
De praktijk van systematische reviews. IV. Het combineren van de resultaten van afzonderlijke onderzoeken

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

In twee vorige artikelen uit deze reeks over systematische reviews kwamen het zoeken en het selecteren van publicaties en het vaststellen van de methodologische kwaliteit ervan uitvoerig aan de orde.^{1,2} In dit artikel geven wij een praktische handleiding voor het combineren van de resultaten van de afzonderlijke onderzoeken.

Indien de binnen een systematische review getraceerde onderzoeken voldoende vergelijkbaar (homogeen) zijn en van een aanvaardbare methodologische kwaliteit, kan men de resultaten van de afzonderlijke onderzoeken combineren tot één samenvattende schatting ('pooling') van het bestudeerde effect.^{3,4} De op deze wijze verkregen schatting van het effect biedt een objectieve samenvatting van alle tot dan toe beschikbare empirische evidentie. Aangezien deze schatting is gebaseerd op een groter aantal waarnemingen, is deze preciezer dan de schattingen verkregen in de afzonderlijke onderzoeken. Bovendien maakt het grotere aantal waarnemingen subgroepanalyses mogelijk die binnen de afzonderlijke onderzoeken wellicht niet uitvoerbaar waren. Aan de hand van dergelijke subgroepanalyses kunnen mogelijk verklaringen gevonden worden voor uiteenlopende (heterogene) resultaten van de verschillende onderzoeken. Combineren van resultaten en analyse van verschillen tussen onderzoeken worden tezamen meta-analyse genoemd.⁵ Een meta-analyse behelst derhalve de analyse van de gegevens die in het kader van een systematische review zijn vergaard.

Subgroepanalyses en de analyse van heterogeniteit tussen onderzoeken zijn onderwerp van het volgende artikel in deze reeks. In dit artikel gaan wij in op het combineren van de in de afzonderlijke onderzoeken verkregen resultaten (pooling). Allereerst geven wij een algemeen statistisch model voor pooling, dat bruikbaar is voor het combineren van alle mogelijke effectmaten. Vervolgens komen verschillende toepassingen van het algemene model aan de orde, die geïllustreerd worden aan de hand van enkele uitgewerkte voorbeelden. Daarna volgen enkele suggesties voor onderzoek naar de robuustheid van de verkregen samenvattende effect-schatting (sensitiviteitsanalyses). Wij besluiten met enkele algemene opmerkingen met betrekking tot poolen.

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

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

samenvatting

- Een meta-analyse behelst de analyse van de gegevens die in het kader van een systematische review zijn vergaard.
- Meta-analyse is het samenvoegen van de resultaten van vergelijkbare onderzoeken tot één samenvattende schatting van het in deze onderzoeken bestudeerde effect, in combinatie met de analyse van de invloed van verschillen tussen de onderzoeken op deze schatting.
- Voor het combineren van effectschattingen ('pooling') is een universeel toepasbaar statistisch model beschikbaar.
- Met behulp van pooling wordt een schatting van het effect van een interventie of blootstelling verkregen, die gebaseerd is op het grootst mogelijke aantal patiënten waardoor een zo groot mogelijke precisie verkregen wordt (smalle betrouwbaarheidsintervallen), zodat ook kleine effecten gedetecteerd kunnen worden.
- De methode van Mantel-Haenszel is eveneens een uitstekende methode voor pooling van effectmaten van dichotome uitkomsten.
- Door middel van pooling kan de beschikbare kennis op een efficiënte wijze worden samengevat.

combineren van resultaten

Het doel van poolen is het verkrijgen van een samenvattende schatting van het effect van bijvoorbeeld een interventie of van blootstelling aan een bepaalde risicofactor ten opzichte van het effect van een andere interventie of van de afwezigheid van blootstelling. Per onderzoek is het resultaat in de groep met de bestudeerde factor (indexgroep) in relatie gebracht met het resultaat in de referentiegroep door het berekenen van een effectmaat. Voorbeelden van effectmaten zijn het risicoverschil (RV), het relatieve risico (RR) en de oddsratio (OR). De schattingen van de effecten in de afzonderlijke onderzoeken worden vervolgens gecombineerd tot één samenvattende schatting via een gestratificeerde analyse. Op deze wijze wordt gecorrigeerd voor verschillen tussen onderzoeken die als potentieel verstorende variabele kunnen optreden, zoals verschillende achtergrondrisico's (de kans op de bestudeerde uitkomst in de referentiegroep).⁶ Voor gestratificeerde analyse zijn meerdere methoden beschikbaar.⁷⁻¹¹ Het principe van al deze methoden is dat het effect geschat wordt aan de hand van een gewogen gemiddelde van de effectschattingen van de afzonderlijke onderzoeken.

