

RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEU  
BILTHOVEN

Rapport nr. 263610 002

**Meta-analyse van observationeel onderzoek**

Mogelijkheden en beperkingen bij toepassingen ten behoeve van het kwantificeren van gezondheidsrisico's

A.E.M. de Hollander, E.A. Preller, S.H. Heisterkamp,  
J. Jansen

oktober 1996

Dit onderzoek werd verricht in opdracht en ten laste van het Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer, Directoraat-Generaal Milieubeheer, alsmede het Ministerie van Volksgezondheid, Welzijn en Sport, de Inspectie voor de Gezondheidszorg, en de directie Gezondheidsbeleid, afd. Gezondheidsbevordering en Ziektepreventie.



**VERZENDLIJST**

- 1-5 Directie Gezondheidsbeleid, afdeling Gezondheidsbevordering en Ziektepreventie van het Ministerie van Volksgezondheid, Welzijn, en Sport
- 6 Directeur-Generaal van de Volksgezondheid
- 7 Plv. Directeur-Generaal van de Volksgezondheid
- 8-11 Directie Algemeen en Internationaal Volksgezondheidsbeleid van het Ministerie van Volksgezondheid, Welzijn en Sport
- 12-15 Directie Informatiebeleid van het Ministerie van Volksgezondheid, Welzijn en Sport
- 16-20 Directie Voeding en Veiligheid van Producten van het Ministerie van Volksgezondheid, Welzijn en Sport
- 21-25 Geneeskundige Hoofdinspectie van de Volksgezondheid
- 26 Geneeskundige Hoofdinspectie voor de Geestelijke Volksgezondheid
- 27 Drs. J.J.L. Pieters, Inspectie voor de Gezondheidszorg
- 28 Dr. K.R. Krijgsheld, VROM/DGM
- 29 Dr. J.H.M. Nieuwenhuis, Veterinaire Hoofdinspectie
- 30 Dr. T. Staarink, Hoofdinspectie Gezondheidsbescherming
- 31 Dr. J.A. van Zorge, VROM/DGM
- 32 Dr. C.J.M. van den Boogaard, Hoofdinspectie Milieuhygiëne
- 33 Mw. Ir. G. van 't Bosch, VWS/PAO
- 34 Mw. Drs. H. Emanuel-Vink, VWS, Staatstoezicht
- 35 Prof. dr. ir. B. Brunekreef, Vakgroep Humane Epidemiologie en Gezondheidsleer L.U., Wageningen
- 36 Dr. Th. Stijnen, Vakgroep epidemiologie, Erasmus Universiteit, Rotterdam.
- 37 Prof. dr. J.C. van Houwelingen, Rijksuniversiteit Leiden
- 38 Prof. dr. L.M. Bouter, EMGO-Instituut, Faculteit Geneeskunde, V.U., Amsterdam
- 39 Prof. dr. L.J. Gunning-Schepers, Instituut Sociale Geneeskunde, A.M.C., Amsterdam
- 40 Prof. dr. D.E. Grobbee, Vakgroep Epidemiologie, Erasmus Universiteit, Rotterdam
- 41 Prof. dr. A. Hofman, Vakgroep Epidemiologie, Erasmus Universiteit, Rotterdam
- 42 Dr. G. Jambroes, Medisch Directeur Nederlandse Hartstichting, Den Haag
- 43 Dr. J.J.P. Kastelein, Afdeling Haematologie, A.M.C., Amsterdam
- 44 Prof. dr. ir. F.J. Kok, Vakgroep Humane Epidemiologie en Gezondheidsleer, L.U., Wageningen
- 45 Dr. Ir. F. van Leeuwen, Nederlands Kanker Instituut, Amsterdam
- 46 Prof. dr. J.P. Mackenbach, Faculteit Geneeskunde, Erasmus Universiteit, Rotterdam
- 47 Prof. dr. W.R.F. Notten, directeur, T.N.O. Preventie en Gezondheid, Leiden

- 48 Dr. P.H.A. Peeters, Vakgroep Algemene Gezondheidszorg en Epidemiologie, R.U. Utrecht
- 49 Prof. dr. Ph. Quanjer, Laboratorium Fysiologie, R.U. Leiden
- 50 Prof.dr. J.P. Vandenbroucke, Afd. Klinische Epidemiologie, Acad Ziekenhuis Leiden
- 51 Dr. W.F. Passchier, Gezondheidsraad, Den Haag
- 52 Ir P. de Mik, Sector Gezondheid en Maatschappelijk Welzijn, Centraal Bureau voor de Statistiek, Voorburg
- 53 Depot van Nederlandse Publikaties en Nederlandse Bibliografie
- 54 Directie RIVM
- 55 Prof. ir. N.D. van Egmond, directeur Milieu, RIVM
- 56 Prof. dr. G Elzinga, directeur Volksgezondheid, RIVM
- 57 Prof. dr. ir. D. Kromhout, directeur Sector Volksgezondheidsonderzoek, RIVM
- 58 Ir. F. Langeweg, directeur Sector Milieu-onderzoek, RIVM
- 59 Dr. ir. G. de Mik, directeur Sector Stoffen en Risico's, RIVM
- 60-68 Laboratoriumhoofden Sector Volksgezondheidsonderzoek, RIVM
- 69 Dr. P.G.N. Kramers, Bureau Volksgezondheidstoekomstverkenning, RIVM
- 70 Drs. D. Ruwaard, Bureau Volksgezondheidstoekomstverkenning RIVM
- 71 Drs. A. van den Berg Jeths, Bureau Volksgezondheidstoekomstverkenning, RIVM
- 72 Dr. P.W. Achterberg, Bureau Volksgezondheidstoekomstverkenning, RIVM
- 73 Dr. H. Verkleij, Bureau Volksgezondheidstoekomstverkenning, RIVM
- 74 Ir. I.A.M. Maas, Bureau Volksgezondheidstoekomstverkenning, RIVM
- 75 Ir. W. van Duyvenbooden, Lab voor Bodem- en Grondwateronderzoek, RIVM
- 76 Dr. M.A.E. Conijn-van Spaendonck, Centrum Infectieziekten Epidemiologie, RIVM
- 77 Dr. H. van Loveren, Laboratorium voor Pathologie, RIVM
- 78 Dr. ir. B. Bloemberg, Centrum Chronische ziekten en Milieu Epidemiologie, RIVM
- 79 Dr. ir. E.J.M. Feskens, Centrum Chronische ziekten en Milieu Epidemiologie, RIVM
- 80 Dr. ir. E. Lebret, Centrum Chronische ziekten en Milieu Epidemiologie, RIVM
- 81 Ir. D.J.M. Houthuijs, Centrum Chronische ziekten en Milieu Epidemiologie, RIVM
- 82 Ir PH Fischer, Centrum Chronische ziekten en Milieu Epidemiologie, RIVM
- 83 Drs. C. de Lezenne Coulander, IMA, RIVM
- 84 Hoofd Bureau Voorlichting & Public Relations, RIVM
- 85-93 Auteurs
- 94-95 Bibliotheek RIVM
- 96 Bureau Rapportenregistratie
- 97-160 Reserve exemplaren t.b.v. Bureau Rapportenbeheer

**INHOUDSOPGAVE**

Verzendlijst	3
Inhoudsopgave	5
Summary	7
Samenvatting	9
Voorwoord	11
1. INLEIDING	13
2. META-ANALYSE	15
2.1. Wat is meta-analyse?	15
2.2. Waarom?	17
2.3. Inhoudelijke en praktische aspecten	18
3. RICHTLIJNEN VOOR META-ANALYSE	25
3.1. Onderzoeks- en begeleidingsgroep	25
3.2. Onderzoeksprotocol	26
3.3. Verzamelen en selecteren van afzonderlijke onderzoeken	27
3.4. Extractie van gegevens	30
3.5. Data-analyse	32
3.6. Presentatie, interpretatie en communicatie van resultaten.	34
4. STATISTISCHE TECHNIKEN BIJ META-ANALYSE	37
4.1. Inleiding	37
4.2. Praktische aanbeveling	38
4.3. Meta-analyse zonder extra informatie	39
4.4. Een uitgewerkt literatuurvoorbeeld	40
4.5. Meta-analyse met extra informatie: opnemen van studiekekenmerken	41
4.6. Ongelijke variantie tussen studies en afhankelijkheid tussen studie uitkomsten	44
4.7. Een uitgewerkt voorbeeld	44
5. SLOTOPMERKINGEN	47
6. LITERATUUR	49
APPENDIX I. NIET-ITERATIEVE EB SCHATTER EN ITERATIEVE EB- EN ML-SCHATTERS VOOR $\theta$ , BETROUWBAARHEIDS- EN TOLERANTIE-INTERVALLEN	53
APPENDIX II. BAYES, EB- EN ML-SCHATTERS VOOR $\mu$ EN TOLERANTIE-INTERVALLEN BIJ GEBRUIK VAN CO-VARIABLEN	55
APPENDIX III. SAS-PROCEDURES	58
APPENDIX IV. S-PLUS PROCEDURES	60



## SUMMARY

In this report the applicability of meta-analysis is evaluated as a tool for quantitative review of observational (non-experimental) studies involving risk factors and health effects. These quantitative reviews are required for the assessment of the effectiveness of preventive policies that will be performed in the framework of next Dutch Public Health Status Forecast reports.

The report describes briefly the 'art' of meta-analysis, recent methodological developments and the objectives of application. Subsequently the limitations and possibilities of meta-analysis of observational studies are evaluated. Based on this evaluation and review of the literature, guidelines are outlined for sound meta-analysis of observational studies. A statistical procedure for data aggregation is presented, providing possibilities for meta-regression analysis, taking relevant study characteristics into account, as well as possibilities for Bayesian updating of individual studies (posterior estimates) 'borrowing strength' from a set of other comparable study results.

The report concludes that meta-analysis can be a useful tool for systematic, consistent, comprehensive, and reproducible quantitative review of observational studies. However, the main objective should be critical, comparative analysis, learning from heterogeneity, identifying patterns among studies, not data aggregation in itself.





## SAMENVATTING

In dit rapport worden de toepassingsmogelijkheden van meta-analyse bekeken als hulpmiddel bij kwantitatieve review van observationeel (niet-experimenteel) onderzoek. Dit onderzoek betreft meestal de relatie tussen risicofactoren (milieuverontreiniging, leefstijl e.d.) en gezondheidseindpunten, soms ook de werkzaamheid van bepaalde vormen van gezondheidszorg. Deze kwantitatieve overzichten zijn noodzakelijk om tot schattingen te komen van de effecten van preventiebeleid, die onder meer in het kader van toekomstverkenningen voor de volksgezondheid of voor de toestand van het milieu worden gemaakt.

Het rapport beschrijft zowel het wat, hoe en waarom van meta-analyse, als recente methodologische ontwikkelingen. Daarnaast worden mogelijkheden en beperkingen van meta-analyse van observationeel onderzoek kritisch bekeken. Op basis hiervan en van literatuuronderzoek worden globale richtlijnen gegeven voor het uitvoeren van meta-analyses. In het rapport wordt een statistische procedure voor meta-analyse voorgesteld, die de mogelijkheid biedt om naast blootstelling en respons andere populatie- of studiekenmerken bij de analyse te betrekken ook wel analytische meta-analyse of 'meta-regressie-analyse' genoemd. Zo kan beter inzicht verkregen worden in de rol van kenmerken als leeftijd, sociaal-economische status, gevoelige groepen (bijvoorbeeld aan de hand van gezondheidsstatus), geografische locatie of type epidemiologisch onderzoek. Tevens biedt de procedure de mogelijkheid een 'Bayesiaanse' post-hoc herwaardering ('update') van afzonderlijke studieresultaten te maken, gegeven de uitkomsten van een verzameling vergelijkbaar uitgevoerde studies.

Het rapport concludeert dat meta-analyse een nuttig hulpmiddel kan zijn bij het maken van systematische, consistente, omvattende en navolgbare, kwantitatieve overzichten van observationeel onderzoek. Daarbij dient de nadruk te liggen op kritische vergelijking, het 'leren' van verschillen tussen studies en patroonherkenning, en niet zozeer op het afleiden van geaggregeerde effectmaten.



## VOORWOORD

Op verzoek van het Directoraat-Generaal Milieubeheer, Directie Lucht en Energie, is door een post-doc medewerker van de Nederlandse Onderzoeksschool voor Gezondheidswetenschappen (NIHES) onderzoek gedaan naar de toepassingsmogelijkheden van meta-analyse bij het kwantificeren van gezondheidsrisico's van milieuverontreiniging. De medewerkster, Mw. dr. ir. E.A. Preller, werd gedetacheerd bij het Centrum voor Chronische ziekte en Milieu-epidemiologie van het Rijksinstituut voor Volksgezondheid en Milieubeheer (CCM-RIVM).

Aan de hand van literatuurstudie en verschillende voorbeeld-onderzoeken zijn richtlijnen ontwikkeld voor het uitvoeren van meta-analyses van observationeel onderzoek naar de gezondheidseffecten van milieuverontreiniging. Hierbij is bijzondere aandacht besteed aan de specifieke problemen van toepassing bij observationeel epidemiologisch onderzoek (als niet-experimentele tegenvoeter van gerandomiseerd interventie-onderzoek). Tevens is een statistische procedure ontwikkeld om zo veel mogelijk kwantitatieve informatie te ontlenen aan het systematisch aggregeren van de resultaten van afzonderlijke onderzoeken. Deze beide producten worden in voorliggend rapport gerapporteerd. De voorbeeldstudies zullen afzonderlijk worden gerapporteerd.

Ten behoeve van inhoudelijke afstemming en ondersteuning werd besloten het onderzoek te integreren in de activiteiten van de RIVM werkgroep 'meta-analyse'. Deze werkgroep, bestaande uit de auteurs van dit rapport, brengt ten behoeve van de komende Volksgezondheids Toekomst Verkenning de mogelijkheden van meta-analyse in kaart als een van de instrumenten om de effecten van preventie en zorg te kwantificeren (Prepare-project, MAP Volksgezondheid, 266750). Ook bij dit onderzoek gaat het veelal om observationeel epidemiologisch onderzoek naar gezondheidseffecten van risicofactoren als bron van kwantitatieve informatie over aan bepaalde risicofactoren verbonden gezondheidsverlies.

De in dit rapport beschreven statistisch procedure voor meta-analyse is ontwikkeld door dr. S.H. Heisterkamp en drs. C. de Lezenne Coulander in nauw overleg met Mw. E.A. Preller (allen werkzaam bij het RIVM)

Dr. Th. Stijnen (Erasmus Universiteit) en prof. dr. ir. B Brunekreef (Landbouwniversiteit Wageningen) hebben als externe deskundigen geadviseerd over opzet en uitvoering van het onderzoek. Over de richting van het onderzoek is op gezette tijden overleg gevoerd in de begeleidingscommissie, bestaande uit dr. K.R. Krijgsheld (directie Lucht en Energie, DGM-VROM), dr C. van den Boogaard (Inspectie voor de Milieuhygiëne, DGM-VROM), drs J.J.L. Pieters, Inspectie voor de Gezondheidszorg, VWS, dr. ir. E.A. Preller (NIHES), ir. P.J.A. Rombout, dr. M. Oortgiessen, dr. ir. E. Lebret en drs. A.E.M. de Hollander (allen RIVM).

Bijzondere dank is verschuldigd aan prof. dr. J.C. van Houwelingen van de Rijksuniversiteit Leiden die als adviseur van de werkgroep 'Meta-analyse' een belangrijke inbreng heeft gehad



## 1. INLEIDING

Bij het beleid gericht op de bescherming van de gezondheid, of het nu tot het volksgezondheid of het milieubeleid wordt gerekend, is doelmatigheid een steeds belangrijker punt geworden. Dit betekent dat bij beslissingen over de inzet van middelen vaker de potentieel te behalen gezondheidswinst betrokken wordt, bij voorkeur in termen van vermeden ziektelast.

In het veld van de volksgezondheid gaat het dan om de effecten van preventie, zoals gezondheidsvoorlichting en opvoeding, of wet- en regelgeving. Belangrijke vragen daarbij zijn bijvoorbeeld: ‘wat is het met een leefstijlfactor als rookgedrag samenhangende gezondheidsverlies, uitgedrukt in verloren levensjaren of, nog liever, in verloren gezonde levensjaren?, en ‘met welke instrumenten, gericht op welke doelgroep wordt de meeste winst geboekt?’. Bij de effecten van zorg draait het er vaak om welk typen van geneeskundig handelen, bij welke patiënten, onder welke voorwaarden het meest doeltreffend zijn.

Ook in het domein van het milieubeheer bestaat een toenemende behoefte aan kwantificering van het gezondheidsverlies ten gevolge van blootstelling aan milieuverontreiniging. Zeker als risico-reducerende maatregelen gepaard gaan met hoge maatschappelijke kosten, zullen betrokken partijen de potentieel te behalen gezondheidswinst bij hun afwegingen willen betrekken. Het belang van resultaten van epidemiologisch onderzoek bij gezondheidskundige risico-analyse is geleidelijk toegenomen en daarmee de behoefte aan kwantitatieve methoden om de resultaten ervan samen te vatten.

Om de gezondheidswinst van maatregelen gericht op de vermindering van blootstelling aan gezondheidsschadelijke factoren in kaart te brengen, is kwantitatieve (epidemiologische) informatie onontbeerlijk, zoals over

- de verdeling van blootstelling aan risicofactoren in de populatie
- het verband tussen blootstelling aan een risicofactor en respons (bijvoorbeeld uitgedrukt als morbiditeits- of mortaliteitsrisico)
- het effect van interventie in termen van verandering van blootstelling aan risicofactoren in de populatie.

Deze informatie wordt in de regel ontleend aan verschillende soorten van epidemiologisch onderzoek, waaronder gezondheidsmonitoring, etiologisch cohort- of patiënt-controle-onderzoek, en (‘community’) interventie-onderzoek. Naast methodologische aspecten die te maken hebben met opzet en kwaliteit van het onderzoek, is hierbij ook aan de orde in hoeverre onderzoeksresultaten kunnen worden vertaald naar de Nederlandse doelpopulatie.

In het medische domein is meta-analyse van resultaten van afzonderlijke ‘gerandomiseerde interventie-onderzoeken’ (‘Randomized Clinical Trials’, RCTs) naar de werkzaamheid van geneeskundig handelen al behoorlijk ingeburgerd. In deze tussenrapportage worden de mogelijkheden en beperkingen van meta-analyse van observationeel (‘niet-experimenteel’) epidemiologisch onderzoek verkend als hulpmiddel bij het kwantitatief karakteriseren van het verband tussen blootstelling/prevalentie en respons (kwantitatieve review). Meta-analyse is daarbij gedefinieerd als het met behulp van statistische technieken combineren van kwantitatieve informatie uit afzonderlijke (epidemiologische) onderzoeken.

In dit rapport wordt allereerst een algemene beschrijving gegeven van het 'wat', 'waarom' en 'hoe'. Vervolgens wordt een aantal belangrijke beperkingen van meta-analyse besproken. In hoofdstuk 3 worden, rekening houdend met deze beperkingen, globale richtlijnen gegeven voor de uitvoering van meta-analyse. Het rapport besluit met een voorstel voor statistische procedures die daarbij betrokken kunnen worden. In dit rapport zal de nadruk liggen op het onderzoek naar de gezondheidseffecten van blootstelling aan of prevalentie van risicofactoren, veel minder op onderzoek naar de werkzaamheid van medisch handelen (effecten van zorg). Het voor meta-analyse geschiktste stramen is echter evenzeer van toepassing op meta-analyse van laatstgenoemde typen van onderzoek, zeker als het van observationele aard is.

In het vervolg van dit project zal aan de hand van enkele voorbeeldstudies de bruikbaarheid van het hulpmiddel meta-analyse verder in kaart gebracht worden.

## 2. META-ANALYSE

### 2.1. Wat is meta-analyse?

Het aantal wetenschappelijke artikelen dat wereldwijd wordt gepubliceerd is zeer sterk toegenomen, niet het minst in de gezondheidkundige vakgebieden (er bestaan in 1990 meer dan 20.000 biomedische tijdschriften). Dat heeft niet alleen voordelen. Het wordt steeds moeilijker ‘kaf van koren’ te scheiden, of door ‘de bomen een bos’ te zien. De kanalen voor verspreiding van belangrijke, nieuwe inzichten dreigen te verstoppelen en het wetenschappelijke systeem van kwaliteitscontrole raakt overbelast. De tijd lijkt rijp voor geavanceerde systemen voor de verwerking van het grote informatie-aanbod, naast het klassieke review of het rapport van een commissie van deskundigen. Meta-analyse is er wellicht een voorbeeld van (Olkin, 1995).

