

# Klinische epidemiologie

H. Burger en A. Hofman

1

## Inleiding

Aan de opbouw van medische kennis hebben zowel culturele, politieke, economische, religieuze, filosofische als wetenschappelijke elementen bijgedragen. Vooral in de westerse wereld wordt de arts tegenwoordig meer en meer geacht wetenschappelijk te handelen; dat wil zeggen dat zijn handelingen gebaseerd zijn op wetenschappelijke kennis. Traditioneel komt een groot deel van deze kennis uit de medisch-biologische deelwetenschappen zoals anatomie, celbiologie, genetica, fysiologie en biochemie. Dit heeft ertoe geleid dat artsen hoofdzakelijk handelen op basis van deterministisch, causaal, pathofysiologisch redeneren, uiteraard in combinatie met eigen ervaringen en die van de opleider.

In de loop der tijd, maar vooral gedurende de afgelopen decennia, is in wisselende mate het besef gegroeid dat deze sterk theoretische basis ontoereikend is. Onder andere door de complexiteit van ziekte, de relatief geringe nauwkeurigheid van veel klinische waarnemingen en de beperkte waarde van individuele ervaringen wordt de medicus practicus geconfronteerd met een grote mate van onvoorspelbaarheid van vrijwel alle facetten van ziekte. De epidemiologie, en vooral de klinische epidemiologie, stelt zich ten doel door formele en systematische kwantificatie die kennis te leveren die de hieruit voortvloeiende onzekerheid moet verkleinen.

Het medisch-praktisch belang van empirische en statistische kennis in plaats van theoretische en deterministische kennis wordt zeer sterk benadrukt door een stroming die in 1992 ontstond uit de klinische epidemiologie. Volgens de aanhangers van deze stroming, de *evidence-based medicine*, behoort eigenlijk elke medische handeling gebaseerd te zijn op evidence uit klinisch-epidemiologisch onderzoek. Hoewel de resultaten van klinisch-epidemiologisch onderzoek in toenemende mate het patiëntenbeleid beïnvloeden, menen veel artsen dat de beide vormen van kennis elkaar aanvullen en daarom beide essentieel zijn voor de medische praktijk.

Terwijl de algemene epidemiologie de frequentie van optreden van ziekte en haar determinanten bestudeert, richt de klinische epidemiologie zich op de vragen in de klinische praktijk. Deze vragen betreffen risico, diagnose, prognose en interventie. Voordat deze vragen apart wor-

den besproken, volgen eerst drie algemene paragrafen over het begrip kans in de geneeskunde, de verschillende vormen van klinisch-epidemiologisch onderzoek en de betekenis van precisie en validiteit.

### Kernpunt

- Evidence-based medicine is het zorgvuldig, expliciet en oordeelkundig gebruik van het huidige beste bewijsmateriaal om beslissingen te nemen voor individuele patiënten. De praktijk van evidence-based medicine vereist de integratie van individuele klinische expertise met het beste externe bewijsmateriaal dat vanuit klinisch-epidemiologisch onderzoek beschikbaar is.

## 1.1 Kans

Zoals gezegd is veel handelen in de geneeskunde handelen in onzekerheid. Zelfs als een arts bij een bepaalde patiënt precies volgens de meest recente wetenschap handelt, blijft onzekerheid bestaan over diagnose, risico of prognose van ziekte en het succes van een eventuele behandeling. Zo is het bijvoorbeeld onmogelijk met zekerheid te voorspellen of een patiënt met hoge bloeddruk in de loop van de tijd een hersenbloeding zal krijgen. Evenzo is het bij het diagnosticeren vrijwel altijd onmogelijk bij bepaalde symptomen en laboratoriumuitslagen – bijvoorbeeld maagklachten, bloedarmoede en een afwijking op de maagfoto – absoluut zeker te zijn over de aan- of afwezigheid van ziekte, in dit geval bijvoorbeeld maagcarcinoom. Deze onzekerheid of onvoorspelbaarheid betreft zeker ook het beloop van ziekte, bijvoorbeeld de ontwikkeling van spraakstoornissen na de diagnose multiple sclerose. Ten slotte is er altijd onzekerheid over de mate waarin het beloop van ziekte ten gunste kan worden gewijzigd, over de effectiviteit van een bepaalde interventie, bij aids bijvoorbeeld door antiretrovirale therapie.