een algemeen statistisch model voor pooling
Uitgangspunt. In tabel 1 wordt een universeel toepasbaar statistisch model gepresenteerd voor het combineren van de in de afzonderlijke onderzoeken geschatte effecten.^{8,9,11} Uitgangspunt van dit model is dat de afzonderlijke onderzoeken schattingen geven van één vas-

te onderliggende waarde van het effect en dat de verschillen in uitkomsten tussen de onderzoeken louter op toeval berusten ('fixed effects'-model).

Voorwaarde voor toepassing van dit model is dat de beschouwde schatters van het effect (bij benadering) normaal verdeeld zijn. Dit geldt bijvoorbeeld voor het risicoverschil. Bij niet normaal verdeelde schatters (zoals die voor het relatieve risico en de oddsratio) kan (bij benadering) normaliteit bewerkstelligd worden door middel van een logaritmische transformatie. Hiertoe wordt de natuurlijke logaritme (ln) genomen (de logaritme met als grondtal e). De verdere berekeningen worden dan gedaan met de getransformeerde parameters. Tot slot wordt het uiteindelijke resultaat weer teruggetransformeerd naar de oorspronkelijke schaal.

Puntschatting en betrouwbaarheidsinterval. Stel dat ons een aantal onderzoeken ter beschikking staan die alle hetzelfde type effectmaat rapporteren (bijvoorbeeld het risicoverschil) of waarbij deze effectmaat telkens uit de gepresenteerde resultaten is af te leiden. Uit ieder onderzoek wordt de schatting van het effect, zo nodig na log-transformatie, genomen en de variantie daarvan. Daarna wordt voor ieder onderzoek een wegingsfactor berekend. Als wegingsfactor wordt de inverse van de variantie van het binnen dat onderzoek berekende effect genomen. Elk type effectmaat heeft zijn eigen variantie. De gecombineerde schatting van het effect wordt vervolgens berekend aan de hand van een gewogen gemiddelde van de afzonderlijke effectschattingen (zie tabel 1). Hoe groter het onderzoek, des te kleiner de variantie en derhalve des te groter het gewicht. Voor het kwan-

tificeren van de precisie van de gecombineerde schatting van het effect wordt een 95%-betrouwbaarheidsinterval berekend (zie tabel 1). Hoe nauwer het interval, des te preciezer (betrouwbaarder) de schatting.

Toetsen. Voor het toetsen van de nulhypothese 'de uitkomsten in de index- en de referentiegroepen zijn gelijk' wordt de toetsingsgrootte U berekend (zie tabel 1). Deze toetsingsgrootte volgt een χ^2 -verdeling met 1 vrijheidsgraad. De bijbehorende overschrijdingskans (p-waarde) wordt opgezocht in een tabel met χ^2 -verdelingen of berekend door middel van een computerprogramma. Bij lage p-waarden (gewoonlijk wordt 0,05 als afkappunt genomen) wordt de nulhypothese van gelijke uitkomsten verworpen. De nulhypothese kan uiteraard ook worden getoetst aan de hand van het eerder berekende betrouwbaarheidsinterval. Bevat dit interval niet de neutrale waarde (0 voor het risicoverschil en 1 voor het relatieve risico en de oddsratio), dan wordt de nulhypothese verworpen.

Het combineren van de effectschattingen van de afzonderlijke onderzoeken is alleen zinvol als men mag veronderstellen dat deze ongeveer gelijk zijn (ofwel homogeen). Voor het toetsen van de nulhypothese 'de effecten van alle onderzoeken zijn gelijk' wordt de toetsingsgrootte Q berekend (zie tabel 1). Q volgt een χ^2 -verdeling met (aantal onderzoeken - 1) vrijheidsgraden. Bij een lage p-waarde ($p < 0,05$) wordt de nulhypothese van homogeniteit verworpen, hetgeen impliceert dat de resultaten van de afzonderlijke onderzoeken dermate uiteenlopen dat van heterogeniteit gesproken kan worden. De waarde van de toets voor homogeniteit is echter beperkt. Hierop wordt in het volgende artikel uit deze reeks ingegaan.

toepassingen en uitgewerkte voorbeelden

Het hier geschetste statistische model is universeel toepasbaar. De benodigde ingrediënten zijn telkens de afzonderlijke schattingen van het effect met bijbehorende wegingsfactoren.