In 1976 introduceerde Glass de term ‘meta-analyse’<sup>1</sup> voor het met behulp van statistische technieken combineren van kwantitatieve informatie uit afzonderlijke onderzoeken (Glass, 1976). De naam was nieuw, de activiteit allerminst. Al in het begin van deze eeuw werden formele meta-analyses gedaan<sup>2</sup>. Vanaf de jaren 30 zijn statistische combinatietechnieken toegepast en beschreven, in het bijzonder in de sociale wetenschappen (sociologie, psychologie en onderwijskunde), waar vaak niet-significante resultaten van grote aantallen kleine studies moesten worden beoordeeld.

In het medische domein deed zich een ware explosie van meta-analyses voor om de resultaten van ‘gerandomiseerd interventie-onderzoek’ (RCT) naar de werkzaamheid van medisch ingrijpen kwantitatief samen te vatten. Een RCT is een vorm van vooruitblikkend cohortonderzoek bij patiëntgroepen die het experimenteel onderzoek enigszins benadert. De onderzoekspopulatie wordt ‘blindelings’, toevalsgewijs in groepen gesplitst om vertekening door bias of versturende variabelen zoveel mogelijk te vermijden. Een of enkele groepen worden onderworpen aan een bepaalde interventie, bijvoorbeeld een geneeskundige verrichting of een gezondheidsprogramma; een andere groep functioneert als referentiegroep. De groepen worden gevolgd om na te gaan in hoeverre de deelnemers baat hebben bij de interventie. Bij voorkeur blijven zowel de onderzoekers als de deelnemers aan het onderzoek in het ongewisse over welke groep welke interventie krijgt om vertekening door vooringenomen waarneming of rapportage (‘informatiebias’) uit te sluiten.

In dit verband dient tevens gewezen te worden op de ‘Cochrane Collaboration’. Dit internationale samenwerkingsverband probeert het pleidooi dat haar naamgever, de Britse epidemioloog Archie Cochrane, al meer dan 20 jaar geleden hield zo breed mogelijk vorm te geven, te weten het systematisch bijeenbrengen van informatie uit gerandomiseerd onderzoek naar ‘werkzaamheid’ om het medisch handelen te verbeteren (‘evidence based medicine’). Voor verschillende geneeskundige vakgebieden zijn *Cochrane fields* gedefinieerd, waarbin-

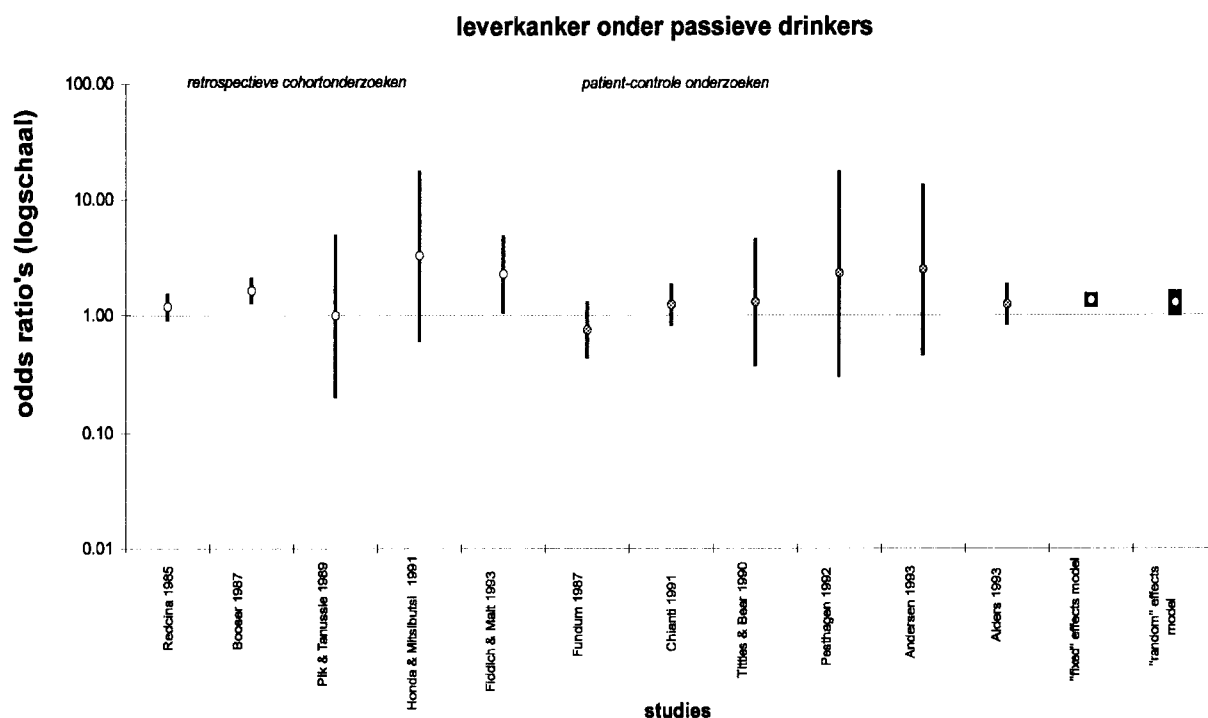
---

<sup>1</sup> Het Griekse voorvoegsel meta kan hier worden vertaald als ‘overstijgend’.

<sup>2</sup> Al in 1904 beschreef Pearson een meta-analyse van zes studies van de associatie tussen de aanwezigheid van vaccinatielittekens en sterfte aan pokken (Pearson, 1904).

nen systematische overzichten worden gemaakt van het relevante onderzoek (bijvoorbeeld RCTs). Deze activiteiten worden gecoördineerd en ondersteund door enkele *Cochrane centres* en begeleid door *Cochrane review groups*. Meta-analyse is hierbij een belangrijk instrument. Tevens zijn er *Cochrane method working groups* die zich met technische en methodologische problemen bezighouden. Er bestaan (nog) geen Cochrane initiatieven met betrekking tot het observationeel onderzoek naar gezondheidseffecten van risicofactoren (Kleijnen et al., 1995).

Bij de snelle groei van meta-analyse van RCTs lag de nadruk vooral op het vergroten van de betrouwbaarheid (precisie). Ook hier ging het aanvankelijk vaak om betrekkelijk kleine onderzoeken die afzonderlijk onvoldoende onderscheidingsvermogen hadden om het effect van behandeling aan te tonen. Door de resultaten van verschillende RCT te 'poolen' kon, althans in theorie, een preciezer schatting van het effect worden gemaakt, met bijbehorende kleinere betrouwbaarheidsintervallen. Figuur 1 laat de karakteristieke presentatie van de resultaten van een fictieve meta-analyse zien.



*Figuur 1. Voorbeeld van een karakteristieke grafische presentatie van de resultaten van een meta-analyse (van Olde Hartevelde, 1994).*

Al vanaf het begin heeft toepassing van meta-analyse aanleiding gegeven tot heftige discussies in de vakliteratuur. In de ogen van de opposanten waren de door het (variantie)gewogen middelen van afzonderlijke resultaten verkregen effectschattingen ogenschijnlijk preciezer, maar welbeschouwd vaak betekenisloos. Eysenck noemde meta-analyse 'an exercise in megasilliness'; hoe geavanceerd ook de techniek van meta-analyse, als er 'rommel' ingaat, komt er 'rommel' uit (Eysenck, 1978). Als een afzonderlijke studie vanwege de onbeheersbare invloed van allerlei vormen van bias en confounding niet ondubbelzinnig in staat is een effect



te onderscheiden van veel ruis, mag niet verwacht worden dat dat wel gelukt in een meta-analyse (Shapiro, 1994). In de kritiek staat grote twijfel over de mogelijkheden van meta-analyse het oplossend vermogen van de 'epidemiologische microscoop' wezenlijk te vergroten centraal. Winst valt hoofdzakelijk te behalen uit het verbeteren van de primaire epidemiologische onderzoeken, zo meent men.

Grootste bron van zorg is de heterogeniteit in beschikbare onderzoeken, bijvoorbeeld met betrekking tot opzet van de studie, de onderzoekspopulatie (leeftijdverdeling, gezondheidsstatus, sociaal-economische status e.d.), de definitie van blootstelling, de mate waarin men erin geslaagd is confounders te meten en ervoor te corrigeren, geografie of tijdstip. Een achteloos combineren van de resultaten van duidelijk ongelijksoortige onderzoeken kan tot niets anders leiden dan betekenisloze effectschattingen (Greenland, 1994). Het betrouwbaarheidsinterval moge na combinatie kleiner geworden zijn, de aanzienlijke onzekerheden verbonden aan allerlei vormen van bias en ongecontroleerde confounding zijn dat zeker niet. Heterogeniteit van studiekekenmerken is in het bijzonder een probleem bij review van observationeel onderzoek. Maar ook bij RCTs is, ondanks steeds nauwkeuriger omschreven protocollen en randomisatie, aanzienlijke heterogeniteit te verwachten.

Ondanks deze kritische geluiden klinken er steeds vaker positieve geluiden als het om de zegeningen van meta-analyse gaat. Greenland meent dat meta-analyse moet worden beschouwd als een onderzoek waarin het primaire onderzoek de eenheid van studie is (zoals het individu dat in een epidemiologisch onderzoek is), en *niet* primair als een manier om preciezere effectschattingen te maken (Greenland, 1994). Dat betekent onder meer dat heterogeniteit in studieresultaten niet weggemoffeld moet worden door toepassing van minder geclaussuleerde 'random' modellen, maar juist als onderwerp van analyse het inzicht kan doen verhogen. De waarde van meta-analyse is in die optiek vooral gelegen in het vergelijken van de afzonderlijke onderzoeken, patroonherkenning en het zoeken naar verklaringen voor verschillen en tegenstrijdigheden (Greenland, 1994; Petitti, 1994; Dickersin and Berlin, 1992; Berlin, 1995, Olkin, 1995). Maclure gaat nog een stapje verder en stelt, in de geest van Popper, dat meta-analyse het expliciet toetsen van een reeks van 'wedijverende' hypothesen moet omvatten om geloofwaardig te zijn. Dat houdt in dat ook zoveel mogelijk relevante 'aanpalende' informatie uit andere typen van wetenschappelijk onderzoek in de analyse betrokken kunnen worden (Maclure, 1993, 1995). Een dergelijke, grondige analyse zal evenwel vaak stuiten op het beperkte aantal beschikbare studies.

## 2.2. Waarom?

Er zijn verschillende redenen om een meta-analyse van afzonderlijke studies te doen. In de literatuur is een verschuiving waar te nemen van meta-analyse waarbij het accent ligt op het maken van een nauwkeuriger effectschatting naar vergelijkende analyse (leren van verschillen en tegenstrijdigheden). Onderstaand overzicht van motieven is, enigszins aangepast, overgenomen uit een recent rapport van de werkgroep 'Meta-analysis in environmental epidemiology' van het ILSI Risk Science Institute (Blair et al., 1995).

- Het vaststellen van een samenvattende maat ('statistische grootheid') om de relatie tussen 'blootstelling' (aan risicofactoren, maar ook preventie of zorg) en gezondheidseindpunt te beschrijven, eerst nadat uitvoerige analyse van heterogeniteit van de uitkomsten van de beschikbare studies heeft uitgewezen dat zoiets verdedigbaar is.
- Het doen toenemen van het statistisch onderscheidingsvermogen in het geval de afzonderlijke studies niet in staat zijn effecten te detecteren, omdat ze van geringe omvang zijn en/of omdat het bestudeerde effect zeldzaam is.
- Het ontwikkelen, verfijnen en testen van hypothesen met betrekking tot het blootstelling-responsverband en de rol van andere kenmerken van de studies; bijvoorbeeld door middel van gevoeligheidsanalyse van veronderstellingen met betrekking tot het blootstelling-responsmodel, de invloed van andere (verstorende) variabelen, of de invloed van bepaalde individuele studies
- Evaluatie van de 'robuustheid' van definities van blootstelling aan risicofactoren
- Beperken van de subjectiviteit van het klassieke review door gebruik van systematische, expliciete en navolgbare procedures
- Identificatie van hiaten in de huidige kennis en het stellen van prioriteiten voor nader onderzoek.

Meta-analyse dient erop gericht te zijn optimaal gebruik te maken van de in een reeks van afzonderlijke onderzoeken verborgen kwantitatieve informatie. Analyse van heterogeniteit, het genereren en testen van hypothesen kan een belangrijk middel zijn om inzicht te verwerven in de invloed van medebepalende, soms verstorende variabelen, zoals leeftijdsopbouw, gezondheidsstatus van populaties, klinische kenmerken van subgroepen, (andere) leefstijlfactoren, en sociaal-economische status.

### **2.3. Inhoudelijke en praktische aspecten**

#### **2.3.1. RCT's versus observationeel onderzoek**

Een Hollands gezegde waarschuwt ervoor appels met peren te vergelijken, zelfs onder de noemer fruit. Sommige auteurs menen dat gerandomiseerde interventie-onderzoeken wellicht voldoende vergelijkbaar zijn voor meta-analyse, maar achten meta-analyse van observationeel onderzoek in het algemeen een brug te ver (Spitzer, 1991).

RCTs worden in toenemende mate volgens omschreven protocols verricht, waarbij bovendien steeds meer naar internationale consensus en eenvormigheid wordt gestreefd. Wellicht nog belangrijker is de randomisatie, het toevalsgewijs (en 'dubbelblind'<sup>3</sup>) toewijzen van deelnemers aan interventiecategorieën, die wordt toegepast om zoveel mogelijk aan het *ceteris paribus* beginsel van experimenteel onderzoek te kunnen voldoen. Met deze randomisatie kan men verschillende bronnen van bias (vertekening) en confounding redelijk afdoende elimineren, die in het 'natural experiment' van observationeel onderzoek moeilijk te controle-

---

<sup>3</sup>

Deelnemer, noch onderzoeker weet om welke interventie categorie eerstgenoemde behoort

ren zijn. Tijdens een internationale consultatie in Potsdam in 1994 zijn door vooraanstaande deskundigen richtlijnen vastgesteld voor het naar behoren uitvoeren van meta-analyses van RCTs (Spitzer, 1995).

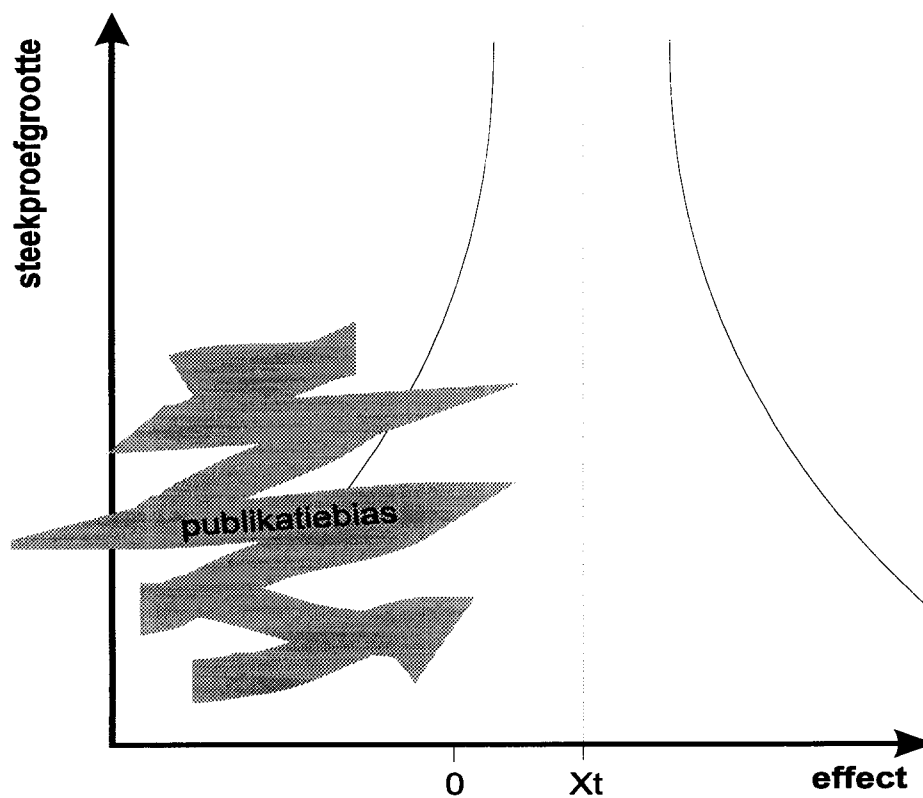
In vergelijking met meta-analyse van RCTs zullen meta-analyses van observationeel onderzoek dan ook veel meer onzekerheden kennen, minder uitsluitel kunnen geven, kortom veel 'rommeliger' zijn. De vraag of observationeel onderzoek al dan niet meta-analyseerbaar is, is nog niet definitief te beantwoorden. Zeker is dat meta-analyse van observationeel onderzoek strenge eisen stelt aan de zuiverheid van het protocol, de helderheid van de vraagstelling, het omgaan met heterogeniteit, de gevoeligheidsanalyse en de terughoudendheid bij de interpretatie en presentatie (Blair et al., 1995).

### **2.3.2. *Vertekening door selectieve publikatie ('publikatiebias')***

Selectieve publikatie van alleen die resultaten die een duidelijk effect van blootstelling of behandeling laten zien, kan bij meta-analyse van gepubliceerde resultaten leiden tot overschatting van het effect. Redacteuren van wetenschappelijke tijdschriften kunnen geneigd zijn manuscripten die een effect beschrijven eerder te accepteren dan manuscripten waarin hiervoor geen aanwijzingen zijn gevonden. Zeker als het gaat om zwakkere, statistisch niet-significante resultaten van onderzoek. Ook onderzoekers zelf kunnen bij het toetsen, beschrijven of publiceerbaar maken van reeds verzamelde gegevens vooral gericht zijn op aanwijzingen voor een effect. Overigens zullen studies die duidelijk wijzen op de afwezigheid van een effect bij controversiële onderwerpen, zoals extreem laag-frequente elektromagnetische straling (ELF) of passief roken, juist weer gemakkelijker gepubliceerd raken.

In de praktijk van de observationele epidemiologie zal men vaak verschillende blootstellingsdefinitieën toetsen, zowel kwantitatief (categorie-indeling, afkappunten) en soms zelfs kwalitatief (relaties met verschillende componenten). Hierbij zullen niet-significante associaties niet of op zijn best minder uitvoerig gerapporteerd worden. Vooral in de observationele epidemiologie speelt deze presentatiebias een niet te onderschatten rol. In dit type onderzoek is minder vaak sprake van a priori en strikt gedefinieerde onderzoeksvragen, zodat de onderzoeker de gelegenheid heeft in publikaties het accent op sterke, significante associaties te leggen, en de zwakke, niet-significanten 'achter de kiezen' te houden.

*Het probleem van publikatiebias is overigens niet uniek voor meta-analyse, maar geldt ook het klassieke review.* In de loop van de tijd zijn verschillende methoden geopperd om bij meta-analyse resultaten grafisch te toetsen op publikatiebias. Als men de in afzonderlijke onderzoeken gemeten effectgrootte grafisch uitzet tegen de variantie of steekproefgrootte, dient de puntenwolk een trechtervorm aan te nemen. De resultaten van grotere studies zijn nauwkeurig en concentreren zich symmetrisch rond de 'werkelijke' effectgrootte, naar mate de studies kleiner en minder nauwkeurig zijn, wordt de spreiding van de resultaten ten opzichte hiervan steeds groter. Het ontbreken van een deel van de trechter, wijst op het niet gepubliceerd raken van studies. Er is dan sprake van publikatiebias. Deze toets is echter voorbehouden aan meta-analyses waarvoor een redelijk grote hoeveelheid primaire studies beschikbaar zijn. Bij kleinere aantallen beschikbare studies, krijgt het al dan niet waarnemen van een trechter het karakter van een Rohrschach-test (Vandenbroucke, 1989, Greenland, 1994).



Figuur 2. Geïdealiseerd trechterplot van 'scatter' van studieresultaat tegen studie grootte rond  $X_t$ , het 'ware effect'; 0 staat voor geen effect.

De enige ware remedie tegen publikatiebias zou het wereldwijd registreren van epidemiologisch onderzoek zijn, en het vastleggen van duidelijk richtlijnen met betrekking tot volledigheid van rapportage van geanalyseerde associaties (modellen). Bij RCTs mag enige hoop gekoesterd worden dat dit bereikt wordt, bij observationeel onderzoek is dit waarschijnlijk nog een ver station, wellicht met uitzondering van omvangrijke multi-centre-studies.

### 2.3.3. Rapportage van afzonderlijke onderzoeken

Het moge duidelijk zijn dat (vergelijkende) meta-analyse veel meer informatie vereist dan slechts de puntschatting van het effect met bijbehorende betrouwbaarheidsintervallen. Soms, vooral in wat oudere publikaties, staan zelfs die niet vermeld en moet men het doen met p-waarden. Zo al mogelijk zal het achterhalen van ontbrekende informatie via communicatie met de auteurs vaak een tijdrovende en moeizame weg blijken. Het veld van epidemiologen is bovendien nog niet onverdeeld enthousiast over het verschijnsel meta-analyse.

