Capita selecta

De praktijk van systematische reviews. III. Methodologische beoordeling van onderzoeken

w.j.j.assendelft, r.j.p.m.scholten, j.th.m.van eijk en l.m.bouter

Iedere onderzoeker en clinicus zal aanvoelen dat de methodologische kwaliteit van een onderzoek van invloed kan zijn op de resultaten en de conclusies ervan. In een systematische review worden op een reproduceerbare, objectieve en expliciete manier de resultaten van alle onderzoeken betreffende een duidelijk geformuleerde vraagstelling samengevat.[1] De beoordeling van de methodologische kwaliteit is een vast onderdeel van een systematische review.[2-4] Er is echter weinig empirisch onderzoek verricht naar de optimale opzet, uitvoering en toepassing van methodologische beoordeling en de weinige beschikbare publicaties betreffen vooral de beoordeling van gerandomiseerde klinische onderzoeken (RCT's).

Evenals voor zoeken en selecteren geldt, is het bij de aanpak en de uitvoering van methodologische beoordeling van RCT's voor de inzichtelijkheid van het reviewproces van belang om hierin duidelijk beschreven, beredeneerde keuzen te maken.[5] In dit artikel beschrijven wij hoe hierbij te werk te gaan. Een eerste stap is helder te formuleren wat men onder methodologische kwaliteit verstaat. Op grond hiervan dient men een beoordelingslijst te construeren of een keuze te maken uit de bestaande lijsten. Vervolgens moet men besluiten wie er beoordeelt, of dit geblindeerd gebeurt en of er nog aanvullende gegevens bij de oorspronkelijke auteurs worden opgevraagd. Tenslotte dient men vast te stellen hoe per onderzoek tot een conclusie over de kwaliteit wordt gekomen (bijvoorbeeld of er een gewogen somscore voor de methodologische kwaliteit wordt berekend), en hoe kwaliteit wordt ingepast in een eventuele statistische 'pooling' (dat is het combineren van de resultaten van de afzonderlijke onderzoeken tot één totaal resultaat).

Over de methodologische beoordeling van etiologische, diagnostische en prognostische onderzoeken bestaan nog weinig conventies. Een aantal principes en problemen zijn hetzelfde als bij de beoordeling van therapeutische RCT's. Voor een verdere toelichting op de specifieke aspecten van beoordeling van niet-gerandomiseerde onderzoeken verwijzen wij naar andere literatuur.[6-11]

Vrije Universiteit, Faculteit der Geneeskunde, Instituut voor Extramuraal Geneeskundig Onderzoek, Amsterdam.
Dr.W.J.J.Assendelft, huisarts-epidemioloog (thans: Dutch Cochrane Centre, Academisch Medisch Centrum, Postbus 22.700, 1100 DE Amsterdam); dr.R.J.P.M.Scholten, arts-epidemioloog; prof.dr.J.Th.M. van Eijk en prof.dr.L.M.Bouter, epidemiologen.
Correspondentieadres: dr.W.J.J.Assendelft (e-mail: w.j.assendelft@amc.uva.nl).

samenvatting

– De methodologische kwaliteit van de in een systematische review opgenomen primaire onderzoeken kan van invloed zijn op de resultaten en de conclusies van die review.
– Methodologische kwaliteit kan op vele manieren worden gedefinieerd. Mede hierdoor bestaat er een veelheid aan beoordelingslijsten.
– De belangrijkste dimensie van kwaliteit is interne validiteit, gedefinieerd als het vertrouwen dat opzet, uitvoering en presentatie van de trial systematische vertekening (bias) in de uitkomsten minimaliseren of uitsluiten. Slechts van een klein aantal items met betrekking tot interne validiteit is in empirisch onderzoek een vertekenende invloed aangetoond: blindering van de randomisatie en van patiënten en effectbeoordeling.
– Kwaliteitsbeoordeling geschiedt bij voorkeur door 2 of meer reviewers onafhankelijk van elkaar. Er is geen consensus of het onderzoek vóór de beoordeling moet worden geblindeerd voor auteurs, tijdschrift, instituut, resultaten en conclusies.
– Interne validiteit kan op verschillende manieren in de statistische pooling worden betrokken: als toelatingscriterium, als weegfactor en als manier om de onderzoeken grafisch te ordenen.
– Om in de toekomst duidelijkere richtlijnen voor methodologische beoordeling van onderzoeken te kunnen geven is er behoefte aan goed opgezette vergelijkende onderzoeken.