In tabel 2 staan de formules vermeld voor de verschillende effectmaten die van toepassing zijn bij onderzoeken met dichotome uitkomstmaten. Dit zijn uitkomstmaten die niet meer dan 2 waarden kunnen aannemen (bijvoorbeeld wel of niet genezen, wel of niet overleden et cetera). Het risicoverschil voldoet (bij benadering) aan de voorwaarde van normaliteit en kan als zodanig in het model worden opgenomen. De berekeningen voor het relatieve risico en de oddsratio worden uitgevoerd op de log-getransformeerde waarden daarvan.

In tabel 3 staan uitgewerkte voorbeelden van het combineren van effectmaten behorende bij dichotome uitkomsten (relatief risico, oddsratio en risicoverschil). Bij het risicoverschil wordt tevens een afgeleide effectmaat geïntroduceerd, te weten het 'number needed to treat' (NNT).^{12 13} Het NNT geeft aan hoeveel patiënten met de beschouwde interventie behandeld moeten worden om 1 behandelingsucces meer te verkrijgen dan met de controlebehandeling mogelijk zou zijn geweest. Het NNT wordt berekend als de inverse van het risicover-

TABEL 1. Een algemeen statistisch model voor het combineren van vergelijkbare, (bij benadering) normaal verdeelde schattingen van het effect uit een aantal onderzoeken

gecombineerde schatting van het effect met 95%-betrouwbaarheidsinterval:

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

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

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

toets voor homogeniteit van de effecten ($H_0: \theta_1 = \dots = \theta_K$):

$$Q = \sum w_i (y_i - y)^2 = \sum y_i^2 w_i - U$$

Σ = som

$i = 1, \dots, K$

K = aantal onderzoeken

θ = werkelijke waarde van het effect

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

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

w_i = wegingsfactor voor het i -de onderzoek = $1/s_i^2$

U = de toetsingsgrootte voor het toetsen van de nulhypothese dat het werkelijke effect nul is; deze volgt een χ^2 -verdeling met 1 vrijheidsgraad

Q = de toetsingsgrootte voor het toetsen van de nulhypothese dat de effecten van alle onderzoeken gelijk zijn; deze volgt een χ^2 -verdeling met $K - 1$ vrijheidsgraden

TABEL 2. Punt-schatting en variantie voor verschillende effectmaten van dichotome uitkomstvariabelen voor toepassing in het algemene statistische model van tabel 1

notatie van een 2x2-tabel	uitkomst		
	ongunstig	gunstig	totaal
indexgroep	a	b	a + b
referentiegroep	c	d	c + d
totaal	a + c	b + d	n

effectmaten (y) en bijbehorende variantie (s²)

relatief risico (RR):

$$y = \ln[RR] = \ln[(a/(a+b))/(c/(c+d))]$$

$$s^2 = 1/a - 1/(a+b) + 1/c - 1/(c+d)$$

oddsratio (OR):

$$y = \ln[OR] = \ln[ad/bc]$$

$$s^2 = 1/a + 1/b + 1/c + 1/d$$

risicoverschil (RV):

$$y = RV = a/(a+b) - c/(c+d)$$

$$s^2 = ab/(a+b)^2 + cd/(c+d)^2$$

schil (1/RV), waarbij het minteken wordt genegeerd. Voor het afleiden van het NNT uit schattingen van het relatieve risico en de oddsratio zijn eenvoudige tabellen beschikbaar.¹³

Het voorbeeld uit tabel 3 betreft fictieve gegevens en zou kunnen slaan op een meta-analyse van drie gerandomiseerde onderzoeken waarin het effect van een nieuwe behandeling bij patiënten met een bepaalde vorm van kanker vergeleken wordt met dat van een gangbare behandeling. De verschillende effectmaten zijn berekend voor de ongunstige uitkomst (overlijden). Zoals eerder opgemerkt, worden relatieve risico's en oddsratio's gecombineerd via een (natuurlijke) logtransformatie. Het verkregen resultaat wordt daarna teruggetransformeerd teneinde het oorspronkelijke type effectmaat weer te verkrijgen.

