

Meta-analyse: principes en valkuilen

Wim Opstelten en Rob J.P.M. Scholten

Bij een meta-analyse worden resultaten van verschillende onderzoeken kwantitatief samengevat. Dit geeft een veel preciezere effectschatting dan in de afzonderlijke onderzoeken mogelijk is.

Voorafgaande aan een meta-analyse moeten alle relevante onderzoeken worden getraceerd, ook onderzoeken met minder gunstige resultaten. Uitsluiting van 'negatieve' onderzoeken leidt tot een te gunstige effectschatting.

Omdat de uitkomsten van een meta-analyse afhankelijk zijn van de kwaliteit van de daarin opgenomen onderzoeken, dient ieder onderzoek systematisch op zorgvuldigheid van uitvoering en de kans op vertekening te worden beoordeeld.

Kwantitatieve samenvatting van onderzoeksresultaten is alleen mogelijk wanneer de verschillende onderzoeken homogeen zijn (ofwel voldoende op elkaar lijken) met betrekking tot onderzoekspopulatie, interventie of blootstelling, uitkomstmaat en meetmoment.

Bij klinische heterogeniteit kan een meta-analyse worden uitgevoerd in klinisch homogene subgroepen.

Statistische heterogeniteit kan verdisconteerd worden door gebruikmaking van het 'random effects'-model.

Meta-analyses zijn toepasbaar bij onderzoeksvragen over etiologie, diagnostiek, therapie en prognose.

Een aanbeveling in de gezondheidszorg is zelden gebaseerd op een enkel onderzoek. Willen we bijvoorbeeld weten welk analgeticum de voorkeur verdient bij acute lagerugpijn, dan zal 1 onderzoek met een vergelijking tussen paracetamol en NSAID's onvoldoende zijn voor het formuleren van een advies. De uitkomsten daarvan kunnen immers op toeval berusten of alleen gelden voor een selecte groep patiënten. Daarom is een aanbeveling of richtlijn meestal door meerdere onderzoeken onderbouwd. Wanneer we beschikken over verschillende onderzoeken met gelijke vraagstelling, dan kunnen we met grotere precisie het effect van bijvoorbeeld een behandeling schatten dan in de afzonderlijke onderzoeken mogelijk is. Daartoe is het belangrijk om op systematische wijze alle uitgevoerde onderzoeken te inventariseren. Wanneer deze zogenoemde systematische review onderzoeken oplevert die met betrekking tot patiëntenpopulatie, behandeling en uitkomstmaat voldoende op elkaar lijken, is het mogelijk om de resultaten kwantitatief samen te voegen. Deze kwantitatieve samenvatting van onderzoeksresultaten noemen we meta-analyse.

Nederlands Huisartsen Genootschap, Utrecht.

Dr. W. Opstelten, huisarts.

Dutch Cochrane Centre, Universitair Medisch Centrum Utrecht, Utrecht.

Prof.dr. R.J.P.M. Scholten, arts-epidemioloog.

Contactpersoon: prof.dr. R.J.P.M. Scholten (r.j.p.scholten@umcutrecht.nl).

SELECTIE VAN LITERatuur

De literatuurselectie bij meta-analyses volgt een vast stramien: de formulering van een expliciete onderzoeksvraag, een systematische en transparante zoekactie naar relevante oorspronkelijke onderzoeken, een kwaliteitsbeoordeling van die onderzoeken en tot slot een heldere presentatie van de resultaten. De onderzoeksvraag wordt meestal geformuleerd met het bekende PICO-systeem, waarbij alle genoemde elementen (patiënten, interventie, vergelijking ('comparison') en uitkomst ('outcome')) in de vraag worden vermeld.¹ Vervolgens is het belangrijk om alle onderzoeken die betrekking hebben op de onderzoeksvraag te achterhalen. Gepubliceerde onderzoeken zijn meestal te vinden in databases zoals Embase en Medline, maar ongepubliceerde onderzoeken zijn lastiger te traceren. Toch is het belangrijk om ook die te betrekken in een meta-analyse. Onderzoeken die een minder gunstig effect van een interventie laten zien zullen namelijk minder vaak gepubliceerd worden. Deze zogenaamde publicatiebias vormt dan ook een bedreiging voor de betrouwbaarheid van een meta-analyse, omdat uitsluiting van 'negatieve' onderzoeken meestal leidt tot een te gunstige effectschatting.²

