

---

## De praktijk van systematische reviews. V. Heterogeniteit tussen onderzoeken en subgroepanalyses

r.j.p.m.scholten, w.j.j.assendelft, p.j.kostense en l.m.bouter

In het vorige artikel uit deze reeks over systematische reviews<sup>1-3</sup> is uitvoerig stilgestaan bij het combineren van de resultaten van de afzonderlijke onderzoeken tot één samenvattende schatting van het effect van een interventie of blootstellingsfactor ('pooling').<sup>3</sup> De in dat artikel gepresenteerde methode van pooling is alleen valide als de in de afzonderlijke onderzoeken gevonden effecten als ongeveer gelijk beschouwd kunnen worden, ofwel homogeen zijn. Lopen deze effecten (sterk) uiteen, dan spreekt men van heterogeniteit. Zou men in geval van het bestaan van heterogeniteit de resultaten van de afzonderlijke onderzoeken toch aan de hand van de eerder gepresenteerde methode combineren, dan heeft de samenvattende schatting van het effect geen direct interpreterbare betekenis.

In dit 5e artikel uit deze reeks gaan wij dieper in op heterogeniteit tussen onderzoeken. De eerste paragrafen betreffen een beschouwing over de mogelijke oorzaken van heterogeniteit en methoden voor het vaststellen ervan. Vervolgens worden suggesties gegeven voor het omgaan met heterogeniteit. Wij presenteren daartoe het zogenoemde 'random effects'-model (een variant van het in het vorige artikel behandelde algemene statistische model), waarmee men eventueel aanwezige heterogeniteit tussen onderzoeken kan verdisconteren. Ook zullen wij ingaan op het uitvoeren van subgroepanalyses.

### oorzaken van heterogeniteit

Heterogeniteit kan berusten op een ongelukkige keuze van het type effectmaat, verschillen in methodologische kwaliteit tussen onderzoeken of werkelijke verschillen tussen onderzoeken (klinische heterogeniteit).

*Type effectmaat.* Er kunnen 2 typen biologische effecten onderscheiden worden: additieve en multiplicatieve. Figuur a geeft een voorbeeld van een additief biologisch effect. Naarmate men ouder wordt, neemt de kans op het krijgen van ziekte A toe en rokers hebben een verhoogd risico op het verkrijgen van de ziekte (de lijn voor rokers ligt hoger). Het verschil tussen rokers en niet-rokers is echter constant, ongeacht de leeftijd (de lijnen

---

### samenvatting

- Een meta-analyse behelst de analyse van de gegevens van vergelijkbare onderzoeken die in het kader van een systematische review zijn vergaard.
- Indien de resultaten van de verschillende onderzoeken (sterk) uiteenlopen, spreekt men van heterogeniteit.
- Heterogeniteit kan berusten op een ongelukkige keuze van het type effectmaat, verschillen in methodologische kwaliteit tussen onderzoeken of werkelijke verschillen tussen onderzoeken.
- Een grondige beschouwing van de verschillen in onderzoekskenmerken in combinatie met een visuele inspectie van de mate van overlap van de betrouwbaarheidsintervallen van de effectschattingen in de afzonderlijke onderzoeken is de beste manier om na te gaan of er heterogeniteit tussen onderzoeken bestaat.
- Door toepassing van het 'random effects'-model voor 'pooling' kan niet-systematische heterogeniteit tussen onderzoeken verdisconteerd worden.
- In geval van het bestaan van heterogeniteit tussen onderzoeken is verdere analyse van mogelijke bronnen daarvan (subgroepanalyse) geïndiceerd.
- Subgroepanalyse dient subgroepen van patiënten uit de afzonderlijke onderzoeken te behelzen en te worden gebaseerd op strikte, bij voorkeur vooraf opgestelde, criteria.

lopen evenwijdig). Het risicoverschil is derhalve in deze situatie de juiste effectmaat om het verband tussen de kans op ziekte A bij rokers en bij niet-rokers weer te geven. In dit voorbeeld neemt het relatieve risico af bij toenemende leeftijd (de verhouding van de kans op ziekte A bij rokers en bij niet-rokers wordt steeds kleiner) en het relatieve risico is derhalve 'heterogeen'. Figuur b geeft een voorbeeld van een multiplicatief biologisch effect. In deze situatie is de verhouding van de kans op de ziekte bij rokers en bij niet-rokers constant, ongeacht de leeftijd, en hier is het relatieve risico de effectmaat die de situatie juist weergeeft. Zou men een risicoverschil berekenen en dus een effectmaat op een additieve schaal gebruikt hebben, dan zouden 'heterogene' risicoverschillen gevonden worden (de verschillen nemen toe bij toenemende leeftijd; de lijnen lopen niet evenwijdig). Homogeniteit op een multiplicatieve schaal impliceert derhalve heterogeniteit op een additieve schaal en vice versa.<sup>4</sup> Stelt men in geval van pooling van een effectmaat van de ene schaal heterogeniteit vast, dan kan overgaan naar een effectmaat van een andere schaal mogelijk een oplossing bieden.