Volledigheid van rapportage is strijdig met de eis van kort en bondigheid die aan wetenschappelijke publikaties gesteld wordt. Sommige auteurs hebben aangedrongen op overeenstemming en afspraken over minimaal te rapporteren kenmerken van studies onder redacteurs van tijdschriften. Hiermee zal zowel de meta-analyse als het klassieke review gebaat zijn (Olkin, 1994).

### 2.3.4. *Vergelijkbaarheid van studies*

Bij de meeste meta-analyses van observationeel onderzoek zal men te maken krijgen met een zekere onvergelijkbaarheid van studies, bijvoorbeeld met betrekking tot:

- definitie van blootstelling, interventie
- definitie van respons
- onderzoekspopulatie (leeftijdsopbouw, geslacht, algehele gezondheid, al dan niet 'geschoond'<sup>4</sup>)
- associatiemaat (relatief risico, odds ratio, risicoverschil, attributief risico etc.)
- type studie (cohort, dwarsdoorsnede, of patiënt-controle onderzoek, tijdserie-analyse)
- data-analyse
- meting en wijze van omgaan met confounders en vormen van vertekening (bias)

In sommige gevallen kunnen onvergelijkbare kenmerken op een noemer gebracht worden. Zo kan men in de milieu-epidemiologie trachten de ene blootstellingsmaat om te rekenen in een ander op grond van inzichten over luchtverontreinigingspatronen in ruimte en tijd (Preller et al., 1996). Verder kan men door heranalyse proberen voor een afzonderlijke studie tot een betere controle van confounders te komen (Greenland, 1987).

Soms gaat het om inherente onvergelijkbaarheid en ontbreken formele mogelijkheden om kenmerken op een noemer te brengen, bijvoorbeeld in geval van verschillende typen van epidemiologische studie. De hypothese dat men de resultaten van verschillende studies als schatters van 'hetzelfde' effect mag beschouwen zal dan in de meta-analyse getoetst moeten worden. Bij een meta-analyse van de associatie tussen passief roken en longkanker presenteerden Wald en medewerkers effectschattingen op basis van patiënt-controle onderzoeken en retrospectieve cohortonderzoeken zowel afzonderlijk als gecombineerd (Wald et al., 1985). Bij wijze van kritisch commentaar lieten Fleiss en Gross zien dat indien alleen de kwalitatief goede Amerikaanse cohortstudies in de meta-analyse werden betrokken er geen sprake meer was van een significant effect (Fleiss en Gross, 1991).

### 2.3.5. *Fixed versus random effect en heterogeniteit*

Bij meta-analyse wordt van twee verschillende modellen gebruik gemaakt: 'fixed effects' en 'random effects'. Bij het 'fixed effects model' gaat men (impliciet) uit van de veronderstelling dat alle studies een vaste functie schatten van de gemeten studievariabelen (zoals blootstelling en respons), alsof het een fysische wetmatigheid zou betreffen (denk aan valversnelling 'g'). Dit betekent dat verschillen in resultaten tussen studies uitsluitend te wijten zouden zijn aan toevalsfluctuaties die samenhangen met het feit dat elke afzonderlijke studie bestaat uit een eindige steekproef. Zeker bij observationeel onderzoek met de vele onbeheersbare bronnen van variatie lijkt deze veronderstelling weinig realistisch. Daarom beschouwt men in een 'random effects model' de resultaten van afzonderlijke studie als 'trekking' uit een verdeling van functies. Daartoe voegt men aan het 'fixed effect regressie model' een term  $\epsilon$  (met

---

<sup>4</sup>

'Geschoond': een selectie van voorbeeldige patiënten, weinig co-morbiditeit, grote compliance.

verwachting nul) toe, waarin de ‘ongecontroleerde’ heterogeniteit is verdisconteerd, zoals verschillen in distributie van effect modifiers of niet gemeten confounding. In formule, respectievelijk ‘fixed effects’ en ‘random effects’:

$$\theta_{true} = \alpha + \beta_1 x_1 + \dots + \beta_k x_k$$

$$\theta_{true} = \alpha + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$$

waarin  $\theta$  de risicomaat is en  $x_k$  een gemeten studiekarakteristiek met bijbehorende regressie coëfficiënt  $\beta_k$ .

In de meest recente literatuur geeft men meestal de voorkeur aan het ‘random effects model’ boven het ‘fixed effects model’ en houdt men bij het berekenen van een samenvattende effectschatting naast ‘sampling error’ dus ook rekening met de haast vanzelfsprekende tussenstudie variatie. Wel zij hierbij opgemerkt dat de resultaten van beide modellen juist dan significant uiteen gaan lopen als er sprake is van aanzienlijke ‘onverklaarde’ heterogeniteit (zie ook hoofdstuk 4).

In een eerdere paragraaf werd aangegeven dat combinatie van studies pas mogelijk zou zijn als zoveel mogelijk voor variatie in studiekekenmerken is gecorrigeerd, bijvoorbeeld door stratificatie of regressie. Of in de woorden van Greenland (1994): *‘If the use of random effects makes a difference, the analysis is incomplete: the analyst should carefully search for the source of discrepancy between fixed and random effect interval estimates. The random effect summary is merely a last resort, to be used only if one cannot identify the predictors of causes of the between-study heterogeneity.’*

In een recente publikatie geven Berlin et al. een aantal voorbeelden van meta-analyses, waarin heterogeniteit in studieresultaten bij heranalyse kon worden verklaard op grond van bepaalde demografische of klinische eigenschappen van de onderzoekspopulatie (op studieniveau, Berlin et al., 1993).

Zo bleek de associatie tussen de duur van het gebruik van orale contraceptie en een toenemend risico op borstkanker alleen zichtbaar in patient-controle-onderzoeken met overwegend *premenopausale* gevallen (< 45 jaar). Binnen deze groep van studies was de resterende heterogeniteit gering. De studies waarin ook postmenopausale gevallen werden betrokken lieten geen duur-responsverband zien, terwijl bovendien sprake was van aanzienlijke heterogeniteit. Tevens was, op studieniveau, sprake van een verband met het *kalenderjaar* waarin de gevallen werden geconstateerd (wellicht als gevolg van veranderende samenstelling van de ‘pil’).

Bij verschillende meta-analyses van studies naar de toename van het risico op kanker in relatie tot blootstelling aan carcinogenen van werknemers (asbest, petroleum, benzeen), bleek dat ‘grove’ stratificatie naar gemiddeld niveau van blootstelling de moeite loont: de heterogeniteit neemt af, de schattingen worden nauwkeuriger, bovendien wordt vaak een (soms kwalitatieve) blootstelling-responsrelatie zichtbaar (Frumkin & Berlin, 1988; Wong & Raabe, 1989; Colditz et al., 1995). Hetzelfde gold bij een meta-analyse van het verband tussen het roken van sigaretten en het risico op een beroerte. Indeling (op studieniveau) naar ‘lichte’, ‘matige’ en zware’ rokers maakte een fraai blootstelling-responsverband zichtbaar (relatieve risico’s respectievelijk 1,37; 1,45 en 1,82; Shinton & Beevers, 1989).

Een ander voorbeeld van het verwerven van inzicht uit analyse van heterogeniteit tussen studies betreft de beschermende werking van streptokinase bij een (vermoed) acuut hartinfarct. Bij voorwandinfarcten was die duidelijk, bij achterwandinfarcten in ieder geval aanzienlijk geringer, zo al aanwezig (Midgette et al., 1990).

Wederom moet erop gewezen worden dat een dergelijke analyse van heterogeniteit door middel van stratificering of regressie-analyse een behoorlijke beschikbaarheid van individuele studies veronderstelt. Dat valt in de praktijk wel eens tegen.

### **2.3.6. Gebruik van kwaliteitsscores**

Bijna even omstreden als meta-analyse van observationele studies zelf is het wegen van studies op basis van scores voor de kwaliteit van de opzet en uitvoering. Belangrijkste bezwaar is dat het vermengen van subjectieve, arbitraire oordelen met objectieve informatie de mogelijkheden tot patroonherkenning en studie van heterogeniteit aanzienlijk kan verstoren. Dit bezwaar doet zich extra gelden als in één kwaliteitsoordeel verschillende kwaliteitsaspecten zijn betrokken. De ILSI werkgroep stelt voor geen 'overall' kwaliteitsscores te hanteren, maar met behulp van gevoeligheidsanalyse of stratificering per aspect na te gaan hoe kwaliteit doorwerkt in het eindresultaat, bijvoorbeeld nauwkeurigheid van blootstellingsschatting, 'design', vergelijkbaarheid van controlepopulatie (Blair et al., 1995). Zo vonden Morris en medewerkers bij meta-analyse van het verband tussen drinkwaterchloring en kanker dat het effect groter werd naar mate de blootstellingsschattingen in de afzonderlijke studies preciezer waren. De kwaliteit van de data-analyse had daarentegen geen invloed op de overall schattingen (Morris et al., 1992).

### **2.3.7. Aantallen beschikbare studies**

Afhankelijk van de vraagstelling kan de beschikbaarheid van studies gaan 'knellen'. Uiteraard is het altijd mogelijk studies te combineren om tot een soort gemiddelde schatting van de effectmaat te komen. Gaat het echter om het toetsen van hypothesen met betrekking tot de rol van andere studiekenmerken, of het testen van robuustheid van definities dan is er een minimum aantal bruikbare studies nodig, waarmee nog op zinvolle wijze gemeta-analyseerd kan worden. Een vuistregel is dat er voor regressie-analyse ongeveer 10 maal zoveel observaties nodig zijn als onafhankelijke variabelen in het model. Bij toepassing van de in het kader van dit project ontwikkelde procedure voor (Empirische Bayesiaanse) Meta-regressie-analyse lijkt die verhouding wat gunstiger te liggen. Bij de meta-analyse van de relatie tussen verhoogde ozonniveaus in de buitenlucht en ziekenhuisopnamen, waarvoor 21 effectschattingen beschikbaar waren, werden met vijf onafhankelijke variabelen in het model nog stabiele resultaten verkregen (Preller et al., 1996).





### 3. RICHTLIJNEN VOOR META-ANALYSE

Hoewel meta-analyse bij het maken van systematische overzichten van gerandomiseerd interventie-onderzoek (RCTs) algemeen wordt toegepast en ook bij observationeel onderzoek steeds meer ingeburgerd raakt, bestaan er nog geen algemeen geaccepteerde regels waaraan zich men bij meta-analyse aan kan houden. Is meta-analyse voor de een het eenvoudigweg middelen van effectschattingen, voor de ander vormt deze middeling slechts een klein onderdeel van de gehele procedure en hoort de nadruk te liggen bij de keuzen die gemaakt worden in verschillende stadia van die procedure. Juist die bepalen in sterke mate de waarde die gehecht kan worden aan het resultaat van de analyse. In dit hoofdstuk worden globale richtlijnen voorgesteld die betrekking hebben op het hele proces van meta-analyse, vanaf het samenstellen van de onderzoeks- en begeleidingsgroep tot de presentatie van de resultaten. Een en ander zal worden verduidelijkt aan de hand van de ervaringen opgedaan bij de meta-analyse van de relatie tussen ozonconcentraties in de buitenlucht en ziekenhuisbezoek en van de relatie tussen blootstelling aan dieselemissie in het beroep en toename van het risico op longkanker (Preller et al., 1996).

Op het gebied van het experimentele klinische onderzoek (RCTs) zijn er begin 1995 richtlijnen gepubliceerd naar aanleiding van een internationale expert consultatie in Potsdam (Spitzer, 1995). Richtlijnen voor observationeel onderzoek zijn onlangs geformuleerd door een werkgroep van deskundigen (Blair et al. 1995). Beide publikaties vormen de basis voor de hier gepresenteerde richtlijnen voor meta-analyses van observationeel onderzoek, zoals die onder andere uitgevoerd zullen worden in het kader van VTV-taken (volksgezondheids toekomst verkenningen). Daarnaast vormen ze de weerslag van de ervaringen die zijn opgedaan met de hierboven genoemde door Preller uitgevoerde meta-analyses van milieu-en arbeidsepidemiologische studies.

De hier gepresenteerde globale richtlijnen beschrijven wenselijke of juist ongewenste aspecten van het proces van meta-analyse. Ze moeten vooral gezien worden als een leidraad aan de hand waarvan de validiteit van een meta-analyse kan worden verhoogd. De richtlijnen zijn meestal zowel op experimenteel als op niet-experimenteel onderzoek van toepassing. Soms betreffen ze uitsluitend één van beide typen van onderzoek.

#### 3.1. Onderzoeks- en begeleidingsgroep

Het verdient om verschillende redenen de voorkeur om inhoudelijke deskundigen bij de meta-analyse te betrekken. Hierbij denken we in de eerste plaats aan epidemiologen die zich met de bestudeerde relatie hebben beziggehouden. Zij zijn vertrouwd met de methodologische knelpunten van het primaire onderzoek op hun terrein en kunnen als beste de kwaliteit van de afzonderlijke studies beoordelen. Bovendien hebben zij inzicht in de meest gangbare hypothesen over factoren die de blootstelling-responsrelatie kunnen beïnvloeden, en die dus geëvalueerd dienen te worden bij het verklaren van heterogeniteit in studie-uitkomsten. Ook deskundigen op het gebied van de blootstelling aan risicofactoren of met betrekking tot werkingsme-

chanismen kunnen een belangrijke bijdragen leveren, vooral als het gaat om het definiëren van de onderzoeksvraag en de interpretatie van de resultaten van meta-analyse.

Specialisten zijn bovendien vertrouwd met de belangrijkste onderzoeksgroepen, kennen het onderzoek dat gepubliceerd is, en mogelijk ook dat wat nog niet gepubliceerd is of wat nog loopt. Bovendien is het waarschijnlijk dat eventueel contact met auteurs van primaire studies vruchtvoller verloopt wanneer voor hen duidelijk is dat de meta-analyse wordt ondersteund door erkende specialisten. Inzet van niet-specialisten kan gewenst zijn om in de verschillende stadia zorg te dragen voor de objectiviteit van de beoordeling van primaire studies.

### 3.2. Onderzoeksprotocol

Als eerste stap in het proces van meta-analyse dient de onderzoeksvraag helder en zo 'nauw' mogelijk te worden omschreven, waarbij aspecten als definitie van blootstelling of prevalentie, gezondheidsuitkomst, versturende variabelen, eventueel interventie, en populatie zo veel mogelijk worden gespecificeerd<sup>5</sup>. In dit stadium is de inbreng van (specialisten uit) de begeleidingsgroep essentieel om snel tot een praktische afbakening te komen. Overzichtsartikelen met betrekking tot etiologie, werkingsmechanisme en epidemiologie kunnen hierbij behulpzaam zijn. De veronderstelde vorm van de blootstellings-effect relatie bijvoorbeeld (lineair, exponentieel, U-vormig), bepaalt mede de manier waarop blootstelling idealiter gedefinieerd zou moeten worden voor de meta-analyse (zie ook rapport werkgroep PAR-berekeningen, De Hollander et al., 1996). Ook analyses van subgroepen om bepaalde hypothesen met betrekking tot studiekenmerken te toetsen, stelt men bij voorkeur a priori vast.

De methoden voor het zoeken, verzamelen, beoordelen en uiteindelijk selecteren van afzonderlijke onderzoeken dient zo veel mogelijk vooraf in een protocol te worden vastgelegd, evenals de wijze van (statistische) analyse.

Het moge duidelijk zijn dat een zo strikt mogelijke, a priori definitie van het onderzoek de beste remedie is tegen allerlei vormen van onderzoekersbias, 'data torture' (het 'martelen' van gegevens tot ze het gewenste antwoord geven) of 'visexpedities' (het zonder a priori hypothesen zoeken naar correlaties in databestanden). Aan de andere kant is het in de praktijk vaak onmogelijk om tot een goede definitie te komen zonder nader inzicht in de kenmerken van de afzonderlijke studies. De meta-analist heeft per definitie geen invloed op de opzet van de afzonderlijke studies, hooguit in geval van multi-center studies. In het voorgaande hoofdstuk (2.2.4) is al een indicatie gegeven tot welk soort 'improvisaties' de vaak aanzienlijke variatie in studiekenmerken kan nopen. Dit betekent dat in de praktijk het protocol pas gaandeweg de analyse wordt aangepast en verfijnd. Zo bleek bij de meta-analyse van tijdserie-analyses van fotochemische luchtverontreiniging naast ziekenhuisopname, soms ook de relatie met poliklinische behandeling beschreven te worden. Hierbij deed de vraag zich voor of ook dat eindpunt in de meta-analyse betrokken zou mogen worden. Bovendien bleken relaties

<sup>5</sup> Zo luidde voor de voorbeeld meta-analyse van de effecten van fotochemische luchtverontreiniging de onderzoeksvraag: *in welke mate nemen de aantallen ziekenhuisopnames en -bezoeken aan de polikliniek voor de verzameling van respiratoire aandoeningen toe bij een verandering van 50 ppb in de 8-uursgemiddelde concentratie van ozon in de buitenlucht? In welke mate zijn andere variabelen hierop van invloed, zoals leeftijd, achtergrondniveau van verontreiniging, en het type studie (de manier van analyse)?*

veelal niet voor de hele groep van respiratoire aandoeningen te zijn weergegeven. De onderzoeksvraag kan dan uitgebreid worden met de vraag: is de grootte van het effect afhankelijk van het type aandoening?

Dergelijke aanpassingen en verfijningen dienen wel zo goed mogelijk gemotiveerd en gedocumenteerd te worden, opdat ze navolgbaar zijn ('open to critical review'). Indien mogelijk, dient de gevoeligheid van keuzen expliciet en kwantitatief te worden geëvalueerd, bijvoorbeeld door stratificatie of door analyse van een homogene selectie van de beschikbare effectschattingen.

### **3.3. Verzamelen en selecteren van afzonderlijke onderzoeken**

#### **3.3.1. Zoekstrategie**

Het identificeren van studies die in beginsel in de meta-analyse betrokken kunnen worden vereist een zoekstrategie. Om publicatie bias te vermijden dient de 'search' zo veelomvattend mogelijk te zijn. De juiste wijze van zoeken zal afhangen van het veld van onderzoek (uitgebreid, diffuus, incoherent of juist beperkt en goed gedefinieerd) en de vraagstelling waartoe de meta-analyse is opgezet. De zoekstrategie bij de meta-analyse naar de effecten van ozon kon heel beperkt blijven, gebruik makend van slechts enkele gerichte termen. Deze methode leverde maar weinig ballast in de zin van (veel) ongewenste, niet ter zake doende literatuurverwijzingen op. Voor de relatie tussen blootstelling aan uitlaatgassen van dieselmotorvoertuigen en longkanker daarentegen, bleek het veel moeilijker een goede zoekstrategie samen te stellen, met name wat betreft een juiste definitie van de blootstelling. Het is dan aan te raden de hulp in te roepen van een specialist op het gebied van literatuuronderzoek.

In de ideale wereld bestaat een volledig overzicht van het epidemiologisch onderzoek van de desbetreffende blootstelling-responsrelatie of interventie. Gerandomiseerde interventiestudies (RCT's) worden in toenemende mate wereldwijd geregistreerd, waardoor er op termijn een redelijk volledig overzicht van studies zou kunnen ontstaan. Bij observationeel-epidemiologische studies is hier zelfs nog geen begin mee gemaakt en zullen alle mogelijke bronnen zoveel mogelijk gebruikt moeten worden, zoals uiteraard geautomatiseerde bibliografische literatuurbestanden van gepubliceerd en niet-gepubliceerd onderzoek, review artikelen, abstracts, verslagen van wetenschappelijke bijeenkomsten, 'case-reports', proefschriften, handboeken, maar ook informatie van experts, subsidiërende instanties en het bedrijfsleven.

Om de identificatie van de studies navolgbaar te maken en bij voorbaat zo veel mogelijk systematische bias te vermijden, dient voor aanvang van de meta-analyse de procedure van identificatie beschreven te worden: welke data-bases, welke jaartallen, zoektermen, taal enzovoorts. In een kort geleden verschenen RIVM-rapport worden aspecten van kwaliteitsborging bij literatuuronderzoek beschreven (van Loon et al., 1996)

#### **3.3.2. Selectie van studies en effectschattingen**

Waar de identificatiefase gericht is op het binnenhalen van zoveel mogelijk studies, gaat het er in de volgende fase van selectie om alleen die studies over te houden die aan tevoren op-

gestelde, heldere criteria voor 'deelname' voldoen. Deze kunnen de relevantie voor de onderzoeksvraag betreffen, de consistentie en methodologische kwaliteit van de studie, zowel als de volledigheid van rapportage van voor meta-analyse benodigde gegevens. Men kan eerst selecteren op basis van de informatie die met de zoekstrategie verkregen en eventueel daarna op basis van informatie van de volledige publikaties. Criteria worden voor aanvang van de meta-analyse gedefinieerd om bias te vermijden en met het oog op de consistentie en reproduceerbaarheid van de selectie. Bij de meta-analyse van de relatie tussen ozon en ziekenhuisbezoek deed zich een interessant dilemma voor. Bij de vroegste onderzoeken zijn statistische technieken gehanteerd die voor het desbetreffende soort relaties inmiddels als achterhaald worden beschouwd. Betreft men de resultaten daarvan nu wel of niet bij de analyse?

