mjuka data

Vad är viktigt?

Av stort intresse och otillräckligt undersökt är hur patienterna mår av behandlingen.

Exempel: överaktiv blåsa

I en Cochranegenomgång av antikolinerga läkemedel mot överaktiv blåsa var livskvalitet en primär effektvariabel bara i 4 av 49 studier (sammanlagt 11 332 vuxna), men i ingen av dessa fanns det användbara data (9).

Exempel: diabetes

Insulinanaloger har kommit till stor användning, både som kort- och långverkande preparat. Två mycket omfattande systematiska översikter av lång- och kortverkande insulinanaloger vid typ 2- och typ 1-diabetes publicerades för ett par år sedan (36,37). Många av de granskade studierna höll tyvärr en metodologisk låg kvalitet och underlaget bedömdes otillräckligt för att värdera effekten på livskvalitet, diabeteskomplikationer och mortalitet, angelägna frågor vid jämförelser mot äldre preparat.

Syfte

Syftet med en klinisk läkemedelsprövning är att, med utgångspunkt från ett begränsat antal väldefinierade patienter, försöka påvisa en skillnad mot standardbehandling

eller att det nya medlet eller metoden inte är sämre än det man jämför med (s k non-inferiority study). Avsikten ska alltid definieras före studiestart. Ett patientmaterial måste av etiska och praktiska skäl vara begränsat. Metoderna i den kliniska läkemedelsprövningen ska så mycket som möjligt försöka kontrollera slumpen och preferenser för vissa slutsatser (bias).

Studietyper

Bestäm om studien är en "handhavandestudie" (management study, pragmatic study) eller en "förklarande studie" (explanatory study). I det första fallet – vanligast för klinisk läkemedelsprövning – undersöker man vilken medicinsk ändamålsenlighet behandlingen har. I en studie som försöker visa att en behandling är bättre än en annan – en överlägsenhetsstudie – ska alla patienter räknas med vid bedömning av utfallet. Notera således att en analys enligt "intention to treat" (ITT) räknar in alla i ursprungsgrupperna, oavsett hur de lyckats genomföra studien. Denna analys ska alltid presenteras.

I det andra fallet (förklarande studie) undersöker man hur en behandling egentligen fungerar. Här ska bara resultat räknas från dem som verkligen genomfört studien på ett korrekt sätt. Ett typexempel är en laboratorieundersökning av ett läkemedels kinetik. I ett sådant fall blir den individ som inte kan genomföra studien ersatt av en annan. Analysen sker då enligt principen "per protocol" (PP). Det är lätt att missa skillnaderna mellan de två studietyperna och tro att resultaten av en vanlig prövning tillhör den andra typen i stället för den första. Om emellertid båda analyserna presenteras kan läsarna själva värdera resultaten.

Statistik

Här ska skälen till dimensionering av studien (antal patienter/försökspersoner/händelser) anges och därmed de statistiska förutsättningarna för att kunna påvisa skillnader/likheter i den primära effektvariabeln. Dessa bygger dock på olika antaganden. Det är till sist ändå precisionen i det faktiska utfallet som är avgörande för bedömning av studiens betydelse. Hur stor är den skillnad man vill försöka visa mellan två behandlingar? Läsaren bör fråga sig om den

skillnad man påvisat är av tillräcklig klinisk betydelse. Användningen av s k surrogatvariabler – indirekta mätpunkter i en studie – dvs laboratorievärden i stället för stroke, hjärtinfarkt och plötslig död, måste beaktas.

Med vilken precision har man kunnat visa skillnaden, eller med andra ord: Hur brett är konfidensintervallet för skillnaden i relation till skillnadens storlek? Fortfarande publiceras många artiklar som inte redovisar preciserade statistiska krav på studiens utformning. Under rubriken hör även de metoder som använts för statistiska beräkningar av materialet – metodval om data inte är normalfördelade, om justering gjorts för multipla jämförelser, vilka typer av bortfallsanalyser som gjorts, vilka interimanalyser som planerats och hur detta hanterats med avseende på kravnivå för statistisk signifikans etc.

Resultat

Huvudresultat

Här binds resultaten ihop med den primära frågeställningen och dess statistiska förutsättningar och data redovisas. Kontrollera om studiens resultat verkligen bygger på en välbeskriven frågeställning och att det inte är undergrupper eller andra utfallsmått än det primära som framhåvts som "viktiga". Svårvärderat kan det bli om studien bygger på frågeställningen "sammansatt primärt effektmått" ("composite endpoints"), speciellt i de fall då de sammansatta effektmåten är beroende av varandra (38). Fenomenet har blivit ganska vanligt men kritiserats mycket ur EBM-synpunkt; ur aspekterna rimlighet och statistisk styrka måste varje effektmått kunna studeras i detalj.

Ingen skillnad – inte detsamma som likhet

Efter ett resultat, där den skillnad man förväntat inte kunnat påvisas, försöker ibland författarna (och sponsorerna) visa att behandlingarna är lika bra. Statistiskt går det dock aldrig att visa att två terapier är lika bra. Däremot kan man visa att skillnaden sannolikt inte är större än vad som definieras som kliniskt relevant. Kvalitetskraven på en sådan, s k non-inferiority-studie, är emellertid högre än för en studie som avser att visa överlägsenhet eftersom kvali-

tetsbrister tenderar att dölja verkliga skillnader mellan experiment- och kontrollgruppen. Så kallad per protokollanalys ska presenteras i tillägg till ITT-analysen. Noggrann beskrivning av de överväganden som ligger till grund för fastställandet av "non-inferiority"-marginalen är fundamental.

"En bild säger mer än tusen ord"

Grafiska presentationer av en studie kan vara alltifrån upplysande till direkt bedrägliga. Huvudprincipen är att data i bild kompletterar tabeller och text. Inte sällan får grafer representera hela studieresultatet. Det är ofta en förenkling – kliniska läkemedelsprövningar är mer komplicerade än så. Se upp med vanligt förekommande suggestiva presentationer med "manipulerad" skalstorlek, avskurna axlar i koordinatsystem, data som inte är normalfördelade, avsaknad av eller fel spridningsmått etc.

Sekundära resultat

Vid sidan av den primära frågeställningen presenteras ofta resultaten av olika sekundära frågeställningar och resultat från olika undergrupper. Vilka utförda beräkningar är redovisade och vilka saknas? En frågeställning av betydelse kan lätt skapas av en relation som egentligen bara är ett slumpfynd. I de situationer utfallet av den primära frågeställningen är negativt är sekundära resultat nästan bara hypotesgenererande och måste undersökas i nya studier med en ny primär frågeställning. Om utfallet av den primära frågeställningen är positivt finns emellertid anledning att närmare analysera om den visade skillnaden mellan experiment- och kontrollgrupp i första hand beror på utfallet i någon subgrupp, alternativt att skillnad saknas i någon subgrupp.

Exempel. När den första jämförande studien av en angiotensin II-antagonist (losartan) och en ACE-hämmare (kaptopril) gjordes på patienter med hjärtsvikt fann man en lägre mortalitet hos dem som behandlats med losartan. Resultatet byggde på en sekundär frågeställning. Klokt nog gjordes studien om med ett större patientmaterial och mortalitet var då en primär effektvariabel. Man fann då ingen skillnad mellan behandlingarna (39).

Det måste starkt betonas att slumpens inflytande är stort. Man bör betrakta en studie som det den verkliga är – ett stickprov!