De schattingen van de relatieve risico's voor overlijden van de afzonderlijke onderzoeken lopen uiteen van 0,20 tot 0,50. De gecombineerde schatting bedraagt 0,45; patiënten die de nieuwe behandeling ondergingen, hebben 0,45 keer zoveel kans om te overlijden als patiënten die de gangbare behandeling kregen. De nieuwe behandeling 'beschermt' derhalve tegen overlijden. Het 95%-betrouwbaarheidsinterval van dit relatieve risico reikt van 0,30 tot 0,68 (door de noodzakelijke logtransformatie en terugtransformatie is dit betrouwbaarheidsinterval niet symmetrisch als wij naar de verschillen kijken, maar wel als wij naar de verhoudingen kijken). De waarde voor de toetsingsgrootte U is 14,63 (1 vrijheidsgraad; p < 0,001). Daar het 95%-betrouwbaarheidsinterval de neutrale waarde '1' niet omvat (of omdat de p-

waarde van de toetsingsgrootte U kleiner is dan 0,05), concluderen wij dat dit resultaat statistisch significant is. De toets voor homogeniteit van de relatieve risico's over de onderzoeken tenslotte levert een Q-waarde van 1,15 (2 vrijheidsgraden; p = 0,563; geen aanwijzingen voor heterogeniteit). Op grond van deze fictieve meta-analyse kan geconcludeerd worden dat de nieuwe behandeling een gunstiger effect lijkt te hebben op het ziektebeloop bij patiënten met de onderzochte vorm van kanker dan de gangbare behandeling.

In de figuur worden de resultaten van deze meta-analyse grafisch (op een logaritmische schaal) weergegeven. De horizontale lijnen vertegenwoordigen de puntschattingen en de 95%-betrouwbaarheidsintervallen van het relatieve risico van de afzonderlijke onderzoeken en het samenvattende resultaat. De verticale lijn geeft de neutrale waarde (in dit geval 1) aan. Men ziet in één oogopslag dat alle puntschattingen van het relatieve risico links van deze lijn liggen en derhalve een beschermend effect van de nieuwe behandeling vertegenwoordigen. Alleen het 95%-betrouwbaarheidsinterval van onderzoek 1 overschrijdt de neutrale lijn; het resultaat in dit onderzoek was als enige 'niet significant'. Het 95%-betrouwbaarheidsinterval van het samenvattende relatieve risico (de onderste horizontale lijn) is uiteraard het smalst aangezien dit op de gegevens van alle patiënten gebaseerd is.

Het combineren van oddsratio's en risicoverschillen levert vergelijkbare resultaten op (zie tabel 3). Het NNT bedraagt 16,7; er moeten gemiddeld (afgerond) 17 patiënten met de nieuwe therapie behandeld worden om 1 patiënt meer van de dood te redden dan met de gangbare behandeling mogelijk zou zijn geweest.

andere methoden voor pooling

Voor pooling van effectmaten van dichotome uitkomsten is ook de methode van Mantel-Haenszel beschikbaar.⁷⁻¹⁰ Deze klassieke methode is destijds ontworpen voor gestratificeerde analyse van oddsratio's en later uitgebreid voor gestratificeerde analyse van relatieve risico's en risicoverschillen.¹⁴ Het principe van de Mantel-Haenszel-methode is hetzelfde als dat van het algemene model. Er wordt een gewogen gemiddelde berekend van het effect, waarbij echter andere wegingsfactoren gebruikt worden. De berekening van de variantie van de gecombineerde schatting van het effect (en dus ook van het betrouwbaarheidsinterval) is erg bewerkelijk; voor deze methode zijn echter diverse softwarepakketten beschikbaar.¹⁵⁻¹⁶ De 'Statistical methods working group' van de Cochrane Collaboration acht de methode van Mantel-Haenszel de methode van eerste keus voor pooling van effectmaten van dichotome uitkomsten.³ De eerder geschetste algemene methode voor pooling voldoet echter ook uitstekend en in de praktijk zijn de verschillen tussen beide methoden verwaarloosbaar klein.