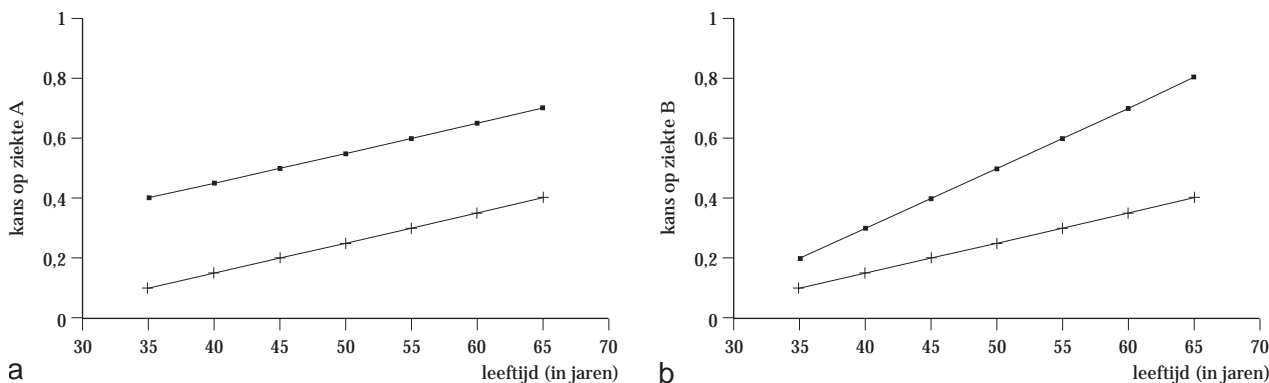
*Methodologische kwaliteit.* Een andere potentiële bron van heterogeniteit is de invloed van onderzoeken van slechte methodologische kwaliteit. In een voor-

---

Vrije Universiteit, Faculteit der Geneeskunde, Instituut voor Extramuraal Geneeskundig Onderzoek, Van der Boechorststraat 7, 1081 BT Amsterdam.

Dr.R.J.P.M.Scholten, arts-epidemioloog; dr.W.J.J.Assendelft, huisarts-epidemioloog (thans: Academisch Medisch Centrum, Dutch Cochrane Centre, Amsterdam); dr.P.J.Kostense, epidemioloog-statisticus (tevens afd. Klinische Epidemiologie & Biostatistiek); prof.dr.L.M.Bouter, epidemioloog.

*Correspondentieadres:* dr.R.J.P.M.Scholten (e-mail: rjpm.scholten.emgo@med.vu.nl).



Kans op ziekte voor rokers (■) en niet-rokers (+) naar leeftijd: fictief voorbeeld van (a) een additief biologisch model, waarbij het risicoverschil tussen rokers en niet-rokers constant is, ongeacht de leeftijd; het relatieve risico neemt af bij toenemende leeftijd en is dus heterogeen voor patiënten van verschillende leeftijd; (b) een multiplicatief biologisch model, waarbij de verhouding van de kans op de ziekte bij rokers en niet-rokers constant is, ongeacht de leeftijd. Het relatieve risico is hierbij de effectmaat die de situatie juist weergeeft; het risicoverschil is heterogeen omdat het toeneemt met de leeftijd.

gaand artikel uit deze reeks werd beschreven dat een gebrekkige methodologische kwaliteit kan samengaan met systematisch afwijkende resultaten.<sup>2,5</sup> Uitsluiting van methodologisch slechtere onderzoeken kan de heterogeniteit wellicht opheffen.

**Klinische heterogeniteit.** Werkelijke verschillen tussen onderzoeken wat betreft de opzet, onderzoekspopulatie, interventie (aard, dosis, duur) of blootstelling (smeting), uitkomstmaten en/of follow-upduur vormen doorgaans de belangrijkste bron van heterogeniteit. Deze vorm van heterogeniteit wordt door Thompson 'klinische heterogeniteit' genoemd.<sup>6</sup> Het moge duidelijk zijn dat in een meta-analyse bijna altijd wel sprake is van op zijn minst enige verschillen tussen de afzonderlijke onderzoeken. Stel, dat 'leeftijd' de relatie tussen de bestudeerde determinant en de uitkomst beïnvloedt (leeftijd is dan een effectmodifier van de bestudeerde relatie). Als een deel van de gevonden onderzoeken voornamelijk ouderen betref, een ander deel voornamelijk jongeren en de overige onderzoeken alle leeftijdscategorieën, dan zullen heterogene resultaten verkregen worden. Selecteert men van alle onderzoeken telkens patiënten van een bepaalde leeftijdscategorie en combineert men de resultaten van deze homogene subgroepen, dan zullen de verschillen tussen onderzoeken verdwijnen. Bovendien kan in een dergelijke subgroepenanalyse een nadere analyse van het leeftijdeffect uitgevoerd worden. Subgroepenanalyses zijn derhalve een belangrijk onderdeel van een meta-analyse.