CONSORT

– att redovisa en klinisk prövning

CONSORT (Consolidated Standards of Reporting Trials) innehåller standardiserade metoder med en checklista och ett flödeschema, vars avsikt är att förbättra rapportkvaliteten hos en RCT. CONSORT (nu CONSORT 2010) är en överenskommelse som har undertecknats av många medicinska tidskrifter. Flödesschemat ger läsarna en tydlig bild av "ödet" för studiedeltagarna i rapporten – från tiden innan de blev randomiserade tills de har avslutat studien. Den som vill undersöka data i studien kan med flödesschemats hjälp bättre värdera och förstå studiens validitet. I Figur 4, s 1208, finns ett typiskt CONSORT-diagram.

Vilka var egentligen med i studien?

Tyvärr saknar många publikationer fortfarande ett CONSORT-diagram. En annan, inte ovanlig, brist är att det första steget på redovisningen saknas. Många gånger beskriver artikeln bara att man hade ett visst antal patienter med den och den diagnosen och så randomiserades de till den ena eller den andra terapin. Var kom patienterna ifrån? Hur många var de från början? Även när det finns uppgifter om den screenade populationen ska man betänka att även denna är selekterad via olika mekanismer. Information om rekryteringen är således också viktig.

Några kända exempel

- I två svenska studier över behandling av hypertoni hos äldre (STOP Hypertension) angavs bara att det var ett visst antal patienter från studiens start – inte hur de hade tagits ur den svenska primärvårdens patientflöde av äldre hypertoniker (40,41). Var de representativa? Vilka var inte med?
- I en välkänd studie över en ACE-hämmare (ramipril) effekter hos patienter med kardiovaskulära riskfaktorer

(HOPE-studien) fanns inte heller motsvarande bakgrundsuppgifter (42).

- I en omtalad jämförande studie av klopidogrel och ASA till patienter med kardiovaskulära sjukdomar (CAPRIE) medgav författarna att de inte visste hur stor ursprungspopulationen var, men kunde redovisa hur många som randomiserats (43).
- I en stor studie över effekt av alendronsyra mot osteoporos framgick det tydligt av publikationen att man för att rekrytera äldre kvinnor med möjlig osteoporos tvingades annonsera efter dem – i det här fallet skickade man ut ca 1 miljon brev för att hitta intresserade kvinnor i de lämpliga åldrarna.

Viktig är också bedömningen av validiteten i studier som medfört stora avbrottsfrekvenser. Exempel på sådana studier är undersökningar av läkemedel mot schizofreni, demens, KOL, alkoholism och nikotinberoende. Ju längre studien varar desto fler avbrott.

Validitet

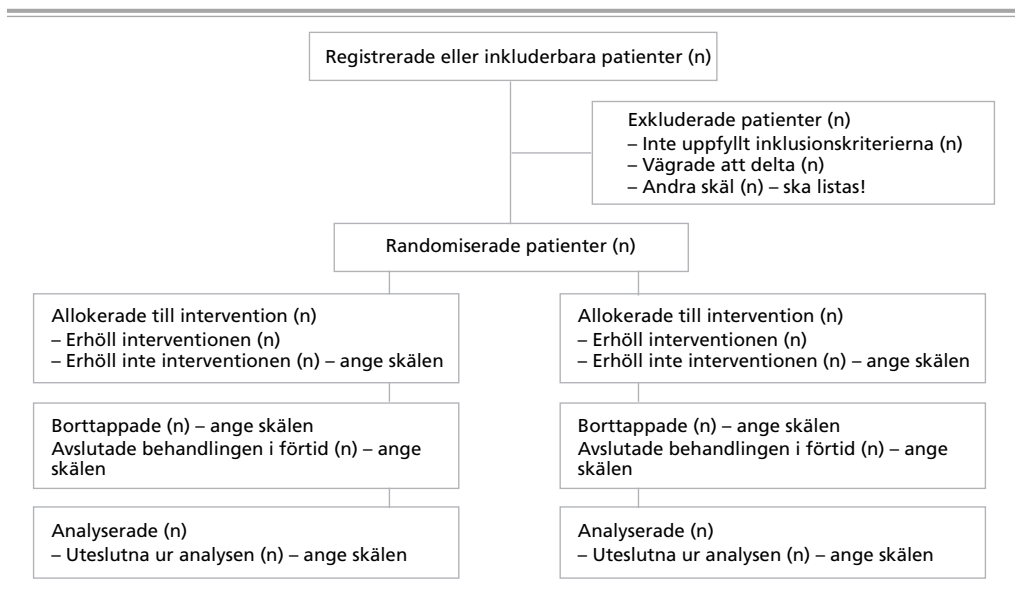
Tillförlitligheten – validiteten – av en undersökning bör bedömas då studieresultat och dess överförbarhet undersöks. Den interna validiteten är tillförlitligheten i själva undersökningen, t ex noggrannheten i metoderna och studiens genomförande.

Den externa validiteten gäller t ex hur resultaten kan överföras till andra grupper med samma sjukdom, dvs hur generaliserbara de är.

Hög intern validitet innebär inte automatiskt hög extern validitet, t ex då undersökningen är noggrann men gäller en starkt selekterad patientgrupp. Eftersom en studie med otillfredsställande intern validitet aldrig kan ha en hög extern validitet är det viktigt att särskilt granska metodologin i studien.

Hur många måste behandlas?

Det antal man måste behandla med ett läkemedel under en viss tid för att undvika en enda händelse (t ex stroke) kan uttryck-



Figur 4. Ett typiskt CONSORT-diagram

as som NNT – number needed to treat. Detta accepteras som ett meningsfullt uttryck för att beskriva ett läkemedels fördelar mot en jämförelse, t ex placebo. Sammanfattningsvis är NNT det inverterade värdet av den absoluta riskminskningen, på engelska ARR (absolute risk reduction) = $1/ARR$. En grafisk metod för att uppskatta NNT finns också (44).

Vilken är evidensen i en studie med alendronsyra?

I en äldre studie, som publicerades 1996, önskade man påvisa en minskning av risken för frakturer med ett osteoporosmedel av bisfosfonattyp (alendronsyra; Fosamax) i jämförelse med placebo (45). Studien gjordes på postmenopausala kvinnor som hade nedsatt bentäthet och som hade haft en kotfraktur. Den planerade behandlingstiden var 3 år. Vid studiens start randomiserades 1 022 kvinnor till aktiv behandling med bisfosfonat och 1 005 till placebo.

Huvudresultatet i studien var att incidensen av kliniska frakturer på alla förutbestämda lokaliseringar minskade signifikant (primär effektvariabel). Multipla kotfrakturer minskade också signifikant. Höftfrakturer – den mest kostsamma frak-

turtypen för sjukvården och samhället i övrigt – uppstod hos 11 patienter (1,076%) som fått aktiv behandling och hos 22 patienter (2,189%) som fått placebo. Skillnaden som uppmättes var statistiskt signifikant, p-värdet angavs till 0,047.

Hur bra är behandlingen med bisfosfonat enligt denna studie?

P-värdet

Resultatbudskapen i medicinska artiklar cirklar mycket kring p-värden (p står för sannolikhet – probability). Vid statistiska jämförelser är alltid grundhypotesen att det inte finns någon skillnad mellan behandlingsgrupperna, den så kallade noll-hypotesen. Det uppnådda resultatet är ett av flera man skulle kunna få om studien gjordes om. Med likadana patientgrupper skulle en del visa en större och andra en mindre skillnad. Man skulle få en viss spridning av skillnaderna beroende på slumpen.

