

Kritisk vurdering af en oversigtsartikel

af Hans Lund, fysioterapeut, ph.d., Parker Institutet, Frederiksberg Hospital

Lund H. Kritisk vurdering af en oversigtsartikel (2000)
(online 2003, 4. november). *Forskning i Fysioterapi* (1. årg.), s. 1-6
URL: <http://www.ffy.dk/sw1250.asp>

Artiklen har tidligere været bragt i *Nyt om Forskning* nr. 1 2000, side 18-22.

Mængden af videnskabelig litteratur er gigantisk. Der publiceres ca. 1-2 millioner medicinsk videnskabelige artikler hvert år, og det ville derfor være nærmest umuligt for en enkelt person at få et overblik. Det bedste alternativ, der findes, er, at du kigger efter en oversigtsartikel (review). Review artikler findes typisk i tre former, og det er vigtigt at skelne mellem disse. Den traditionelle form er oversigtsartiklen, hvor en ekspert sammenskriver sin viden om et emne. Forfatteren understøtter sine påstande med udvalgte artikler, og der er ofte mange referencer til en sådan artikel. Denne type af review artikler kan være en stor hjælp, hvis man ønsker at få en ide om begreber og termer inden for et bestemt felt. I nogle tilfælde er sådanne forfatteres konklusioner ikke altid i overensstemmelse overens med den faktiske evidens, der er på området (2). Sådanne "usystematiske" review artiklers konklusion bør derfor læses som et udtryk for forfatterens holdninger og ikke som en præsentation af den faktiske evidens.

Søger vi dokumentation for en bestemt behandlingsmetode eller lignende, kan man ikke anbefale at fæste lid til de traditionelle review artikler. Den systematiske review artikel er derimod netop udfærdiget efter nogen relative skrappe krav, for at læseren kan stole på den konklusion man kommer frem til. Disse krav er bl.a., at forfatteren fortæller, hvordan vedkommende har søgt efter den videnskabelige litteratur – f.eks. hvilke database, hvilke søgeord o. l. Derudover bør forfatteren også både have gennemført og beskrevet, hvordan de enkelte artiklers videnska-

belige og faglige kvalitet er blevet vurderet. Når sådanne krav er opfyldt, vil forfatteren være bundet af den faktiske evidens, dvs. den forskning der er på området, uanset om det støtter forfatterens egen forståelse eller ikke. Vi kan derfor som læsere føle os mere trygge, når vi læser konklusionen af en sådan artikel, for vi ved at den burde bygge på al den viden, der er på området og ikke kun det, som forfatteren tilfældigvis kender. Selv en systematisk review artikel kan dog stadig være problematisk, fordi konklusionen bygger på en subjektiv vurdering af resultaterne fra de enkelte artikler. Der findes derfor en tredje type af review artikler, som kaldes for metaanalyser, fordi resultaterne her ikke blot vurderes men indgår i en samlet statistisk analyse af alle de fundne artiklers resultater.

Der findes altså tre former for oversigtsartikler:

- 1) Den traditionelle hvor en ekspert samler sin viden. Den er meget anvendelig, når det gælder om at få et overblik over et emnes begreber og rammer, men bør ikke danne grundlag for en evidensbaseret praksis.
- 2) Den systematiske oversigtsartikel. Her skal forfatteren gøre rede for formålet med review artiklen, søgeprocessen og kvalitetsvurderingen af de fundne originale artikler. Kan danne grundlag for en evidensbaseret praksis.
- 3) Metaanalysen. En metaanalyse er en systematisk oversigtsartikel, hvor også resultaterne tolkes ved hjælp af en samlet statistisk analyse af alle de enkelte artiklers resultater. Bedste grundlag for en evidensbaseret praksis.

Uanset om du har fundet en traditionel oversigtsartikel, en systematiske oversigtartikel eller en metaanalyse bør artiklen læses igennem med kritiske briller. Selv den bedste metaanalyse er kun udarbejdet af mennesker, og vi kan alle begå fejl. I de følgende vil jeg derfor gennemgå en checkliste som hjælp til at kunne forholde dig kritisk til en oversigtartikel. Checklisten er udarbejdet af bl.a. D. L. Sackett fra Oxford Universitetet og er oversat til dansk af læge Klaus Witt og Hans Lund. Den samlede checkliste kan du se i faktaboksen på side??.