methodologische kwaliteit

Dimensies van kwaliteit

Methodologische kwaliteit kan op vele manieren worden gedefinieerd.[12] Afhankelijk van de definitie omvat deze de interne validiteit (die wij hier aanduiden als 'validiteit'), de externe validiteit (die wij 'generaliseerbaarheid' noemen), de volledigheid van datapresentatie en algemene aspecten van correct mensgebonden onderzoek ('good clinical practice') (tabel 1). De eerste drie dimensies hebben elk hun eigen plaats in de systematische review, de vierde dimensie ('good clinical practice') heeft geen directe relatie met de uitkomsten van een onderzoek en speelt in een systematische review in de regel geen rol. Een voorbeeld van een beoordelingslijst voor methodologische kwaliteit wordt gegeven in tabel 2.[13-17]

De onderzoekkenmerken die betrekking hebben op de generaliseerbaarheid zijn soms onderdeel van de in- en exclusiecriteria voor de selectie van onderzoeken voor de review,[3] maar kunnen ook gebruikt worden in de subgroep- en sensitiviteitsanalysen van de meta-analyse (zie tabel 1). Een voldoende gedetailleerde datapresentatie in een onderzoek is voor een systematische

TABEL 1. Dimensies van kwaliteit in de methodologische beoordeling van een vergelijkend effectonderzoek

dimensie	*verwante items*	*omschrijving*	*voorbeelden*
interne validiteit	validiteitsitems	kenmerken in de opzet of de uitvoering van een onderzoek waarbij er kans is op systematische vertekening van de uitkomst	randomisatie; blinde uitkomstmeting
generaliseerbaarheid	descriptieve items	de mogelijkheid om te beoordelen bij welke patiënten welke interventie is toegepast en op welke wijze hiervan het effect is gemeten; de mogelijkheid om in geval van heterogeniteit of bij een gerichte vraagstelling relevante subgroepen te kunnen onderscheiden of sensitiviteitsanalysen uit te kunnen voeren	beschrijving van patiëntenpopulatie; voldoende gedetailleerde beschrijving van interventie
kwantitatieve aspecten	kwantitatieve items	voldoende beschrijving van de kwantitatieve gegevens om, onafhankelijk van de oorspronkelijke auteurs van het onderzoek, conclusies te kunnen trekken (onder andere over het statistisch onderscheidingsvermogen) en het onderzoek met andere te kunnen combineren om zodoende een totaal effect te kunnen bepalen	goede beschrijving van aantal patiënten per meting; beschrijving van gemiddelde en standaarddeviatie
overige aspecten	overige items	refereren aan 'good clinical practice'	beschikking over 'informed consent'; toestemming van medisch-ethische commissie

review van belang om de resultaten van overeenkomstige onderzoeken te kunnen combineren tot één totaal resultaat (poolen) of om, in geval er niet gepoold kan worden, in ieder geval het statistisch onderscheidingsvermogen ('power') van de individuele onderzoeken te kunnen rapporteren.[18] Hoewel generaliseerbaarheid en adequate datapresentatie voor een review belangrijke aspecten zijn, ligt het accent in dit artikel echter op de validiteit, gedefinieerd als het vertrouwen dat de opzet, uitvoering en presentatie van de trial systematische vertekening (bias) in de uitkomsten minimaliseren of uitsluiten.[12]

De relatie tussen validiteitsitems en de uitkomst van onderzoeken

Er is nog weinig onderzoek gedaan naar de waarde van individuele validiteitsitems. Van enkele items is in empirisch onderzoek een vertekenende invloed (bias) aangetoond: blindering van de randomisatie ('randomisation concealment') en blindering van patiënten en effectbeoordeling (dubbele blindering).[19-23] Een beoordelingslijst zou derhalve tenminste deze items moeten bevatten.[24]

Bestaande beoordelingslijsten

Voor de keuze van een beoordelingslijst, maar ook voor het begrip van reeds gepubliceerde systematische reviews, is het van belang om te weten dat er een groot aantal beoordelingslijsten beschikbaar is (waarvan de lijst in tabel 2 dus slechts een voorbeeld is).