I alendronsyrastudien blev p-värdet 0,047. Sannolikheten att få vad man har fått eller att få något ännu extremare om noll-hypotesen är sann är 0,047. Vilken slutsats man ska dra av det uppnådda resultatet beror egentligen på betraktaren, som själv ska avgöra frågan. Antingen beror skillnaden på

slumpen (men är ovanligt) eller så är resultatet ett tecken på att det finns en faktisk effekt av alendronsyra. Sannolikhetsgränsen sätts oftast vid 5%, vilket bara är en konvention. Men om man har bestämt sig för 5% som gräns är man också överens om att ett lägre p-värde ($p < 0,05$) betyder att man ska förkasta noll-hypotesen. I ett sådant fall tror vi att grupperna är olika. I vår 20:e studie med ett sådant utfall finns dock ingen verklig skillnad (falskt positiv) utan den betingas av slumpen.

P-värdet är sannolikheten att få ett visst utfall och säger heller inget om skillnaden i absoluta tal. Samma numeriska p-värde kan erhållas om man observerat en liten skillnad mellan mycket stora grupper (statistisk signifikans utan klinisk signifikans) eller en stor skillnad mellan mindre grupper.

Relativ risk (RR)

I exemplet blev utfallet av bisfosfonatbehandlingen (noga räknat) 1,076% höftfrakturer och av placebobehandlingen 2,189% höftfrakturer och förhållandet dem emellan $1,076/2,189 = 0,491$. Den relativa risken för höftfraktur vid bisfosfonatbehandling är 49%.

Relativ riskreduktion (RRR)

En 49-procentig relativ risk för höftfraktur betyder att den relativa risken (RR) för fraktur reducerats med 51%. Man kan nu hävda att behandlingen reducerat antalet höftfrakturer med 51%. Siffran ser imponerande ut, men beror på relationen mellan grupperna. Procentangivelsen lurar oss lätt.

Absolut riskreduktion (ARR)

I exemplet är den absoluta riskreduktionen 2,189% minus 1,076%, vilket blir 1,113% = 0,01113 (med ett konfidensintervall från 0,00011 till 0,02327). Talet förefaller inte så imponerande nu.

NNT

Vi behöver emellertid översätta resultatet till mer lätthanterliga siffror. I vårt exempel är NNT 1/0,01113, dvs 90 (vid exakt beräkning) och detta tal imponerar ännu mindre. För att förhindra ett fall av höftfraktur med denna bisfosfonat måste vi således behandla 90 patienter under i genom-

snitt 3 år med läkemedlet (tillsammans med kalcium).

Konfidensintervall

Vi kan inte bara nöja oss med resultatet från den genomförda undersökningen. Vi behöver veta hur pass osäkert det angivna NNT-värdet är. Precis som för p-värdet kommer vi att få andra NNT-värden om vi genomför fler undersökningar med samma förutsättningar. Genom att redovisa ett konfidensintervall kring NNT-värdet kan vi förmedla den bilden. Men den nedre och övre gränsen för konfidensintervallet placerar sig inte symmetriskt runt NNT-värdet och överraskar nog.

Av hävd anges oftast det 95-procentiga konfidensintervallet, vilket betyder att det har 95% chans att falla så att det täcker det sanna värdet för NNT. Man har alltså 95% konfidens (förtroende) för intervallet.

I vårt exempel kan vi med hjälp av ett dataprogram räkna ut att konfidensintervallet ligger mellan 43 och 8 949 patienter.

Vilken beskrivning tilltalar mest i studien med alendronsyra?

Uppgifterna om evidensen i alendronsyrastudien kan nu jämföras. Vilken ska vi välja för att bäst beskriva studieresultatet?

- Höftfrakturer minskas signifikant av alendronsyra, $p = 0,047$.
- Alendronsyra minskar höftfrakturincidensen med 51%.
- Den absoluta riskreduktionen för höftfraktur är 1,1%.
- För att förhindra en höftfraktur måste 90 patienter behandlas med alendronsyra i 3 år.

Siffran 90 är osäker – antalet ligger någonstans mellan 43 och 8 949 patienter (95% konfidensintervall).

Varje studie bär sina egna NNT-värden

För att rätt kunna tolka resultaten måste man ha tillgång till många andra uppgifter i studierna, t ex kriterierna för de patienter som deltagit i undersökningarna. NNT-värdet för olika studier kan därför sällan jämföras direkt.

NNH – number needed to harm

För att beskriva hur biverkningarna skiljer sig mellan två terapier kan man på motsvarande sätt som för NNT beskriva hur många man måste behandla tills en patient drabbas av en viss negativ effekt, NNH.

NNH – exempel 1

Effekten av klopido­grel (Plavix) som tillägg till ASA vid instabil angina pectoris studerades i en randomiserad, placebokontrollerad studie (46). En sammansatt primär effektvariabel var död av kardiovaskulära orsaker samt icke-dödlig hjärtinfarkt eller stroke. Behandling gavs i medeltal i 9 månader. I den grupp som fick kombinationsbehandling (n = 6 259) uppnådde 582 patienter (9,3%) denna händelse mot 719 patienter (11,4%) av dem som enbart fått ASA (n = 6 303). Absolut riskreduktion var 2,1%, dvs NNT blir 48 individer (32–96 individer med 95% konfidensintervall). Utöver den kvantitativa uppskattningen av nytta och risk(er) måste självklart respektive allvarlighetsgrad vägas in.

Behandlingen ledde dock till en ökad frekvens blödningar. Antalet allvarliga blödningar inträffade hos 231 individer med kombinationsbehandlingen (avrundat 3,7%), mot 169 individer som enbart fått ASA (avrundat 2,7%). NNH blir 99 individer (61–251 individer med 95% konfidensintervall).

Balansen mellan effekt och biverkningar av behandlingen kan beskrivas på följande sätt: Behandlar man 96 patienter (2 x 48) med klopido­grel i kombination med ASA i medeltal under 9 månader förhindras 2 patienter, utöver vad enbart ASA åstadkommer, att dö av en kardiovaskulär orsak, få en icke-dödlig hjärtinfarkt eller stroke. Samtidigt drabbas en patient av en allvarlig blödningskomplikation.

NNH – exempel 2

Behandling med erythropoietin kan förbättra hemoglobinvärdena hos olika patienter med anemi. I en systematisk översikt och metaanalys med erythropoietinstimulerande läkemedel som jämfördes med placebo eller annan kontrollbehandling till patienter med cancerrelaterad anemi, förbättrades hemoglobinvärdena medan mortalitet och venös tromboembolism ökade efter en be-

handlingstid på 6 till 52 veckor (47). För tromboemboliska händelser uppgick NNH till 37 individer (24–67 individer med 95% konfidensintervall). Denna tillsammans med flera studier har lett till revision av bedömning och användning av erythropoietinläkemedel.

Brist på evidens

På många områden finns det inte tillräckligt med studier för att dra några praktiska slutsatser om vad som är bra eller dåligt. ”Det finns ingen evidens.” Ibland misstolkas detta som att de metoder som används är felaktiga eller värdelösa. Så är oftast inte fallet – brist på evidens är inte liktydigt med evidens på bristande effekt – ”absence of evidence is not evidence of absence”.

EBP i stället för EBM?