#### vaststellen van heterogeniteit

Heterogeniteit van de resultaten tussen onderzoeken kan statistisch onderzocht worden door middel van de Q-toets.<sup>3</sup> De p-waarde van de Q-toets geeft aan hoe toevallig de gevonden verschillen in resultaten tussen de onderzoeken zijn als deze in werkelijkheid als gelijk beschouwd moeten worden (ofwel 'onder de nulhypothese van homogeniteit'). De nulhypothese van homogeniteit wordt al dan niet verworpen op grond van de groot-

te van de toetsingsgrootte Q. Het is echter riskant om louter op deze toets voor homogeniteit te varen. Een bekend probleem van deze toets is namelijk het gebrek aan onderscheidingsvermogen.<sup>7,8</sup> Heeft men een groot aantal onderzoeken tot zijn beschikking en/of zijn de patiëntenaantallen in de afzonderlijke onderzoeken relatief klein, dan heeft de Q-toets weinig onderscheidingsvermogen en is er een grote kans op een type-II-fout: het ten onrechte niet verwerpen van de nulhypothese, ofwel een grote kans om ten onrechte te concluderen dat er sprake is van homogeniteit. Een pragmatische oplossing voor dit probleem is het ruimer stellen van de grens voor het verwerpen van de nulhypothese door bijvoorbeeld de kritieke p-waarde op 0,10 te stellen in plaats van op 0,05, maar ook dan kan de kans op een type-II-fout aanzienlijk zijn. Heeft men daarentegen een relatief klein aantal onderzoeken tot zijn beschikking of zijn de patiëntenaantallen in de afzonderlijke onderzoeken aanzienlijk, dan kunnen subtiele verschillen tussen de resultaten al leiden tot het verwerpen van de nulhypothese en zouden triviale verschillen al als heterogeniteit geduid kunnen worden. In dat geval kan men de uitslag van de toets voor homogeniteit beter negeren. Vanwege de gevoeligheid van de Q-toets voor de hier geschetste omstandigheden is deze dan ook van beperkte waarde voor het vaststellen van heterogeniteit tussen onderzoeken. Bovendien is de Q-toets een indicator voor de aanwezigheid van statistische heterogeniteit, terwijl in het kader van een meta-analyse het bestaan van klinische heterogeniteit veel meer belang heeft.

De beste methode voor het vaststellen van (klinische) heterogeniteit is het gebruik van gezond verstand.<sup>6</sup> Men dient zich op basis van eerder onderzoek of klinische ervaring af te vragen of de te combineren onderzoeken voldoende vergelijkbaar zijn (geen al te grote verschillen vertonen) wat betreft de patiënten, de centrale determinant (interventie, blootstelling) en de uitkomsten. Concludeert men dat de onderzoeken onvoldoende vergelijkbaar zijn, dan verdient het aanbeveling pooling van

alle onderzoeken achterwege te laten en alleen de resultaten van vergelijkbare subgroepen te combineren (zie ook verder).

Als men na nauwkeurige beschouwing van de kenmerken van de te combineren onderzoeken tot de conclusie is gekomen dat deze voldoende vergelijkbaar lijken, dan kunnen andere, buiten beschouwing gebleven onderzoekskenmerken nog steeds tot heterogene resultaten leiden. Visuele inspectie van de afzonderlijke punt-schatters van de effectmaten en hun betrouwbaarheidsintervallen kan dan een goede indruk geven van de overeenkomsten of verschillen van de onderzoeksresultaten. Men beoordeelt heterogeniteit dan 'op het oog'. Als de betrouwbaarheidsintervallen van de afzonderlijke onderzoeken elkaar in aanzienlijke mate overlappen, zijn er niet of nauwelijks aanwijzingen voor heterogeniteit.<sup>9</sup>

Een weloverwogen vergelijking van de onderzoekskenmerken in combinatie met een visuele inspectie van de mate van overlap van de betrouwbaarheidsintervallen van de afzonderlijke onderzoeken is derhalve de beste manier om na te gaan of er heterogeniteit tussen onderzoeken bestaat.