Aanwijzingen voor publicatiebias kunnen zichtbaar worden in een 'funnelplot'.³ Hierin wordt van alle onderzoeken de grootte van de effectschatting uitgezet tegen de precisie daarvan (figuur). Aan de basis staan kleine onderzoeken, die door hun omvang minder precieze effectschattingen geven en sterker onderhevig zijn aan toeval dan de effectschattingen in grote onderzoeken, die aan de top van de figuur staan. Zo krijgt de figuur de vorm van een omgekeerde trechter (fuik, of 'funnel'). Wanneer de trechter asymmetrisch is, bijvoorbeeld door het ontbreken van kleine onderzoeken met een minder

gunstig effect, kan dit een aanwijzing zijn voor publicatiebias. Bij inclusie van slechts enkele onderzoeken, is het lastig om publicatiebias op deze wijze waar te nemen.

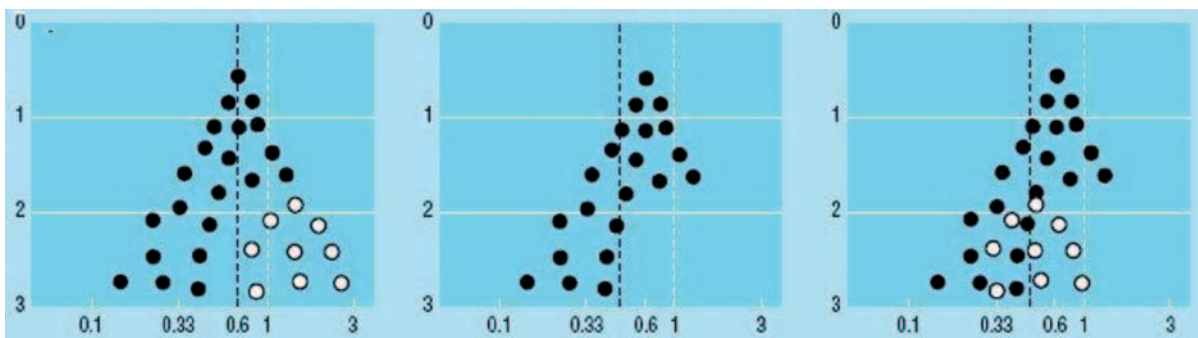
Asymmetrie van de trechter kan ook optreden wanneer de zoekactie niet goed is uitgevoerd (zogenoemde 'retrieval bias'). Tegenwoordig worden veel onderzoeken voor aanvang aangemeld bij nationale en internationale registers (zoals www.trialregister.nl en www.clinicaltrials.gov).⁴ Raadpleging hiervan kan onderzoeken opsporen, waarvan de resultaten niet of nog niet gepubliceerd zijn.

Kwaliteitsbeoordeling

De validiteit van de conclusies van een meta-analyse is afhankelijk van de kwaliteit van de daarin opgenomen onderzoeken. Daarom is het essentieel om ieder onderzoek op kwaliteit te beoordelen. Hiervoor bestaan verschillende beoordelingslijsten voor verschillende typen onderzoek.⁵⁻⁷ Als een onderzoek niet zorgvuldig is opgezet of uitgevoerd, is er een grote kans op vertekening van de resultaten. Deze leiden meestal tot een overschatting van het effect. Bij meta-analyses is een beoordeling van de methodologische kwaliteit (ofwel kans op vertekening) van de ingesloten onderzoeken dan ook een vast onderdeel. Bij lage kwaliteit van onderzoek kan men besluiten om deze niet in de meta-analyse op te nemen. Als alternatief kunnen 2 meta-analyses worden uitgevoerd: met en zonder inclusie van kwalitatief slechte onderzoeken. De kwaliteit van de systematische review zelf en de meta-analyse kan ook weer beoordeeld worden, met het zogenaamde AMSTAR-instrument.⁸

POOLEN

Als de gevonden onderzoeken voldoende op elkaar lijken met betrekking tot patiëntpopulatie, interventie (of