Efter gennemgangen af checklisten, vil jeg kort beskrive hvor du især kan finde disse oversigtsartikler / metaanalyser (se nederst side 5).

Studiets troværdighed

Før oversigtsartiklens konklusion overvejes, og man beslutter, hvad man skal gøre med sine patienter, bør det vurderes, om der er et acceptabelt grundlag for overhovedet at se på konklusionen. Hvor troværdigt er studiet, er metoderne forfatteren har anvendt, acceptable osv.

Jeg har som eksempel på en oversigtsartikel har jeg fundet en fra Cochrane Library, der hedder "Bed rest for acute low back pain and sciatica". Den er udarbejdet af Kåre Birger Hagen, fysioterapeut fra Oslo (1).

Spørgsmål 1: *Er oversigtsartiklen rettet mod et velafgrænset klinisk problem?*

Mængden af videnskabelig litteratur er enorm. Alene i databasen MEDLINE er der mere end 10 millioner referencer, og MEDLINE dækker oven i købet kun 4000 af de mere end 20.000 videnskabelige medicinske tidsskrifter, der er i verden. Hvis du ikke har et meget præcist spørgsmål, vil du drukne i data. Du vil heller ikke kunne konkludere særlig præcist, hvis du har ikke har et præcist spørgsmål. Kåre B Hagen et al stiller spørgsmålet: Hvad er effekten af sengeleje for patienter med akutte lændesmerter (LBP) eller udstråling til benet. Deres spørgsmål bliver yderligere specificeret ved, at de fortæller, at følgende sammenligninger vil blive foretaget: sengeleje versus forblive aktiv, sengeleje versus andre behandlinger og kortere perioder (2-4 dage) sammenlignet med længere perioder (mere end 4 dage) med sengeleje. De har ikke alene spurgt om effekten, men også forklaret på

hvilken måde de vil vurdere effekten og hvad de vil sammenligne med.

Spørgsmål 2: *Blev artiklerne inkluderet efter eksplicitte og velegnede kriterier og er søgestrategien beskrevet i detaljer?*

Dette spørgsmål er i virkeligheden to spørgsmål. Det første drejer sig om, hvorvidt artiklerne blev inkluderet efter eksplicitte og velegnede kriterier. Hvis man ønsker at koble resultater fra flere forskellige studier sammen, bør man være sikker på, at alle disse studier undersøger det samme problem, eller ser på de samme patienter. Det sker somme tider, at et studie viser ingen effekt, alene fordi den gruppe af patienter, der var med i studiet, i virkeligheden bestod af to forskellige grupper. En gruppe som havde god effekt af behandlingen, og en gruppe som fik det værre. Hvis man lægger resultaterne fra den samlede gruppe sammen, vil resultatet blive nul dvs. ingen effekt. Der er næsten større risiko for, at det samme problem opstår, når man kobler flere studier sammen. Det er derfor vigtigt, at forfatterne til review artiklen har været meget præcise i deres beskrivelse af, hvilke studier, dvs. hvilke patientgrupper, hvilken behandling og hvilke effektmål, der anvendes i de fundne studier. I Kåre B Hagens metaanalyse fra Cochrane er alle disse kriterier beskrevet meget omhyggeligt. F.eks. skal patienterne, der er med i de enkelte studier, have akut LBP i et område afgrænset af det nederste ribben og glutealfolden. Akut defineres som varende under 4 uger. De inkluderer dog også studier, hvor kroniske LBP forværres, men forværringen må ikke have varet længere end 4 uger. På samme præcise måde blev både typer af behandling (interventioner) og effektmål (outcome measures) beskrevet.

Det andet spørgsmål drejer sig om selve søgningen af den videnskabelige litteratur. Hvilke databaser søges der i, hvilke søgeord anvendes, hvor langt tilbage er der søgt o.l. Kåre B Hagen beskriver, hvordan de har søgt i 6 forskellige databaser som f.eks. MEDLINE og EMBASE. Derudover har de set igennem de fundne artiklers referenceliste for at se, om de henviste til andre relevante studier. De tog også kontakt til nøgleforfattere for at høre, om de havde yderligere referencer. Desuden beskrives, hvilke ord

og formuleringer de anvender i de forskellige databaser og henviser til Cochrane Ryggruppens anbefalinger, når det gælder litteratur-søgning. De har vist gjort, hvad man kunne forvente.