Moher et al. bespreken in een overzicht uit 1995 34 dergelijke lijsten.[25] Het aantal items per lijst varieert van 3 tot 34. De gemiddelde invulduur loopt van minder dan 10 min tot 45 min per onderzoek. De bestaande beoordelingslijsten bevatten in wisselende mate items die naast de interne validiteit ook betrekking hebben op de andere dimensies van kwaliteit (generaliseerbaarheid, kwantitatieve aspecten en good clinical practice). Er zijn lijsten die specifiek bedoeld zijn voor een bepaald domein, zoals farmacologie of het bewegingsapparaat, terwijl andere juist meer een generiek karakter hebben.[25] De keuze van de lijst kan van grote invloed zijn op de uiteindelijke absolute somscore die een onderzoek krijgt. Ook de relatieve rangschikking in methodologische kwaliteit van onderzoeken ten opzichte van elkaar kan per beoordelingslijst aanzienlijk verschillen.

Moher et al. lieten een team van ervaren reviewers 12 trials naar de effectiviteit van trombolyse bij acuut myocardinfarct met 6 verschillende lijsten beoordelen.[12] Voor hetzelfde onderzoek kon de score variëren van 29-74%. Dit heeft zwaarwegende consequenties wanneer somscore als inclusiecriterium voor pooling wordt gebruikt (zie verder). Ook de onderlinge rangschikking van de onderzoeken bleek, afhankelijk van de gebruikte lijst, aanzienlijk te verschillen.[12] Ook dit kan consequenties hebben voor het bepalen van een uiteindelijke conclusie voor de review.

Vrijwel alle beschikbare lijsten zijn door de reviewers zelf op basis van theoretische uitgangspunten uit leerboeken samengesteld. Onlangs zijn er echter 2 lijsten ontwikkeld die volgens formelere procedures totstandkwamen. Jadad et al. kwamen via een empirische selectie uit 49 items uiteindelijk tot een lijst van 3 items.[13] Verhagen et al. kwamen na een gestructureerde consultatie van internationale experts (Delphi-procedure) over een oorspronkelijke verzameling van 206 onderwerpen tot een selectie van 8 items.[17] Beide lijsten bevatten de genoemde items blindering van de randomisatie en blindering van patiënten en behandelaars (zie tabel 2).[22]

TABEL 2. Te controleren onderwerpen voor kwaliteitsbeoordeling van een vergelijkend effectonderzoek volgens de Amsterdam-Maastricht-consensuslijst, onderverdeeld in validiteitsitems, descriptieve en kwantitatieve items, alsmede mate van overlap met andere lijsten

Amsterdam-Maastricht-consensuslijst[16]	*Jadad-lijst*[13]	*Delphi-lijst*[17]
validiteitsitems		
genereren van een aselecte volgorde voor het toewijzen van interventies	ja	ja
geblindeerde toewijzing van interventies		ja
vergelijkbare interventiegroepen ten aanzien van de belangrijkste uitkomstmaten en prognostische factoren bij aanvang van het onderzoek		ja
blindering van de behandelaar voor de toegewezen interventie; controle op succes van de blindering		ja
controle op co-interventies		
controle op therapietrouw		
relevante, betrouwbare en valide uitkomstmaten		
blindering van de patiënt voor de toegewezen interventie; controle op succes van de blindering	ja	ja
weinig uitvallers en ontbrekende waarden	ja	
blindering van de uitkomstmeting; controle op succes van de blindering	ja	ja
gelijk tijdstip van de uitkomstmeting in de interventiegroepen		
analyse en rapportage volgens het 'intention-to-treat'-principe		ja
descriptieve items		
beschrijving van de selectiecriteria		ja
beschrijving van de interventies		
rapportage van relevante uitkomstmaten		
duur van de follow-up		
beschrijving van bijwerkingen		
kwantitatieve items		
beschrijving van de onderzoekomvang direct na randomisatie en ten tijde van de belangrijkste uitkomstmeting(en)		
beschrijving van puntschattingen en spreidingsmaten of frequentieverdelingen voor de belangrijkste uitkomstmaten		ja

Beoordelingsprocedure

Gangbaar is om de beoordeling door 2 of meer reviewers onafhankelijk van elkaar te laten geschieden. Vervolgens worden de scores op de individuele items vergeleken. Overwogen kan worden om vóór de eigenlijke review een proefbeoordeling te doen van enige artikelen over een andersoortige interventie bij dezelfde aandoening en/of dezelfde interventie bij een andere aandoening. De items worden zo nodig aan de hand van afspraken tussen de reviewers in een appendix bij de review verder geoperationaliseerd. Dit dient om, ook voor de lezers, de beoordelingscriteria te expliciteren en de reproduceerbaarheid van de beoordeling verder te verhogen. De mate van discrepantie, uitgedrukt in percentage overeenstemming en Cohen's kappa, wordt per item van de beoordelingslijst berekend.[26,27] Oordelen die van elkaar afwijken, worden vervolgens besproken. Wanneer de discrepantie blijft bestaan, kan een derde reviewer geconsulteerd worden.