FIGUUR Funnelplot. Links een symmetrische plot in afwezigheid van vertekening (bias); de lichtgekleurde rondjes vertegenwoordigen kleine studies waarin geen positief effect werd gevonden. Midden: asymmetrische plot door publicatiebias; kleinere studies waarin geen effect werd gevonden zijn weggelaten. Rechts: asymmetrische plot door vertekening die wordt veroorzaakt door kleine studies van onvoldoende methodologische kwaliteit waarin er een overschatting is van het effect. De donkere stippellijn geeft de gepoolde oddsratio weer en de witte stippellijn geeft de neutrale waarde 1 aan (geen effect).³

TABEL 1 Forestplot van een meta-analyse naar het effect van behandeling van patiënten met lagerugpijn met NSAID's of paracetamol op algehele verbetering gemeten tot 3 weken na de start van de behandeling, zoals aangegeven door de patiënt⁹

eerste auteur, jaar publicatie	NSAID verbeterd/totaal; n	paracetamol verbeterd/totaal; n	gewicht	relatieve risico (95%-BI)
Hickey, 1982	13/16	7/14	26,2%	1,63 (0,92-2,89)
Milgrom, 1993	16/24	13/24	45,7%	1,23 (0,77-1,96)
Muckle, 1986	7/25	8/25	28,1%	0,88 (0,37-2,05)
totaal			100%	1,23 (0,88-1,73)

blootstelling aan de etiologische of prognostische factor) en uitkomstmaten, kunnen de resultaten daarvan gecombineerd worden tot één samenvattende schatting van het effect van de interventie of blootstelling. Dit statistisch combineren van onderzoeksuitkomsten wordt ‘poolen’ genoemd. Daarbij wordt een gewogen gemiddelde berekend van de resultaten van de onderzoeken die in de meta-analyse zijn opgenomen. De wegingsfactoren zijn gebaseerd op de grootte van de onderzoeken: hoe groter het onderzoek, des te sterker het meeweegt in de berekening van het gewogen gemiddelde. Daarnaast wordt het gewicht van een onderzoek groter, naarmate meer deelnemers de bestudeerde uitkomst (‘event’) hadden. Intuïtief is dat ook logisch want grotere onderzoeken en onderzoeken met meer events leggen meer gewicht in de schaal omdat ze gewoonweg meer informatie bevatten. Het voordeel van een meta-analyse is dat gebruikgemaakt wordt van alle beschikbare resultaten, waardoor een veel preciezere schatting van het effect verkregen wordt: het 95%-betrouwbaarheidsinterval (95%-BI) rond de effectschatting wordt smaller.

VOORBEELD: PIJNSTILLING BIJ ACUTE LAGERUGPIJN

Voor patiënten met acute lagerugpijn zijn verschillende typen pijnstillers beschikbaar. Een Cochrane-review heeft deze met elkaar vergeleken.⁹ Tabel 1 toont een meta-analyse van 3 onderzoeken die een NSAID vergeleken met paracetamol. De uitkomst was de door de patiënt aangegeven algehele verbetering gemeten tot 3 weken na de start van de behandeling. In de rijen staan de resultaten van ieder onderzoek vermeld. In het eerste onderzoek ondervonden 13 van de 16 NSAID-gebruikers een verbetering tegen 7 van de 14 paracetamol-gebruikers, resulterend in een relatief risico (RR) van 1,63. Dit is grafisch

weergegeven door het blauwe vierkantje, waarbij het lijntje het 95%-BI voorstelt. Dit loopt van 0,92 tot 2,89. Omdat het de neutrale waarde 1 (de waarde voor het RR als er geen verschil in effect is) omvat, zien we in een oogopslag dat het waargenomen effect niet statistisch significant is. De grootte van het vierkantje is evenredig met het gewicht van het onderzoek (26,2%). De onderste regel vat de onderzoeksresultaten samen en geeft het ‘gepoolde’ relatieve risico. Dit wordt grafisch weergegeven door een wybertje, waarvan de breedte het 95%-BI bestrijkt. Het gepoolde RR bedraagt 1,23 en is niet statistisch significant. Direct valt op dat het betrouwbaarheidsinterval van het gepoolde effect veel smaller is dan dat in de verschillende onderzoeken. Omdat de grafische weergave van effectschattingen met betrouwbaarheidsintervallen lijkt op een woud van lijnen, wordt dit een ‘forest’-plot genoemd.