### hoe om te gaan met heterogeniteit?

Heterogeniteit kan op meerdere manieren gehanteerd worden.<sup>10</sup> De beste oplossing kan soms zijn te besluiten om de resultaten eenvoudigweg niet te combineren. Negeert men het bestaan van heterogeniteit en zou men ondanks het bestaan van substantiële verschillen tussen onderzoeken de resultaten ervan toch via het in het vorige artikel gepresenteerde 'fixed effects'-model combineren, dan staat dat volgens sceptische onderzoekers gelijk aan 'het vergelijken van appels met peren'. De pooling leidt dan tot een schatting van een grootste gemene deler die echter in de dagelijkse praktijk geen betekenis heeft. Beter is het om in geval van heterogeniteit een statistische methode van pooling toe te passen die de bestaande heterogeniteit kan verdisconteren (het zogenoemde 'random effects'-model) of de heterogeniteit verder te exploreren door subgroepanalyses uit te voeren.

Het 'random effects'-model. Het uitgangspunt bij het in het vorige artikel beschreven algemene statistische model is dat er één vaste onderliggende waarde is voor het werkelijke effect en dat de verschillen in uitkomsten tussen de onderzoeken op toeval berusten (fixed-effectsmodel). De beschikbare onderzoeken vormen de basis voor het maken van de beste schatting van de waarde van dit ene werkelijke effect. Een andere invalshoek wordt geboden door het door DerSimonian en Laird geïntroduceerde 'random effects'-model.<sup>11</sup> Hierbij wordt ervan uitgegaan dat verschillen tussen onderzoeken in de praktijk onontkoombaar en in die zin reëel zijn: ieder onderzoek schat zijn eigen werkelijke effect. De beschikbare onderzoeken worden beschouwd als een steekproef uit een oneindige reeks van mogelijke onderzoeken met oneindig veel verschillende waarden voor het werkelijke effect. Bij toepassing van het random-effectsmodel wordt aan de hand van de beschikbare 'steekproef' van onderzoeken een schatting gemaakt van de kansverdeling van al deze verschillende werkelijk-

ke effecten (en dus niet van één vaste waarde). Deze kansverdeling wordt doorgaans normaal verondersteld en wordt dan gekarakteriseerd door een gemiddelde en variantie. Naast toevalsvariantie binnen de afzonderlijke onderzoeken (aangeduid door  $s_i^2$ ) wordt de variatie tussen onderzoeken ( $\tau^2$ ) in de formules van de wegingsfactoren verdisconteerd (tabel 1). De berekeningen geschieden vervolgens op dezelfde wijze als die voor het fixed-effectsmodel (zie tabel 1), echter met andere wegingsfactoren. Indien werkelijk sprake is van heterogeniteit tussen onderzoeken, leidt het random-effectsmodel doorgaans tot een andere schatting van het effect, een breder betrouwbaarheidsinterval en een lagere waarde voor de toetsingsgrootheid  $U'$ . Ook is de interpretatie van de resultaten die verkregen zijn aan de hand van het random-effectsmodel anders dan die van het fixed-effectsmodel. De met het random-effectsmodel

TABEL 1. Berekening van gewichten voor het 'random effects'-model\* van DerSimonian en Laird<sup>11</sup> en de toepassing ervan in een algemeen statistisch model voor het combineren van (bij benadering) normaal verdeelde effectschattingen uit een aantal onderzoeken

gecombineerde schatting van het gemiddelde effect met 95%-betrouwbaarheidsinterval:

$$y' = (\sum y_i w_i) \pm (1,96 / \sqrt{\sum w_i})$$

toets voor de nulhypothese ' $\theta = 0$ ':

$$U' = (\sum y_i w_i)^2 / \sum w_i$$

$\Sigma$  = som

$i$  = 1, ..., K

K = aantal onderzoeken

$\theta$  = werkelijke waarde van het gemiddelde effect

$y_i$  = schatting van het effect in het i-de onderzoek

$y'$  = gecombineerde schatting van het gemiddelde effect (random-effectsmodel\*)

$s_i^2$  = variantie van  $y_i$  in het i-de onderzoek

$w_i$  = wegingsfactor voor het i-de onderzoek =  $1/s_i^2$  ('fixed effects'-model†)

Q = waarde van de toetsingsgrootheid voor de toets voor homogeniteit in het fixed-effectsmodel†

$$= \sum y_i^2 w_i - (\sum y_i w_i)^2 / \sum w_i$$

$\tau^2$  = variantie tussen onderzoeken

$$= 0, \text{ indien } Q \leq K - 1$$

$$= (Q - (K - 1)) / (\sum w_i - (\sum w_i^2 / \sum w_i)), \text{ indien } Q > K - 1$$

$w_i'$  = wegingsfactor (random-effectsmodel\*) voor het i-de onderzoek =  $1/(s_i^2 + \tau^2)$

$U'$  = de toetsingsgrootheid voor het toetsen van de nulhypothese dat het werkelijke gemiddelde effect nul is; deze volgt een  $\chi^2$ -verdeling met 1 vrijheidsgraad

\*In het random-effectsmodel wordt aan de hand van de beschikbare onderzoeken een schatting gemaakt van de kansverdeling van de verschillende werkelijke effecten. Waargenomen variatie berust op toeval en op verschillen tussen de onderzoeken.