Sommige auteurs bevelen aan om de artikelen vóór de beoordeling te blinderen voor auteurs, tijdschrift, instituut, resultaten en conclusies.[20] Omdat van prestigieuze tijdschriften ook de opmaak en het lettertype bekend zijn, worden soms de oorspronkelijke manuscripten gescand om ze in één opmaak en lettertype te kunnen beoordelen.[28,29] Scannen, controleren en opnieuw vormgeven zijn tijdrovend: in het onderzoek van Berlin et al. kostte dit gemiddeld bijna 11 uur per artikel.[28] Blinderen kan ook met correctievloeistof, een merkstift of met een schaar en lijm. Dit is minder tijdrovend. Nadeel is echter dat het lettertype en de opmaak van het tijdschrift herkenbaar blijven. Het nut van blinderen staat nog volop ter discussie: in 3 onderzoeken bleken de validiteitsscores bij een geblindeerde beoordeling significant lager dan bij niet-geblindeerde beoordeling.[13,23,28] Verhagen et al. vonden geen verschil tussen geblindeerde en niet-geblindeerde beoordeling.[29] Wel of niet geblindeerde beoordeling bleek geen invloed te hebben op de uiteindelijke resultaten van de 5 meta-analysen in het onderzoek van Berlin et al.[28]

Kwantificeren van kwaliteit: weging en somscores

Vaak wordt bij een beoordeling naar een summatie van de verschillende criteria gestreefd om zo tot een somscore van het artikel te komen. Daarbij kan per criterium een gewicht worden toegekend, zodat sommige criteria een groter aandeel in de kwaliteitsscore krijgen dan andere.[4,25] Er is geen enkele empirische grond waarop het gewicht van de items op zinvolle wijze kan worden bepaald. Somscores hebben het voordeel van de eenvoud, maar het methodologische nadeel is de discutabele aanname dat tekortkomingen op bepaalde items blijkbaar gecompenseerd kunnen worden door een positieve score op andere. Summatie van scores van items die verschillende dimensies (zie tabel 1) representeren, is conceptueel discutabel. Beter is het om in dit geval de somscore per dimensie of alleen een validiteitssomscore te rapporteren.

De beslissing om aan een afzonderlijk criterium wel of geen score toe te kennen lijkt eenvoudig, maar behoeft toelichting. Om reden van overzichtelijkheid wordt er doorgaans van uitgegaan dat een onderzoek wel aan een criterium kan hebben voldaan ('+') of hierin duidelijke tekortkomingen heeft ('–'). In de praktijk blijkt echter dat informatie over een bepaald item onvolledig of onduidelijk kan zijn ('?') of geheel ontbreekt ('o'). Onvolledige of ontbrekende informatie kan in principe worden opgevraagd bij de oorspronkelijke auteurs.[24] Dit is echter een arbeidsintensief en tijdrovend karwei, waarvan de resultaten vaak teleurstellend zijn. Veelal wordt een '?' of een 'o' als een '–' geïnterpreteerd. Daarnaast kunnen nog sensitiviteitsanalysen worden gedaan, waarbij een '?' en 'o' verschillende wegingen krijgen.[4]

Validiteit als onderdeel van meta-analyse

Meta-analyse is het kwantitatieve onderdeel van de systematische review. Ze omvat onder andere het samenvoegen van de resultaten van de afzonderlijke onderzoeken tot één totaal resultaat. Validiteit kan op verschillende manieren in deze pooling worden betrokken:

Selectiecriterium. Hierbij worden alleen onderzoeken

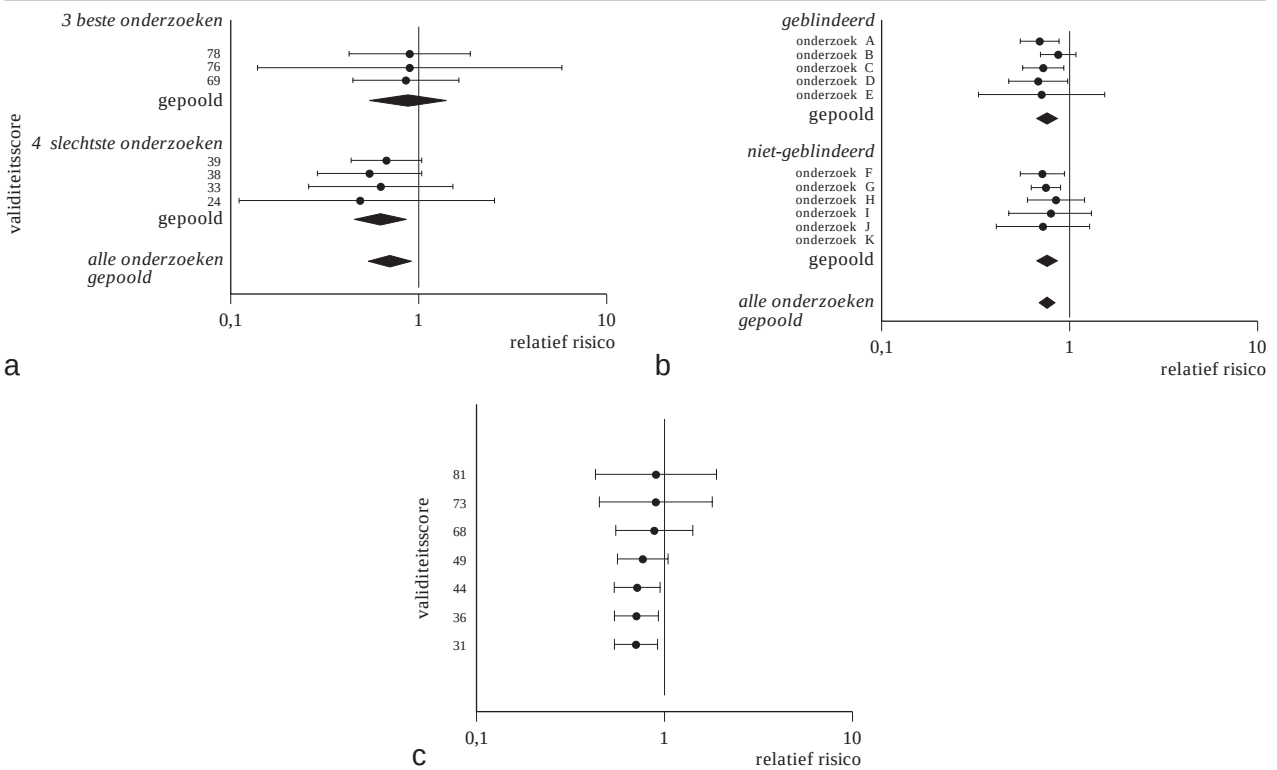
met een validiteitssomscore boven een bepaald afkappunt in de pooling toegelaten. Ook kan een positieve beoordeling op een afzonderlijk validiteitscriterium als voorwaarde voor pooling worden gesteld (zie eerder). Het nadeel van een dergelijke selectie vooraf is dat deze de inzichtelijkheid van het reviewproces aantast. Een oplossing is om de betere en de slechtere onderzoeken beide te includeren, maar ze apart te poolen (figuur, a).

Een variant is om niet uit te gaan van somscores, maar van een of meer afzonderlijke validiteitsitems. Aparte pooling van onderzoeken die wel en die niet aan het betreffende validiteitsitem(s) voldoen, is dan de oplossing (zie de figuur, b). Als blijkt dat de somscore of het betreffende item helemaal geen invloed op de grootte van het effect heeft (dus geen bias introduceert), dan kan overwogen worden om het totale effect, ongeacht de methodologische kwaliteit, voor alle onderzoeken tezamen te berekenen.

In gepubliceerde reviews waarin kwaliteit op deze manier in de pooling wordt ingebracht, blijkt de relatie tussen de kwaliteit van de afzonderlijke onderzoeken en de door deze onderzoeken gerapporteerde effectgrootte niet eenduidig te zijn. De grootte en de richting van de door lage kwaliteit geïntroduceerde bias verschillen in de verschillende reviews: soms wordt er geen relatie gevonden, in andere gevallen blijkt een hogere kwaliteit samen te hangen met een kleinere effectschatting, terwijl ook het tegenovergestelde wordt gevonden.[30-35] Deze verschillen hangen waarschijnlijk zowel samen met het onderwerp van onderzoek als met de itemkeuze (vaak bevatten de gebruikte instrumenten overigens meer dan alleen validiteitsitems).