Tabel 2 – uit dezelfde review – betreft een meta-analyse van bijwerkingen van NSAID's en paracetamol. Omdat niet alle onderzoeken naar de effectiviteit ook bijwerkingen vermeldden, zijn in deze meta-analyse deels andere onderzoeken opgenomen dan in de voorgaande. Het gepoolde RR voor bijwerkingen bedraagt 1,76. Uit de tabel kan direct worden opgemaakt dat dit effect statistisch significant is, omdat het 95%-BI (1,12 tot 1,76) de neutrale waarde 1 niet omvat. Hier wordt de kracht van een meta-analyse duidelijk: de individuele onderzoeken tonen een niet-significant verhoogd risico op bijwerkingen van NSAID's, terwijl na pooling van de onderzoeksresultaten dit risico wel significant verhoogd (en klinisch relevant) blijkt te zijn.

KLINISCHE HETEROGENITEIT

Zoals eerder gezegd moeten om een meta-analyse uit te kunnen voeren de afzonderlijke onderzoeken voldoende op elkaar lijken met betrekking tot de patiëntpopulatie, interventie (of blootstelling) en uitkomstmaten. Is dat het geval, dan spreken we van klinische homogeniteit. Om deze homogeniteit te kunnen beoordelen is voldoende kennis van het onderwerp noodzakelijk. Relevante vragen daarbij zijn of de in- en exclusiecriteria van de onderzoeken die in de meta-analyse zijn opgenomen min of meer gelijk waren, of de patiënten hetzelfde stadium van ziekte hadden en of de onderzochte interventies voldoende vergelijkbaar waren.

Is dat niet het geval en is er dus sprake van klinische heterogeniteit, dan zouden de waargenomen verschillen kunnen leiden tot verschillende effecten van de interventie en is het resultaat van een meta-analyse nietszeggend en wellicht misleidend. Het is dan beter klinisch homogene subgroepen te onderscheiden en een meta-analyse uit te voeren binnen die subgroepen. Zo werd in de eerder genoemde Cochrane-review terecht een onderscheid gemaakt tussen acute en chronische lagerugpijn.⁹

STATISTISCHE HETEROGENITEIT

Als de resultaten van de in de meta-analyse opgenomen onderzoeken sterk verschillen, spreken we van statistische heterogeniteit. Deze kan veroorzaakt zijn door wellicht nog onbekende klinische heterogeniteit of door methodologische tekortkomingen van onderzoeken (methodologische heterogeniteit). De beste manier om statistische heterogeniteit op te sporen is te kijken naar de forest-plot van de meta-analyse, de zogenoemde 'eye-ball'-test: zijn de schattingen van het effect in de verschil-

lende onderzoeken ongeveer gelijk en is er voldoende overlap van de betrouwbaarheidsintervallen? Als dat het geval is, zoals in de tabellen 1 en 2, is statistische heterogeniteit onwaarschijnlijk.

Een andere manier om statistische heterogeniteit te traceren is aan de hand van de zogenoemde I^2 -statistiek. I^2 geeft het percentage variatie tussen de effectschattingen van de verschillende onderzoeken weer dat toegeschreven kan worden aan heterogeniteit. Daarbij worden de volgende arbitraire waarden gehanteerd: bij een $I^2 < 30\%$ is nauwelijks sprake van heterogeniteit, bij een $I^2 > 60\%$ is er aanzienlijke heterogeniteit. Bevat de meta-analyse echter enkele grote onderzoeken, dan zal dit sowieso leiden tot een hogere I^2 -waarde. De eye-ball-test is dan betrouwbaarder om heterogeniteit te beoordelen.

OPTIES BIJ HETEROGENITEIT

Bij heterogeniteit kunnen de resultaten van de verschillende onderzoeken niet zomaar worden samengevoegd. Het zou een vertekend beeld geven van het werkelijke effect waarnaar we op zoek zijn. Er zijn dan 3 mogelijkheden.