†In het fixed-effectsmodel wordt ervan uitgegaan dat er één werkelijk effect bestaat en dat de verschillen tussen de onderzoeksuitkomsten op toeval berusten.

verkregen schatting van het gemiddelde effect vertegenwoordigt het gemiddelde van een hele 'populatie' van mogelijke effecten onder verschillende omstandigheden, terwijl de met het fixed-effectsmodel verkregen waarde de beste schatting is van het enige 'ware' effect (onder de aanname dat dit bestaat).

In tabel 2 staat de uitwerking van het combineren van risicoverschillen van de in het vorige artikel beschouwde fictieve meta-analyse volgens het random-effectsmodel.<sup>3</sup> Het fixed-effectsmodel leverde een risicoverschil van -0,06 met een 95%-betrouwbaarheidsinterval van -0,09 tot -0,03 en een waarde voor de toetsingsgrootheid U van 15,49 ( $p < 0,001$ ). In dit geval leidt analyse met het random-effectsmodel slechts tot subtiële verschillen: het risicoverschil bedraagt nu -0,07 met een 95%-betrouwbaarheidsinterval van -0,11 tot -0,02; de waarde van de toetsingsgrootheid U' is 9,14 ( $p = 0,003$ ). De interpretatie van deze resultaten is als volgt: -0,07 is een schatting van het gemiddelde van een reeks van mogelijke risicoverschillen; het 95%-betrouwbaarheidsinterval van dit gemiddelde reikt van -0,11 tot -0,02.

Analyse van het relatieve risico en de oddsratio met behulp van het random-effectsmodel levert in dit voorbeeld dezelfde resultaten als die met het fixed-effectsmodel. De waarden van de toetsingsgrootheid voor homogeniteit (Q) waren respectievelijk 1,15 en 1,29. In beide gevallen is dat lager dan 2 (= het aantal onderzoeken, in ons voorbeeld 3, minus 1). De variatie tussen onderzoeken ( $\tau^2$ ) wordt dan op 0 gesteld (zie tabel 1) en de gewichten  $w'$  van het random-effectsmodel worden daardoor gelijk aan de gewichten  $w$  van het fixed-effectsmodel.

Een hinderlijke eigenschap van het random-effectsmodel is dat kleine onderzoeken bij de berekening van het gecombineerde effect bijna evenveel gewicht kunnen krijgen als grote.<sup>12</sup> De waarde van  $\tau^2$  is immers voor alle onderzoeken gelijk (zie tabel 1) en indien  $\tau^2$  groot is ten opzichte van  $s_i^2$ , zullen de gewichten ( $w_i'$ ) relatief weinig verschillen. Dit betekent dat kleine onderzoeken de pooling ten onrechte kunnen gaan domineren. Het random-effectsmodel is derhalve geenszins een panacee.

*Exploreren van heterogeniteit.* Als de resultaten van de verschillende onderzoeken heterogeen zijn, kan dit ons op het spoor zetten van mogelijke oorzaken daarvan. Potentiële bronnen van heterogeniteit kunnen onderzocht worden door middel van subgroepenanalyses en dit is wellicht de beste oplossing voor het omgaan met (klinische) heterogeniteit. Subgroepenanalyses dienen bij voorkeur subgroepen van patiënten uit de afzonderlijke onderzoeken te behelzen en niet een vergelijking van afzonderlijke onderzoeken met en zonder bepaalde kenmerken (subgroepenanalyse tussen onderzoeken).<sup>10-13</sup> In het laatste geval kunnen andere belangrijke onderzoekskenmerken de resultaten beïnvloeden, waardoor vertekende resultaten verkregen kunnen worden.

### subgroepenanalyses

Subgroepenanalyses zijn in het bijzonder aangewezen bij (klinische) heterogeniteit, maar ook in geval van homogene resultaten verschaffen subgroepenanalyses verder inzicht in het effect van verschillende andere factoren op de bestudeerde relatie.