Redenen om van vooropgestelde criteria af te wijken, dienen gemotiveerd te worden. Redenen voor het uitsluiten van studies moeten expliciet worden gemaakt, vooral waar het gaat om de 'grensgevallen'. Niet betrokken studies kunnen overigens waardevolle informatie bevatten over kwalitatieve aspecten. Soms kunnen ze een rol spelen bij gevoeligheidsanalyse, bijvoorbeeld om de vraag te beantwoorden of het eindresultaat sterk verandert als de methodologisch mindere studies meegenomen worden in de analyse.

Bij de selectie van de artikelen zijn de volgende aspecten van belang:

#### *Onderzoeksopzet*

Het type epidemiologisch onderzoek is op zichzelf geen reden tot uitsluiting. In beginsel kunnen resultaten van bijvoorbeeld cohort- en patiënt-controle-onderzoeken in dezelfde meta-analyse opgenomen worden. Eventuele met de onderzoeksopzet samenhangende verschillen in de associatie tussen blootstelling en effect kunnen met behulp van gevoeligheidsanalyses onderzocht worden. De meest eenvoudige vorm hiervan is het vergelijken van de uitkomsten van analyses al dan niet 'gepooled' naar studietype (Fleiss en Gross, 1991, Smith, 1995). Indien er sprake is van een systematische verschil, verdient het de voorkeur resultaten van meta-analyse per onderzoeksopzet te rapporteren.

#### *Verschillende studies bij dezelfde populatie*

Als voor een bepaalde onderzoekspopulatie verschillende herzieningen met betrekking tot de te evalueren blootstellings-effect relatie zijn gepubliceerd, dient een keuze gemaakt te worden welke er gebruikt gaat worden in de meta-analyse. Het opnemen van verscheidene herzieningen geeft een te zwaar gewicht aan de desbetreffende populatie, en leidt bovendien tot onderschatting van de variatie tussen de studies. Vaak zal de meest recente publikatie de voorkeur verdienen, omdat hierin onvolkomenheden in eerdere analyses rechtgezet zijn en de meest recente opvattingen over de optimale data-analyse (inclusief het omgaan met confounders) zijn meegenomen. De keuze voor een eerdere publikatie kan echter ingegeven worden door een grotere vergelijkbaarheid met de andere studies in de meta-analyse, of door een betere benadering van de gedefinieerde onderzoeksvraag, bijvoorbeeld als het gaat om de gezondheidsuitkomst of de blootstellingsmaat.

Is de relatie tussen verhoogde ozonniveaus in de buitenlucht en ziekenhuisbezoek in een bepaalde populatie in een eerste artikel beschreven voor alle respiratoire aandoeningen, en in een tweede voor alleen astma, dan bepaalt de onderzoeksvraag (en de responsdefinitie in de

andere studies) welke relatie wordt opgenomen. Bij een herziening kan het verband tussen blootstelling aan huisstofmijtallergeen en astma bij een cohort kinderen op geavanceerde wijze nader gedefinieerd zijn met behulp van allergeenbepalingen in verzameld stof. Als daarentegen in alle andere beschikbare studies de blootstelling is gekarakteriseerd aan de hand van de aanwezigheid van stoffen vloerbedekking, kan de herziening niet zinvol meegenomen worden in de meta-analyse.

#### *Kwaliteit van onderzoek*

Alleen wanneer er zeer ernstige fouten in de methoden van onderzoek, de opzet en de data-analyse worden geconstateerd, kunnen studies bij voorbaat uitgesloten worden. Het verdient aanbeveling door middel van gevoeligheids-analyses de invloed van kwalitatief 'slechtere' studies op de resultaten van de meta-analyse te beoordelen.

#### *Beschikbaarheid van informatie*

De beschikbaarheid van noodzakelijke informatie kan een reden zijn om bepaalde studies wel te gebruiken, en andere niet. *A priori* dienen criteria opgesteld te worden in welke situaties en op welke manier getracht wordt ontbrekende informatie van de auteurs te verkrijgen. Men kan afspreken auteurs wel te benaderen om direct beschikbaar materiaal, zoals een exacte p-waarde of de standard error, maar niet voor informatie waarvoor (naar verwachting) aanvullende analyses nodig zijn, bijvoorbeeld wanneer de associatie gecorrigeerd zou moeten worden voor een andere set van versturende variabelen.

#### *Statistische eigenschappen*

Statistische significantie of statistisch onderscheidingsvermogen ('power') zijn geen reden voor de selectie van studies. Een van de problemen van (kwantitatieve) review is juist het niet gepubliceerd geraken van niet-statistisch significante resultaten, waardoor de wel gepubliceerde studies een vertekend beeld kunnen geven. Het uitsluiten van statistisch niet-significante studies zou publikatiebias alleen nog maar doen toenemen.

#### *Publikatie-bias*

Het bij de meta-analyse betrekken van studies die niet in gerefereerde tijdschriften zijn gepubliceerd, zou in beginsel publikatiebias kunnen verminderen. Methodologische gebreken pleiten tegen opname, maar kwalitatief goede studies die een klein, eventueel niet-significant effect of juist een onverwacht effect laten zien, moeten juist bij voorkeur worden meegenomen.

Om gevoel te krijgen voor de potentiële rol van publikatiebias, kan men nagaan hoe groot een studie met tegenstrijdige uitkomsten zou moeten zijn om het resultaat bij opname in de meta-analyse ingrijpend te veranderen (Washburn et al., 1994). Uiteraard kan bij voldoende beschikbaarheid van effectschattingen een zogenoemd trechterplot (zie paragraaf 2.3.2) wat 'gevoel' geven voor de mate van bias.

#### *Presentatie-bias*

Auteurs zullen vaak geneigd zijn in hun publikaties maar een selectie van de analyse-resultaten weer te geven, namelijk de sterkste en meest opzienbarende associaties. Dit leidt in een meta-analyse tot een overschatting van het effect, en dient daarom zoveel mogelijk be-

perkt te worden. Zeker in de observationele epidemiologie kan presentatiebias een grote rol spelen doordat er een grote vrijheid is bij analyse en rapportage van data.

Bij onderzoek naar de relatie tussen ozonblootstelling en ziekenhuisopname, bijvoorbeeld, is variatie mogelijk in de definitie van de gezondheidsuitkomst, confounders, de periode waarover analyses gedaan worden, de 'lag-time' (tijd tussen blootstelling en effect), en de blootstellingsmiddelingsstijd. Gegeven de stand van kennis kan geen uitspraak gedaan worden over het biologisch meest plausibele model voor blootstelling en respons. Hiermee staat het de auteurs vrij de sterkste verbanden te presenteren. Lang niet altijd wordt vermeld welke andere modellen zijn geanalyseerd (Preller et al., 1996). Presentatie-bias is alleen uit te sluiten door de auteurs van de afzonderlijke studies om ontbrekende gegevens te verzoeken.

Op verschillende plaatsen in dit hoofdstuk wordt aanbevolen de auteurs van de oorspronkelijke studies te benaderen om aanvullende informatie te verkrijgen. Men dient er echter rekening mee te houden dat dit lang niet altijd binnen redelijke termijn zal lukken, afhankelijk van de moeite die het de auteurs kost gegevens boven water te halen, en wellicht ook afhankelijk van hun houding ten opzichte van het 'verschijnsel meta-analyse'.

### 3.4. Extractie van gegevens

In ieder geval dienen uit de publikaties de (kwantitatieve) blootstelling-effectmaat en de bijbehorende standaardfout (of andere informatie waaruit dit te berekenen is) geëxtraheerd te worden. Vaak zijn er verschillen in de manier waarop blootstelling gedefinieerd is of hoe er met confounding is omgegaan. Het vergelijkbaar maken van de relaties is dan een onderdeel van de data-extractie.

Naast de blootstelling-effect maat dient de volgende informatie genoteerd te worden, voor zover van toepassing:

- steekproefgrootte (omvang studie-populatie)
- periode van de studie
- type epidemiologische studie
- populatie-kenmerken
- geografische locatie
- statistische analyse
- exacte gezondheidsuitkomst
- exacte blootstellingsmaat
- meetmethoden
- confounders
- niveau van blootstelling,

en alle andere kenmerken die invloed kunnen hebben op de effectschatting.

Uit overwegingen van efficiëntie is het beter al bij de eerste data-extractie zoveel mogelijk kenmerken te noteren, ook al lijken deze in eerste instantie overbodig.

Bij de data-extractie kunnen fouten gemaakt worden en kan subjectieve beoordeling een rol spelen. Bij voorkeur dient dit daarom onafhankelijk door ten minste twee personen te gebeuren. Als er keuzen gemaakt worden die afwijken van de oorspronkelijke vraagstelling of

die anderszins afwijken van vooropgezette criteria (bijvoorbeeld in de keuze tussen twee vergelijkbare gezondheidsuitkomsten) dient dit vastgelegd te worden.

Een aantal aspecten komt bij de data-extractie aan de orde.

#### *Blinderen voor o.a. auteurs en studie-uitkomst*

Het blinderen voor studie-aspecten als auteurs en studie-uitkomst van degenen die de data extraheren, kan de objectiviteit ten goede komen. Over de praktische haalbaarheid en het nut hiervan is twijfel mogelijk. Uit onherkenbaar gemaakte informatie van 'abstracts' op papier kan soms het tijdschrift al afgeleid worden. Bovendien zullen experts aan de wel gegeven informatie vaak nog steeds de studie kunnen herkennen (zie o.a. Washburn, 1994).

#### *Mate van volledigheid en vergelijkbaarheid van informatie*

Als de in de publikatie gepresenteerde informatie niet volledig of vergelijkbaar, dienen auteurs benaderd te worden voor de ontbrekende informatie op een vooraf vastgestelde manier.

#### *Vergelijkbaarheid van blootstellingsdefinitie*

Idealiter hanteren de afzonderlijke studies dezelfde blootstellingsdefinitie. Wanneer dit niet het geval is, dient men na te gaan of de maten op één noemer gebracht kunnen worden. Bij observationeel epidemiologisch onderzoek wordt vaak gebruik gemaakt van zeer afgeleide blootstellingsindicatoren ('surrogaten'), zoals het al dan niet hebben van licht, matig of zwaar rokende partners bij het onderzoek naar effecten van passief roken, het uitoefenen van bepaalde typen van functies in de chemische industrie ('job-characteristics') in de beroepsepidemiologie, of afstand van de woning tot de bron van milieuverontreiniging (verkeersemissies, extreem laag frequente magnetische velden). Soms zullen dergelijke surrogaten voor blootstelling niet via indirecte informatie over blootstellingsniveaus in relatie tot blootstellingsattributen op één noemer te brengen zijn. In dat geval kan men een gestratificeerde analyse uitvoeren of de blootstellingsdefinitie als variabele in een meta-regressiemodel opnemen.

Bij studies naar de relatie tussen ozonniveaus in de buitenlucht en ziekenhuisbezoek hangt de grootte van het effect af van de middelingstijd van de ozonconcentratie. De correlatie tussen de maximale 1-uurs waarde en de daggemiddelde waarde blijkt vaak redelijk hoog te zijn, waarbij het daggemiddelde echter wel een factor 2-3 beneden de maximale 1-uurs waarde ligt<sup>6</sup>. Het relatief risico berekend voor een toename van 50 ppb is dan groter wanneer met het daggemiddelde gerekend wordt. In deze situatie kunnen beide blootstellingsmaten vergelijkbaar gemaakt worden op grond van fenomenologisch inzicht in de verhouding tussen beiden. Andere voorbeelden van het vergelijkbaar maken van blootstellingsindicatoren bij observationeel onderzoek zijn te vinden in Wald et al., 1986 (passief roken), Schwartz, 1994 (deeltjesvormige luchtverontreiniging) en Hasselblad, 1992 (NO<sub>2</sub> door gasapparaten in huis).

Het effect van de aannames op de geaggregeerde blootstellings-effect maat kan door gevoeligheids-analyse geëvalueerd worden. Bij de meta-analyse van de ozonstudies verdiende het omrekenen van blootstellingsdefinitie op basis van 'valide' aannames met betrekking tot

<sup>6</sup> Overigens dient juist bij ozon rekening gehouden te worden met aanzienlijke verschillen in fenomenologie tussen verschillende lokaties. Zo komen in de VS en Nederland piekconcentraties van ozon vooral in de late middag en vroege avond voor, terwijl in Athene in de ochtend al hoge ozonconcentraties worden gemeten die aanhouden tot tegen zonsondergang.

de verhouding tussen 1-uursmaxima, 8- en 24-uursgemiddelden de voorkeur boven opname als co-variabele in het meta-regressiemodel. Opname van co-variabelen gaat in sterke mate ten koste van het statistische onderscheidingsvermogen (Preller et al., 1996).

Ook de indeling in blootstellingscategorieën beïnvloedt de effectschatting en de bijbehorende mate van statistische significantie. Dit geldt zowel voor blootstellingscategorieën die gebaseerd zijn op de blootstellingsverdeling (zoals de mediaan, of quantielen) als voor vastgestelde (bijvoorbeeld wettelijke) grenswaarden. Aan de hand van aannames over de blootstellingsverdeling kan middels heranalyse getracht worden op overeenkomstige indelingen uit te komen (Berlin, 1993; Greenland, 1987).

#### *Vergelijkbaarheid van de typen effectschattingen*

De vergelijkbaarheid van effectschattingen kan in het geding zijn als verschillende typen van analyses gehanteerd worden. Zo werd de relatie tussen fotochemische luchtverontreiniging en ziekenhuisbezoek beschreven aan de hand van Poisson regressiemodellen (exponentieel verband tussen opnames en blootstelling) of aan de hand van lineaire regressiemodellen. Uit Poisson modellen volgt een RR (met standard error) voor de toename van de blootstelling met een bepaalde concentratie. Bij lineaire modellen is het RR afhankelijk van het referentieniveau. Voor de lineaire modellen is er gezocht naar een benadering van het RR en de daarbij behorende standard error, zodat studies die gebaseerd zijn op verschillende analyse-methoden met verschillende effectschattingen toch beide in dezelfde meta-analyse verwerkt kunnen worden (Preller et al., 1996). Door middel van een gevoeligheids-analyse kan beoordeeld worden of er verschil is in deze effectschattingen.

### **3.5. Data-analyse**

Onder data-analyse wordt hier verstaan het zoveel mogelijk aggregeren van de in afzonderlijke studies waargenomen associaties tussen blootstelling en effect, bijvoorbeeld door toepassing van een statistisch model of het toetsen van hypothesen. Het is hierbij de vraag in hoeverre resultaten van verschillende studies nog wel op een zinvolle wijze samengevat kunnen worden in geaggregeerde effectschatters. Dit hangt af van de mate van heterogeniteit in effectschattingen, het verschil in grootte van deze effectschattingen, maar evenzeer van de wetenschappelijke context. Beschrijven de verschillende studies hetzelfde verschijnsel in vergelijkbare termen? Vaak kunnen verschillen in effectschattingen worden toegeschreven aan variatie in studiekenmerken, zoals variatie in de definitie van blootstelling of van respons.

Bij meta-analyses wordt heterogeniteit in de blootstellings-effectrelaties vaak geëvalueerd aan de hand van op  $X^2$ -statistiek gebaseerde testen (Colditz et al., 1995). Een dergelijke test geeft globaal aan of de variatie in effectschattingen op toeval zou kunnen berusten. De beslissing of effectschattingen worden gecombineerd hangt behalve van deze tests ook af van de wetenschappelijke context. Soms kan een verschil in RR van 25% als relatief klein worden beschouwd, in andere gevallen juist als een groot verschil (zie hoofdstuk 4).

Met verklaarde heterogeniteit dient rekening gehouden te worden bij de data-analyse door middel van stratificatie of toepassing van multivariate regressiemodellen. In het eerste geval wordt er voor elke groep een aparte geaggregeerde effectschatting gegeven, bijvoorbeeld voor



ouderen of voor kinderen. In het tweede geval wordt de effectschatting beschreven als functie van een of meer aspecten die als bron van heterogeniteit worden gezien (bijvoorbeeld niveau van blootstelling).

Recentelijk wordt het gebruik van een 'random effects model' (REM) gepropageerd als beste aggregatiemodel, gegeven de heterogeniteit tussen studies. Bij zo'n model wordt er één geaggregeerde blootstellings-effect schatting geconstrueerd, waarbij de variatie tussen de studies wordt vertaald in een groot betrouwbaarheidsinterval voor de effectschatting. Dit doet geen recht aan de 'ware' situatie, en draagt het gevaar in zich dat de geaggregeerde schatting gehanteerd wordt voor situaties waarin deze sterk afwijkend is. Daarnaast is het gebruik van een REM feitelijk pas toegestaan als de variatie zuiver 'random', toevallig, is, en niet als de variatie wordt veroorzaakt door verschillen in populatie of studiedesign en dergelijke. In andere woorden: het random effects model mag eigenlijk pas toegepast worden als het resultaat daarvan niet al te veel afwijkt van dat van toepassing van een 'fixed effects model' (Greenland, 1995).

Potentiële bronnen van heterogeniteit dienen geëvalueerd te worden. Welke die bronnen zijn, is afhankelijk van de specifieke blootstellings-effect relatie. In aanmerking komen onder andere: 'study-design', de manier waarop blootstelling is gekarakteriseerd (surrogaat of werkelijke meting, cumulatief of gemiddelde blootstelling etc.), gezondheidsuitkomst, controle voor confounding, geografie, populatie-kenmerken, meetmethoden etc.

Als heterogeniteit wordt veroorzaakt door interne validiteitsproblemen, door bias of confounding in de individuele studies, dan dienen betreffende studies niet opgenomen te worden in de meta-analyse.

Voor de statistische procedures zij verwezen naar hoofdstuk 4.

### *Blootstellingskarakterisering*

Studies waarin zeer uiteenlopende blootstellingsniveaus zijn onderzocht, kunnen sterk variërende effectschattingen geven. Dit kan samenhangen met een niet-lineair blootstellings-effect verband, maar ook met een verschil in gevoeligheid van de populatie (denk aan werknemers versus de algemene bevolking). Combineren van studies kan dan ongeoorloofd zijn.

De blootstellingsmaat die de voorkeur verdient voor gebruik in de meta-analyse is die maat die in veruit de meeste studies gehanteerd wordt. Dit hoeft niet in alle gevallen de meest nauwkeurig gedefinieerde maat te zijn, of de maat waarop interventie gepleegd kan worden. Door aannames te doen over de relatie tussen ongelijke maten (vloerbedekking aanwezig en hoeveelheid huisstofmijtallergeen) kan het aantal studies dat opgenomen kan worden in de meta-analyse vergroot worden.

Blair et al (1995) bevelen aan die blootstellingsmaat te kiezen die de persoonlijke blootstelling of lichaamsbelasting het dichtst benadert. Hierbij dient men wel na te gaan of een dergelijke maat 'terug te vertalen' is naar preventieve risico-reducerende maatregelen, bijvoorbeeld gericht op het terugdringen van concentraties van een schadelijke stof in lucht of in voedingsmiddelen. Zo is een risicoverhoging gerelateerd aan bloedserumwaarden van een bepaalde metaboliet van een milieuverontreinigende stof niet altijd even makkelijk te vertalen naar concentratie van die stof in een of enkele milieucompartimenten en vervolgens naar

emissiereducerende maatregelen. De concentratie van de stof in een milieucompartiment kan dan een geschiktere blootstellingsmaat zijn.

Indien studies duidelijk verschillen wat betreft duur van de blootstelling, intensiteit, frequentie of route van blootstelling, dient geëvalueerd te worden of dit een bron van heterogeniteit is. In dit geval kan het gebruik van ordinale gradiënten (hoog, midden, laag) overwogen worden. Het combineren van studies waarin de blootstelling juist alleen 'gekwantificeerd' wordt op basis van dit soort indelingen is in het algemeen echter een punt van discussie. Laag in het ene geval kan hoog in het andere zijn. De noodzaak om dit soort studies op te nemen zal onder meer bepaald worden door het aantal bruikbare studies met meer kwantitatieve blootstellingsmaten. Bij het wel combineren zal de onzekerheid in de blootstellings-effect schatting altijd sterk benadrukt moeten worden.

#### *Gebruik van kwaliteitsscores*

In meta-analyse van RCTs en observationeel onderzoek hanteert men soms kwaliteitsscores om het resultaat van de betere studies zwaarder mee te wegen in de geaggregeerde blootstellings-effect schatting (Morris, 1992).