De onzekerheid of een bepaalde gebeurtenis zal optreden, kan worden uitgedrukt als een kans op die gebeurtenis, een waarschijnlijkheid. Kennis van de grootte van kansen op medische uitkomsten en hoe deze afhangen van bepaalde

determinanten is essentieel bij de beantwoording van vragen die in de klinische praktijk naar voren komen. Het kwantificeren van de kans op het ontstaan van ziekte (het risico), de kans op aanwezigheid van een bepaalde ziekte (de diagnose), de kans op een bepaald beloop van ziekte (de prognose) en de kans op een bepaald therapeutisch effect is de belangrijkste opdracht van de klinische epidemiologie.

1

### Kernpunten

- Vrijwel al het medisch handelen is handelen in onzekerheid.
- Onzekerheid kan worden uitgedrukt als een kans.
- Het kwantificeren van kansen op ziekte(gevolgen) is de belangrijkste opdracht van de klinische epidemiologie.



#### Intermezzo 1.1

##### Verre voorvaders van de klinische epidemiologie

Aangenomen wordt dat de term 'klinische epidemiologie' voor het eerst werd gebruikt in 1938 door de internist John Paul (1893-1971), overigens voor het type activiteit dat we nu zouden omschrijven als sociale geneeskunde. Toch is klinisch-epidemiologisch onderzoek al veel ouder. Een van de eerste klinisch epidemiologen was waarschijnlijk James Lind (1716-1794). Hij was scheepsarts en werd op lange reizen regelmatig geconfronteerd met zeelieden die aan scheurbuik leden. Op 20 mei 1747 vormde hij een groep van twaalf, zoveel mogelijk gelijke patiënten met scheurbuik en verdeelde die in zes groepen van twee. Elke groep kreeg verplicht, zo nodig met geweld, een andere dagelijkse aanvulling op het zeemansdieet toegediend: cider, elixer 'vitriol', azijn, zeewater, twee sinaasappels en een citroen, of enkele muskaatnoten. Het bleek dat de patiënten die de citrusvruchten kregen, al na zes dagen zodanig waren opgeknapt dat zij fit genoeg werden bevonden om het werk te hervatten! De overige zeelieden verbleven nog in een nagenoeg zelfde slechte toestand als voor dit experiment *avant la lettre*. De conclusie van Lind was dan ook dat het eten van citrusvruchten scheurbuik geneest. Het duurde nog tot 1932 voordat vitamine C als het werksame bestanddeel werd aangetoond.

Een andere grondlegger van de klinische epidemiologie was de Franse arts Pierre Charles Alexandre Louis (1787-1872). Hij was teleurgesteld over de beperkingen van de geneeskunde en besloot op 32-jarige leeftijd te stoppen met praktisch werk en zich geheel te wijden aan onderzoek. In die tijd waren diagnose en therapie vooral gebaseerd op theorie, ervaring en intuïtie van prominente artsen en niet op evidence. Dit stoorde Pierre Louis en hij ontwikkelde een systeem dat hij aanduidde als *la méthode numérique* en dat neerkwam

op het zorgvuldig observeren, registreren en kwantitatief analyseren van gegevens van alle patiënten die werden opgenomen in het Hôpital de la Charité in Parijs. Zo bestudeerde hij de effectiviteit van aderlaten bij patiënten met pneumonie op een manier die wij nu klinisch-epidemiologisch zouden noemen. Hij kwam tot de conclusie dat aderlaten de duur van pneumonie verkort als de patiënten vroeg in het ziektebeloop worden behandeld, maar dat deze interventie op korte termijn een toename van de mortaliteit teweeg kan brengen. Een 'moderne' heranalyse van de oorspronkelijke gegevens laat deze conclusie ongewijzigd.

Ook Ignaz Philippe Semmelweis (1818-1865) is een van de 'founding fathers' van de klinische epidemiologie. Hij liet zien dat de incidentie van kraamvrouwenkoorts – een vorm van sepsis – in het universitaire ziekenhuis waar hij werkte, veel groter was dan in de naburige niet-universitaire vroedvrouwenkliniek. Uiteindelijk wist hij dit verschil toe te schrijven aan de gebrekkige hygiëne rond obducties die alleen in het universitaire ziekenhuis werden uitgevoerd. Na invoering van maatregelen die tegenwoordig vanzelfsprekend zijn – zoals na een obductie handen wassen en instrumenten steriliseren – daalde de frequentie van kraamvrouwenkoorts drastisch, tot beneden het niveau van de vroedvrouwenkliniek.

Andere personen die meer recent hebben bijgedragen aan de ontwikkeling van de klinische epidemiologie zijn Alvan R. Feinstein (1925-2001), onder andere bekend door zijn werk op het gebied van de klinimetrie, en Sir Austin Bradford Hill (1897-1991), die veel betekend heeft voor de *clinical trial*.