Spørgsmål 3: *Blev artiklerne inkluderet uanset sprog?*

Hvis man skal lave sin egen søgning, bør man holde sig til de sprog, man kan læse selv, men skal man udarbejde en oversigtsartikel/metaanalyse, bør sprog ikke være nogen hindring. Hvis der er artikler på sprog, som forfatteren ikke kan læse, kan denne sende den til nogle kompetente personer, der kan læse det sprog eller få artiklen oversat. Engelsk er det internationale sprog, og alle de førende medicinske videnskabelige tidsskrifter er på engelsk. Det betyder imidlertid, at man risikerer det, der hedder publikationsbias. Det vil sige, at kun en bestemt slags artikler publiceres, mens en anden slags afvises af redaktionerne. Typisk optages artikler, som viser et positivt resultat, altså at den og den behandling er god, mens artikler, der fortæller, at noget ikke virker, ofte ikke vil blive optaget af tidsskriftet. Den slags artikler bliver så i stedet publiceret i forfatterens hjemland og vil derfor være på det lokale sprog. Man kan altså risikere, hvis man ikke tager alle sprog med, at man får alle de positive artikler om en behandling på engelsk, mens alle de, sikkert ligeså gode artikler, på det lokale sprog ikke kommer med i oversigtsartiklen.

Spørgsmål 4: *Er det sandsynligt, at alle vigtige og relevante studier er fundet?*

Det kan man selvfølgelig ikke vide med sikkerhed, men ved at vurdere om alle relevante databaser er med, om de væsentlige søgeord er anvendt, om de fundne artiklers referencelister anvendes, om der er taget personlig kontakt til nøglepersoner inden for området o.l., kan man få et indtryk af, om det er sandsynligt eller ikke. Kåre B Hagen har f.eks. ikke søgt i CINAHL, hvilket kan undre lidt. CINAHL er en database for sygeplejersker og andre paramedicinske grupper (som f.eks. fysioterapi) og der er erfaring for, at der ikke altid er så stort overlap mellem MEDLINE, EMBASE og CINAHL, som man måske skulle tro. Spørgsmålet er vigtigt, fordi man som læser bør træde et skridt tilbage og vurdere, om det lyder sandsynligt, at forfatterne

har fået alle relevante artikler med. Kåre B Hagen har kun fundet 11 artikler, hvoraf 2 af dem blev ekskluderet, og man ville måske umiddelbart reagere og sige: det lyder ikke af så mange. Man må dog huske, at forfatterne netop også havde opstillet en række inklusionskriterier (hvilke patienter, hvilke interventioner og hvilke effektmål), som skulle være opfyldt.

Spørgsmål 5: *Er der en tabel med de ekskluderede artikler med forklaring på eksklusionen?*

Det er netop karakteristisk for den traditionelle oversigtsartikel, at man som læser får det indtryk, at forfatteren har fundet alt, hvad der er relevant, mens han i virkeligheden har udelukket en række studier (måske ubevidst), som burde have været med. Ved at forlange at forfatteren skal forklare, hvorfor nogle studier er ekskluderede, sikrer man, at de som måske taler imod forfatterens mening, alligevel kommer med. Kåre B Hagen har f.eks. forklaret, at et studie blev ekskluderet, fordi det ikke kunne opfylde inklusionskriterierne, når det gjaldt interventionen. Et andet studie blev ekskluderet, fordi det pga. studiets design ikke var muligt at vurdere effekten af sengeleje.

Spørgsmål 6: *Blev validiteten af de inkluderede artikler vurderet?*

Det er ikke nok, at finde alle relevante artikler, man er også som forfatter til en review artikel, forpligtet til at vurdere den videnskabelige og faglige kvalitet af artiklen. Kåre B Hagen forklarer meget omhyggeligt, hvordan to af forfatterne uafhængigt af hinanden vurderede kvaliteten ud fra nogle i forvejen opstillede kriterier. Derefter forsøgte man at diskutere sig frem til enighed om, hvilken kvalitet artiklen havde. Hvis man ikke kunne blive enige, fungerede en tredje forfatter som en slags opmand, og man fandt så frem til en enig kvalitetsvurdering af artiklen. Kriterierne er i princippet identiske med de kriterier, der danner grundlag for checklisten til at vurdere en behandlingsartikel (Se DF 14/99).