Het gebruik van kwaliteitsscores als weging voor afzonderlijke studieresultaten is twijfelachtig, omdat er geen objectieve maten zijn om de feitelijke kwaliteit te meten en de beoordeling bovendien afhankelijk is van rapportage van studie-aspecten in publikaties. Veronderstellingen over verschillen in kwaliteit van studies kunnen beter beoordeeld worden op basis van gevoeligheidsanalyse (bijvoorbeeld 'quality component analysis', Greenland, 1994, Morris et al., 1992, zie ook paragraaf 2.3.6).

#### *Aggregatie van de studie uitkomsten*

Na evaluatie van bovenstaande kunnen samenvattende maten berekend worden. Bij duidelijke, verklaarbare, heterogeniteit dienen er ofwel gestratificeerde effectschattingen berekend te worden, ofwel er dient meta-regressie-analyse te worden gedaan, waarbij 'verklarende' variabelen in het model worden opgenomen. In hoofdstuk 4 wordt een voorstel gedaan voor daarbij te volgen statistische procedures.

#### *Gevoeligheidsanalyses*

De mate waarin bepaalde aannames, bepaalde kwaliteitsaspecten en de invloed van bepaalde afzonderlijke studies doorwerken in het resultaat van de meta-analyse wordt beoordeeld aan de hand van gevoeligheidsanalyses.

### **3.6. Presentatie, interpretatie en communicatie van resultaten.**

Een meta-analyse moet gezien worden als een kwantitatief en 'objectief' review. Uit de presentatie zal dit moeten blijken. Deze dient zo uitgebreid te zijn dat de meta-analyse in beginsel reproduceerbaar is. De opzet van de presentatie wijkt dus niet af van die van een ander onderzoek en dient dus sleutelaannames, belangrijke onzekerheden (bias en confounding) en de uitkomsten van gevoeligheidsanalyses zoveel mogelijk te expliciteren. Een aantal elementen is specifiek voor meta-analyse.

Bij de presentatie zal aandacht besteed moeten worden aan de methodologische beperkingen van de afzonderlijke studies en de doorwerking daarvan in de meta-analyse. Hoe wordt vermeden dat meta-analyse tot mega-bias leidt? Hoe waterdicht moeten exclusiecriteria zijn? Welke bias introduceert meta-analyse extra? Publikatiebias kan de uitkomst van kwantitatieve review sterk beïnvloeden en dient dus expliciet te worden beoordeeld.

Een ander heikel punt is in hoeverre een geaggregeerde schatter met ‘vernauwd’ betrouwbaarheidsinterval als *de* samenvattende kwantitatieve uitkomst van de meta-analyse kan worden gepresenteerd. Men zou zo veel mogelijk willen vermijden dat ten behoeve van de onderbouwing van beleid met dergelijke schatters ‘aan de haal gegaan wordt’ zonder dat enige notie wordt genomen van de nauwgezette, onderliggende kwantitatieve analyse. Een belangrijke remedie tegen dergelijk gebruik van de resultaten van meta-analyse is in ieder geval een expliciete presentatie van de analyse van de invloed van studiekenmerken, bijvoorbeeld de resultaten van stratificering of regressie-analyse. Er wordt dan een reeks van effectschatters gepresenteerd, die meer gevoel geeft voor de bijbehorende onzekerheden en variabiliteit (zie bijvoorbeeld uitwerking van de meta-analyse ozon en ziekenhuisbezoek in paragraaf 4.6).

De volgende aspecten dienen beschreven te worden:

#### *Introductie*

- achtergrond van de blootstellings-effect relatie (literatuuroverzicht)
- onderzoeksvraag
- mogelijke bronnen van heterogeniteit, eventuele te toetsen hypothesen met betrekking tot de invloed van co-variabelen (karakteristieken op studieniveau)

#### *Methoden*

- zoekstrategie
- in- en exclusiecriteria van studies
- benaderen van auteurs van primaire artikelen
- statistische procedures

#### *Resultaten*

- overzicht van studies die wel en niet in de meta-analyse zijn opgenomen met alle belangrijke kenmerken
- keuze voor bepaalde blootstellings-effectschattingen voor zover die van de vooropgezette criteria afwijken
- herberekening van effectschattingen en variantie
- indien geoorloofd een geaggregeerde effectschatting, eventueel gestratificeerd weergegeven of als resultaat van multi-variante regressie-analyse
- resultaten invloed/gevoeligheidsanalyses

#### *Tabellen en figuren*

- overzicht studie met belangrijkste kenmerken, puntschatting en spreiding
- tabel en figuur met resultaten van meta-analyses
- eventueel trechterplot of varianten daarvan om publikatiebias zichtbaar te maken

*Discussie*

- gevonden kwantitative verband(en) tussen blootstelling en effect, en de mate waarin variabelen die hierop van invloed zijn
- mogelijke publikatie-bias
- afwijkingen van de onderzoeksvraag
- verklaringen voor heterogeniteit, bijvoorbeeld op grond van studiekekenmerken
- beoordeling gevoeligheid van aannames en de invloed van specifieke studies op uitkomsten

## 4. STATISTISCHE TECHNIEKEN BIJ META-ANALYSE

### 4.1. Inleiding

In dit hoofdstuk worden de achtergronden en algoritmen besproken voor aggregatie van blootstelling-responsverbanden, gebaseerd op de a priori aanname van heterogeniteit tussen de studies (zie paragraaf 2.3.5). In de hier gepresenteerde analyses zal onderscheiden worden naar de (log) Normale verdeling, voor studies gebaseerd op de (log)-odds ratio<sup>7</sup>. Tevens zal de mogelijkheid van opname van karakteristieken van de studie besproken worden en de eventueel beschikbare software voor dergelijke analyses.

Bij analyses wordt uitgegaan van heterogeniteit tussen onderzoeken. De reden hiervoor is deels theoretisch, deels praktisch. Het is uit filosofisch oogpunt uiterst onwaarschijnlijk dat een verzameling van onderzoeken, waarin verschillende populaties (of soms zelfs species) onder uiteenlopende omstandigheden met verschillende methoden worden onderzocht, telkens één en dezelfde kwantitatieve relatie zal opleveren, alsof het een soort natuurwet betreft. Het ligt veel meer voor de hand dat er een range van waarden zal blijken te gelden voor verschillende populaties, of verscheidene omstandigheden. Dus zelfs als de meta-analist een onuitputtelijke reeks van studies tot zijn beschikking had, dan nog zou hij eerder een *verdeling* van het blootstelling-responsverband in kwestie vinden, dan één ware waarde. Een meer praktisch argument is dat de beschikbare tests op heterogeniteit van een verzameling uitkomsten een zeer laag onderscheidend vermogen hebben.

Methoden om ten behoeve van meta-analyse afzonderlijke studie-uitkomsten te aggregeren, kunnen worden afgeleid met de frequentistische statistische theorie van gemengde lineaire modellen of de (empirisch) Bayesiaanse statistische theorie. De in dit hoofdstuk behandelde 'schatters' leiden, bij voldoende grote omvang van de meta-analyse, tot vrijwel identieke uitkomsten, zij het dat de interpretatie strikt genomen zal verschillen. Dit geldt met name voor de schatters van het betrouwbaarheidsinterval. De strikte frequentistische definitie van een 95% betrouwbaarheidsinterval is dat het interval met een kans van 0.95 de 'ware' waarde zal bevatten. Er wordt geen enkele uitspraak gedaan *waar* in het interval de meest waarschijnlijke waarde ligt: alle waarden in het interval zijn even waarschijnlijk. De Bayesiaanse interpretatie van datzelfde interval is geheel anders: de beide eindpunten zijn een schatter voor het 0,025 respectievelijk 0,975 quantiel van de posteriori verdeling van de schatter, het midden van het interval is het 50% punt in geval van de normale verdeling. Een dergelijke interpretatie van een betrouwbaarheidsinterval sluit beter aan bij de intuïtie.

Een ander fundamenteel voordeel van de Bayesiaanse aanpak is dat de uitkomst van een 'volgende' studie voorspeld kan worden in het licht van de reeds gevonden uitkomsten. Men denke hierbij aan de klassieke situatie, waarin, ondanks het ontbreken van gegevens over ziekte-incidentie voor één geografische regio, toch een aannemelijk schatting gemaakt kan

<sup>7</sup>

Hierbij zij opgemerkt dat vooral bij meta-analyse van RCTs soms afwijkende statistische procedures noodzakelijk zullen blijken, afhankelijk van de gehanteerde uitkomstmaten (niet-Normaal-verdeeld, ordinaal, proportioneel) en steekproefomvang.

worden op grond van de wel beschikbare gegevens voor de omringende regio's. Deze denktrant voortzettend, kan vervolgens de betrouwbaarheid van de incidentieschatters voor alle afzonderlijke regio's verbeterd worden door betrouwbaarheid te 'lenen' bij buurregio's (het zogenaamde 'smoothen' bij het karteren van gezondheidsgegevens, zie Heisterkamp, 1993). Een soortgelijk verhaal geldt ook voor reeds uitgevoerde epidemiologische studies. Stel, men is geïnteresseerd in de resultaten van een in Nederland uitgevoerde studie. De studie wordt op zichzelf niet voldoende geacht. Gegeven de andere studieresultaten, hetzij van de buitenlandse, hetzij van andere Nederlandse studies of van beiden, kan men met behulp van Bayesiaanse technieken een herwaardering uitvoeren met behulp van de uitgevoerde meta-analyse.

Anders gezegd, de meta-analist kan zich gedragen als een voorspeller: 'Wat zou, met een gegeven betrouwbaarheid, de uitkomst zijn van een nieuw onderzoek (gegeven de uitkomsten van de reeds uitgevoerde studies)?', of, 'Wat zou, in het licht van alle hier beschouwde studies, de uitkomst zijn van een bestaande studie, als deze opnieuw zou worden uitgevoerd?'. In de statistische literatuur spreekt men wel van 'borrowing strength from other studies', of in geval van het maken van gezondheidskaarten, 'het gladstrijken' van de waarnemingen. Deze uitkomsten worden voorzien van een voorspellingsinterval, of met een andere naam 'tolerantieinterval' (Heisterkamp et al., 1993).

Voor de praktische analyse maakt het vervolgens niet uit of men dit doet volgens de Empirisch Bayesiaanse technieken (EB), of dat men gebruik maakt van de technieken voor gemengde lineaire modellen met behulp van ML (maximum 'likelihood') of REML ('restricted maximum likelihood'). Alleen technieken gebaseerd op een volledig Bayesiaanse analyse geven iets afwijkende uitkomsten, in hoofdzaak omdat deze rekening houden met de eindigheid van de omvang van de meta-analyse, ofwel het aantal studies. Hiervoor heeft men speciale programmatuur nodig, bijvoorbeeld BUGS. In het vervolg zullen we daar naar verwijzen voor volledig Bayesiaanse analyses. Voor de andere typen van analyses heeft men of geen speciale programmatuur nodig, of kan men met wat handigheid een en ander zelf programmeren, bijvoorbeeld met SAS Proc mixed en SAS-IML. Berkey geeft een SAS-IML procedure voor het uitvoeren van een meta-analyse met een random-effects model, gebaseerd op een algoritme beschreven door Morris (Berkey et al., 1995; Morris, 1983). Voor de hier beschreven analyses zijn ook functies geschreven in de taal S-Plus voorhanden.

## 4.2. Praktische aanbeveling

In het vervolg van dit rapport wordt uitgegaan van het gebruik van het random-effects model voor meta-analyse. In de literatuur wordt dan vaak gewerkt met de methode van Laird, als eerste orde benadering (Laird, 1986; DerSimonian en Laird, 1986). De REML en ML methode wijken slechts van Laird af in de wijze van schatten van de (rest) variantie, waarbij de REML methode een zuivere schatter geeft. De EB methode is rekentechnisch identiek aan de REML methode. De volledig Bayesiaans methode tenslotte doet recht aan de onzekerheid die in de rest-variantie aanwezig is, en weegt deze mee. De laatste drie methoden staan het gebruik van extra informatie over de studies toe met behulp co-variabelen. Aangezien de REML

methode ofwel standaard in statistische pakketten aanwezig is, of eenvoudig is bij te programmeren gaat hier de voorkeur naar uit. Met de gebruikte methoden kan men bij bestaande studies het gemiddelde en de variantie corrigeren in het licht van de overige studies zowel zonder als met gebruik van co-variabelen; 95% betrouwbaarheidsintervallen volgen daaruit (posteriori schattingen). Het kan echter ook zo zijn dat men als het ware de uitkomst van een nieuw op te zetten studie wil voorspellen, hetgeen in het bijzonder voor gezondheidsverkenningen van belang kan zijn. In dat geval kan men een schatting geven van de te verwachten waarde  $\theta_{\text{true}}$  en de voorspelde variantie  $\sigma_{\text{true}}$ , hieruit kan een 95% tolerantieinterval worden berekend (a priori schatting). In het geval er geen co-variabelen worden gebruikt, is een schatting voor  $\theta_{\text{true}}$  te geven voor  $\mu_{\text{true}}$  en wordt het tolerantie interval  $\theta_{\text{true}} \pm 1.96 \sigma_{\text{true}}$  verkregen door deze te schatten met  $\mu_{\text{true}} \pm 1.96 \sqrt{(\sigma_{\text{true}}^2 + \sigma_{\mu}^2)}$ .

Voor de gebruikte algoritmen wordt verwezen naar appendix I en II. In appendices III en IV staan source codes van de gebruikte programmatuur voor SAS-IML respectievelijk S-Plus.

### 4.3. Meta-analyse zonder extra informatie

In deze paragraaf gaat het over meta-analyse van studies, waarbij eventueel ‘verklarende’ kenmerken van onderzoeken niet in beschouwing worden genomen. In een volgende paragraaf zal ook het opnemen van co-variabelen behandeld worden. Aangenomen wordt dat de uitkomstmaten van de studies bij benadering normaal verdeeld zijn met gemiddelden en variantie zoals binnen elke studie vastgesteld. Dit hoeft niet per sé zo te wezen. Voor studies met betrekking tot prevalenties, gestandaardiseerde mortaliteitsratios en dergelijk zij verwezen naar literatuur over ‘Disease mapping’ (Clayton et al., 1987; Heisterkamp et al., 1993).

Veronderstel, er zijn  $n$  studies met een uitkomstmaat die bij benadering normaal verdeeld is, bijvoorbeeld de log-oddsratio of een andere verschilmaat. De uitkomst van de  $i$ -de studie, uitgedrukt in deze verschilmaat, wordt aangeduid met de letters  $y_i$  en  $E(y_i) = \mu_i$ ,  $i = 1 \dots n$ . Tevens wordt verondersteld dat van elke studie de standaardfout van die maat is gerapporteerd als  $\sigma_i$ ,  $i = 1 \dots n$ . Voorts wordt verondersteld dat ‘de ware waarde’ een normale verdeling volgt met gemiddelde  $\mu_{\text{true}}$  en standaardfout  $\sigma_{\text{true}}$ . Het doel is nu deze laatste waarden te schatten en een tolerantie interval op te stellen voor  $\mu_{\text{true}}$ , daarbij onderscheid makend tussen de variantie binnen studies en variantie tussen studie-uitkomsten. De variantie tussen de studies kan veroorzaakt worden door verschillen in populaties, verschillen in het uitvoeren van de studies of in de gebruikte analysemethode. Hierop wordt nog nader teruggekomen.

Indien de  $n$  studies allemaal betrekking hebben op *de* populatie waarvoor betere informatie gewenst is, of indien alle studies betrekking hebben op ‘buitenlandse’ populaties, waarbij geen enkele de bedoelde populatie exact representeert, kan worden volstaan met het geven van  $\mu_{\text{true}}$ ,  $\sigma_{\text{true}}$  en een tolerantie interval.

Het kan echter ook zijn dat een (of meer) studies betrekking hebben op bijvoorbeeld Nederland. In dat geval kunnen uitkomsten van Nederlandse studies, of van studies die de bedoelde populatie het beste representeren, gebruikt worden om tot een betere puntschatting en schatting van het betrouwbaarheids-interval te komen, in het licht van de overige studies. Ge-

geven deze uitkomsten, wordt als het ware een gecorrigeerde voorspelling voor de studie-uitkomst gemaakt.

*Tabel 1 Resultaten van Berlin (1993) en van verschillende methoden van meta-analyse.*

Bron (eerste auteur)		variantiegecorrigeerde regressiecoëfficiënt	standaard fout
1	Paffenbarger	0.07520	0.07298
2	Sartwell	-0.09076	0.16481
3	Lees	0.07705	0.09146
4	Ravnihar	-0.06213	0.18726
5	Paffenbarger	0.12678	0.05555
6	Kelsey	-0.22474	0.14268
7	Harris	-0.05963	0.09844
8	Brinton	0.06292	0.04149
9	Janerich	0.06388	0.04055
10	Pike	0.18606	0.13185
11	Vessey	0.02443	0.04597
12	Rosenberg	0.04043	0.04278
13	Miller	0.07081	0.10570
14	Meirik	0.14802	0.06204
15	Ellery	-0.04248	0.11683
16	La Vecchia	0.03451	0.11228
17	Paul	-0.11584	0.01081
18	Lee	0.07065	0.08961
19	CASH	0.00252	0.01927
20	McPherson	0.09147	0.06910
21	McPherson	-0.04075	0.05640
22	Miller	0.28717	0.08282
23	Jick	-0.01308	0.14784
24	Olsson	0.18536	0.07594

Methode	$\mu_{true}$	se van $\mu_{true}$	$\sigma_{true}$
Fixed effects model	-0.0434	0.0079	nvt
Laird	0.0450	0.0262	0.1001
REML	0.0433	0.0215	0.0737
Maximum Likelihood	0.0429	0.0211	0.0711

#### 4.4. Een uitgewerkt literatuurvoorbeeld

Deze studie is ontleend aan meta-analyse van blootstelling-responsrelaties met betrekking tot het gebruik van de 'pil' en het risico op borstkanker, zoals beschreven in een publikatie van Berlin et al. (1993). De gegevens zijn ter illustratie opgenomen in tabel 1.

De auteurs berekenen 0.0450 als geaggregeerde effectmaat onder heterogeniteits veronderstelling met 0.0262 als standaardfout. Dezelfde schatting onder niet-heterogeniteits veronderstellingen ('fixed effects') is -0.0434 met 0.0079 als standaardfout, hetgeen een kwalitatief andere uitspraak oplevert. De schatting voor het 'random effects' is naar alle waarschijnlijkheid gedaan met de niet-iteratieve schatter van DerSimonian en Laird (1986). In dezelfde tabel zijn de ML-schatter en de REML-schatter voor het random effect vergeleken met die van DerSimonian en Laird.

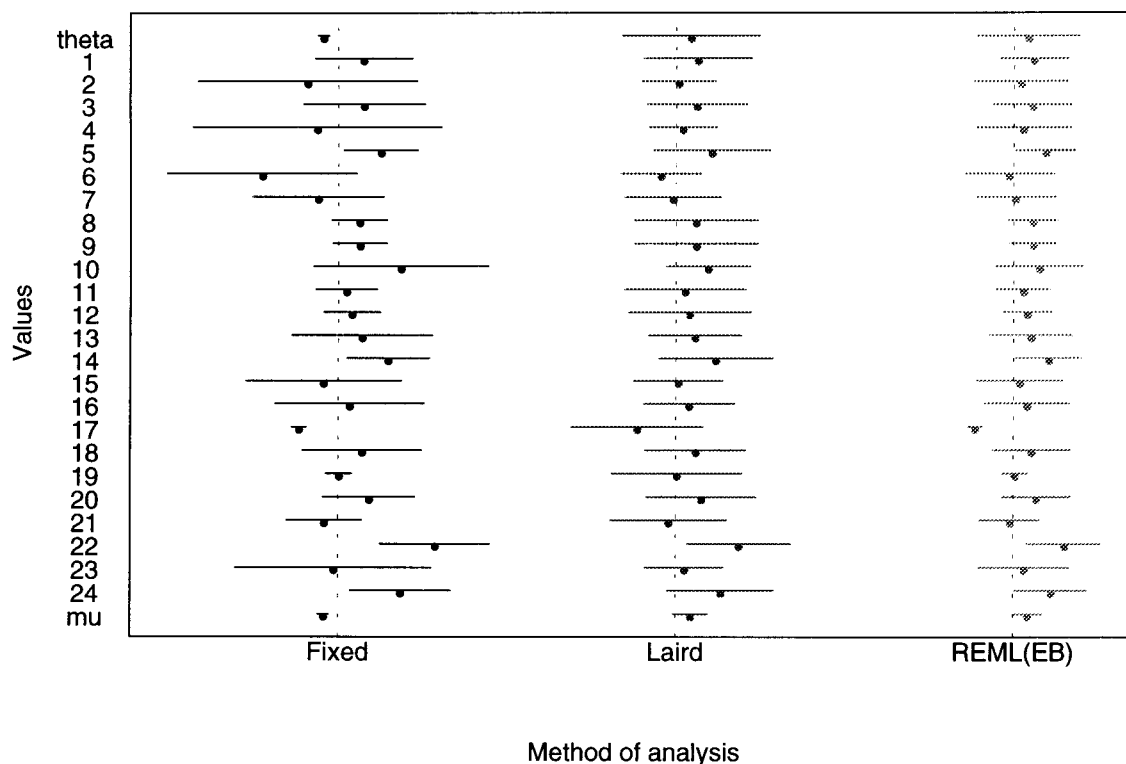
In figuur 3 zijn de studieresultaten weergegeven met hun betrouwbaarheidsintervallen, en de 'posteriori'<sup>8</sup> schattingen van de studieresultaten met de betrouwbaarheidsintervallen. Het

<sup>8</sup> Posteriori wil hier zeggen: gegeven de resultaten van de overige studies.



meer gebruikelijke tolerantie-interval voor de ware waarde (de geaggregeerde schatting) is ook weergegeven. Duidelijk is te zien dat de posteriori schattingen dicht bij het algemeen gemiddelde liggen naarmate de onzekerheid in de studie-uitkomst in kwestie groter is. Gaat het om betrouwbare studie-uitkomsten, dan heeft de informatie uit de overige studies veel minder invloed en wijkt de posteriori schatting nauwelijks af. De puntschattingen voor de posterior studie-uitkomst van REML (EB) en Laird ontlopen elkaar niet veel.