## 1.2 Vormen van klinisch-epidemiologisch onderzoek

Enkele vormen van klinisch-epidemiologisch onderzoek zijn vermeld in tabel 1.1. De onderzoeksvormen die in de klinische epidemiologie worden gebruikt, kunnen worden ingedeeld op basis van vier dichotomieën: voorspellend versus causaal, experimenteel versus observationeel, longitudinaal versus transversaal, en cohort- versus patiënt-controleonderzoek.

### Kernpunten

- Onderzoeksvormen die in de klinische epidemiologie worden gebruikt, kunnen ingedeeld worden op basis van vier dichotomieën: voorspellend versus causaal, experimenteel versus observationeel, longitudinaal versus transversaal, en cohort- versus patiënt-controleonderzoek.
- De onderzoeksvorm heeft belangrijke consequenties voor de uitvoering van het onderzoek en voor de interpretatie van de resultaten.

**Tabel 1.1 Voorbeelden van vormen van klinisch-epidemiologisch onderzoek.**

onderzoeksvorm	vraagstelling	voorbeeld
voorspellend observationeel transversaal cohortonderzoek	diagnose	de waarde van het inspannings-ecg voor de diagnose coronaire hartziekte
causaal observationeel longitudinaal cohortonderzoek	risico	effect van regelmatige lichaamsbeweging op het risico van een heupfractuur
voorspellend observationeel longitudinaal cohortonderzoek	prognose	kans op een myocardinfarct als functie van de dikte van de wand van de halslagader, klassieke risicofactoren voor hart- en vaatziekten, en leeftijd en geslacht
causaal observationeel longitudinaal patiënt-controleonderzoek	risico	relatief risico op de ziekte van Alzheimer door NSAID-gebruik
causaal experimenteel longitudinaal cohortonderzoek (clinical trial)	interventie-effectiviteit	radioactief jodium bij hyperthyroidie

De eerste dichotomie betreft het doel van het onderzoek: voorspellen of verklaren. Voorspellend onderzoek moet het mogelijk maken een gefundeerde inschatting te maken van de kans op aanwezigheid van ziekte (diagnose) of de kans op een bepaald beloop van ziekte (prognose), een kernactiviteit van elke arts. Deze schattingen worden gemaakt op basis van demografische kenmerken, zoals leeftijd en geslacht, gegevens uit de anamnese, bevindingen bij lichamelijk onderzoek en uitslagen van diagnostische tests zoals bloedbepalingen en beeldvormende technieken. Het is daarbij niet relevant of een bepaald kenmerk causaal met de ziekte of prognose verband houdt of niet. Zo wordt de diagnose angina pectoris gesteld op basis van haar risicofactoren, zoals roken en diabetes, maar ook op basis van haar gevolgen, zoals ischemische pijn en ecg-afwijkingen. Bij osteoporose kan de kans op een heupfractuur (prognose) onder andere worden geschat aan de hand van wervelinzakkingen. Deze afwijkingen worden niet gezien als oorzaak van de heupfractuur, maar weerspiegelen de slechte toestand van het botweefsel en zijn daarmee slechts indicators van de kans op fracturen.

Aan onderzoek dat de evidence moet leveren voor het louter schatten van kansen, zoals in bovenstaande voorbeelden, worden heel andere eisen gesteld dan aan onderzoek waarin men probeert een oorzakelijk verband aannemelijk te maken. Het belangrijkste verschil is dat bij voorspellend onderzoek confounding (vertekening) niet aan de orde is. Dit in tegenstelling tot causaal onderzoek, waarin het probleem van confounding juist centraal staat. Men wil dan immers zo zeker mogelijk weten dat een associatie tussen blootstelling en uitkomst niet door een derde externe factor kan worden verklaard (zie ook paragraaf 1.3.2). Al het diagnostische onderzoek is voorspellend van aard; prognostisch onderzoek kan zowel

voorspellend als causaal zijn. Het onderscheid tussen voorspellend en causaal heeft overigens grote consequenties voor alle fasen van onderzoek: onderzoeksontwerp, gegevensanalyse en rapportage.

Het tweede fundamentele onderscheid is experimenteel versus observationeel onderzoek. Bij experimenteel onderzoek bepaalt de onderzoeker, meestal *at random* (door het lot bepaald) welke patiënt aan de te onderzoeken factor wordt blootgesteld en wie niet. Deze vorm van onderzoek vindt zijn belangrijkste toepassing in het onderzoek van de effectiviteit van interventies, de clinical trial. Bij observationeel onderzoek heeft de onderzoeker geen invloed op de blootstelling. Het grote nadeel van observationeel ten opzichte van experimenteel onderzoek is dat causaliteit van een bepaald verband moeilijker aannemelijk kan worden gemaakt. Hierop komen we bij het begrip confounding nog terug. Voor de rapportage van observationeel onderzoek is de *STrengthening the Reporting of OBServational studies in Epidemiology (STROBE) statement* ([www.strobe-statement.org](http://www.strobe-statement.org)) opgesteld. Hierin worden veel aanbevelingen gedaan voor niet alleen de rapportage, maar ook de data-analyse van observationeel onderzoek.