De eerste is dat besloten kan worden om niet te poolen en alleen de resultaten van de afzonderlijke onderzoeken te beschrijven in de systematische review.

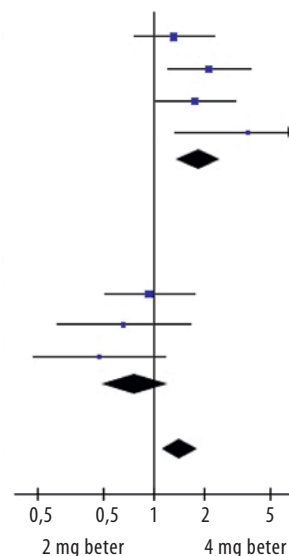
Een tweede optie is om de analyse aan te passen aan het gegeven van heterogeniteit. Toepassing van het 'random effects'-model kan heterogeniteit verdisconteren. Dit model gaat er vanuit dat de verschillende effecten die in onderzoeken worden gevonden niet alleen berusten op toevalsvariatie, maar dat de interventie per onderzoek ook een daadwerkelijk verschillend effect toont. De gevonden verschillende effecten vormen dan een random

TABEL 2 Forestplot van een meta-analyse naar het effect van behandeling van patiënten met lagerugpijn met NSAID's of paracetamol op door de patiënt ervaren bijwerkingen bij maximaal 3 weken follow-up⁹

eerste auteur, jaar publicatie	NSAID	paracetamol	gewicht	relatieve risico (95%-BI)
	bijwerkingen/ totaal; n	bijwerkingen/ totaal; n		
Evans, 1980	20/30	13/30	68,8%	1,54 (0,95-2,49)
Hickey, 1982	2/16	1/14	5,6%	1,75 (0,18-17,29)
Nadler, 2002	11/106	5/113	25,6%	2,35 (0,84-6,53)
totaal			100%	1,76 (1,12-2,76)

TABEL 3 Forestplot van een meta-analyse naar het effect van kauwgum met 4 mg of 2 mg nicotine op het stoppen met roken, met een subgroepanalyse voor sterk en minder sterk verslaafde rokers*¹⁰

eerste auteur, jaar publicatie	dosis 4 mg	dosis 2 mg	gewicht; %	relatieve risico (95%-BI)
	gestopt met roken/ totaal; n	gestopt met roken/ totaal; n		
sterk verslaafde rokers				
Garvey, 2000	24/116	18/115	22,2	1,32 (0,76-2,30)
Herrera, 1995	30/87	13/81	16,5	2,15 (1,21-3,82)
Kornitzer, 1987	24/73	16/86	18,0	1,77 (1,02-3,06)
Tonnesen, 1988	12/27	4/33	4,4	3,67 (1,33-10,08)
subtotaal			61,2	1,85 (1,36-2,50)
minder sterk verslaafde rokers				
Garvey, 2000	16/87	17/87	20,9	0,94 (0,51-1,74)
Hughes, 1990	5/19	8/20	9,6	0,66 (0,26-1,66)
Kornitzer, 1987	5/17	5/8	8,4	0,47 (0,19-1,17)
subtotaal			38,8	0,77 (0,49-1,21)
totaal			100	1,43 (1,12-1,83)



* I^2 voor respectievelijk alle onderzoeken, sterk verslaafde rokers en minder sterk verslaafde rokers 62%, 13% en 0%.

verdeling rond een gemiddeld globaal effect. Tegenover dit model staat het 'fixed effect'-model dat er vanuit gaat dat de gevonden verschillen alleen op toeval berusten en dat alle onderzoeken in de meta-analyse in werkelijkheid hetzelfde (fixed) effect schatten. Dit model kan alleen worden toegepast bij homogeniteit. Het random-effects-model geeft bredere betrouwbaarheidsintervallen en daardoor minder snel een significant resultaat dan het fixed-effect-model. Een fixed-effect-model geeft dus de beste schatting van hét effect van een interventie, terwijl een random-effects-model de beste schatting geeft van het gemiddelde van alle mogelijke effecten die een interventie kan hebben in verschillende situaties. Een random-effects-meta-analyse vraagt dan ook eigenlijk altijd om een nadere exploratie om specifiekere uitspraken te kunnen doen: kunnen subgroepen of factoren worden onderscheiden die leiden tot een hoger of lager effect?