### Berlin data



*Figuur 3 Metaanalyse van tabel 1. Gemiddelden en 95%-betrouwbaarheidsintervallen van studies en het algemeen gemiddelde (mu). Voor de ware waarde (theta) zijn voor Laird en REML 95% tolerantie intervallen gegeven. De studiegemiddelden behorende bij REML (EB) en Laird zijn posteriori gemiddelden De cijfers verwijzen naar de regelnummers 1 t/m 24 in de tabel.*

#### 4.5. Meta-analyse met extra informatie: opnemen van studiekenmerken

Heterogeniteit in de uitkomsten van afzonderlijke studies zal vaak voor een deel worden ‘verklaard’ door verschillen in studiekenmerken. Door het meenemen van ‘verklarende’ studiekenmerken als co-variabele in het analyse-model kan men in beginsel de nauwkeurigheid van de geaggregeerde schatter vergroten. Bovendien kan men de geaggregeerde schatter, zo men dat wil, toespitsen op de Nederlandse situatie door de relatie af te leiden voor de meest hiermee overeenkomende studiekenmerken.

Co-variabelen kunnen alleen meegenomen worden als ze voor elke afzonderlijke studie op gelijke wijze kunnen worden gedefinieerd. Dit is echter bijna nooit het geval. Daarom worden hier co-variabelen op meta-niveau geïntroduceerd, dat wil zeggen co-variabelen die de studies als steekproefeenheid ('sampling unit') karakteriseren (geen co-variabelen die direct betrekking hebben op de steekproefeenheden van de afzonderlijke studie).

Van Houwelingen (1993) onderscheidt de volgende typen van co-variabele op meta-niveau:

- Co-variabelen op meta-niveau die betrekking hebben op de populatie, bijvoorbeeld leeftijd(-sopbouw), geslacht(-sverhoudingen), gemiddelde niveau van 'binnen-studie co-variabele(n)', 'klinisch kenmerken' van patientgroepen of zelfs species (mens, rat, muis).
- Co-variabelen op meta-niveau die betrekking hebben op de studie, bijvoorbeeld: regio (of land), jaar of periode van onderzoek, type onderzoek (RCT, non-randomized, ander onderzoek), en kwaliteit (in categoriën). Deze zou men eventueel als vaste effecten kunnen beschouwen.

In principe kan dan een meta-model worden beschreven als een regressiemodel met design matrix  $\mathbf{X}$ , een vector van random regressie coëfficiënten  $\mu$  en een extra toevalscomponent die de variatie rondom de 'ware' parameter waarden beschrijft. Principieel is de schattingsmethode onder vooronderstellingen van (log-) normaliteit van de onderhavige verdelingen voor zowel coëfficiënten als de extra toevalsbijdrage uit te voeren met elke procedure die op REML is gebaseerd (o.a. ProcMixed voor SAS) en waarbij men de co-variantiematrix kan specificeren, zonder dat de volledige set van waarnemingen nodig is.

Onderstaande schets van het algoritme is ontleend aan Dumouchel en Harris (1983, 1990). Zij analyseerden verscheidene studies van blootstellings-responsrelaties met betrekking tot de carcinogeniteit van chemische verbindingen. Het betrof zowel mensen als verschillende proefdierspecies. Door de karakteristieken van deze studie onder te brengen in een design matrix (van volle rang) zoals gebruikelijk is bij een meer-weg variantie analyse, konden zij meta-analyses uitvoeren waarbij aggregatie van op het oog zeer ongelijksoortige studies toch een betere effect schatting opleverde. Hier wordt één EB-methode besproken waarbij de ware variantie als een vrije parameter geschat zal worden. Ook deze analyse is in principe geheel met op REML gebaseerde programmatuur uit te voeren. Eerst worden het model en het algoritme beschreven, daarna volgt een voorbeeld.

Bij de in 1983 gepubliceerde EB-benadering van DuMouchel wordt er van uitgegaan dat de studies onderling onafhankelijk zijn. In een meer recentere publikatie van (1990), wordt ook afhankelijkheid toegelaten. Dit laatste is van belang bij het gebruik van uitkomstmaten, die weliswaar afkomstig zijn van één studie, maar bij sub-populaties zijn waargenomen, en die bijvoorbeeld een gemeenschappelijke rest variantie gebruiken. Voor uitgewerkte algoritmen verwijzen we naar appendix II.

In principe is er geen verschil met de analysemethode zonder co-variabelen. De designmatrix  $\mathbf{X}$  van het regressiemodel zonder co-variabelen bestaat uit een kolom van louter 1-en, de te schatten regressieparameter is het algemeen gemiddelde  $\mu$  en de te schatten heterogeniteit

is de variantie tussen studiegemiddelden  $\sigma^2$ . De varianties binnen studies zijn weergegeven als uitkomsten, evenals de studiegemiddelden. Een eenvoudige co-variabele is er een die onderscheid maakt tussen bijvoorbeeld twee regio's, Noord-Amerika en Zuid-Amerika. Men kan de van het regressiemodel zo kiezen dat interpretatie gemakkelijk is. Zo kan men de eerste kolom met 1-en vullen en de tweede kolom met een 0 als het een studie uit Noord-Amerika betreft en een 1 als het om een studie in Zuid-Amerika gaat. Het algemeen gemiddelde  $\mu$  en het 'verschil' tussen Noord- en Zuid-Amerika  $\delta$  vormen dan de regressieparameters. Bij gebalanceerde studies (gelijke varianties) is dit een juiste interpretatie; bij meta-analyses wordt deze interpretatie echter vaak verstoord door verschillende weggingen tussen studies. Daarom kan een andere parametrisering soms beter interpreteerbaar zijn. Een alternatief is er een waarbij men in de eerste kolom 1-en zet voor studies uit Noord-Amerika, en 0-en elders, en in de tweede kolom 0-en voor studies uit Noord-Amerika en 1-en elders. In dat geval is de eerste parameter het gemiddelde van een studie in Noord-Amerika  $\mu_1$ , en de tweede het gemiddelde in Zuid-Amerika  $\mu_2$ . Beide modellen zijn uit oogpunt van modelaanpassing identiek, alleen de interpretatie van de parameters is verschillend. Heeft men meer dan twee categorieën dan gaat men in principe op dezelfde wijze te werk, dat wil zeggen er worden uitsluitend 0-en en 1-en gebruikt om categorieën aan te wijzen, ofwel men creëert dummy-variabelen.

Het is ook mogelijk een continue co-variabele te gebruiken, bijvoorbeeld de gemiddelde leeftijd per studiepopulatie of het gemiddelde niveau van de buitenluchtconcentratie van ozon in elk van de afzonderlijke studies. Dit laatste kan handig zijn omdat in studies vaak wordt gekeken naar de toename van een bepaald effect ten opzichte van een gemiddelde niveau.

De gegeven variantie van de uitkomstmaat binnen de studies (de kwadraten van de standaardafwijkingen) wordt voorgesteld door een matrix met alleen op de diagonaal waarden ongelijk nul en genoteerd als de matrix  $C$ . Indien de studies niet statistisch onafhankelijk zijn, dient een covariantiematrix te worden gebruikt. Indien alle studies in wezen hetzelfde fenomeen met dezelfde nauwkeurigheid beschrijven dan is de co-variantie matrix  $V$  met *ware* tussenstudie variantie ook weer een diagonaal matrix met op de diagonaal de onbekende waarde  $\sigma^2$ . Bij praktische toepassing wordt op de diagonaal dan 1 gezet, de SAS procedure schat dan deze ware tussenstudie-variantie. De gebruikte SAS procedure is overigens niet gebaseerd op de exacte vergelijkingen in appendix II, maar op een benadering van Morris (1983) beschreven in Berkey et al. (1995). De door Berkey et al. gepubliceerde procedure is echter gewijzigd met betrekking tot de benadering van de schatter voor  $\sigma^2$ . Een benadering gebaseerd op REML, zoals voorgesteld door Morris verdient de voorkeur boven de door hen gebruikte benadering gebaseerd op ML. Berkey et al. gebruikten bewust de ML-benadering omdat 'de REML schatter een te lage variantie opleverde', we kunnen deze bevinding niet bevestigen. Tevens wordt een 'goodness of fit' criterium geïntroduceerd: AIC, Akaike's Information Criterion, gedefinieerd met behulp van de (EB) loglikelihood, zie appendix II voor details. AIC levert niet het model op met de beste 'fit', maar een model dat het beste voorspelt. AIC geeft een (relatieve) indruk van de mate waarin het model past; hoe kleiner het AIC hoe beter de voorspelling. AIC corrigeert voor onzekerheid in de parameters die een rol spelen bij het voorspellen. Anders dan bij het gebruik van  $R^2$  kan de AIC een optimale waarde bereiken

zonder dat alle variabelen zijn opgenomen, men wordt als het ware gestraft voor over-fitten van het model door te veel parameters op te nemen. Aangezien de AIC gedefinieerd is met behulp van de log-likelihood, die op een constante na vast ligt, is er geen absolute waarde van de AIC aan te geven. Een vuistregel is dat modellen niet van elkaar te onderscheiden zijn als de AIC niet meer verbeterd (of verslechterd) met een waarde groter dan 2 *per vrijheidsgraad* (uitgaande van een Chi-kwadraat benadering).

De benadering wordt in SAS gebruikt omdat het op dit moment niet mogelijk is om de loglikelihood zoals gedefinieerd in appendix II rechtstreeks te maximaliseren.

Tenslotte kan ook hier weer een gecorrigeerd studiegemiddelde en een 95% betrouwbaarheidsinterval worden berekend. Bij het opzetten van een nieuwe studie, of bij het doen van een voorspelling, kan men gegeven de co-variabelen een regressie schatting maken van het nieuw te verwachten studie gemiddelde en een 95% tolerantieinterval berekenen.

#### **4.6. Ongelijke variantie tussen studies en afhankelijkheid tussen studie uitkomsten**

Soms heeft men een goede reden om aan te nemen dat sommige studies een veel grotere variantie hebben dan andere (bijvoorbeeld afhankelijk van de toegepaste detectiemethode). Een apriori schatting zou dan kunnen zijn dat voor een aantal studie de variantie wel tweemaal zo groot is. In dat geval vult men op de desbetreffende plaats van de diagonaal van  $V$  een 2 in. Vaker zal voorkomen dat men weliswaar een aantal studieuitkomsten meeneemt, maar dat de studieresultaten niet onafhankelijk zijn. Bijvoorbeeld, twee uitkomsten in een publikatie zijn waargenomen bij dezelfde algemene studiepopulatie, maar hebben betrekking op twee sub-populaties. Meestal worden de uitkomstmaten, en hun bijbehorende varianties, uitgerekend met dezelfde *restvariantie* uit de studie. Indien de studie een co-variantie opgeeft tussen beide uitkomsten, dan kan men de covariantie tussen beide studie uitkomsten in de matrix  $C$  opnemen, door op de desbetreffende '*off-diagonal*' plaats die covariantie in te vullen. Indien deze onbekend is kan men volstaan het uitvoeren van de meta-analyses met verschillende waarden tussen 0 en 1, en afhankelijk van het resultaat deze als een range weer te geven.

#### **4.7. Een uitgewerkt voorbeeld**

Bij wijze van voorbeeld worden hier kort enkele resultaten weergegeven van een door Preller uitgevoerde meta-analyse van tijdserie-analyses met betrekking tot de relatie tussen ozonconcentraties in de buitenlucht en aantallen ziekenhuisopnames of bezoeken aan de polikliniek in verband met luchtwegklachten. De data zijn afkomstig van observationele studies, die tamelijk heterogeen zijn voor wat betreft opzet, analyse methode en uitkomst variabele (zie tabel 2).

In tabel 3 zijn de resultaten van diverse meta-analyses weergegeven, tezamen met het Akaikes Information Criterion, een maat voor de fit van het model. Het AIC geeft een indruk of het zinvol is specifieke co-variabelen mee te nemen in het model. Met andere woorden: of het model met die co-variabele meer van de variatie verklaart dan zonder. Een waarde die, per

vrijheidsgraad, minimaal 2 minder negatief is dan in het model zonder die co-variabele geeft aan dat het model beter is.

Tabel 2. Overzicht van bruikbare effectschattingen voor meta-analyse van de relatie tussen ozonconcentraties in de buitenlucht en ziekenhuisbezoek (Preller et al., 1996)

Nr	auteur	locatie	analyse	aandoening	populatie	M	type	logRR	se
2	Schwartz	Detroit	Poisson	COPD	> 65	24	opn	0.27	0.10
4	Schwartz	Birmingham	Poisson	COPD	> 65	1	opn	0.07	0.06
6	Romieu	Mexico	Poisson	astma	kind	1	poli	0.36	0.07
7	White	Atlanta	Poisson	astma	kind	1	poli	0.29	0.19
9	Schwartz	NewHaven	Poisson	alle	> 65	24	opn	0.14	0.07
10	Schwartz	Tacoma	Poisson	alle	> 65	24	opn	0.36	0.13
12	Schwartz	Spokane	Poisson	alle	> 65	1	opn	0.44	0.22
13	Castellsague	Barcelona	Poisson	astma	volw	1	poli	-0.04	0.11
27	Leon	Londen	Poisson	alle	kind	8	opn	0.05	0.03
28	Leon	Londen	Poisson	alle	volw < 65	8	opn	0.12	0.03
29	Leon	Londen	Poisson	alle	> 65	8	opn	0.10	0.03
35	Dab	Parijs	Poisson	COPD	alle	8	opn	0.11	0.06
36	Dab	Parijs	Poisson	astma	alle	8	opn	-0.13	0.05
55	Schouten	Nederland	Poisson	alle	alle	8	opn	0.07	0.01
37	Delfino	Montreal	lineair	alle	alle	8	opn	0.08	0.06
39	Thurston	Toronto	lineair	alle	alle	1	opn	0.13	0.08
41	Thurston	Buffalo	lineair	alle	alle	1	opn	0.12	0.06
42	Thurston	New York	lineair	alle	alle	1	opn	0.04	0.01
45	Burnett	Ontario	lineair	alle	alle	1	opn	0.02	0.01
46	Cody	New Jersey	lineair	astma	alle	8	poli	0.25	0.08

Het model zonder co-variabelen levert een 'gemiddelde' relatief risico (RR) op van 1.12, met een 95% betrouwbaarheidsinterval van 1.07 tot 1.18, en is dus statistisch significant. Overigens wordt de meta-analyse gedaan op de log van de RR's die uit de oorspronkelijke studies komen. De berekeningen geven dan als resultaat de log van het RR. Voor de berekening van de betrouwbaarheidsgrenzen geldt hetzelfde.

Tabel 3. Resultaten meta-analyses van de relatie tussen ozonconcentraties in de buitenlucht en ziekenhuisbezoek (Preller et al., 1996)

	DF	rest.	AIC	subgroep	RR	95% bthi	
algemeen gemiddelde	20	0.00865	-56.7		1.12	1.07	1.18
blootstellings-middelingstijd	18	0.00857	-44.6	1-uur	1.12	1.03	1.21
				24-uur	1.26	1.08	1.47
				8-uur	1.09	0.98	1.23
analyse <sup>1</sup>	17	0.00946	-37.0	Poisson	1.09	1.01	1.18
				lineair	1.11	0.99	1.24
aandoening <sup>1</sup>	16	0.00959	-31.0	alle	1.08	0.99	1.17
				astma	1.15	1.03	1.29
				COPD	1.09	0.94	1.28
opname <sup>1</sup>	17	0 <sup>2</sup>	-39.9	opname	1.07	1.05	1.09
				poli.	1.32	1.21	1.43

<sup>1</sup> resultaten gecorrigeerd voor de verschillen in middelingstijd; weergegeven is het voor het verschil van 50 ppb in 8-uursgemiddelde concentratie.

<sup>2</sup> fixed effects model

DF=aantal vrijheidsgraden, rest.= rest variantie AIC=Akaike's Information Criterion, RR=relatief risico met het 95%-betrouwbaarheidsinterval (bthi).

Dat opname van de blootstellingsmiddelingstijd in het model tot een betere 'fit' leidt, ligt voor de hand. Immers, in termen van blootstelling is een stijging van het 24-uursgemiddelde voor de ozonconcentratie met 50 ppb aanzienlijk belangrijker dan eenzelfde stijging voor de maximale 1-uurswaarde.

Bij opname van de middelingstijd verandert het AIC met 11.1, bij afname van het aantal vrijheidsgraden met 2. Dus het verschil is 5.5 per vrijheidsgraad. Dit model is dus duidelijk beter dan het model zonder de blootstellingsmiddelingstijd. In alle volgende modellen is de blootstellingsmiddelingstijd daarom gehandhaafd. Het toevoegen van het kenmerk 'analyse' maakt het model beter dan met alleen blootstellingsmiddelingstijd: de AIC verandert met bijna 8 (met 1 vrijheidsgraad verschil). Ook het model met type aandoening naast blootstellingsmiddelingstijd is beter dan dat met middelingstijd alleen. Dit geldt in iets mindere mate ook voor het kenmerk 'opname'; er is een geringe verbetering in de voorspellende waarde van het model. Wel lijkt daarbij sprake te zijn van een duidelijk verschil in RR voor opname versus poliklinische behandeling.

## 5. SLOTOPMERKINGEN

De kwestie of meta-analyse ook bij observationeel epidemiologisch onderzoek kan worden toegepast, is eigenlijk een al gepasseerd station. Men doet het. Belangrijker is welk oogmerk men ermee heeft. Meta-analyse is een nuttig hulpmiddel om epidemiologisch onderzoek systematisch, kwantitatief en in samenhang te beoordelen en samen te vatten, ook als het observationeel onderzoek betreft. Het is echter *niet* in de eerste plaats een instrument om een preciezere, geaggregeerde uitkomstmaat af te leiden. De grote mate van heterogeniteit in studiekekenmerken, waarvan bij observationeel onderzoek meestal sprake is, maakt aggregatie soms zelfs onzinnig. Bovendien is de suggestie van precisie van de nauwere betrouwbaarheidsintervallen rond de geaggregeerde schatter misleidend; de nauwere betrouwbaarheidsintervallen brengen de veelal aanzienlijke, inherente beperkingen van observationeel onderzoek niet in rekening. Meta-analyse moet veeleer gezien worden als een middel om te leren van verschillen tussen studies, patronen te herkennen en tot synthese te komen, waar dat mogelijk, zinvol en inhoudelijk verantwoord is. Op grond van analyse van de rol van studiekekenmerken kan inzicht verkregen worden in die factoren of omstandigheden die gezondheidsschade of juist het succes van zorg mede bepalen. Om met enige nuance terug te komen op de in hoofdstuk geformuleerde redenen om naar het 'paardemiddel' te grijpen, meta-analyse kan een hulpmiddel zijn bij:

- Systematische, consistente, omvattende en navolgbare literatuurevaluatie.
- Het verfijnen van de schatting van het kwantitatieve verband tussen blootstelling aan risicofactoren en gezondheidseffecten (of de inzet van zorgelementen en gezondheidswinst) door statistische aggregatie van uitkomsten van afzonderlijke studies; daarbij krijgen grote, preciezere studies een groter gewicht. Nauwgezette analyse moet uitwijzen of dergelijke aggregatie verantwoord is; statistische zowel als inhoudelijke analyse van heterogeniteit tussen studies vormt daarom een belangrijk onderdeel van meta-analyse.
- Het analyseren van tegenstrijdigheden tussen studie-uitkomsten, niet om enig pleit te beslechten, maar meer om het inzicht te vergroten in die studiekekenmerken die ervoor verantwoordelijk zijn. Analyse van de invloed van deze studiekekenmerken kan waardevolle informatie opleveren met betrekking tot effectmodificerende factoren of omstandigheden. Hierbij kan gedacht worden aan duur of intensiteit van blootstelling, klinische of demografische kenmerken van (patiënt)groepen, geografische of temporele verschillen, interactie met andere (risico)factoren, definities van blootstelling en respons, enzovoorts. Een dergelijke analyse kan van belang zijn voor de beoordeling van effecten van preventief beleid of zorgelementen in, bijvoorbeeld, de Nederlandse situatie, zowel als voor het opzetten van nieuw onderzoek.