Det är väl etablerat att det vi benämner evidensbaserad medicin (EBM) utgör en metod för bedömning av pålitligheten av presenterade uppgifter i den medicinska litteraturen.

Många kliniker, med rätta, tycker det är svårt att diskutera förutsättningar för enskilda patienter med utgångspunkt från exempelvis NNT-siffror. Patientens fråga är förstås: ”Gäller det mig?” Patienten har sina högst personliga bedömningar, egna bekymmer, förväntningar, föreställningar och värderingar.

Hur kan man föra över ett resultat från en studie till en enskild patient du träffar på mottagningen med samma diagnos (sjukdom)? För ett rimligt svar på den frågan måste du fråga dig följande: Skiljer sig din patient från dem som ingick i studien på något avgörande sätt? Ålder, kön, annan samtidig sjukdom, andra samtidiga läkemedel? Kan du bedöma om din patient har samma risk för sjukdom eller komplikationer som i den behandlade gruppen? Och hur är det med risk för biverkningar?

Antalet patienter som måste behandlas för att en patient ska ha nytta av behandlingen – NNT – är ett slags genomsnitt för effekt. Patienter med stor risk har en bättre möjlighet att dra nytta av behandlingen, patienter med liten risk har mindre nytta av behandlingen och kan kanske ”slippa”.

Nationella behandlingsråd kan inte ta hänsyn till denna aspekt och kan leda till kontroverser om vad som är "rätt eller fel", inte minst vid jämförelser mellan olika läkare, olika specialister (organspecialister mot allmänläkare), olika sjukvårdsområden och olika landsting.

EBP – Evidence Based Practice – har utvecklats för att göra vardagens patientmöte och de kliniska frågorna mer hanterliga. Evidensbaserad praktik innebär att den professionelle väger samman sin expertis med bästa tillgängliga kunskap, den enskildes situation, erfarenheter och önskemål vid beslut om insatser (48,49). En handledning för EBP kan nås via www.hsl.unc.edu/Services/Tutorials/EBM/welcome.htm.

Vad gäller för din patient?

Etiken får naturligtvis inte glömmas bort. Är det rimligt att din patient ska få den här behandlingen? En vis kollega uttryckte saken så här: "Du måste bekämpa det terapeutiska imperativet!" Allting behöver inte göras men du bör känna till evidensen för den behandling du avstår ifrån!

Störst risk (mindre effekt) löper din patient om du skriver ut ett läkemedel som inte är väl evidensbaserat. Tron på nyare läkemedel är ofta överdriven. Det kommer inte så många "genombrott" per år! Skynda långsamt med nya behandlingsmetoder (50)! Men kom då även ihåg att evidens inte bara behövs för att införa utan också för att skrota metoder (51).

Det är också viktigt att beakta skillnader i samtidig medicinsk behandling som patienterna har. Genom en övergång till andra terapitraditioner kommer patientpopulationerna inte att vara jämförbara med varandra – inom landet eller mellan länder – om tillräcklig tid förlöper mellan de olika studierna. NNT-värden är då ännu svårare att jämföra!

Ett problem är förstås att uppgifter av det här slaget inte finns till hands då man behöver dem. En möjlighet är att införa NNT-data (eller evidensgraderade uppgifter) i rekommendationslistor från läkemedelskommittéerna. Vid diskussioner där en behandling anges vara bättre än en annan bör man samtidigt tvingas att svara på hur stort NNT-värdet är. Bäst är att ange ett konfi-

densintervall så att man förstår osäkerheten i uppgifterna!

Risker med behandling med svag evidens

Vi har ofta en mycket optimistisk syn på olika terapiers effektivitet. Även om vi behandlar på allra bästa sätt med ett läkemedel, som t ex ska fungera förebyggande (t ex ASA), kommer många patienter ändå att drabbas av det vi försöker förebygga! Ett misslyckande (t ex stroke trots ASA-behandling) leder ofta till slutsatsen att behandlingen nu behöver bytas mot ett "modernare" medel (t ex klopido-grel). Patienten ser sig sannolikt bättre omhändertagen av åtgärden, men gott vetenskapligt stöd för behandlingsändringen kanske saknas. Det är ändå troligt att den behandlande läkaren uppfattar bytet som effektivt då det kanske dröjer länge innan en ny stroke inträffar. Bytet till den nya behandlingen uppfattas ske enligt "beprovad erfarenhet" men saknar evidens. Se vidare kapitlet Cerebrovaskulära sjukdomar, avsnittet Trombocythämmare efter TIA/ischemisk stroke, s 386.

Riktlinjer

– varför gör de inte som vi säger?

Behandling av lipidrubbnig

Hos patienter med stor risk för kardiovaskulär sjukdom eller med kardiovaskulär sjukdom som tagits fram ur ett danskt projekt om diabetesscreening, undersökte man användningen av lipidsänkande läkemedel (52). Undersökningen gjordes på 139 allmänläkarmottagningar. Trots tydliga riktlinjer fann man en stor underförskrivning av de lipidsänkande medlen med en tydlig tröghet vid nyförskrivning.

Kirurgi

En retrospektiv värdering av omfattningen av kirurgi hos patienter med mitralisinsufficiens som enligt gällande riktlinjer borde ha opererats gjordes i Michigan (53). I en kohort med 300 patienter med allvarlig mitralisinsufficiens var det bara hälften som hade genomgått ingreppet trots att indikation förelåg. En tänkbar orsak kan ha varit en förhöjd operationsrisk, men data visade inte detta.

Läkemedel efter hjärtinfarkt

Efter hjärtinfarkt säger riktlinjer att patienterna ska behandlas med 4 olika farmaka – ASA, ACE-hämmare, betablockerare och statiner. I en studie ville man undersöka användningen och hur förändringarna av dessa läkemedel ser ut i ett populationsperspektiv (54). Studien beskrev användningen efter hjärtinfarkt i förhållande till ålder, kön och under olika tidsperioder. Totalt följdes 6 334 patienter (män och kvinnor) med akut hjärtinfarkt på 11 sjukhus under vartannat år mellan 1995 och 2005 i ett storstadsområde i Massachusetts. Inte oväntat kunde man se en stor ökning av dessa terapier, både hos män och kvinnor. Under de sista 2 åren av undersökningsperioden var det dock mindre än hälften som fått alla 4 läkemedelstyperna. Trots att trenden var positiv ansågs det finnas ett behov av att identifiera patienter som är underbehandlade, särskilt kvinnor och äldre, och öka utbildningen för att förbättra behandlingen.

Förklaringar till varför man inte följer riktlinjer

Frågan om varför läkare inte följer riktlinjer försökte man belysa vid en studiegenomgång (55). Vid litteraturgenomgång identifierade man > 5 000 artiklar i frågan. Av dessa valdes 76 ut för närmare granskning. Genomgången visade på att det var svårt att använda ett förslag på lösning från ett område till ett annat.

En omfattande analys av exempel på avvikelser från följsamhet beskrevs av chefredaktören i *American Journal of Medicine* (56):

”Idag finns det hundratals och åter hundratals riktlinjer. De har alla noggrant utformats av en stor mängd olika vetenskapliga sällskap/föreningar. En del av dessa är tämligen omfattande, det finns exempel på riktlinjer som överstiger 300 textsidor. Om man betänker antalet riktlinjer och deras storlek är det inte förvånande att jäktade läkare inte har tid att gå igenom, än mindre att ta till sig det presenterade materialet. Vidare kan man lätt förstå att de rekommenderade åtgärderna kommer att förbises i en läkares arbetsdag med dess upprepade störningar.