Voor het berekenen van een gecombineerde oddsratio is nog een derde, in de meta-analyseliteratuur tot nu toe vaak toegepaste, methode voorhanden: de methode van Peto.¹⁷ Deze is ook gebaseerd op het algemene sta-

TABEL 3. Uitwerking van het combineren van verschillende typen effectmaten voor dichotome uitkomsten via het algemene statistische model uit tabel 1 (fictieve gegevens)

relatief risico (RR)

onderzoek	nieuwe behandeling			oude behandeling			RR _i (95%-BI)	y _i = ln(RR _i)	s _i ²	w _i = 1/s _i ²	ln(RR _i) × w _i	ln(RR _i) ² × w _i
	overleden	in leven	totaal	overleden	in leven	totaal						
1	1	24	25	5	20	25	0,20 (0,03-1,59)	-1,61	1,12	0,89	-1,44	2,31
2	5	95	100	15	85	100	0,33 (0,13-0,88)	-1,10	0,25	4,05	-4,45	4,89
3	25	475	500	50	450	500	0,50 (0,31-0,80)	-0,69	0,06	17,86	-12,38	8,58
Σ										22,80	-18,27	15,79

gecombineerd:

$$RR = e^{(\Sigma y_i w_i / \Sigma w_i \pm 1,96 / \sqrt{\Sigma w_i})} = e^{(-18,27 / 22,80 \pm 1,96 / \sqrt{22,80})} = 0,45 (0,30-0,68)$$

toets voor H₀: θ = 0: U = (Σy_iw_i)²/Σw_i = (-18,27)²/22,80 = 14,63 (1 vrijheidsgraad; p < 0,001)

toets voor H₀: θ₁ = θ₂ = θ₃: Q = Σy_i²w_i - U = 15,79 - 14,63 = 1,15 (2 vrijheidsgraden; p = 0,563)

oddsratio (OR)

onderzoek	nieuwe behandeling			oude behandeling			OR _i (95%-BI)	y _i = ln(OR _i)	s _i ²	w _i = 1/s _i ²	ln(OR _i) × w _i	ln(OR _i) ² × w _i
	overleden	in leven	totaal	overleden	in leven	totaal						
1	1	24	25	5	20	25	0,17 (0,02-1,55)	-1,79	1,29	0,77	-1,39	2,49
2	5	95	100	15	85	100	0,30 (0,10-0,86)	-1,21	0,29	3,46	-4,19	5,07
3	25	475	500	50	450	500	0,47 (0,29-0,78)	-0,75	0,06	15,55	-11,62	8,68
Σ										19,78	-17,19	16,23

gecombineerd:

$$OR = e^{(\Sigma y_i w_i / \Sigma w_i \pm 1,96 / \sqrt{\Sigma w_i})} = e^{(-17,19 / 19,78 \pm 1,96 / \sqrt{19,78})} = 0,42 (0,27-0,65)$$

toets voor H₀: θ = 0: U = (Σy_iw_i)²/Σw_i = (-17,19)²/19,78 = 14,94 (1 vrijheidsgraad; p < 0,001)

toets voor H₀: θ₁ = θ₂ = θ₃: Q = Σy_i²w_i - U = 16,23 - 14,94 = 1,29 (2 vrijheidsgraden; p = 0,525)

risicoverschil (RV)

onderzoek	nieuwe behandeling			oude behandeling			RV _i (95%-BI)	y _i = RV _i	s _i ²	w _i = 1/s _i ²	RV _i × w _i	RV _i ² × w _i
	overleden	in leven	totaal	overleden	in leven	totaal						
1	1	24	25	5	20	25	-0,16 (-0,33-0,01)	-0,16	0,0079	126,01	-20,16	3,23
2	5	95	100	15	85	100	-0,10 (-0,18--0,02)	-0,10	0,0018	571,43	-57,14	5,71
3	25	475	500	50	450	500	-0,05 (-0,08--0,02)	-0,05	0,0003	3636,36	-181,82	9,09
Σ										4333,80	-259,12	18,03

gecombineerd:

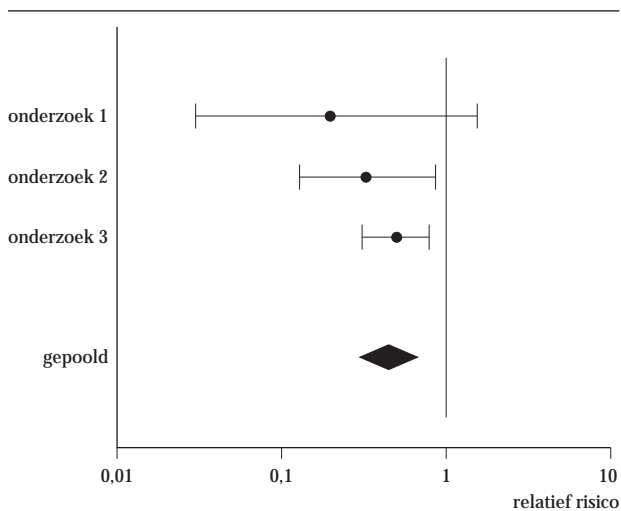
$$RV = (\Sigma y_i w_i / \Sigma w_i \pm 1,96 / \sqrt{\Sigma w_i}) = (-259,12 / 4333,80 \pm 1,96 / \sqrt{4333,80}) = -0,06 (-0,09--0,03)$$

toets voor H₀: θ = 0: U = (Σy_iw_i)²/Σw_i = (-259,12)²/4333,80 = 15,49 (1 vrijheidsgraad; p < 0,001)

toets voor H₀: θ₁ = θ₂ = θ₃: Q = Σy_i²w_i - U = 18,03 - 15,49 = 2,54 (2 vrijheidsgraden; p = 0,281)

'number needed to treat' (NNT) = 1/|RV| = 1/|-0,06| = 16,7 (11,2-33,3)

95%-BI = 95%-betrouwbaarheidsinterval.



Puntschatters en 95%-betrouwbaarheidsintervallen (op een logaritmische schaal) van het relatieve risico van 3 afzonderlijke fictieve onderzoeken en de samenvattende schatting.

tistische model, maar onderscheidt zich door een afwijkende (benaderende) berekening van de oddsratio. Dit laatste vormt direct het grootste nadeel van de Peto-methode: de waarde van de oddsratio wordt overschat als deze ver verwijderd ligt van de neutrale waarde 1 ($< 0,2$ of $> 5,0$).¹⁸ Vooralsnog lijken er geen goede redenen te zijn om de Peto-methode te gebruiken voor het berekenen van gecombineerde oddsratio's.^{9 18}

sensitiviteitsanalysen

Na de berekening van een samenvattend effect is het zinvol na te gaan hoe gevoelig de uitkomst is voor veranderingen in de insluitcriteria voor de meta-analyse.³ Als de waarde van de schatting hierdoor sterk verandert, dan dient het resultaat met grote voorzichtigheid geïnterpreteerd te worden.

Sensitiviteitsanalysen kunnen uitgevoerd worden aan de hand van de methodologische kwaliteit van de onderzoeken.¹⁻³ Men berekent hiertoe bijvoorbeeld het effect op grond van alleen onderzoeken die aan bepaalde methodologische criteria voldoen. Indien het aldus verkregen resultaat sterk verschilt van het oorspronkelijke, na pooling van alle onderzoeken verkregen resultaat, dan verdient het aanbeveling de methodologisch slechtere onderzoeken van de gecombineerde analyse uit te sluiten en de schatting van het effect louter te baseren op de kwalitatief betere. Door wijziging van de inclusiecriteria voor toelating van een onderzoek tot de gecombineerde analyse kunnen verdere sensitiviteitsanalysen uitgevoerd worden. Het gaat hier in feite om post-hocsubgroepanalysen, waarbij het doel is de robuustheid van de schatting van het effect te onderzoeken en niet het opsporen van subgroepen zelf. Verandert deze schatting onder invloed van dergelijke wijzigingen, dan dient men het resultaat met grote voorzichtigheid te presenteren of pooling zelfs achterwege te laten.

beschouwing

In het voorgaande gaven wij een algemene handleiding voor het combineren van de resultaten van afzonderlijke onderzoeken. Voor het uitvoeren van de statistische pooling kunnen de benodigde formules op eenvoudige wijze geprogrammeerd worden in een 'spreadsheet'. Ook kan men bestaande software gebruiken.^{16 19}

De hier geschetste methode van pooling is breed toepasbaar. Ze is niet alleen geschikt voor het combineren van de resultaten van experimentele onderzoeken ('randomized clinical trials'), maar ook van die van niet-experimentele onderzoeken (cohort- en patiënt-controle-onderzoek). De hier beschreven algemene methode kan ook gebruikt worden voor pooling van effectmaten van continue uitkomsten, bijvoorbeeld bloeddruk.²⁰ De details van de feitelijke berekeningen hiervan vallen buiten het bestek van dit artikel.