Subgroepenanalyses dienen niet lukraak uitgevoerd te worden, maar bij voorkeur aan de hand van vooraf geformuleerde (biologisch) plausibele hypothesen.<sup>10-15</sup> Zo kan het om inhoudelijke redenen zinvol zijn te onderzoeken of het effect van een bepaalde behandeling anders is bij mannen, ouderen of patiënten met een ernstiger ziektebeeld. Bij het uitvoeren van subgroepenanalyses dient men het aantal ervan te beperken. Hoe meer subgroepenanalyses men uitvoert, des te groter is immers de kans op een fout-positief resultaat (type-I-fout: het ten onrechte verwerpen van een juiste nulhypothese ofwel ten onrechte concluderen dat een behandeling in een bepaalde subgroep effectief is). Dergelijke fout-positieve resultaten kunnen leiden tot onjuiste aanbevelingen ten aanzien van bijvoorbeeld de indicatiestelling van een bepaalde behandeling.<sup>16</sup>

Subgroepenanalyses op grond van achteraf geformuleerde, door de bevindingen gestuurde, hypothesen dienen achterwege te blijven of duidelijk als tentatief ('hypothese genererend') te worden geduid, omdat ook in

TABEL 2. Uitwerking van het combineren van risicoverschillen (RV) via het 'random effects'-model van tabel 1 (fictieve gegevens)

onderzoek	nieuwe behandeling			gangbare behandeling			$y_i = RV_i$	$s_i^2$	$w_i = 1/s_i^2$	$w_i^2$	$w_i' = 1/(s_i^2 + \tau^2)$	$RV_i \times w_i'$
	overleden	in leven	totaal	overleden	in leven	totaal						
1	1	24	25	5	20	25	-0,16	0,0079	126,01	15 878	119,29	-19,09
2	5	95	100	15	85	100	-0,10	0,0018	571,43	326 530	455,20	-45,52
3	25	475	500	50	450	500	-0,05	0,0003	3 636,36	13 223 140	1 385,37	-69,27
$\Sigma$									4 333,80	13 565 549	1 959,87	-133,08

$$Q = 2,54 > K - 1 = 2$$

$$\tau^2 = (Q - (K - 1)) / (\Sigma w_i - (\Sigma w_i^2 / \Sigma w_i)) = (2,54 - (3 - 1)) / (4333,80 - 13.565.549 / 4333,80) = 0,000449$$

gecombineerd risicoverschil (95%-betrouwbaarheidsinterval):

$$RV = \Sigma y_i w_i' / \Sigma w_i' \pm 1,96 / \sqrt{\Sigma w_i'} = -133,08 / 1959,87 \pm 1,96 / \sqrt{1959,87} = -0,07 (-0,11 - -0,03)$$

$$\text{toets voor } H_0: \theta = 0: U' = (\Sigma y_i w_i')^2 / \Sigma w_i' = (-133,08)^2 / 1959,87 = 9,14 \text{ (1 vrijheidsgraad; } p = 0,003)$$



deze situatie weer een aanmerkelijke kans is op fout-positieve resultaten.<sup>10-17</sup> Voor een uitvoerige behandeling van de richtlijnen voor het op een juiste wijze uitvoeren en interpreteren van subgroepanalyses wordt naar anderen verwezen.<sup>13-15</sup>

Een praktisch probleem bij het doen van subgroepanalyses is dat de benodigde gegevens vaak niet in de publicaties voorhanden zijn. Men zal dan pogingen in het werk moeten stellen om de ontbrekende gegevens alsnog van de oorspronkelijke auteurs te krijgen. Een recente ontwikkeling is het uitvoeren van meta-analyses aan de hand van (een selectie uit) de oorspronkelijke gegevensbestanden van de afzonderlijke onderzoeken.<sup>18</sup> Dergelijke meta-analyses, die veel tact en geduld vergen en uiterst arbeidsintensief zijn, zijn uiteraard bij uitstek geschikt voor het uitvoeren van gerichte subgroepanalyses in het kader van een meta-analyse.

### beschouwing

Onderzoek van (bronnen van) heterogeniteit is wellicht het belangrijkste onderdeel van meta-analyse. Zoals eerder opgemerkt, is er in een meta-analyse altijd wel enige heterogeniteit. Dit geldt in het bijzonder voor meta-analyse van niet-experimentele onderzoeken (cohort- en patiënt-controleonderzoek). De methodologie van meta-analyse van niet-experimenteel onderzoek staat echter in menig opzicht nog in de kinderschoenen en is momenteel sterk in ontwikkeling.<sup>19-23</sup> In experimenteel onderzoek ('randomized clinical trials') is de interne validiteit van de onderzoeken in principe gewaarborgd door de randomisatie. De variatie tussen niet-experimentele onderzoeken kan echter aanzienlijk zijn, bijvoorbeeld in de manier waarop en de mate waarin in de analyse gecorrigeerd is voor potentieel verstoringe variabelen ('confounders'). Bovendien bestaan doorgaans, evenals bij experimenteel onderzoek, grote verschillen tussen de onderzoeken ten aanzien van de onderzoekspopulatie, de wijze waarop blootstelling en uitkomsten gemeten zijn, en de follow-upduur. In sommige gevallen is het mogelijk verschillen tussen onderzoeken te elimineren door correcties uit te voeren op de gepresenteerde resultaten.<sup>22</sup> Voor een uitvoerige beschrijving van meta-analyse van niet-experimentele onderzoeken verwijzen wij naar andere literatuur.<sup>22-26</sup>