Slutligen – stora, randomiserade, globala, dubbelblinda studier ger resultat som oftast ger en bild av resultat för majoriteten, men inte för alla patienter som deltagit i dessa; det finns förstås en minoritet som inte uppför sig som majoriteten. Läkare kan tänka att deras speciella patienter inte passar in i mönstret som man sett i dessa studier. I viss utsträckning kan de ha rätt i det antagandet. Attityden leder en del läkare till att utveckla ett eget terapeutiskt protokoll som avviker från huvudresultaten.”

Riktlinjer bygger inte alltid på bästa evidens

Riktlinjer försöker ge en rimlig bild av hantering och metodbeskrivning av specifika medicinska frågor. Vem arbetar med riktlinjer och hur oberoende är dessa personers åsikter från tex läkemedelsindustri och medicintekniska företag? Hur är den bakomliggande kunskapen?

Exempel. De amerikanska kardiologförbunden, ACC (American College of Cardiology) och AHA (American Heart Association), har i mer än 20 år publicerat kliniska riktlinjer som är rekommendationer för vård av patienter med kardiovaskulära sjukdomar. Råden har blivit betydelsefulla dokument (även internationellt) för hur kardiologisk handläggning ska ske och utgör rättesnören för vårdens kvalitet inom ämnesområdet.

År 2009 publicerades en analys av dessa riktlinjer från åren 1984 till 2008 (57). För 53 riktlinjer med över 22 ”frågeställningar”, sammanlagt 7 196 rekommendationer, studerades förändringar över åren samt evidensgraden i de nuvarande versionerna.

Man visade bl a att endast 245 av 1 305 klass-1-rekommendationer grundades på den högsta evidenskategorin, evidensnivå A, som innebär rekommendationer baserade på evidens från multipla RCT- eller metaanalyser. Undersökningen visade att påtagligt många riktlinjer är dåligt underbyggda och till stor del utgår från en lägre evidensnivå eller (ännu sämre) från expertåsikter. Undersökningen visade också att andelen rekommendationer som bygger på icke konklusiva resultat ökar.

Undersökningen belyser behovet av förbättringar i arbetsprocessen, både vad gäller sättet att skriva riktlinjer på samt det faktum att basen för evidensbedömning behöver utökas.

Hur rätt och fort går processen?

I ett annat exempel påpekade nyligen författaren av ett brev till BMJ att amerikanska riktlinjer för behandling av diabetes (ADA, 2010) inte tagit hänsyn till en metaanalys i BMJ om utebliven effekt av ASA vid diabetes (21). I riktlinjerna föreslog man således fortfarande (sedan flera år) användning av ASA, trots en betydande publicerad studie.

Utveckling av riktlinjer

Processen att utveckla riktlinjer/rekommendationer (guidelines) för utredning och behandling av särskilda, definierade diagnoser är en grannliga procedur. De ”regler” som omger processen är mycket detaljerade, vilket framgår av dokumenten publicerade av det amerikanska National Guideline Clearinghouse (www.guideline.gov) och National Quality Measures Clearinghouse (www.qualitymeasures.ahrq.gov).

AGREE

AGREE (58) är ett internationellt samarbetsprojekt mellan forskare och policyansvariga i flera europeiska länder samt USA, Kanada och Nya Zeeland för att förbättra kvalitet och effektivitet i kliniska riktlinjer.

Syftet

Syftet med instrumentet för bedömning av riktlinjer för forskning och utvärdering är att ge en ram för den kvalitativa bedömningen av kliniska riktlinjer.

Kliniska riktlinjer definieras: ”Systematiskt utformade utlåtanden som ska verka som ett stöd för läkare och patienter vid beslut om lämplig hälso- och sjukvård vid specifika kliniska förhållanden.”

1. Syftet är att ”ge explicita rekommendationer med tydlig avsikt att påverka klinikernas arbete”.

2. Med kvalitet i kliniska riktlinjer avses förvisningen om att eventuell bias i arbetet med riktlinjerna har blivit korrekt hanterad, att rekommendationerna har både hög intern och extern validitet, samt att de är användbara i det praktiska arbetet.

Detta innebär att hänsyn tas till fördelar, nackdelar och kostnader av rekommendationerna samt därtill hörande praktiska aspekter. Därför innehåller bedömningen även ställningstaganden om de metoder som har använts vid utformningen av riktlinjerna, de slutliga rekommendationerna samt faktorer som berör implementeringen.

AGREE-instrumentet bedömer kvaliteten både i rapporteringen och i vissa aspekter av rekommendationerna. Det bedömer den avsedda validiteten i en riktlinje, dvs sannolikheten att önskat resultat uppnås. Instrumentet bedömer däremot inte riktlinjens validitet vad gäller resultatet för den enskilda patienten. De flesta av AGREE-instrumentets kriterier baserar sig snarare på teoretiska antaganden än på empiri. Kriterierna är resultatet av diskussioner mellan forskare från olika länder, vilka har stor erfarenhet av och kunskap om kliniska riktlinjer. Därför bör AGREE-instrumentet uppfattas som en återspeglning av kunskapsläget inom området.

Vilka riktlinjer kan

bedömas med AGREE-instrumentet?

AGREE-instrumentet är utformat för att bedöma utarbetade riktlinjer från lokala, regionala, nationella eller internationella arbetsgrupper eller av berörda statliga organisationer. Detta inkluderar:

1. nya riktlinjer,
2. befintliga riktlinjer,
3. uppdatering av befintliga riktlinjer.

AGREE-instrumentet är övergripande och kan appliceras på riktlinjer inom alla slags sjukdomskategorier, inklusive vid diagnostik, hälsofrämjande arbete, vård och behandling eller olika slags interventioner.

Socialstyrelsen arbetar med framtagning av nationella riktlinjer och ett med AGREE innehållsmässigt likartat sätt användes vid det framtagna riktlinjearbetet för diabetes

2010 (59). Användning av AGREE tycks inte ha använts vid andra frågeställningar.

Vilka borde använda AGREE-instrumentet
AGREE-instrumentet är avsett att användas av följande grupper:

1. Policyansvariga, som ett stöd i deras beslut om vilka riktlinjer som bör rekommenderas i verksamheten. På denna nivå bör instrumentet vara en del av den formella bedömningen.
2. Personer som själva utarbetar riktlinjer, som ett sätt för dem att arbeta metodiskt, strukturerat och noggrant samt använda verktyget för självskattning, så att riktlinjerna blir välgenomtänkta.
3. Vårdgivare, som först vill göra en egen bedömning innan de antar rekommendationerna.
4. Utbildningsansvariga eller lärare för att bidra till en bättre kompetens hos sjukvårdspersonalen att göra kritiska bedömningar.

Kanadensiska erfarenheter

I en ledarartikel i tidskriften "Canadian Journal of Cardiology" (maj 2010) redovisades en beskrivning av hur man omformulerat tillvägagångssättet för framtagning av riktlinjer, inte minst pga brister i evidens bakom tidigare överenskommelser, som tidigare varit starkt USA-influerade (60), jämför (57). Man avser arbeta för att minst 51% av medlemmarna av expertgrupper inte har några intressekonflikter eller industrianknytningar, liksom att arbetet är öppet för omfattande diskussioner och att ett godkännande krävs innan publicering.