Als er onvoldoende onderzoeken van voldoende kwaliteit beschikbaar zijn, kan ook worden besloten om in het geheel niet te poolen.[4 36] In dat geval zal bij de beschrijvende samenvatting van de onderzoeken doorgaans meer gewicht worden gegeven aan onderzoeken met een hogere interne validiteit.

Grafische presentatie van de relatie tussen effectgrootte en somscore. Het voordeel hiervan is dat de lezer di-



Weergave van de resultaten van een fictieve meta-analyse op een logaritmische schaal, met inbreng van de validiteitsscore in de statistische pooling (combinatie) van onderzoeken op 3 manieren. De horizontale lijnen en de wybertjes vertegenwoordigen de puntschattingen en de 95%-betrouwbaarheidsintervallen van het relatieve risico van de afzonderlijke onderzoeken en van het gepoolde resultaat. De verticale lijn geeft de neutrale waarde (in dit geval: relatief risico = 1) aan, waarbij de onderzochte behandelingen hetzelfde effect hebben. Puntschattingen links van de neutrale lijn duiden op een gunstiger effect van de onderzochte behandeling ten opzichte van de controlebehandeling.

(a) Rangschikking van de onderzoeken op volgorde van afnemende methodologische kwaliteit met aparte pooling van de methodologisch betere en slechtere onderzoeken. Het totale resultaat van alle onderzoeken lijkt voornamelijk bepaald door de 4 slechtste onderzoeken, die tezamen een sterk effect laten zien. Het gecombineerde resultaat van de 3 beste onderzoeken is minder overtuigend: de puntschatting ligt veel dichter in de buurt van de neutrale waarde. (b) Aparte pooling van onderzoeken die voldoen aan het validiteitscriterium 'geblindeerde effectmeting'. (c) Cumulatieve pooling: het bovenste lijnstukje is gebaseerd op het methodologisch beste onderzoek, het tweede lijnstukje op de beste 2 onderzoeken, het derde op de beste 3, enzovoorts.

rect zelf de invloed van de somscore kan beoordelen. In de figuur (a) zijn de onderzoeken naar somscore geordend.

Cumulatieve pooling van de effectgrootte met afnemende somscore. Deze aanpak stelt de lezer in staat om zelf te bepalen welk afkappunt wordt gekozen (zie de figuur, c). Ook kan worden bepaald of het totale effect bij afnemende somscore begint af te wijken van dat van de methodologisch betere onderzoeken. In figuur (c) is het bovenste lijnstukje gebaseerd op het methodologisch beste onderzoek, het tweede op de beste 2 onderzoeken, enzovoorts.

Somscore als weegfactor voor de individuele onderzoeken. Hierbij wordt kwaliteit (mede) als weegfactor voor de pooling gebruikt. Dit wordt zelden in meta-analysen toegepast.[37]

beschouwing

In de inleiding werd reeds opgemerkt dat iedere onderzoeker en clinicus zal aanvoelen dat de methodologische kwaliteit van een onderzoek onderdeel van een systematische review behoort te zijn. De praktijk leert echter dat er veel variatie is in de aanpak en ook in de uitkomsten van methodologische beoordeling. Kwaliteit omvat zowel de interne validiteit als ook de generaliseerbaarheid en de volledigheid van datapresentatie van een onderzoek. Deze dimensies worden elk op een andere manier in een systematische review ingepast.

Empirisch onderzoek ter onderbouwing van keuzen in de procedure en de uitvoering van methodologische beoordeling is helaas nog schaars. Voor een belangrijk deel zullen reviewers moeten afgaan op 'face validity' van de te gebruiken lijst of individuele items. De keuze van de lijst of items wordt voor een deel bepaald door het indicatiegebied; sommige lijsten zijn aandoening- of domeinspecifiek.[25] Enkele items (blindering van de randomisatie en van patiënten en behandelaars) zijn door onderbouwend onderzoek min of meer standaard geworden.[24] Momenteel ontbreken nog duidelijke richtlijnen voor de weging van verschillende items ten opzichte van elkaar en voor de manier waarop somscores of afzonderlijke items het beste in de pooling kunnen worden ingebracht.