Met meta-analyse heeft men een krachtig instrument in handen om alle tot dan toe beschikbare evidentie op een precieze en overzichtelijke wijze samen te vatten. Men dient zich er echter te allen tijde van te vergewissen dat de patiënten, de centrale determinant (interventie, blootstelling) en de uitkomsten in de te combineren onderzoeken voldoende vergelijkbaar zijn. Mocht dit niet het geval zijn en is er derhalve sprake van heterogeniteit, dan ligt het voor de hand juist gebruik te maken van die heterogeniteit door aparte analyses uit te voeren om klinisch relevante subgroepen te identificeren. Dit geldt uiteraard ook in geval van homogene onderzoeksresultaten. Door subgroepanalyses kan men nagaan of bepaalde categorieën patiënten wellicht meer baat hebben bij de onderzochte interventie of groter risico lopen bij blootstelling aan een bepaalde factor. De

in opkomst zijnde methode van meta-analyse die gebaseerd is op individuele patiëntgegevens biedt in dit kader veel perspectieven.<sup>18</sup> De kans op slagen van een dergelijke benadering zal afhangen van de, bij voorkeur vooraf uitgesproken, bereidheid van onderzoekers hun gegevens hiervoor beschikbaar te stellen.

Een zorgvuldig en weloverwogen toegepaste meta-analyse vat de beschikbare kennis op een controleerbare en reproduceerbare wijze samen en maakt deze op een efficiënte wijze toegankelijk. Meta-analyse kan daardoor een belangrijke bijdrage leveren aan de individuele patiëntenzorg en kan dienen als beleidsondersteuning.

---

### abstract

*The practice of systematic reviews. V. Heterogeneity between studies and subgroup analyses*

– Meta-analysis comprises the analysis of the data of similar studies gathered in a systematic review.

– If the results of the studies differ (strongly), this is called heterogeneity.

– Possible causes of heterogeneity are: a wrong choice of type of measure of treatment effect, differences of methodological quality between studies, or real differences between studies.

– Heterogeneity between studies can be assessed by thorough examination of the differences between study characteristics in combination with a visual inspection of the degree of overlap of the confidence intervals of the estimates of effect of the different studies.

– By applying the random effects model for pooling, (non-systematic) heterogeneity between studies can be addressed.

– In case of heterogeneity between studies, sources of heterogeneity should be explored by means of subgroup analysis according to strict, preset criteria.

– Subgroup analysis should address subgroups of patients within studies.

---

### literatuur

<sup>1</sup> Assendelft WJJ, Tulder MW van, Scholten RJPM, Bouter LM. De praktijk van systematische reviews. II. Zoeken en selecteren van literatuur. Ned Tijdschr Geneesk 1999;143:656-61.

<sup>2</sup> Assendelft WJJ, Scholten RJPM, Eijk JThM van, Bouter LM. De praktijk van systematische reviews. III. Methodologische beoordeling van onderzoeken. Ned Tijdschr Geneesk 1999;143:714-9.

<sup>3</sup> Scholten RJPM, Kostense PJ, Assendelft WJJ, Bouter LM. De praktijk van systematische reviews. IV. Het combineren van de resultaten van afzonderlijke onderzoeken. Ned Tijdschr Geneesk 1999; 143:786-91.

<sup>4</sup> Greenland S, Rothman KJ. Introduction to stratified analysis. In: Rothman KJ, Greenland S. Modern epidemiology. Philadelphia: Lippincott-Raven; 1998. p. 253-79.

<sup>5</sup> 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.

<sup>6</sup> Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994;309:1351-5.

<sup>7</sup> Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. Stat Med 1991;10:1665-77.

<sup>8</sup> Fleiss JL. The statistical basis of meta-analysis. Stat Methods Med Res 1993;2:121-45.

<sup>9</sup> Walker AM, Martin-Moreno JM, Artalejo FR. Odd man out: a graphical approach to meta-analysis. Am J Public Health 1988;78: 961-6.

<sup>10</sup> Lau J, Ioannidis JPA, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997;127:820-6.