Det mest spännande i de nya rekommendationerna är att man kommer att följa GRADE (se nedan) i stället för det amerikanska systemet, vilket blir en naturlig övergång från evidens till rekommendation.

Framtid och nutid

– Är riktlinjer förhandlingsbara?

I samma tidskrift presenterades en uppdatering av de kanadensiska hypertoni-riktlinjerna. Rekommendationerna hade framröstats av de 63 medlemmarna i en av arbetsgrupperna. Alla framlagda rekommendationer nådde konsensus i minst 80%.

Evidensgradering

Undersökningsresultat (t ex från läkemedelsstudier) och deras betydelse kan behöva grupperas efter sin styrka, benämnt som t ex evidensstyrka. Ett vanligt sätt utgår från en gradering av studier från en högsta nivå (bäst) – RCT eller systematiska översikter av flera väldegnade RCT – ner till en lägsta nivå (sämst) – uppfattningar av "respekterade auktoriteter". Så har SBU tidigare använt en gradering av det vetenskapliga underlaget som benämnts "starkt" (evidensstyrka 1), "måttligt" (evidensstyrka 2) eller "begränsat" (evidensstyrka 3). I en fjärde kategori fanns underlag som befunnits otillräckligt eller motsäggande.

Evidensgradering och rekommendationer – GRADE

Sedan år 2000 har en internationell arbetsgrupp arbetat med att ta fram ett systematiskt och tydligt sätt att värdera kvaliteten på evidens och givna behandlingsrekommendationer (61,62). Systemet benämns GRADE (Grading of Recommendations, Assessment, Development and Evaluations). Arbetsgruppen ger fortlöpande ut råd (guides) i form av systematiska översikter och HTA-värderingar (Health Technology Assessment). Dessa publiceras på en hemsida, tillgängliga i fullformat och pdf-format utan kostnad (www.gradeworkkinggroup.org). För närvarande (juli 2013) finns där ett 15-tal publikationer.

GRADE används av flera internationella och nationella aktörer. GRADE-systemet har införts i dokument från Socialstyrelsen (t ex nationella riktlinjer) och SBU.

GRADE försöker värdera utfall i specifika, kliniska studier. Metoden bedömer metodologiska svagheter i sammanlagda studier, konsistensen i resultaten, deras överförbarhet till aktuell patientgrupp och hur effektiva behandlingarna har visat sig vara.

Vad innebär GRADE?

Evidensgradering innebär att man bedömer hur starkt det vetenskapliga underlaget är. För att bedöma den sammanlagda evidensstyrkan utgår man enligt GRADE från att initialt tilldela slutsatser baserade på RCT-studier 4 poäng, 3 poäng för s k pseudoran-

domiserade studier, 2 poäng för evidens baserad på observationella studier och 1 poäng för evidens grundad på fallstudier. Därefter kan poängen sänkas eller ibland höjas baserat på åtta olika bedömningsgrunder: studiekvalitet, överförbarhet, överensstämmelse, precision i data, publikationsbias, effektstorlek, dos-/respons samband och "omvända confounders" (störfaktorer).

Studiekvalitet

Studiekvaliteten (risk för bias) för de studier som slutsatsen baseras på bedöms. De delmoment som bedöms är selektionsbias, behandlingsbias, bedömningsbias, bortfallsbias, rapporteringsbias och intressekonflikter. Risk för bias kan sänka evidensgraden med 1, eller vid stor risk, 2 poäng.

Överensstämmelse

Man minskar med 1 poäng om relevanta studier till stor del är motsägelsefulla i sina utfall. Vid påtaglig heterogenitet blir avdraget 2 poäng.

Överförbarhet

Upp till 2 poäng kan dras ifrån av följande skäl:

- studier som är begränsat generaliserbara då ingående population är för snäv eller för bred
- för utfall som är svåra att använda för generalisering
- då jämförelsealternativen inte varit i överensstämmelse med dem som normalt används.

Precision i data

Vid oprecisa data blir avdraget 1 eller, vid stor osäkerhet 2 poäng.

Publikations- och/eller rapporteringsbias

Vid klar risk för publikationsbias blir avdraget 1 poäng.

Effektstorlek

Om effekten är stor, > 2 gånger eller mycket stor > 5 gånger kan 1 respektive 2 poäng adderas.

Dos-/respons samband

Man kan lägga till 1 poäng om det finns evidens för en dos-/responseffekt.

Omvända confounders

Om det finns s k confounders (störfaktorer) som tydligt borde minska den uppmätta effekten eller sambandet alternativt faktorer som tydligt kan förklara en utebliven effekt eller samband kan 1 poäng läggas till.

Summering av evidensstyrka

När samtliga faktorer tagits i beaktande görs en sammanvägning av poängen vilket ger den slutliga evidensgraden för ett utfall eller en slutsats. Vid tveksamhet kan man tänka sig att addera eller dra ifrån delar av poäng för olika defektorer. Huvudsaken är att evidensgraderingen redovisas på ett tydligt och konsekvent sätt. För evidensgradering se Tabell 1, s 1216.

Värdering av kostnadseffektivitet

Värdering av kostnadseffektivitet bedöms separat av hälsoekonomisk expertis.

Källor om EBM

Ett bra exempel på hur man kan sammanställa evidensbaserad kunskap är BMJ Clinical Evidence, nedan kallad CE (på Internet och som "Handbook"). Innehållet och antal ämnesområden i CE utvidgas snabbt och belyser vardagsnära medicinska problem. Tillgången är helt kostnadsfri för bl a engelska allmänläkare. Några svenska landsting har prenumererat på servicen.

Numera finns en annan möjlighet som heter Best Practice, också denna under BMJ. Detta är ett helt nytt koncept vad gäller information för beslutshjälp. Den kombinerar den senaste evidensen från forskningen hämtad från CE tillsammans med granskning av experter. Den utgår från patientkonsultationen och täcker diagnos, prognos och prevention. I Best Practice finns referenser med direktlänkar till dokument eller till sammanfattningar i PubMed. Frånsett en gratis provomånad kan man enskilt eller i grupp prenumerera på tjänsten, som sker över Internet. Databanken Best Practice finns nu som resurs via landstingens litteratursökningstjänst EIRA.

Tabell 1. Evidensgradering enligt Grade

Evidensgrad	Kommentar
Starkt (++++)	Det är mycket osannolikt att framtida forskning kommer att ha betydelse för skattningen av effekt och dess tillförlitlighet.
Måttligt starkt (+++)	Det är sannolikt att framtida forskning kommer att ha betydelse för skattningen av effekt och dess tillförlitlighet. Det är möjligt att skattningen kommer att ändras.
Begränsat (++)	Det är högst sannolikt att framtida forskning kommer att ha betydelse för skattningen av effekt och dess tillförlitlighet. Det är mycket möjligt att skattningen kommer ändras.
Otillräckligt vetenskapligt underlag (+)	Skattningen av effekten är mycket osäker (inget uttalande om effekt). Det vetenskapliga underlaget saknas eller är bristfälligt.

En hel del publicerade, systematiska översikter kvalitetsbedöms och kan sökas på Internet hos CRD (Centre for Reviews and Dissemination). Hos CRD kan man även hämta in analyser av hälsoekonomiska bedömningar och andra rapporter från många HTA-organisationer (SBU är en av dessa).