De voordelen van pooling zijn duidelijk: de conclusies zijn gebaseerd op alle tot dan toe beschikbare empirische evidentie, samengevat op een objectieve en reproduceerbare manier. De schatting van het effect van een interventie of blootstelling wordt gebaseerd op een accumulatie van patiëntenaantallen van verschillende onderzoeken, waardoor een grotere precisie wordt verkregen (smalle betrouwbaarheidsintervallen), zodat ook kleinere effecten gedetecteerd kunnen worden. Bovendien worden subgroepanalysen mogelijk, zodat bijvoorbeeld optimaal onderzocht kan worden of bepaalde categorieën patiënten wellicht meer baat hebben bij de onderzochte interventie of groter risico lopen bij blootstelling aan een bepaalde factor.

Er zijn echter ook kanttekeningen te plaatsen bij de pooling van resultaten. Poolen dient weldoordacht en weloverwogen te geschieden. Vaste richtlijnen hiervoor zijn niet voorhanden, maar duidelijk is dat hiervoor nuchter verstand en klinisch inzicht onontbeerlijk zijn.⁴ Men dient zich er in ieder geval van te vergewissen of 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 (klinische) heterogeniteit, dan ligt het voor de hand de heterogeniteit door middel van subgroepanalysen verder te exploreren. Heeft men slechts de beschikking over onderzoeken van slechte methodologische kwaliteit, dan geeft poolen ten onrechte de indruk van objectiviteit en precisie; in dergelijke gevallen dient men het achterwege te laten.

Een groot en vooralsnog onoplosbaar probleem bij pooling (en dus ook bij meta-analyse) is de mogelijkheid van de aanwezigheid van publicatiebias.^{21 22} In zo'n geval hebben onderzoeken met extreme ('positieve') resultaten, die wellicht voor een groot deel aan het toeval toe te schrijven zijn, een hogere kans om gepubliceerd te worden dan die met minder spectaculaire resultaten. Dit kan dan leiden tot een vertekening van de samenvattende effectschatting. De beste oplossing voor het probleem van publicatiebias zou bestaan uit de preventie ervan door het oprichten van een register waarin alle geïnitieerde onderzoeken opgenomen worden.²¹ Op deze wijze kan men in voorkomende gevallen achterhalen

wat de oorzaak is van het nooit in de medische literatuur verschijnen van de resultaten van sommige onderzoeken, of kunnen de resultaten van dergelijke onderzoeken (indien die beschikbaar zijn) alsnog in de pooling betrokken worden.

Door middel van pooling kan de beschikbare kennis op een efficiënte wijze worden samengevat. Voorwaarde is echter dat de te combineren onderzoeken voldoende vergelijkbaar (homogeen) zijn. Hoe te handelen bij heterogeniteit van onderzoeken is onderwerp van het volgende artikel in deze reeks.

abstract

The practice of systematic reviews. IV. Pooling of results of individual studies

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

– Meta-analysis can be defined as the combination of pooling results of similar studies into one overall estimate of the effect on the one hand and the exploration of sources of heterogeneity between studies on the other.

– A general statistical model is available for pooling the estimates of various types of effect parameters.

– Pooling enables the estimation of the effect of an intervention or exposure as precisely as possible (with narrow confidence intervals) which allows detection of minor effects.

– The Mantel-Haenszel method is also an excellent method for pooling effect parameters of dichotomous outcomes.