- 11 DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
- 12 Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989;8:141-51.
- 13 Mulrow CD, Oxman AD, editors. *The Cochrane Collaboration handbook*. Version 3.0.2 [compiled September 1997]. The Cochrane Collaboration. The Cochrane Library [database on CDROM]. Oxford: Update Software [updated quarterly].
- 14 Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991;266:93-8.
- 15 Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78-84.
- 16 Antiplatelet Trialists' Collaboration. Secondary prevention of vascular disease by prolonged antiplatelet treatment. *Br Med J* 1988;296:320-31.
- 17 Peto R. Statistical aspects of cancer trials. In: Halnan KE, editor. *Treatment of cancer*. Londen: Chapman and Hall; 1982.
- 18 Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Cochrane Working Group. Stat Med* 1995;14:2057-79.
- 19 Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994;140:771-8.
- 20 Greenland S. Can meta-analysis be salvaged? *Am J Epidemiol* 1994;140:783-7.
- 21 Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 1995;48:71-9.
- 22 Greenland S. Meta-analysis. In: Rothman KJ, Greenland S. *Modern epidemiology*. Philadelphia: Lippincott-Raven; 1998. p. 643-73.
- 23 Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998;316:140-4.
- 24 Eddy DM, Hasselblad V, Shachter R. Meta-analysis by the confidence profile method. *The statistical synthesis of evidence*. Boston: Academic Press; 1992.
- 25 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.
- 26 Hollander AEM de, Preller EA, Heisterkamp SH. Meta-analyse van observationeel onderzoek. Rapportnummer 263610002. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu; 1996.

Aanvaard op 17 november 1998

## Voor de praktijk

# Fysische diagnostiek – de spierrekkingsreflexen

j.stam

Het testen van de spierrekkingsreflexen is een onmisbaar onderdeel van het neurologisch onderzoek. Afwijkingen van deze reflexen zijn vooral van belang bij het anatomisch lokaliseren van afwijkingen in het zenuwstelsel, een belangrijke tussenstap in de neurologische diagnostiek.<sup>1</sup>

## anatomie en pathofysiologie

Door een tikje met de reflexhamer op de juiste plaats ontstaat een kortdurende rekking van de in de spier gelegen receptoren (spierspoelen), die daardoor actiepotentialen genereren in de zenuw die de spier innerveert. Via de plexussen en de dorsale wortels bereikt dit signaal het ruggenmerg. Daar maken de zogenaamde afferente Ia-zenuwvezels monosynaptisch contact met de  $\alpha$ -motoneuronen van dezelfde spier. Indien er voldoende actiepotentialen tegelijk aankomen, wordt een deel van deze motoneuronen gedepolariseerd en wordt er een signaal teruggestuurd naar de spier, via de ventrale wortels, de plexus en de zenuw. De efferente actiepotentialen veroorzaken via de neuromusculaire synaps depolarisatie van de spiercelmembraan, hetgeen leidt tot contractie van de spiervezels. Op deze wijze veroorzaakt een kortdurende rek van de spier een reflectoire verkorting van dezelfde spier (figuur).

De spierrekkingsreflex die met de reflexhamer wordt opgewekt is een fysiologisch artefact, veroorzaakt door

## samenvatting

– Het onderzoek van de spierrekkingsreflexen kan helpen bij het lokaliseren van een afwijking in het zenuwstelsel, een belangrijke tussenstap bij de diagnostiek van neurologische aandoeningen.

– Een probleem bij de interpretatie van het onderzoek is de grote intra- en interindividuele variatie in reflexhoogte. Onderzoek naar de overeenstemming tussen verschillende waarnemers geeft tegenstrijdige resultaten.

– Zowel bij de diagnostiek van perifere als van centrale aandoeningen is de specificiteit van duidelijke reflexafwijkingen redelijk tot goed (70-95%, afnemend met de leeftijd) en is de sensitiviteit duidelijk lager (50-70%).

– Derhalve geeft het vinden van een duidelijke reflexafwijking vaak nuttige diagnostische informatie, terwijl aan het ontbreken van afwijkingen minder waarde kan worden gehecht.

een niet-fysiologische stimulus. Het systeem dat met de reflexhamer wordt gestimuleerd heeft echter wel degelijk een belangrijke functie, namelijk die van lengtestabilisatie van de spieren. Passieve verlenging door een uitwendige verstoring leidt reflectoir tot verkorting van de spier, waardoor veel regelwerk dat nodig is om een bepaalde houding te handhaven op ruggenmergniveau kan worden afgehandeld.

Bij het klinische onderzoek gaat het erom vast te stellen of een spierrekkingsreflex relatief te sterk is ('levendig') of te zwak of zelfs afwezig. Een te zwakke respons is te verwachten wanneer de signaaltransmissie ergens in de reflexboog gestoord is, dus bij een gestoorde geleiding in de zenuw, de plexus of de dorsale of ventrale

Academisch Medisch Centrum/Universiteit van Amsterdam, afd. Neurologie, Postbus 22.700, 1100 DE Amsterdam.  
Prof.dr.J.Stam, neuroloog.