Dit brengt ons op de derde optie om heterogeniteit te verklaren, de subgroepanalyse. Dat kan door per onderzoek de resultaten van een meer homogene groep te verzamelen en deze apart te poolen. Zo laat tabel 3 zien dat een hoog gedoseerde nicotinebevattende kauwgum (4 mg) effectiever is bij het stoppen met roken dan een lager gedoseerde kauwgum (2 mg) (RR = 1,36).¹⁰ Over het

geheel genomen is er echter aanzienlijke heterogeniteit ($I^2 = 62\%$). Deze verdwijnt bij de uitsplitsing van de patiënten in 2 subgroepen, namelijk sterk en minder sterk verslaafde rokers. Beide subgroepen zijn nu homogeen (I^2 respectievelijk 13% en 0%) en het wordt direct zichtbaar dat de hoge dosering vooral effectief is voor sterk verslaafde rokers (RR = 1,85). Het kan overigens lastig en arbitrair zijn om subgroepen te kiezen. Bovendien: wanneer subgroepen niet vooraf in de oorspronkelijke onderzoeken waren gedefinieerd, hebben de resultaten van deze subgroepanalyse altijd een post-hoc-karakter. De gevonden uitkomst kan dan op toeval berusten en zal in nader onderzoek geïnterpreteerd moeten worden. De waarde van een post-hocanalyse neemt toe wanneer deze wordt uitgevoerd op de primaire uitkomstmaat van de onderzoeken.

PERSPECTIEVEN

Door de grote precisie hebben meta-analyses een belangrijke plaats in de klinische praktijk. Ook bij het voorbereiden van wetenschappelijk onderzoek zijn meta-analyses waardevol. Ze zijn toepasbaar in alle domeinen van de klinische epidemiologie: etiologie, diagnostiek, therapie en prognostiek. Inmiddels zijn ook nieuwe technie-

LEERPUNTEN

- In een meta-analyse worden resultaten uit meerdere onderzoeken kwantitatief samengevoegd, waardoor effectschattingen veel preciezer zijn dan in de individuele onderzoeken.
- Door deze grote precisie hebben meta-analyses een belangrijke plaats bij het formuleren van aanbevelingen in de gezondheidszorg.
- Een meta-analyse is alleen mogelijk wanneer de individuele onderzoeken voldoende op elkaar lijken met betrekking tot de patiëntpopulaties, interventies en uitkomstmaten.
- Vooral publicatiebias, slechte kwaliteit van geïncludeerde onderzoeken en heterogeniteit zijn bedreigingen voor de validiteit van de conclusies van een meta-analyse.

ken ontwikkeld, zoals de netwerkmeta-analyse.^{11,12} Waar in gebruikelijke meta-analyses de vergelijkingen steeds slechts 2 alternatieven betreffen, kunnen in een netwerkmeta-analyse gelijktijdig meerdere opties met elkaar vergeleken worden. Op deze wijze kunnen – onder stringente aannames – effecten van verschillende behandelopties tegen elkaar worden afgezet, ook al zijn die nooit in één onderzoek direct met elkaar vergeleken. Zo werd onlangs een onderzoek gepubliceerd waarin het risico op veneuze trombose bij gebruik van orale anticonceptiva (OAC) werd bepaald.¹³ Hoewel directe vergelijkingen tussen de verschillende generaties OAC vaak ontbraken, konden door netwerkmeta-analyse de risico's van eerste-, tweede- en derdegeneratie OAC toch met elkaar vergeleken worden.

Daarnaast komen meta-analyses op basis van individuele patiëntgegevens steeds meer in zwang.^{14,15} Deze techniek

biedt onder meer de mogelijkheid van analyse van subgroepen op een manier die met een conventionele meta-analyse onmogelijk is.