– Pooling provides an efficient means of summarizing the available knowledge.

literatuur

- 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.
- 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.
- 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].
- Lau J, Ioannidis JPA, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997;127:820-6.
- Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. Epidemiol Rev 1992;14:154-76.
- Simpson EH. The interpretation of interaction in contingency tables. J R Stat Soc (Br) 1951;13:238-41.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719-48.
- Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. Stat Med 1991;10:1665-77.
- Fleiss JL. The statistical basis of meta-analysis. Stat Methods Med Res 1993;2:121-45.
- Greenland S, Rothman KJ. Introduction to stratified analysis. In: Rothman KJ, Greenland S, editors. Modern epidemiology. Philadelphia: Lippincott-Raven; 1998. p. 253-79.
- Greenland S. Meta-analysis. In: Rothman KJ, Greenland S, editors. Modern epidemiology. Philadelphia: Lippincott-Raven; 1998. p. 643-73.
- Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988;318:1728-33.
- McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. Ann Intern Med 1997;126:712-20.

¹⁴ Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. Biometrics 1985;41:55-68.

¹⁵ Dean AG, Dean JA, Coulombier D, Brendel KA, Smith DC, Burton AH, et al. Epi Info. Version 6. Atlanta, Ga.: Centers for Disease Control and Prevention; 1994.

¹⁶ Review Manager (RevMan) [computer program]. Version 3.1 for Windows. Oxford, Engl.: The Cochrane Collaboration; 1998.

¹⁷ Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog Cardiovasc Dis 1985;27:335-71.

¹⁸ Greenland S, Salvan A. Bias in the one-step method for pooling study results. Stat Med 1990;9:247-52.

¹⁹ Normand SL. Meta-analysis software: a comparative review. Am Stat 1995;49:298-309.

²⁰ Shadish WR, Haddock CK. Combining estimates of effect size. In: Cooper H, Hedges L, editors. Handbook of research synthesis. New York: Russell Sage; 1994. p. 261-81.

²¹ Dickersin K. Why register clinical trials? – revisited. Control Clin Trials 1992;13:170-7.

²² Egger M, Smith GD. Bias in location and selection of studies. BMJ 1998;316:61-6.

Aanvaard op 17 december 1998

Bladvulling

Narcose breekt door

Vijftig jaren ruim genieten patiënten en operateurs van den chloroform en den aether; ons gevoel voor de lijders wordt ten volle bevredigd, de rust bij het opereeren ten volle gewaarborgd, waarmede een groote stoot aan de verbetering der techniek werd gegeven. Dat echter de gevaren van een vergif, dat in zoo weinige oogenblikken het centrale zenuwstelsel op zoo krasse wijze aanpakt, niet uitbleven, mag niemand verwonderen.

Die kennis werd niet zonder offers verkregen, en mag men nu na 50-jarige ervaring de gevaren te voren grootendeels voorzien en voorkomen, nog altijd is de narcose-dood, hoe zeldzaam ook, niet voor goed vermeden. Doch trooste ons de overtuiging, dat het aantal sterfgevallen over een gelijk aantal grotere operatiën zonder narcose bij in gelijke mate verzwakte patiënten door noodlottige reflexwerking veel groter zijn zou. In elk geval wordt de later verkregen kennis van slechts plaatselijk verdoovende middelen ten volle gewaardeerd.

(Ned Tijdschr Geneesk 1899;43I:575.)

Boontje komt om zijn loontje

De Ortskrankenkas heeft geen zijde gesponnen bij de proef om geneeskundigen tegen een vast inkomen aan te stellen. Zij had vijf artsen met een inkomen van 5000 Mk. aangesteld, maar toen haar de medewerking ontviel van de oude artsen, die zij nog niet aan den dijk had gezet, en de werklieden met het nieuwe stelsel geen genoegen namen, moest zij de nieuwe artsen ontslaan, doch hen schadeloos stellen voor den tijd waarover hun contract had geloopt. Nadat zij ruim 16000 Mk. betaald had, bleef zij nog ruim 36000 Mk. schuldig. De bevoegde overheid liet echter niet toe, het geld van het ziekenfonds hiervoor te gebruiken. De vijf artsen stelden zich nu tevreden met 22000 Mk., waartoe de Barmensche artsenvereniging 8000 Mk. wil bijdragen, mits de niet-onfortuinlijke 'blacklegs' de stad met 1 October zouden verlaten. Voor dit voorstel moet de toestemming der overheid worden afgewacht.

(Berichten Buitenland. Ned Tijdschr Geneesk 1899;43II:838.)