Spørgsmål 7: *Er vurderingen af artiklen reproducerbar?*

Det er ikke nok, at forfatteren af en review artikel fortæller, at kvaliteten er vurderet. Som læser må vi få vurderingsmetoden beskrevet så omhygge-

ligt, at vi ville kunne foretage den samme kvalitetsvurdering. Pointen er netop at jo mere åbenhed jo mere troværdigt. Kåre B Hagen beskriver metoden meget omhyggeligt og fortæller, hvilke kriterier, hvordan man svarer på kriterierne (f.eks. opfyldt, uklar, ikke-opfyldt) osv. Cochrane gruppen stiller meget præcise krav op, som alle Cochrane metaanalyser bør følge.

Spørgsmål 8: *Lignede resultaterne hinanden fra studie til studie?*

Dette spørgsmål kan lyde lidt mærkelig, da man netop søger alle relevante artikler for at få forskellene med i opgørelsen. Ideen er imidlertid, at hvis nu behandlingen har en bestemt effekt, burde denne effekt vise sig i alle tilfælde. Måske ikke i lige høj grad, men tendensen burde være den samme. Hvis der er meget stor forskel på effekten, skyldes det højst sandsynligt, at de inkluderede studier ikke bør kobles sammen. Det kan være, at der er forskellig effekt, fordi der er forskellige patienter i de forskellige studier. Eller det kan være, der er forskel, fordi det er forskellige behandlinger, der gives. Hvis man ser en meget stor forskel på effekten fra studie til studie, bør man derfor straks lade de røde lamper lyse og gå tilbage og prøve at undersøge, om det nu virkelig er den samme slags studie, der sammenlignes. Kåre B Hagen beskriver f. eks. et sted i resultatafsnittet, at tre studier så på smerteresponset efter sengeleje versus almindelig aktivitet. To viste ingen effekt, mens et tredje viste, at sengeleje syntes at være smertelindrende. Men hvis man så på kvaliteten af de tre studier, var de to gode studier dem, som viste ingen effekt, mens det dårlige studie viste effekt. Kåre B Hagen tillod sig derfor at konkludere, at sengeleje ikke har nogen reel smertelindrende effekt i forhold til at forblive aktiv.

Spørgsmål 9: *Var materialet egnet til en meta-analyse? Er den gennemført?*

En metaanalyse er, som nævnt, en statistisk metode til at koble resultaterne fra flere studier sammen så man får et samlet resultat fra alle studierne. Forudsætningen for at kunne gøre dette, er først og fremmest, at de studier der kobles sammen ligner hinanden. Det skal være de samme slags patienter. Det skal være den samme intervention, og det skal være de sam-

me effektmål, man beregner den samlede effekt ud fra. Ofte gennemføres en test for homogenitet, inden selve metaanalysen. Kåre B Hagen fandt i nogle tilfælde så stor variation (heterogenitet), at der ikke kunne gennemføres en metaanalyse, men i de fleste tilfælde kunne de gennemføre en metaanalyse.

Hvis vi nu mener, at oversigtsartiklen var af en acceptabel kvalitet (troværdig), kan vi roligt gå videre og se på hvad resultatet egentlig var. Kåre B Hagens metaanalyse synes at opfylde de fleste krav, så vi går videre.