Alleen in de geest van Münchhausen trekt men zich aan de eigen haren uit het moeras; de kwaliteit van een meta-analyse staat of valt bij de kwaliteit van de afzonderlijke studies. Geen statistische procedure van aggregatie zal ooit in staat zijn de fundamentele beperkingen van de afzonderlijke observationele studies te compenseren. Bovendien vereisen de hierboven be-

schreven procedures veelal een aanzienlijk hoeveelheid studies (effectschattingen) om een zinvolle vergelijkende analyse te kunnen maken.

De in hoofdstuk 4 beschreven procedure voor meta-regressie-analyse biedt aantrekkelijke mogelijkheden voor vergelijkende analyse. Bovendien kan met de procedure in beginsel, afhankelijk van de kwantiteit en kwaliteit van beschikbare input, een op de Nederlandse situatie toegesneden effectschatter worden berekend. Alleen door het uitvoeren van verschillende voorbeeldanalyses kan men evenwel werkelijk gevoel krijgen voor de betekenis van de geaggregeerde uitkomsten. In het vervolg van het project zullen daarom nog enkele voorbeeldstudies worden uitgewerkt, die afzonderlijk worden gerapporteerd.



## 6. LITERATUUR

- Anello Ch, Fleiss JL. Exploratory or analytic meta-analysis: should we distinguish between them? *J Clin Epidemiol* 1995; 48: 109-16.
- Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. *Epidemiol* 1993;4:218-28.
- Berkey CS, Hoaglin DC, Mosteller F, Colditz G.A., A random-effects regression model for meta-analysis, *Stat Med* 1995; 14: 395-411.
- Bernardinelli L, Montomoli C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Stat Med* 1992; 11: 983-1007.
- Blair A, Burg J, Foran J et al. Guidelines for application of meta-analysis in environmental epidemiology. *Regulatory toxicology and pharmacology* 1995; 22: 189-197.
- Clayton DG, Kaldor J. Empirical Bayes Estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; 43: 671-81.
- Colditz-GA; Burdick-E; Mosteller-F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary [see comments]. *Am J Epidemiol* 1995; 142(4): 371-82.
- Cook DJ, Sackett DL, Spitzer WO. Methodological guidelines for systematic review of randomized control trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol* 1995; 48: 167-71.
- DerSimonian R, Laird N. *Meta-analysis in Clinical trials, Controlled Clinical trials* 1986; 7: 177-88. New York, Elsevier
- Dickerson K, Berlin JA. *Meta-analysis: State-of-the-Science*. *Epidemiol Rev* 1992; 14: 154-76.
- Dumouchel WH, Harris JE. Bayes Methods for Combining the results of cancer studies in humans and other species (with discussion). *JASA* 1983; 78: 293-315.
- Dumouchel WH. Bayesian meta analysis. In: *Statistical methodology in the pharmaceutical sciences*. Berry DA (eds) 1990; 509-529, New York.
- Dumouchel WH. Meta-analysis for dose-response models. *Stat Med* 1995; 14: 679-85.
- Eysenk HJ. An exercise in mega-silliness. *Am Psychol* 1978; 33: 517.
- Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J-Clin-Epidemiol* 1991; 44: 127-39.
- Frumkin H, Berlin J. Asbestos exposure and gastrointestinal malignancy: review and meta-analysis. *Am J Ind Med* 198; 14: 79-95.
- Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res* 1976; 5: 3-8.
- Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol-Rev* 1987; 9: 1-30.
- Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994; 140: 290-6.

- Greenland S. Invited commentary: Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994; 140: 300-1.
- Greenland S. Invited commentary: can meta-analysis be salvaged? *Am J epidemiol* 1994; 140: 783-7.
- Greenland, S. Point Counterpoint: Meta-analysis of observational studies - Reply. *Am J Epidemiol* 1995; 142: 781-782.
- Hasselblad V, Eddy DM, Kotchmar DJ. Synthesis of environmental evidence: nitrogen dioxide epidemiology studies. *J Air Waste Manage Assoc* 1992; 42(5): 662-71.
- Hasselblad V. Meta-analysis of environmental health data. *Sci Total Environ* 1995; 160/161: 545-58.
- Heisterkamp S.H., Doornbos G., Gankema M., Disease mapping using empirical Bayes and Bayes Methods on mortality statistics in the Netherlands, *Stat Med* 1993, 12, pp1895-1913
- Hollander AEM de, Hoogenveen RT, Jansen J. Rekenen aan effecten van preventie met behulp van het populatie attributieve risico. Bilthoven: RIVM rapportnummer 266750 001 (in voorbereiding).
- Houwelingen JC van, R. Brand, J. Hermans, M.J.N.C. Kreise (eds), Meta-analyse, Boerhave Commissie voor Post Academisch Onderwijs, Leiden, 1995
- Kleijnen J, De Vet HCW, Rinkel GJE, Keirse MJN. De Cochrane Collaboration; systematische overzichten van kennis uit gerandomiseerd onderzoek. *Ned Tijdschr Geneesk* 1995; 139: 1478-82.
- Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; 38: 963.
- Li-Y, Powers TE, Roth HD. Random-effects linear regression meta-analysis models with application to the nitrogen dioxide health effects studies. *J-Air-Waste-Manage-Assoc* 1994; 44: 261-70.
- Loon EML van, Melis PHAM, Polder MD, Swartjes FA, Zwan CW van der. Kwaliteitsborging binnen literatuuronderzoek. Bilthoven: RIVM, rapportnummer 980301 004, 1996.
- Maclure M. Demonstration of deductive meta-analysis: ethanol intake and risk of myocardial infarction. *Epidemiol Rev* 1993; 15:328-51.
- Maclure M. Karl Popper and his unending quest: an epidemiologic interpretation. *Epidemiol* 1995; 6: 331-3.
- Midgette AS, O'Connor GT, Baron JA, et al. Effects of intravenous streptokinase on early mortality in patients with suspected acute myocardial infarction. A meta-analysis by anatomic location of infarction. *Ann Intern Med* 1990; 113: 961-8.
- Morris CN. Parametric Empirical Bayes Inference: Theory and Applications, *J Am Stat Assoc* 78, 381, p 47-55, 1983
- Morris RD; Audet AM; Angelillo IF; Chalmers TC; Mosteller F. Chlorination, chlorination by-products, and cancer: a meta-analysis. *Am J Public Health* 1992; 82(7): 955-63 [erratum in *Am J Public Health* 1993;83(9):1257].
- Morris RD. Meta-analysis in cancer epidemiology. *Environ Health Perspect* 1994;102: 61-6.
- Olde Hartevelde P van. Passief drinken en leverkanker; een kritische beschouwing van het bewijs. *Nieuw Tijdsch Geneesk* 1994; 1: 32-5.
- Olkin I. Invited commentary: Re: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994; 140: 297-9.
- Pearson K. Report on certain enteric fever inoculation statistics. *Br Med J* 1904; 3: 1243-6.

- Petitti DB. Of babies and bathwater. *Am J Epidemiol.* 1994; 140: 779-82.
- Preller EA, Hollander AEM de, Heisterkamp S, Lezenne Coulander C de. A protocol for meta-regression analysis of studies in environmental epidemiology: ozone and hospital admissions [abstract]. *Epidemiol* 1996; 7: S90.
- Preller EA, Hollander AEM de, Heisterkamp S, Lezenne Coulander C de. Ziekenhuisopname en polyklinische behandeling in relatie tot ozonconcentraties in de buitenlucht; een meta-regressie analyse. Bilthoven: RIVM, rapportnummer 263610 003, 1996.
- Preller EA, Hollander AEM de. Beroepsmatig blootstelling aan uitlaatgassen van dieselmotoren in relatie tot het risico op longkanker; een voorbeeld van meta-regressie-analyse. Bilthoven: RIVM rapportnummer 263610 004 (in voorbereiding).
- Schwartz J. Air pollution and daily mortality: a review and meta analysis. *Environ-Res.* 1994; 64(1): 36-52.
- Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol.* 1994; 140(9): 771-8
- Shapiro S. Re: Is there or is there ain't no baby? *Am J Epidemiol* 1994; 140: 788-91.
- Shapiro S. Point Counterpoint: Meta-analysis of observational studies - Reply. *Am J Epidemiol* 1995; 142: 780-1
- Shinton R, Beevers G. Meta-analysis of the relation between cigarette smoking and stroke. *Br Med J* 1989; 298: 789-94.
- Spitzer WO. Meta-analysis: unanswered questions about aggregating data. *J-Clin-Epidemiol* 1991; 44: 103-7.
- Spitzer WO (eds). The Potsdam International Consultation on Meta-analysis. *J Clin Epidemiol* 1995; 48.
- Smith, A.H.; Lopipero, P.A.; Barroga, V.R. Meta-analysis of studies of lung cancer among silicotics. *Epidemiol* 1995; 6: 617-624.
- Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* 1994; 309:1351-55.
- Wald NJ, Nanchahal K, Thompson SG, Cuckle HS. Does breathing other people's tobacco smoke cause lung cancer? *Br Med J* 1986; 293: 1217-22.
- Washburn EP, Orza MJ, Berlin JA, Nicholson WJ, Todd AC, FrumkinH, Chalmers TC. Residential proximity to electricity transmission and distribution equipment and risk of childhood leukemia, childhood lymphoma, and childhood nervous system tumors: systematic review, evaluation, and meta-analysis. *Cancer Causes Control* 1994; 5(4): 299-309.
- Wong O, Raabe GK. Critical review of cancer epidemiology in petroleum industry employees, with quantitative meta-analysis by cancer site. *Am J Ind Med* 1989; 15: 283-310.



## APPENDIX I. NIET-ITERATIEVE EB SCHATTER EN ITERATIEVE EB- EN ML-SCHATTERS VOOR $\theta$ , BETROUWBAARHEIDS- EN TOLERANTIE-INTERVALLEN

Hieronder volgt het veel geciteerde niet-iteratieve algoritme van DerSimonian en Laird (1986) voor de bepaling van het algemeen gemiddelde, en een iteratief algoritme voor de ML-schatter. Het essentiële verschil tussen beide schatters is de wijze waarop  $\sigma_{true}^2$  wordt geschat.

### *Niet-iteratieve schatter voor $\mu$ (DerSimonian en Laird)*

Bereken  $\omega_i = \sigma_i^2$  en  $\mu_{fixed} = \frac{\sum_i y_i / \omega_i}{\sum_i 1 / \omega_i}$ , schat vervolgens de  $\sigma_{true}^2$  met

$$\sigma_{true}^2 = \frac{\sum_i \omega_i (y_i - \mu_{fixed})^2 - (n-1)}{\sum_i \omega_i - \frac{\sum_i \omega_i^2}{\sum_i \omega_i}}$$
 of nul, indien voorgaande uitdrukking kleiner is.

Bereken vervolgens  $\omega_i^* = \sigma_i^2 + \sigma_{true}^2$  en  $\hat{\mu}_{Laird} = \frac{\sum_i y_i / \omega_i^*}{\sum_i 1 / \omega_i^*}$  met  $\hat{\sigma}_{\mu_{Laird}} = \frac{1}{\sqrt{\sum_i 1 / (\sigma_i^2 + \hat{\sigma}_{true}^2)}}$

### *Tolerantie interval voor het gemiddelde*

Een 95% betrouwbaarheidsinterval voor  $\mu$  wordt gegeven door  $\hat{\mu} \pm 1.96 \hat{\sigma}_{\mu}$  en een tolerantie interval voor  $\theta_{true}$  wordt dan gegeven door  $\hat{\mu} \pm 1.96 \sqrt{\hat{\sigma}_{true}^2 + \hat{\sigma}_{\mu}^2}$ .

### *Iteratieve schatter voor $\theta$ (ML)*

Hierbij moet worden gebruik gemaakt van een (numeriek) algoritme om hetzij het maximum van de loglikelihood voor  $\sigma_{true}^2$  te vinden of een nulpunt te zoeken voor de afgeleide van deze log-likelihood.

1. Stap k=0. Zet  $\sigma_{true(0)}^2 = \bar{\sigma}^2$ , with  $\bar{\sigma}^2 = \frac{\sum_j \sigma_j^2}{n}$

2. Stap k=k+1. Bepaal  $\omega_i = \sigma_i^2 + \sigma_{true(k-1)}^2$

3. Bereken  $\mu_{true(k)} = \frac{\sum_i y_i / \omega_i}{\sum_i 1 / \omega_i}$

4. Bereken  $\sigma_{true(k)}^2$  waarvoor geldt dat

$$lik(\sigma_{true}^2) = -\sum_i \frac{(y_i - \mu_{true}^{k-1})^2}{\sigma_i^2 + \sigma_{true}^2} - \sum_i \log(\sigma_i^2 + \sigma_{true}^2).$$
 maximaal wordt. Of als alternatief bereken

$$\text{het nulpunt van : } dlik(\sigma_{true}^2) = \sum_i \frac{(y_i - \mu_{true}^{k-1})^2}{(\sigma_i^2 + \sigma_{true}^2)^2} - \sum_i \frac{1}{\sigma_i^2 + \sigma_{true}^2}$$

5. Als  $|\mu_{true(k-1)} - \mu_{true(k)}| \leq \varepsilon$  ga verder naar 6, anders herhaal vanaf 2
6. Zij  $\hat{\mu}_{true}$  en  $\hat{\sigma}_{true}^2$  de laatst verkregen waarden. De standaardfout behorende bij  $\hat{\mu}_{true}$  wordt gegeven door  $\hat{\sigma}_{\mu} = \frac{1}{\sqrt{\sum_i 1/(\sigma_i^2 + \hat{\sigma}_{true}^2)}}$  en een 95% betrouwbaarheids interval voor

$\mu_{true}$  gebaseerd op de Normale verdeling (een groot aantal studies) wordt gegeven door:  $[\hat{\mu}_{true} - 1.96 \cdot \hat{\sigma}_{\mu}, \hat{\mu}_{true} + 1.96 \cdot \hat{\sigma}_{\mu}]$ .

Het verschil zit hier dus in de wijze waarop  $\hat{\sigma}_{true}^2$  wordt berekend, hetgeen een iteratief algoritme tot gevolg heeft.

De REML of EB schatter verschilt van de ML-schatter in een toevoeging aan de loglikelihood die rekening houdt met de onzekerheid in  $\beta$ . Voor deze schatter dient men in 4 de likelihood te vervangen door:

$$lik_{EB}(\sigma_{true}^2) = -\frac{1}{2} \sum_i \frac{(y_i - \mu_{true}^{k-1})^2}{\sigma_i^2 + \sigma_{true}^2} - \frac{1}{2} \sum_i \log(\sigma_i^2 + \sigma_{true}^2) - \frac{1}{2} \log\left(\sum_i (\sigma_i^2 + \sigma_{true}^2)^{-1}\right)$$

te zoeken van:

$$dlik_{EB}(\sigma_{true}^2) = \frac{1}{2} \sum_i \frac{(y_i - \mu_{true}^{k-1})^2}{(\sigma_i^2 + \sigma_{true}^2)^2} - \frac{1}{2} \sum_i \frac{1}{\sigma_i^2 + \sigma_{true}^2} + \frac{1}{2} \frac{\sum_i (\sigma_i^2 + \sigma_{true}^2)^{-2}}{\sum_i (\sigma_i^2 + \sigma_{true}^2)^{-1}}$$

Indien de binnenstudie variantie verwaarloosbaar klein is gaat de schatter  $\mu_{true}^{EB}$  over in een *ongewogen* studiegemiddelde en wordt  $\sigma_{trueEB}^2$  de gebruikelijke steekproef variantie tussen de studiegemiddelden (delen door n-1), zoals is af te leiden uit de laatste uitdrukking. Onder soortgelijke aannamen wordt de variantie geschat volgens de ML-methode afgeleid door de som van de kwadratische afwijkingen ten opzichte van het gemiddelde te delen door n in plaats van n-1.

### **Een gecorrigeerd studie gemiddelde en bijbehorend betrouwbaarheidsinterval**

Een nieuwe schatting voor het gemiddelde van studie  $j$  kan op de volgende wijze worden berekend. Bepaal het gewicht  $w_j$  voor deze studie met,  $w_j = \frac{\hat{\sigma}_{true}^2}{\hat{\sigma}_{true}^2 + \sigma_j^2}$ .

Het nieuwe gemiddelde is dan:  $\hat{\mu}_j^{new} = \hat{\mu}_{true} \cdot (1 - w_j) + y_j \cdot w_j$  en de bijbehorende standaard fout  $\hat{\sigma}_j^{new} = \sqrt{\hat{\sigma}_{\mu}^2 \cdot (1 - w_j)^2 + \omega_j \cdot w_j^2}$  met  $\omega_j = \sigma_j^2 + \hat{\sigma}_{true}^2$  en het bijbehorende 95% betrouwbaarheidsinterval wordt op de gebruikelijke wijze gevormd met  $[\hat{\mu}_j^{new} - 1.96 \cdot \hat{\sigma}_j^{new}, \hat{\mu}_j^{new} + 1.96 \cdot \hat{\sigma}_j^{new}]$ . Dit betrouwbaarheidsinterval houdt overigens geen rekening met het gegeven dat de gewichten geschat zijn. In de volgende paragraaf is daar wel rekening mee gehouden.

Hierbij kan worden aangetekend dat indien  $\sigma_j^2$  groot is ten opzichte van  $\hat{\sigma}_{true}^2$  het gewicht  $w_j$  klein wordt en de nieuwe schatting voor studie  $j$  naar het algemeen gemiddelde wordt *gekrompen*, in de literatuur worden deze schatters met 'shrinkage estimators' aangeduid. Indien

het omgekeerde het geval is, de studie heeft relatief weinig onzekerheid, dan zal de nieuwe informatie weinig invloed hebben.

Ten opzichte de gebruikelijke methode van schatten bij heterogeniteit is dit laatste in feite de enige uitbreiding in de toepassing. Bij toepassingen in 'Disease mapping' is dit een normale gang van zaken.

## APPENDIX II. BAYES, EB- EN ML-SCHATTERS VOOR $\mu$ EN TOLERANTIE-INTERVALLEN BIJ GEBRUIK VAN CO-VARIABLEN

Laat weer gegeven zijn de studie-uitkomst maten in de vorm van een kolomvector  $\mathbf{Y} = \{y_i\}$ , en de bijbehorende varianties zijn gegeven in de vorm van een diagonaal matrix  $\mathbf{C} = \text{diag}(c_1^2, \dots, c_n^2)$ . We gebruiken nu een afwijkende notatie die het mogelijk maakt een en ander wat algemener te formuleren. De niet geobserveerde parameters noteren we met  $\theta$ ,  $\beta$ , en  $\sigma$ . De parametervector  $\theta$  is de conditionele verwachting van  $\mathbf{Y}$ . De parametervector  $\beta$  zijn de parameters waarin we geïnteresseerd zijn. Veronderstel dat we een designmatrix  $\mathbf{X}$  hebben die deze parametervector  $\beta$  verbindt met de conditionele verwachtingen  $\theta$  van de studies. De matrix  $\mathbf{X}$  heeft  $n$  rijen en  $k$  kolommen, de vector  $\beta$  is van lengte  $k$ . Tevens veronderstellen we dat de variantie-covariantie matrix van  $\beta$ , de matrix  $\mathbf{V}$  bekend is. Het model formuleren we dan als een hiërarchische rij van normale verdelingen:

$$(Y|\theta, \beta, \sigma) \sim N(\theta, \mathbf{C})$$

$$(\theta|\beta, \sigma) \sim N(\mathbf{X} \cdot \beta, \sigma^2 \mathbf{I})$$

De variantie  $\sigma^2$  heeft hier weer dezelfde rol als in de vorige paragraaf, en aangezien we alleen de ML of de EB variant beschouwen hoeven we hierover geen verdere aannamen te doen.

Het verband met de schatter uit de vorige paragraaf is onmiddellijk duidelijk als we voor  $\mathbf{X}$  een matrix nemen bestaande uit één kolom met enen en voor  $\beta$  de (scalair)  $\mu_{true}$  lezen en  $\mathbf{V} = \sigma_{true}^2 \cdot \mathbf{I}$ . De matrix  $\mathbf{C}$  is gelijk aan de diagonaal matrix met als diagonaalelementen  $c_i^2 = \sigma_i^2$ . Met behulp van  $\mathbf{X}$  en  $\mathbf{V}$  kunnen we echter ook structuur brengen in de studies.