Ett känt nätverk av institutioner som lätt hittas på Internet är Cochrane Collaboration (CC) – "We are world leaders in evidence-based health care." Bidrag sker i ett samarbete med sammanställning av systematiska värderingar av hälsovårdsinsatser, de s k Cochrane Reviews, som publiceras online i The Cochrane Library. Många länder har skaffat nationella licenser för direktaccess – i Sverige sponsras åtkomsten av SBU. CC och Cochrane Library kan nu nås direkt via hemsidorna www.cochrane.org och www.thecochranelibrary.com/view/0/index.html.

Nedan följer en lista över några webbplatser som kan vara av värde. Flera av dem innehåller länkar till liknande områden.

- SBU – Statens beredning för medicinsk utvärdering. SBU-Alert via samma länk. www.sbu.se
- Rapporter från det internationella nätverket (etablerat 1993) för SBU-liknande organisationer, International Network of Agencies for Health Technology Assessment – INAHTA – hittas via följande länk: www.inahta.org
- Den kanadensiska motsvarigheten till SBU heter Canadian Agency for Drugs and Technologies in Health – publicerar en stor mängd lättöverskådliga EBM-rapporter. www.cadth.ca
- NICE – National Institute for Health and Clinical Excellence. Organisation i England som bedömer medicinska behandlingar. Verksamheten är organiserad under NHS (National Health System). www.nice.org.uk
- The Centre for Evidence Based Medicine; i Oxford under NHS R&D. Flera praktiska länkar. www.cebm.net
- Canadian Institute of Health Research (Toronto). Lär ut att öva och praktisera EBM. <http://ktclearinghouse.ca/cebm>
- Information Resources at SchARR. En utomordentlig lista över EBM-länkar som produceras av Andrew Booth vid School of Health and Related Research (SchARR), Universitetet i Sheffield, England. www.shf.ac.uk/scharr/sections/ir
- Bandolier – webbplats i Oxford med översikter och kritiska bedömningar av olika slags medicinska åtgärder inklusive behandling med läkemedel. Slutade skapa ny information 2010. www.medicine.ox.ac.uk/bandolier
- Therapeutics Initiative – Evidence based information on drug therapy, University of British Columbia, Vancouver, Canada. Ger ut tidskriften Therapeutics Letter. <http://ti.ubc.ca>
- Länk till Clinical Evidence: bestpractice.bmj.com
- Länk till CONSORT: www.consort-statement.org
- Länkar till PRISMA, MOOSE, STARD, REMARK, TREND, STROBE: www.consort-statement.org/?o=1219

- Länk till QUROM:
<http://www.ncbi.nlm.nih.gov/pubmed/10584742>
- CRD Databases:
www.york.ac.uk/inst/crd
- Länk till RATS:
www.biomedcentral.com/info/fora/rats

Nationella och regionala/lokala riktlinjer

Behandlingsrekommendationer ges regelbundet ut av Läkemedelsverket (www.lakemedelsverket.se). Rekommendationer, t ex från läkemedelskommittéer, har "däremot" ofta saknat tydliga beskrivningar av evidens och skäl till rekommendationer av terapier bortsett från ekonomiska skäl. Sådana hänsyn har emellertid tagits av Stockholms läns landsting i Janusinfo med "Kloka Listan" (www.janusinfo.se).

Kvalitetsregister – hur är det med evidensen?

"Det finns ett tilltagande intresse i svensk sjukvård för att bättre värdera vårdens resultat och kvalitet (63). Fokus är inte längre enbart på ekonomi och organisation. Ett uttryck för detta är de s k öppna jämförelser som publicerats senare år av SKL. Syftet med dessa öppna jämförelser är framför allt att stimulera till förbättringar i vården. Men en viktig förutsättning för att en förbättringsprocess ska uppstå är att de mätmetoder som används i jämförelserna är valida och relevanta, dvs att de mäter vad som avses att mätas och att de är uttryck för viktiga aspekter av den jämförda vården. Mätmetoderna måste också vara reliabla, dvs mätfelet måste vara litet. Om dessa förutsättningar inte är uppfyllda kan jämförelserna aldrig uppfattas som rättvisa – därigenom blir de snarast kontraproduktiva.

"Det viktigaste fyndet ... var en oväntat hög andel patienter med felaktiga eller osäkra diagnoser i det patientadministrativa registret (63)." En bedömning av öppna jämförelser har senare publicerats som ger en mer positiv bild (64).

Flera medicinska områden har blivit föremål för datainsamling med avsikt att an-

vändas för kvalitetsutveckling och forskning inom vården. År 2012 fanns ett drygt 70-tal nationella kvalitetsregister i drift med gemensamt ekonomiskt stöd från sjukvårdshuvudmännen och staten, se vidare www.kvalitetsregister.se. På uppdrag av Socialdepartementet och Sveriges Kommuner och Landsting (SKL) har en översyn av de nationella kvalitetsregistren nyligen gjorts för framtida användning.

Referenser

1. Järhult B. På flykt från universitetets förgiftade modersmjölk. Hur undviker vi att bli informationens avfallskvarn?
Läkartidningen. 1997;94:4804–6.
2. www.kvalitetsregister.se
3. Werkö L. Tidningsläsning – fortbildning för aktiva läkare. *Läkartidningen*. 1999;96:4241–3.
4. Liedholm H, Linné AB. Läkemedelsinformationen måste granskas kritiskt. Industrins enda önskade sluteffekt: Receptet från doktorn till patienten. *Läkartidningen*. 2001;98:1892–7.
5. Relman AS. Are we a filter or a sponge?
N Engl J Med. 1978;299:197.
6. Lövttrup M. "Hedersförfattare" förekommer i varannan medicinsk avhandling.
Läkartidningen. 2010;107:164–8.
7. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up?
PLoS Med. 2010 Sep 21;7(9):e1000326.
8. Nakao N, Yoshimura A, Morita H, Takada M, Kayano T, Ideura T. Combination treatment of angiotensin II receptor blocker and angiotensin-converting-enzyme inhibitor in non-diabetic renal disease (COOPERATE): a randomised controlled trial. *Lancet*. 2003;361:117–24.
9. Hay-Smith J, Ellis G, Herbison GP. Which anticholinergic drug for overactive bladder symptoms in adults.
Cochrane Database Syst Rev. 2005;(3): CD005429.
10. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. *Lancet*. 1995;345:669–85.

11. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. *Lancet*. 1988;2:349–60.
12. Maggioni AP, Franzosi MG, Fresco C, Turazza F, Yognoni G. GISSI trials in acute myocardial infarction. Rationale, design, and results. *Chest*. 1990;97(4 Suppl):1465–1505.
13. Toole JF, Malinow MR, Chambless LE, Spence JD, Pettigrew LC, Howard VJ, et al. Lowering homocysteine in patients with ischemic stroke to prevent recurrent stroke, myocardial infarction, and death: the Vitamin Intervention for Stroke Prevention (VISP) randomized controlled trial. *JAMA*. 2004;291:565–75.
14. Bønaa KH, Njølstad I, Ueland PM, Schirmer H, Tverdal A, Steigen T, et al; NORVIT Trial Investigators. Homocysteine lowering and cardiovascular events after acute myocardial infarction. *N Engl J Med*. 2006;354:1578–88.
15. Martí-Carvajal AJ, Solà I, Lathyris D, Karakitsiou DE, Simancas-Racines D. Homocysteine-lowering interventions for preventing cardiovascular events. *Cochrane Database Syst Rev*. 2013;1:CD006612.
16. The perils of journal and supplement publishing (editorial). *Lancet*. 2010;375:347.
17. Vedula SS, Bero L, Scherer RW, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *N Engl J Med*. 2009;361:1963–71.
18. Wang AT, McCoy CP, Murad MH, Montori VM. Association between industry affiliation and position on cardiovascular risk with rosiglitazone: cross sectional systematic review. *BMJ*. 2010;340:c1344.
19. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *JAMA*. 2004;291:2457–65.
20. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev*. 2009;(1): MR000006.
21. Cook GE. Missing published and unpublished data in guidelines. *BMJ*. 2010;340:c1337.
22. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ*. 2003;326:1171–3.
23. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials. A reflection of treatment effect or adverse events? *JAMA*. 2003;290:921–8.
24. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*. 2003;326:1167–70.
25. Jefferson T, Rudin M, Brodney Folse S, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database Syst Rev*. 2007;(2): MR000016.
26. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA*. 2009;302:977–84.
27. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294:2203–9.
28. Zetterström R, Mitelman F, Rydholm A. Vetenskapligt fusk – tidskrifternas dilemma. *Läkartidningen*. 2004;101:2743–4.
29. Response. Steven L Shafer responds to Jeanne Lenzer. *BMJ*. 2010;340:b3920.
30. Wikipedia (tysk version) (sökdatum 13 april 2010).
31. Milerad J. Forskningsfusk – lättare upptäcka än åtgärda. *Läkartidningen*. 2006;103:2112.
32. Shafer SL. Retraction notice. *Anesth Analg*. 2009;108:1351.
33. Taube A. Att sätta P för fusk. *Läkartidningen*. 2006;103:143–4.
34. Spielmanns GI, Parry PI. From evidence-based medicine to marketing-based medicine: Evidence from internal industry documents. *Bioethical Inquiry*. 2010;7:13–29.
35. Anonym. Makt i vit rock. *Om FN:s klimatpanel*. DN. 7 mars 2010.
36. Tran K, Banerjee S, Li H, Cimon K, Daneman D, Simpson SH, et al. Long-acting insulin analogues for diabetes mellitus: meta-analysis of clinical outcomes and assessment of cost effectiveness [Technology Report number 92]. *Ottawa: Canadian Agency for Drugs and Technologies in Health*; 2007.

37. Banerjee S, Tran K, Li H, Cimon K, Daneman D, Simpson S, Campbell K. Short-acting insulin analogues for diabetes mellitus: meta-analysis of clinical outcomes and assessment of costeffectiveness [Technology Report no 87]. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2007.
38. Montori VM, Busse JW, Permyanier-Miralda G, Ferreira I, Guyatt GH. How should clinicians interpret results reflecting the effect of an intervention on composite endpoints: Should I dump this lump? *ACP J Club*. 2005;143:A8.
39. Pitt B, Poole-Wilson PA, Segal R, Martinez FA, Dickstein K, Camm AJ, et al. Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial the Losartan Heart Failure Survival Study ELITE II. *Lancet*. 2000;355:1582–7.
40. Ekblom T, Dahlöf B, Hansson L, Lindholm LH, Scherstén B, Wester PO. Antihypertensive efficacy and side effects of three beta-blockers and a diuretic in elderly hypertensives: a report from the STOP-Hypertension study. *J Hypertens*. 1992;10:1525–30.
41. Hansson L, Lindholm LH, Ekblom T, Dahlöf B, Lanke J, Scherstén B, et al. Randomised trial of old and new antihypertensive drugs in elderly patients: cardiovascular mortality and morbidity the Swedish Trial in Old Patients with Hypertension-2 study. *Lancet*. 1999;354:1751–6.
42. Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med*. 2000;342:145–53.
43. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). CAPRIE Steering Committee. *Lancet*. 1996;348:1329–39.
44. Chatellier G, Zapletal E, Lemaitre D, Menard J, Degoulet P. The number needed to treat: a clinically useful nomogram in its proper context. *BMJ*. 1996;312:426–9.
45. Black DM, Cummings SR, Karpf DB, Cauley JA, Thompson DE, Nevitt MC, et al. Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. Fracture Intervention Trial Research Group. *Lancet*. 1996;348:1535–41.
46. Yusuf S, Zhao F, Mehta SR, Chrolavicius S, Tognoni G, Fox KK; Clopidogrel in Unstable Angina to Prevent Recurrent Events Trial Investigators. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med*. 2001;345:494–502.
47. Bennett CL, Silver SM, Djulbegovic B, Samaras AT, Blau CA, Gleason KJ, et al. Venous thromboembolism and mortality associated with recombinant erythropoietin and darbepoetin administration for the treatment of cancer-associated anemia. *JAMA*. 2008;299:914–24.
48. Socialstyrelsen. Att arbeta evidensbaserat. www.socialstyrelsen.se
49. Haynes RB, Devereaux PJ, Guyatt GH. Physicians' and patients' choices in evidence based practice. *BMJ*. 2002;324:1350.
50. Werkö L. Skynda långsamt med nya behandlingsmetoder! *Läkartidningen*. 2002;99:52–3.
51. Werkö L, Eliasson M. Evidens behövs inte bara för att införa utan också skrota metoder. *Läkartidningen*. 2001;98:4228–9.
52. Graversen L, Christensen B, Borch-Johnsen K, Lauritzen T, Sandbaek A. General practitioners' adherence to guidelines on management of dyslipidaemia: ADDITION-Denmark. *Scand J Prim Health Care*. 2010; 28:47–54.
53. Bach DS, Awais M, Gurm HS, Kohnstamm S. Failure of guideline adherence for intervention in patients with severe mitral regurgitation. *J Am Coll Cardiol*. 2009;54:860–5.
54. Fornasini M, Yarzebski J, Chiriboga D, Lessard D, Spencer FA, Aurigemma P, et al. Contemporary trends in evidence-based treatment for acute myocardial infarction. *Am J Med*. 2010;123:166–72.
55. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. A Framework for improvement. Why don't physicians follow clinical practice guidelines? *JAMA*. 1999;282:1458–1465.
56. Alpert JS. Why are we ignoring guideline recommendations? *Am J Med*. 2010;123:97–8.
57. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA*. 2009;301:831–41. Erratum in: *JAMA*. 2009;301:1544.
58. Bedömning av riktlinjer för forskning och utvärdering – AGREE instrumentet. Socialstyrelsen. 2003. Artikelnummer: 2003-119-1. www.socialstyrelsen.se
59. Nationella riktlinjer för diabetesvården 2010 – Stöd för styrning och ledning. Socialstyrelsen. 2010. ISBN: 9789186301880. www.socialstyrelsen.se

60. Graham MM, Pullen C. Renovating canadian cardiovascular society guidelines and position statements: a new foundation. [Can J Cardiovasc.](#) 2010;5:233–5.
61. Clinical Evidence www.clinicalevidence.bmj.com
62. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. [BMJ.](#) 2008;336:924–6.
63. Rytberg B, Nyström E, Stenberg A, Palm K, Näslund I. "Öppna jämförelser" ger inget säkert mått på infarktvårdens kvalitet. Epidemiologiska registret duger inte för att jämföra landsting. [Läkartidningen.](#) 2009;106:2121–4.
64. Sörman H. Öppna jämförelser analyserade: "Ett av de viktigaste bidragen till vårdens modernisering". [Läkartidningen.](#) 2010;32:1828–9.
-