Tot slot noemen we de meta-regressieanalyse.¹⁶ Dit is een multivariate analysetechniek om in een systematische review de relatie te onderzoeken tussen bepaalde onderzoekskenmerken (bijvoorbeeld gemiddelde leeftijd of geslachtsverdeling) en de grootte van het waargenomen effect. Een recente meta-regressieanalyse van 41 onderzoeken naar het effect van ijzersuppletie op het hemoglobinegehalte illustreert dit: per 10 µg/l toename van het serumferritine nam het effect van ijzersuppletie af met gemiddeld 0,08 g/dl.¹⁷

CONCLUSIE

Een meta-analyse is een waardevol instrument voor het formuleren van aanbevelingen in de gezondheidszorg. Voorwaarde is wel dat een meta-analyse goed wordt uitgevoerd. Vooral publicatiebias, slechte kwaliteit van onderzoeken en heterogeniteit kunnen de validiteit en de resultaten van een meta-analyse bedreigen en de interpretatie ervan bemoeilijken.

In de rubriek Stand van zaken verschijnen regelmatig bijdragen over methoden die gebruikt worden bij het opzetten van wetenschappelijk onderzoek. De artikelen in deze serie illustreren op begrijpelijke wijze wat een bepaalde methode behelst, zonder dat hier uitvoerige methodologische kennis voor nodig is. Zowel oude als nieuwe methodologische principes worden zo inzichtelijk gemaakt voor artsen die klinische onderzoeken goed willen interpreteren.

Belangenconflict en financiële ondersteuning: geen gemeld.

Aanvaard op 17 december 2013

LITERATUUR

- 1 Koopmans RP, van Benthem PPG, Offringa M. De juiste vragen stellen. In: Offringa M, Assendelft WJJ, Scholten RJPM (red). Inleiding in evidence-based medicine. Houten: Bohn Stafleu van Loghum; 2008.
- 2 Scholten RJ, Hooft L. Einde aan selectieve publicatie nog niet in zicht. Ned Tijdschr Geneeskd. 2012;156:A4848.
- 3 Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. BMJ. 2001;323:101-5.
- 4 Hooft L, et al. Een nationaal prospectief trialregister voor gerandomiseerde en gecontroleerde trials: ethisch en praktisch noodzakelijk. Ned Tijdschr Geneeskd. 2004;148:1866-9.
- 5 Higgins JP, Altman DG, Sterne JA. Assessing risk of bias in included studies. In: Higgins JP and Green S (ed). Cochrane Handbook for Systematic Reviews of Interventions version 5.1.0 (updated March 2011) The Cochrane Collaboration; 2011.
- 6 Whiting PF, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-36.
- 7 Wells GA, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp geraadpleegd op 31 augustus 2013.
- 8 Shea BJ, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS ONE. 2007;2:e1350.
- 9 Roelofs PD, et al. Non-steroidal anti-inflammatory drugs for low back pain. Cochrane Database Syst Rev. 2008;23(1):CD000396.

- 10 Stead LF, et al. Nicotine replacement therapy for smoking cessation. *Cochrane Database Syst Rev.* 2012;11(14):CD000146.
- 11 Song F, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ.* 2009;338:b1147.
- 12 Cipriani A, et al. Conceptual and Technical Challenges in Network Meta-analysis. *Ann Intern Med.* 2013;159:130-7.
- 13 Stegeman BH, et al. Different combined oral contraceptives and the risk of venous thrombosis: systematic review and network meta-analysis. *BMJ.* 2013;347:f5298.
- 14 Rovers MM, Reitsma JB. Meta-analyse op basis van individuele-patientengegevens. *Ned Tijdschr Geneeskd.* 2012;156:A4743.
- 15 Van Walraven C, et al. Effect of age on stroke prevention therapy in patients with atrial fibrillation: the atrial fibrillation investigators. *Stroke.* 2009;40:1410-6.
- 16 Green R, et al. The effect of rising food prices on food consumption: systematic review with meta-regression. *BMJ.* 2013;346:f3703.
- 17 Casgrain A, et al. Effect of iron intake on iron status: a systematic review and meta-analysis of randomized controlled trials. *Am J Clin Nutr.* 2012;96:768-80.

Citeer als: *Ned Tijdschr Geneeskd.* 2014;158:A6882

 **KIJK OOK OP WWW.NTVG.NL/KLINISCHEPRAKTIJK**