We geven nu twee schatters voor de parameters, een gebaseerd op de ML-methode, en een EB-schatter volgens Dumouchel & Harris (1983). De reden daarvoor is dat de wijze waarop deze schatters worden berekend deels parallel lopen. Alleen de laatste stap, het schatten van de variantie  $\sigma^2$  is verschillend. Overigens heeft dit invloed op de eerste stappen aangezien deze variantie een essentieel onderdeel van de geschatte gemiddelden is.

We negeren even het probleem voor het schatten van  $\sigma^2$  en laten dit symbool even in de vergelijkingen staan.

Maak een gewichten matrix  $\mathbf{W}_\sigma = (\mathbf{C} + \sigma^2 \cdot \mathbf{I})^{-1}$ . Aangezien de matrices ingeval van onafhankelijke studies diagonaal zijn, is dit in dit geval ook weer een diagonaal matrix, met elementen  $1/(\sigma_i^2 + \sigma^2)$ . De schatters voor (eventuele vector)  $\beta$  en de co-variantie (matrix) verkrijgt men door substitutie van de juiste  $\sigma^2$  in de volgende formules:

$$\begin{aligned}\hat{\beta}_\sigma &= (\mathbf{X}^T \mathbf{W}_\sigma \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_\sigma \mathbf{Y} \\ \hat{V}_\sigma &= (\mathbf{X}^T \mathbf{W}_\sigma \mathbf{X})^{-1}\end{aligned}\quad (1)$$

Waarin we de gewogen kleinste kwadratenschatters in herkennen.

Voor de ML-schatters moeten we de volgende loglikelihood in  $\sigma^2$  maximaliseren:

$$\begin{aligned}\mathbf{S}_\sigma &= \mathbf{W}_\sigma - \mathbf{W}_\sigma \mathbf{X} (\mathbf{X}^T \mathbf{W}_\sigma \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_\sigma \\ \text{dan} \\ l_{ML}(\sigma) &= \frac{1}{2} \log(|\mathbf{W}_\sigma|) - \frac{1}{2} \mathbf{Y}^T \mathbf{S}_\sigma \mathbf{Y} \\ &= \frac{1}{2} \log(|\mathbf{W}_\sigma|) - \frac{1}{2} \mathbf{Y}^T \mathbf{W}_\sigma \mathbf{Y} + \frac{1}{2} \mathbf{Y}^T \mathbf{W}_\sigma \mathbf{X} \hat{\beta}\end{aligned}\quad (2)$$

We kunnen de vergelijkingen (1) en (2) op verschillende wijze oplossen.

- Een methode is iteratief: Uitgaande van het maximaliseren van  $l_{ML}(\sigma)$  zoals deze in de laatste regel van (2) is opgeschreven, als functie van  $\beta$ . (2). In dat geval kiest men een startwaarde voor  $\sigma^2$  lost vervolgens (1) op, en maximaliseert  $l_{ML}(\sigma)$ , waarna men de gevonden waarde in (1) terug substitueert. Men kan doorgaan totdat voldoende nauwkeurigheid is bereikt.
- De tweede methode vereist geen enkele iteratie. Maximaliseer  $l_{ML}(\sigma)$ , zoals deze in (2) in de eerste regel is opgeschreven met behulp van de matrix  $\mathbf{S}$ , en los vervolgens (1) op. In dat geval moet men de matrix  $\mathbf{S}$  uitrekenen, met een statistisch pakket dat matrix-vector berekeningen toestaat, is dat geen probleem.

Dit kan met elk geschikt statistisch pakket. (N.B. de symbolen  $||$  geven de determinant van de matrix aan, voor een diagonaal matrix is dit het produkt van de diagonaal elementen).

Voor de EB schatter dienen we een ander likelihood te maximaliseren, dit hangt samen met het feit dat we ook de onzekerheid in de  $\beta$  meenemen bij het schatten van de variantie, in plaats van een eenvoudige substitutie te plegen.

Voor de EB-schatter van  $\sigma^2$  maximaliseren we:

$$\begin{aligned}l_{EB}(\sigma) &= \frac{1}{2} \log(|\mathbf{W}_\sigma|) - \frac{1}{2} \log(|\mathbf{X}^T \mathbf{W}_\sigma \mathbf{X}|) - \frac{1}{2} \mathbf{Y}^T \mathbf{S}_\sigma \mathbf{Y} \\ &= \frac{1}{2} \log(|\mathbf{W}_\sigma|) - \frac{1}{2} \log(|\mathbf{X}^T \mathbf{W}_\sigma \mathbf{X}|) - \frac{1}{2} \mathbf{Y}^T \mathbf{W}_\sigma \mathbf{Y} + \frac{1}{2} \mathbf{Y}^T \mathbf{W}_\sigma \mathbf{X} \hat{\beta}\end{aligned}\quad (3)$$

Het verschil met (2) is hier een extra term waarin we een functie van de variantie van de geschatte  $\beta$ 's in herkennen. Ook hier kan men voor iteratie of 'n een-staps methode kiezen.

Een tolerantie interval voor  $\beta$  kan men op analoge wijze verkrijgen als in de vorige paragraaf, maar nu doorgebruikmaking van de wortel uit diagonaal element(en) van  $\hat{V}_\sigma$  uit (1). Extra aandacht geven we nog aan het schatten van de verwachte waarden van een studie in het licht van de overige studies.



***Een gecorrigeerd studie gemiddelde en bijbehorend betrouwbaarheidsinterval alsmede een nieuw studiegemiddelde en een tolerantie-interval***

Ingeval van een EB aanpak kunnen we een gecorrigeerd studie gemiddelde en de variantie als volgt schatten:

$$\hat{\theta}_{EB} = \left[ \mathbf{I} + \sigma_{EB}^{-2} \mathbf{C} \left( \mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \right]^{-1} \mathbf{Y}$$

$$\hat{V}_{EB,\theta} = \left[ \mathbf{C}^{-1} + \sigma_{EB}^{-2} \left( \mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \right]^{-1}$$
(4)

Waarna men met behulp van de normale verdeling eenvoudige tolerantie intervallen kan bepalen.

Baseert men zich op de ML-schatters dan zien bovenstaande gelijkheden er als volgt uit:

$$\hat{\theta}_{ML} = \hat{V}_{ML,\theta} \left[ \mathbf{C}^{-1} \mathbf{Y} + \sigma_{ML}^{-2} \mathbf{X} \hat{\beta}_{ML} \right]^{-1}$$

$$\hat{V}_{ML,\theta} = \left[ \mathbf{C}^{-1} + \sigma_{ML}^{-2} \mathbf{I} \right]^{-1}$$
(5)

Ingeval men de uitkomst van een nieuw op te zetten studie willen voorspellen en het bijbehorende tolerantie-interval kunnen we op de gebruikelijke manier de regressievergelijking invullen met  $\tilde{\mathbf{X}}^T \hat{\beta}$  voor zekere waarden van de co-variabelen ingevuld in de kolom vector  $\tilde{\mathbf{X}}$  en de hierbij behorende variantie uitrekenen  $\hat{\sigma}_{res}^2 + \left( \tilde{\mathbf{X}}^T \mathbf{V}(\hat{\beta}) \tilde{\mathbf{X}} \right)$  ten behoeve van het tolerantie interval.

***Goodnes-of-fit: Akaike's Information Criterion (AIC)***

Als Goodness-of-fit criterium gebruiken we het Akaike's Information Criterion (AIC). Dit criterium houdt rekening met het aantal variabelen dat in een regressievergelijking wordt opgenomen en door minimalisatie kiest men het model dat het best voorspelt. Anders dan de gewone  $R^2$  kan een optimum worden bereikt zonder dat alle variabelen zijn opgenomen in het model. Het AIC kan worden gedefiniëerd als:

$$AIC = -2 \cdot \max(\log - \text{likelihood}) + 2 \cdot \# \text{ parameters}$$

Een optimale waarde over een set van parameters wordt dus bereikt bij een vermindering van de (-2maal) de likelihood bijdrage en een verhoging van het aantal parameters, het optimum is dus een minimum. Het AIC is echter eigenlijk op een constante na gedefiniëerd (ook ontbreekt soms de factor -2), daarom moet men het altijd als een relatieve waarde gebruiken. Een vuistregel is dat modellen die per vrijheidsgraad niet meer dan een waarde 2 verschillen in het AIC, niet van elkaar te onderscheiden zijn.

***Een benadering voor de EB variantie bij onafhankelijke studies***

Aangezien het uitrekenen van de variantie door het rechtstreeks maximaliseren van vergelijking (3) soms tot moeilijkheden leidt, kan men ook benaderingen gebruiken in geval van onafhankelijke studies. Berkey et al. (1995) geven een SAS procedure gebaseerd op zo'n benadering afgeleid in Morris (1983). Berkey et al. gebruiken echter de benadering van Morris voor de ML-schatter (vergelijking 2), terwijl in een Empirisch Bayesiaanse context juist de

REML-schatter gebruikt zou moeten worden (vergelijking 3). We geven hier de originele benadering van Morris voor de REML schatter voor het geval de studies onafhankelijk zijn.

De ware variantie kan volgens Morris worden benaderd door:

$$\hat{\sigma}_{EB_{morris}}^2 \approx \frac{\sum W_i^2 \left\{ n/(n-r) (Y_i - \mathbf{X}_i \cdot \hat{\beta})^2 - C_i \right\}}{\sum W_i^2} \quad (6)$$

hierin is  $r$  het aantal parameters in de regressie vergelijking,  $\mathbf{C}$  is een diagonaal matrix evenals  $\mathbf{W}$ .

De posteriori verwachtingen van een nieuwe studie kan nu ook inzichtelijker worden benaderd door

$$\hat{\theta}_i \approx (1 - \hat{B}_i) Y_i + \hat{B}_i (\mathbf{X}_i \hat{\beta}) \quad (7)$$

met posterior variantie

$$\hat{\sigma}_i^2 \approx C_i \cdot \left[ 1 - ((n - \hat{r}_i)/n) \cdot \hat{B}_i \right] + \hat{v}_i^2 \cdot (Y_i - \mathbf{X}_i \cdot \hat{\beta})^2 \quad (8)$$

waarin,

$$\hat{r}_i = n \cdot W_i \left[ \mathbf{X} \cdot (\mathbf{X}^T \cdot \mathbf{W} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \right]_{ii} \quad (9)$$

en

$$\hat{v}_i^2 = (2/(n-r-2)) \cdot \hat{B}_i^2 (\bar{C}^2 + \hat{\sigma}_{EB_{morris}}^2) / (C_i + \hat{\sigma}_{EB_{morris}}^2) \quad (10)$$

en de gemiddelde variantie

$$\bar{C}^2 = \sum W_i C_i / \sum W_i \quad (11)$$

voorstelt. de posteriori tolerantie intervallen worden dan op de gebruikelijke wijze uit (7) en (8) verkregen.

In bovenstaande vergelijkingen geldt dat:

$$\hat{B}_i = (n-r-2)/(n-r) \cdot C_i / (C_i + \hat{\sigma}_{EB_{morris}}^2) \quad (12)$$

### APPENDIX III (SAS)

SAS procedure voor een EB/ random model meta-analyse met co-variabelen en onafhankelijke studies (naar Berkey, aangepast door C. de Lezenne Coulander) en schattingen voor nieuwe studies

```
libname abc '.....';
data a;
  set abc.naam;
  .....;
  ...gewenste aanpassingen...;          /* dataset met o.a. de variabelen logRR, */
  .....;                                  /* varlogRR, x1 en x2 */
run;
proc iml;
```

```

reset noprint;

use a;
read all var {logRR} into y;          /* y is een kolomvector met de te onder- */
                                      /* zoeken grootheid (b.v. de log van een */
                                      /* relatief risico); vsubi is de kolom- */
read all var {varlogRR} into vsubi;  /* vector met de bijbehorende varianties */
read all var {x1} into xcov1;
read all var {x2} into xcov2;

k=nrow(y); one=j(k,1,1); x=one||xcov1||xcov2; /* one is de kolomvector met enen,      */
                                                /* waarmee het intercept wordt geschat; */
                                                /* xcov1 en xcov2 zijn de kolomvectoren */
                                                /* voor de schatting van de covariabelen */
                                                /* (voor meer covariabelen extra termen */
                                                /* xcov3, xcov4, etc toevoegen en als */
                                                /* geen intercept gewenst is de term */
                                                /* one weglaten) */

r=ncol(x); print k r;
a=0;
start b; niters=20;
do iter=1 to niters;
  wsubi=1/(vsubi+a);
  d=diag(wsubi);
  covbeta=inv(x`*d*x);
  beta=covbeta*(x`*d*y);
  sebeta=sqrt(vecdiag(covbeta));
  z=beta/sebeta;
  if iter=1 then print 'FIXED-EFFECTS MODEL' a beta sebeta z;
  newa=(sum((wsubi##2)#((k/(k-r))#(y-x*beta)#(y-x*beta)-vsubi)))/sum(wsubi##2);
  diff=a-newa; a=newa; if (a<0.0) then a=0;
  if(abs(diff)< .00001) then goto exit;
end;
exit: print '**** converged at iteration' iter '****';
print '* between var A not change by more than .00001 between iters *';
print iter a k vsubi wsubi y x covbeta; /* a is de tussen studie variantie      */
                                          /* k is het aantal studies              */
                                          /* covbeta is de var-covar matrix van de */
                                          /* beta coëfficiënten van de meta      */
                                          /* co-variabelen (de eerste waarde ver- */
                                          /* tegenwoordigt het intercept)       */

print ' ***** RANDOM_EFFECTS REGRESSION MODEL *****';
  print beta sebeta z;
finish; run b;
AIC=-log(det(d))+log(det(x`*d*x))+(y`*d*y)-(y`*d*x*beta)+2#r;
  print AIC;
bsubi=((k-r-2)/(k-r))#(vsubi/(vsubi+a));
ckwadr=sum(wsubi#vsubi)/sum(wsubi);
r_i=k#wsubi#vecdiag(x*inv(x`*d*x)*x`);
vkwadr=(2/(k-r-2))#bsubi#bsubi#(ckwadr+a)/(vsubi+a);
ynew=(1-bsubi)#y+bsubi#(x*beta);
vsubinew=vsubi#(1-((k-r_i)/k)#bsubi)+vkwadr#(y-x*beta)##2;
print ynew vsubinew; /* ynew is de kolomvector met de op */

```

```

/* grond van de meta-analyse aangepaste */
/* y-waarden; vsubinew is de kolomvector */
/* met de bijbehorende varianties      */

print '*** IF Y = LOG(RR) THEN USE t WITH DF= k-r-3 *****';
quit;

```

## APPENDIX IV (S-PLUS)

S-Plus functies voor een EB/ random model meta-analyse met of zonder co-variabelen en studies die wel of niet onafhankelijk zijn. (S.H. Heisterkamp)

### Hoofdfunctie f.muX

```

"f.muX"<-
function(y, sigma, X, which, help = F, Cov)
{
#
#       if(help) {
#           cat(... help text...)
#           return(invisible())
#       }
#
#
#----- Algorithm
#       see Dumouchel and Harris
#
#
#       k <- length(y) #
#---- either sigma must be given (uncorrelated studies)
#       or C as a complete covariance matrix
#       if(missing(Cov) & missing(sigma)) {
#           cat("Either Sigma must be given (uncorrelated studies) or Cov as covarinace
matrix\n"
#               )
#           stop()
#       }
#       indep <- T
#       if(missing(Cov)) {
#           Cov <- diag(sigma^2)
#       }
#       else {
#           indep <- F
#           sigma <- sqrt(diag(Cov))
#       }
#       if(missing(X)) {
#           X0 <- rep(1, length(y))
#           which <- c(1)
#           npar <- 1
#       }
#       else {
#           if(missing(which)) {
#               which <- 1:ncol(X)

```

```

    }
    X0 <- X[, which]
    npar <- ncol(X0)
  }
  cat("Dumouchel E-Bayes running\n")
  tau.dum <- optimize(f.dumouchel, lower = 0, upper = 1000 * max(sigma^2), keep.xy = F,
Y.Y
    = y, X = X0, cov.y = Cov)$minimum #
#
#
#
#-- compute AIC
#
  df <- k - npar
  AIC <- f.dumouchel(x = tau.dum, cov.y = Cov, y.y = y, X = X0) + 2 * npar #
#
#
#--- estimate parameters
#
  betahup <- f.beta(y = y, cov = Cov, tau = tau.dum, X = X0) #
#nu volgens formules 3.11
#
#
  hup <- f.ebupdate(tau = tau.dum, cov = Cov, y.y = y, X = X0)
  y.s.dum <- hup$theta
  se.s.dum <- sqrt(hup$vtheta) #
#
#
#
  tau.dum <- sqrt(tau.dum) #
#
#
#---- construct differences between collums and compute ses
#
  mean.diff <- matrix(nrow = npar, ncol = npar, data = NA, dimnames = list(1:npar,
1:npar))
  se.diff <- matrix(nrow = npar, ncol = npar, data = NA, dimnames = list(1:npar,
1:npar)) #
  Z.diff <- matrix(nrow = npar, ncol = npar, data = NA, dimnames = list(1:npar, 1:npar))
  if(npar > 1) {
    diag(mean.diff) <- betahup$beta
    diag(se.diff) <- sqrt(diag(betahup$var))
    for(i in 1:(npar - 1)) {
      for(j in (i + 1):npar) {
        X1 <- rep(0, npar)
        X1[i] <- 1
        X1[j] <- (-1)
        mean.diff[i, j] <- X1 %*% betahup$beta
        mean.diff[j, i] <- mean.diff[i, j]
        se.diff[i, j] <- sqrt(t(X1) %*% betahup$var %*% (X1))
        se.diff[j, i] <- se.diff[i, j]
      }
    }
  }
  Z.diff <- round(mean.diff/se.diff, 3)

```

```

    }
#
  if(indep) {
    input <- list(y = y, sigma = sigma, X = X0, which = which)
  }
  else {
    input <- list(y = y, sigma = sigma, X = X0, which = which, Cov = Cov)
  }
  if(npar == 1) {
df
    dum <- list(sqrt.tau = tau.dum, y.s = y.s.dum, sigma.s = se.s.dum, AIC = AIC,
                = df, beta = betahup$beta, sebeta = sqrt(diag(betahup$var)), varbeta =
                betahup$var)
  }
  else {
df
    dum <- list(sqrt.tau = tau.dum, y.s = y.s.dum, sigma.s = se.s.dum, AIC = AIC,
                = df, beta = betahup$beta, sebeta = sqrt(diag(betahup$var)), varbeta =
                betahup$var, mean.diff = mean.diff, se.diff = se.diff, Z.diff = Z.diff)
  }
  output <- list(input = input, dumouchel = dum)
  return(invisible(output))
}

```

### functie f.dumouchel

```

"f.dumouchel"<-
function(x, y.sigma, y.y, X, cov.y)
{
#
#
#
#----- evaluates log(L*) for the EB method according to DuMouchel & Harris
# for one mu in meta analysis
# (equation 3.11 with X'=[1...1] is default, so finding a mu
#
#
  if(missing(X)) {
    X <- rep(1, length(y.y))
  }
  if(missing(cov.y)) {
    omega <- x + y.sigma^2
    W <- diag(1/omega)
  }
  else {
    omega <- diag(x, nrow = length(y.y)) + cov.y
    W <- solve(omega)
  }
  V <- t(X) %*% W %*% X
  S <- W - W %*% X %*% solve(V) %*% t(X) %*% W
  s3 <- t(y.y) %*% S %*% y.y
  s1 <- prod(eigen(W)$values)
  s2 <- prod(eigen(V)$values)
  s <- log(s1) - log(s2) - s3
  return( - s)
}

```

}

**functie f.ebupdate**

```
"f.ebupdate"<-
function(tau, y.sigma, y.y, X, cov)
{
#
#
#
#----- updates for new studies according to dumouchel 3.11
# DIFFERENT in appearances....
#
  if(missing(X)) {
    X <- rep(1, lenght(y.y))
  }
  if(missing(cov)) {
    W <- diag(y.sigma^2)
    Winv <- diag(1/y.sigma^2)
  }
  else {
    W <- cov
    Winv <- solve(W)
  }
  Ind <- diag(lenght(y.y))
  VV <- Ind - X %*% solve(t(X) %*% X) %*% t(X)
  Veb <- solve(Winv + 1/tau * VV)
  theta <- as.vector(solve(Ind + 1/tau * W %*% VV) %*% y.y)
  output <- list(theta = theta, vtheta = as.vector(diag(Veb)))
  return(output)
}
```

**functie f.beta**

```
"f.beta"<-
function(tau, sigma, cov, y, X)
{
#
#
#
#
  if(missing(X)) {
    X <- rep(1, lenght(y))
  }
  if(missing(cov)) {
    W <- diag(1/(tau + sigma^2))
  }
  else {
    W <- solve(diag(tau, nrow = lenght(y)) + cov)
  }
  Veb <- solve(t(X) %*% W %*% X)
  theta <- as.vector(Veb %*% t(X) %*% W %*% y)
  output <- list(beta = theta, varbeta = Veb)
  return(output)
}
```