De keuzen die in de opzet en de uitvoering van de methodologische beoordeling van onderzoeken worden gemaakt, blijven vaak impliciet, hetgeen de transparantie van het reviewproces niet ten goede komt. Beter is het om de beoordelingsprocedure van tevoren goed te plannen en keuzen expliciet te beredeneren. Hierbij spelen de beschikbare tijd en middelen zeker een rol.

Om in de toekomst duidelijkere richtlijnen voor methodologische beoordeling van onderzoeken te kunnen geven is er behoefte aan goed opgezette vergelijkende onderzoeken. In dergelijke onderzoeken dienen de verschillende facetten van kwaliteitsbeoordeling (keuze van de beoordelingslijst of -items, blindering van de beoordeling, weging en bepalen van somscores) en de manier van inbreng van validiteitsbeoordeling verder te worden onderzocht.

abstract

The practice of systematic reviews. III. Assessment of methodological quality of studies

– The methodological quality of the primary studies included in a systematic review may influence its results and final conclusions.
– Methodological quality may be defined in various ways. Partially because of this there are many different assessment lists.
– The most important dimension of quality is internal validity, defined as the confidence that the design, performance and report of a trial prevent or reduce systematic errors (bias) in the outcomes. For only a limited number of internal validity items a relationship with bias has been proven in empirical studies: concealment of randomisation and blinding of patients and outcome assessors.
– Preferably, quality should be assessed by at least 2 assessors independently. There is no consensus whether assessment should be done blinded for authors, journal, results and conclusions.
– Internal validity can be incorporated into statistical pooling in various ways: as a selection criterion, to be used as weight or to hierarchically order studies in a presentation.
– Well-designed comparative studies are needed to provide clearer guidelines for methodological assessment in the future.

literatuur

1 Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126:376-80.
2 Assendelft WJJ, Koes BW, Knipschild PG, Bouter LM. The relationship between methodological quality and conclusions in reviews of spinal manipulation. JAMA 1995;274:1942-8.
3 Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. Ann Intern Med 1997;127:531-7.
4 Vet HCW de, Bie RA de, Heijden GJMG van der, Verhagen AP, Sijpkes P, Knipschild PG. Systematic reviews on the basis of methodological criteria. Physiotherapy 1997;83:284-9.
5 Assendelft WJJ, Tulder MW van, Scholten RJPM, Bouter LM. De praktijk van systematische reviews. II. Zoeken en selecteren van literatuur. Ned Tijdschr Geneeskd 1999;143:656-61.
6 Irwig L, Tosteson ANA, Gatsollis C, Lau J, Colditz G, Chalmers TC, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994;120:667-76.
7 Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. JAMA 1994;272:234-7.
8 Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. JAMA 1994; 271:1615-9.
9 Blair A, Burg J, Foran J, Gibb H, Greenland S, Morris R, et al. Guidelines for application of meta-analysis in environmental epidemiology. ISLI Risk Science Institute. Regul Toxicol Pharmacol 1995;22:189-97.
10 Hollander AEM den, Preller EA, Heisterkamp SH, Jomsen J. Meta-analyse van observationeel onderzoek. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu; 1996. p. 1-63.
11 Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. BMJ 1997;315:540-3.
12 Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. Int J Technol Assess Health Care 1996;12:195-208.
13 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1-12.
14 Tulder MW van, Assendelft WJJ, Koes BW, Bouter LM. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. Spine 1997;22:2323-30.