Hvad var resultatet

Spørgsmål 10: *Hvor stor var behandlingseffekten?*

Hvis man læser en systematisk oversigtsartikel, hvor der ikke er gennemført en metaanalyse, er det ikke muligt at svare på dette spørgsmål – i hvert fald ikke med tal. Hvis den systematiske oversigtsartikel indeholder en metaanalyse, kan man få et talmæssigt udtryk for effekten. Ofte anvendes et begreb kaldet odds ratio. Odds kender vi fra f.eks. hestevæddeløb og betyder sandsynligheden for, at noget sker i forhold til, at det ikke sker. Nogle af dem, som bliver lagt i seng med ondt i ryggen, får værre smerter, andre får det bedre og nogen forbliver uændret. Odds for smerter i denne gruppe ville altså være sandsynligheden for at få smerter divideret med sandsynligheden for ikke at få smerter, hvis du lægges i seng. Sådan kan vi også beskrive det for dem, som får lov til at blive oppe. Nogle får det værre, nogle har det uændret og nogle får det bedre. Der er også en odds for smerter i den gruppe. Odds ratio er altså forholdet mellem oddsen for at få smerter, når man bliver lagt i seng i forhold til oddsen for at få smerter, når man bliver oppe. Hvis de har den samme odds, uanset om de ligger i sengen eller er oppe, vil odds ratio blive en. Det er nemlig det samme tal, der står i både tæller og nævner. Hvis den altså bliver mindre end en eller større end en, er der forskel på de to gruppers odds. Normalt opstilles regnestykket sådan, at hvis odds ratio bliver mindre end en, betyder det, at behandlingen har effekt – de får det bedre. Hvis den bliver større end en betyder det, at behandlingen forværrer situationen. Hvis den ligger lige omkring en, er der ingen forskel. F.eks. viste en af metaanalyserne i Kåre B Hagens studie, at

odds ratio var 1,08 for helbredelse ved at ligge få dage i sengen sammenlignet med mange dage. Med andre ord: det gør ingen forskel.

Spørgsmål 11: *Hvor præcis var estimeret af behandlingseffekten? Sikkerhedsgrænser?*

Det er imidlertid ikke nok, blot at fortælle at odds ratio var 1,08. Dette tal er ikke et fast tal, men et udtryk for en sandsynlighed, og der vil altid være en vis usikkerhed ved sandsynligheden. Derfor bør forfatteren fortælle læseren, hvilken sikkerhedsgrænse (konfidensintervallet) der er på resultatet. Hvis nu sikkerhedsgrænsen er fra en odds ratio på -2 til en odds ratio på +3, så virker 1,08 ikke længere så overbevisende. Der kunne nemlig ligeså godt have været god effekt (-2) som det det kunne være blevet meget værre (+3). I Kåre B Hagens studie går sikkerhedsgrænsen for odds ratioen på 1,08 fra 0,60 til 1,95. Der er ikke ligeså stor usikkerhed, som hvis det havde været -2 til +3, men der er alligevel en lille usikkerhed. Den kan man dog næppe fuldstændig slippe for.

Nu har vi vurderet om review artiklen er troværdig og vi har set på, hvilke resultater den viste. Det sidste og vigtige spørgsmål er så, om resultaterne kan brus i behandlingen.

Vil resultaterne hjælpe mig i behandlingen af mine patienter?

Spørgsmål 12: *Kan jeg overføre resultaterne til min aktuelle patient?*

Den eneste måde, man kan vurdere dette på, er at se om de patienter, som indgik i oversigtsartiklen, ligner din patient, man selv behandler. Passer aldersgruppen, passer sygdomsvarighed, symptomer osv. Det er i virkeligheden inklusionskriterierne for artiklerne, der skal tjekkes efter.

Spørgsmål 13: *Tog studiet højde for alle relevante kliniske effektmål?*

Det er imidlertid ikke nok, at patienterne i oversigtsartiklen ligner ens egne patienter. Det er lige så vigtigt, at man overvejer, om forfatteren til artiklen har målt den effekt, som man mener, er relevant sin patient. Kåre B Hagen medtager smerte, funktionsstatus, antal helbredte patienter

og antal af patienter, der vendte tilbage til arbejde. Umiddelbart virker det som relevante og anvendelige mål for en effekt af en behandling til en akut rygpatient, men det kan være, at de patienter, man selv har, er interesserede i andre mål for effekt end de fire nævnt her.

Spørgsmål 14: *Opvejer den gavnlige effekt de potentielle bivirkninger og omkostningerne?*

Spørgsmålet taler for sig selv, men det er uhyre vigtigt, at gøre sig disse overvejelser, inden man kaster sig ud i at anvende eller forkaste en eller behandling.

Hvor finder man oversigtsartiklerne?