15 Heijden GJMG van der, Windt DAWM van der, Winter AF de. Physiotherapy for patients with soft tissue shoulder disorders: a systematic review of randomised clinical trials. BMJ 1997;315:25-30.
16 Heijden GJMG van der, Windt DAWM van der, Berg SGM van den, Bouter LM. Effectiviteit van ultrageluid behandeling bij aandoeningen van het bewegingsapparaat. Hoensbroek: Instituut voor Revalidatievraagstukken; 1997. p. 19-43.
17 Verhagen AP, Vet HCW de, Bie RA de, Kessels AGH, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998;51:1235-41.
18 Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994;272:122-4.
19 Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM, Ascorbic acid for the common cold. A prophylactic and therapeutic trial. JAMA 1975;231:1038-42.
20 Chalmers TC, Celano P, Sacks HS, Smith jr H. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983;309:1358-61.
21 Colditz GA, Hiller SN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. Stat Med 1989;8:441-54.
22 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12.
23 Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609-13.
24 Mulrow CD, Oxman AD. Cochrane Collaboration Handbook [updated September 1997]. The Cochrane Collaboration. The Cochrane Library [database on CDROM]. Oxford: Update Software [updated quarterly].
25 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995;16:62-73.
26 Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37-46.
27 Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ 1992;304:1491-4.
28 Berlin JA, Miles GC, Cirigliano MD, Conill AM, Horowitz DA, Jones F, et al. Does blinding of readers affect the results of meta-analyses? Results of a randomized trial. Online J Curr Clin Trials 1997 [serial online] 29 May 1997 (Doc No 205).
29 Verhagen AP, Vet HCW de, Bie RA de, Kessels AGH, Boers M, Knipschild PG. Balneotherapy and quality assessment: interobserver reliability of the Maastricht criteria list and the need for blinded quality assessment. J Clin Epidemiol 1998;51:335-41.
30 Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. Control Clin Trials 1990;11:339-52.
31 Khan KS, Daya S, Jadad AR. The importance of quality of primary studies in producing unbiased systematic reviews. Arch Intern Med 1996;156:661-6.
32 Nurmohamed MT, Rosendaal FR, Buller HR, Dekker E, Hommes DW, Vandenbroucke JP, et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. Lancet 1992;340:152-6.
33 Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. Control Clin Trials 1989;10:31-56.
34 Brown SA. Meta-analysis of diabetes patient education research: variations in intervention effects across studies. Res Nurs Health 1992;15:409-19.
35 Imperiale TF, McCullough AJ. Do corticosteroids reduce mortality from alcoholic hepatitis? A meta-analysis of the randomized trials. Ann Intern Med 1990;113:299-307.
36 Labelle H, Guibert R, Joncas J, Newman N, Fallaha M, Rivard CH. Lack of scientific evidence tor the treatment of lateral epicondylitis of the elbow. An attempted meta-analysis. J Bone Joint Surg (Br) 1992;74B:646-51.
37 McGeer AJ, Naylor CD, O'Rourke K, Detsky AS. Study quality as a factor in meta-analysis: approaches in the literature. Clin Res 1989;37:320A.

Aanvaard op 17 november 1998

De ziekte van Paget van bot: diagnostiek en behandeling

j.w.g.jacobs, a.m.huisman, h.c.van paassen en j.w.j.bijlsma

De ziekte van Paget (ostitis deformans) is genoemd naar Sir James Paget, die in 1877 deze chronische, deformerende botaandoening beschreef.[1] De aandoening is het laatst in 1987 in dit tijdschrift beschreven.[2] Sedert die tijd zijn er nieuwe laboratoriumbepalingen en nieuwe medicamenten voor botziekten beschikbaar gekomen. In dit artikel bespreken wij in hoeverre deze ontwikkelingen de aanpak van de ziekte hebben veranderd.

Samenvatting: zie volgende bladzijde.

beschrijving, pathofysiologie

De ziekte van Paget treedt op in één of meerdere botten en wordt gekenmerkt door toegenomen botombouw (remodellering) met hypertrofie en een abnormale structuur van het botweefsel met afgenomen stevigheid, hetgeen leidt tot misvormingen en fracturen. Bij deze botombouw staat excessieve botresorptie door een toegenomen aantal meerkernige, sterk vergrote osteoclasten voorop; hiermee samenhangend is er versterkte, maar onregelmatige botaanmaak door osteoblasten. De oorzaak van de ziekte is niet bekend. De vastgestelde verhoogde incidentie bij familieleden van een patiënt met de ziekte van Paget en de sterk wisselende prevalentie in verschillende rassen wijzen erop dat een genetische factor van belang is.[2] Omdat de aandoening in bepaalde streken vaker voorkomt, lijkt daarnaast een omgevingsfactor een rol te spelen. Er zijn virusachtige insluitlichaampjes aangetoond in osteoclasten van het

Academisch Ziekenhuis, afd. Reumatologie en Klinische Immunologie, Heidelberglaan 100, 3584 CX Utrecht.
Dr.J.W.G.Jacobs en prof.dr.J.W.J.Bijlsma, reumatologen.
Sint Franciscus Gasthuis, afd. Reumatologie, Rotterdam.
Mw.dr.A.M.Huisman, internist; dr.H.C.van Paassen, reumatoloog.
Correspondentieadres: dr.J.W.G.Jacobs.