Review artikler og metaanalyser bør først og fremmest søges i Cochrane Library, da denne database sandsynligvis indeholder de bedste reviews/metaanalyser. Man kan få adgang til Cochrane Library via et abonnement hos Update Software. Her vil man enten kunne få en opdateret CD-ROM tilsendt hvert kvartal, eller man kan abonnere på direkte online adgang via internettet. Update Software's adresse er: Update Software Ltd., Summertown Pavillion, Middle Way, Oxford OX2 7LG, UK, e-mail: infor@update.co.uk, telefon: +44 18 65 51 39 02, fax: +44 18 65 51 69 18.

Der er dog gratis adgang til abstrakts over de forskellige reviews /metaanalyser via PubMed (<http://www.ncbi.nlm.nih.gov/PubMed/>). Alle øvrige reviews/metaanalyser publiceres på fuldstændig samme vilkår som enhver anden videnskabelig artikel og kan derfor søges i de samme databaser, som man søger al anden videnskabelig litteratur.

Evidensbaseret medicin (EBM) eller evidensbaseret fysioterapi tager selvsagt udgangspunkt i de forskningsresultater der offentliggøres, og EBM er derfor i høj grad afhængig af forskning, indsigt i forskning og så videre. Det er dog ikke nok udelukkende at fokusere på forskningen. For at EBM skal have den effekt, som vi ønsker, nemlig at patienterne får det bedre, må EBM ses i en større sammenhæng. Den amerikanske medieforsker Neil Postman definerede i forbindelse med en diskussion om informationssamfundet, information på tre niveauer. Den rene information (som vi næsten drukner i) kalder han

for information og kan opfattes som rå og usorterede oplysninger, f.eks. de millioner af videnskabelige artikler, der henvises til i de forskellige databaser. Viden derimod definerer Postman som anvendelig information. Vi kunne i relation til EBM kalde de fagligt og videnskabelige gode artikler, vi har fundet ved en god søgning for viden. Men ifølge Postman bliver al denne viden først til visdom, når vi forstår at vælge den viden ud, som er den bedste. I relation til EBM vil det sige, at den viden, vi finder, bør relateres til patienten, først da bliver det, vi finder, til visdom.

Referencer

1. Hagen, K. B., G. Hilde, G. Jamtvedt, and M. Winnem. *Bed rest for acute low back pain and sciatica (Cochrane Review)*. Cochrane Library 1999.
2. Oxman, A. D. and G. H. Guyatt. *The science of reviewing research*. *Ann N.Y.Acad.Sci* 703: 125-133, 1993.

Checkliste for review-artikler
<p><i>Spørgsmål 1:</i> Er oversigtsartiklen rettet mod et velafgrænset klinisk problem?</p>
<p><i>Spørgsmål 2:</i> Blev artiklerne inkluderet efter eksplicitte og velegnede kriterier og er søgestrategien beskrevet i detaljer?</p>
<p><i>Spørgsmål 3:</i> Blev artiklerne inkluderede uanset sprog?</p>
<p><i>Spørgsmål 4:</i> Er det sandsynligt, at alle vigtige og relevante studier er fundet?</p>
<p><i>Spørgsmål 5:</i> Er der en tabel med de ekskluderede artikler med forklaring på eksklusionen?</p>
<p><i>Spørgsmål 6:</i> Blev validiteten af de inkluderede artikler vurderet?</p>
<p><i>Spørgsmål 7:</i> Er vurderingen af artiklen reproducerbar?</p>
<p><i>Spørgsmål 8:</i> Lignede resultaterne hinanden fra studie til studie?</p>
<p><i>Spørgsmål 9:</i> Var materialet egnet til en metaanalyse? Er den gennemført?</p>
<p><i>Spørgsmål 10:</i> Hvor stor var behandlingseffekten?</p>
<p><i>Spørgsmål 11:</i> Hvor præcis var estimatet af behandlingseffekten? Sikkerhedsgrænser?</p>
<p><i>Spørgsmål 12:</i> Kan jeg overføre resultaterne til min aktuelle patient?</p>
<p><i>Spørgsmål 13:</i> Tog studiet højde for alle relevante kliniske effektmål?</p>
<p><i>Spørgsmål 14:</i> Opvejer den gavnlige effekt de potentielle bivirkninger og omkostningerne?</p>