Een derde onderscheid betreft longitudinaal versus transversaal onderzoek. Bij transversaal onderzoek, ook wel aangeduid als cross-sectioneel onderzoek of dwarsdoorsnedeonderzoek, worden determinant en uitkomst op hetzelfde moment vastgesteld, bij longitudinaal onderzoek wordt de determinant voorafgaand aan de uitkomst gemeten. Al het experimentele onderzoek is natuurlijk longitudinaal. Observationeel onderzoek kan zowel longitudinaal als transversaal zijn. Veel diagnostisch onderzoek is transversaal omdat de vraagstelling geen tijdsverloop kent: wat is de kans op de aanwezigheid van een bepaalde ziekte nu, gegeven de kenmerken van de patiënt en de uitslagen van diagnostische tests op dit moment? Eén van de problemen van transversaal onderzoek is uiteraard dat er geen tijdsrelatie tussen determinant en uitkomst kan worden vastgesteld. De meeste andere dan diagnostische vragen in de klinische epidemiologie zijn inherent longitudinaal en de beantwoording met een transversaal onderzoek kan daarom problematisch zijn.

Een vraag betreffende risico is bijvoorbeeld of een lage botdichtheid bij vrouwen rond de menopauze leidt tot wervelinzakkingen op hogere leeftijd. Stel dat in een transversaal onderzoek bij 80-jarigen de vrouwen met wervelinzakkingen een lagere botdichtheid hebben dan de vrouwen zonder deze afwijkingen. In dat geval is het niet zonder meer duidelijk of de lage botdichtheid een gevolg is van de wervelinzakkingen of juist een oorzaak. Wervelinzakkingen zouden immers kunnen leiden tot een verminderde mobiliteit, die op haar beurt weer kan leiden tot een lagere botdichtheid. Een dergelijke vraag kan daarom beter met een longitudinaal onderzoek worden beantwoord.

Het vierde onderscheid is cohortonderzoek versus patiënt-controleonderzoek en betreft de wijze waarop de onderzoekspopulatie wordt samengesteld.

Bij cohortonderzoek wordt een groep personen onderzocht die een bepaald gemeenschappelijk kenmerk hebben, bijvoorbeeld een geboortjaar, woonplaats, een bepaalde diagnose. Bij longitudinaal cohortonderzoek wordt een populatie met en zonder de te onderzoeken blootstelling gevolgd in de tijd, waarna de frequentie van ziekte of complicatie tussen beide groepen wordt vergeleken. Het longitudinale cohortonderzoek biedt vele voordelen. De belangrijkste zijn dat de blootstelling met het doel van het onderzoek voor ogen gemeten kan worden, dat weinig of geen twijfel bestaat over de tijdsrelatie (blootstelling → ziekte), dat naast het relatieve ook het absolute risico van ziekte kan worden gemeten en dat relatief gemakkelijk verschillende uitkomsten bestudeerd kunnen worden.

Bij patiënt-controleonderzoek wordt een serie patiënten met een bepaalde aandoening verzameld en bij hen wordt de frequentie van blootstelling bepaald. Tevens wordt uit de al of niet denkbeeldige populatie waaruit de patiënten afkomstig zijn een steekproef met 'controles' genomen om de blootstellingsfrequentie in de zogenoemde bronpopulatie te schatten. De selectie van controles is vaak aanleiding voor discussie. In het algemeen is er geen probleem wanneer de controles zodanig worden gekozen dat zij, als zij de ziekte hadden gekregen, ook als patiënten in het onderzoek zouden zijn opgenomen. Een groot voordeel van het patiënt-controleonderzoek is dat verschillende determinanten van zeldzame ziekten met relatief geringe inspanning onderzocht kunnen worden.

De patiëntencasus of -serie kan worden opgevat als patiënt-controleonderzoek waarbij de controles feitelijk niet worden onderzocht, maar in gedachten wel de referentie vormen bij de gevolgtrekkingen. Bij dit onderzoek heeft de arts een aantal patiënten met een bepaalde aandoening opgemerkt en in kaart gebracht. Vervolgens wordt gekeken of er een gemeenschappelijk kenmerk is dat men niet verwacht bij personen zonder de aandoening, bijvoorbeeld het gebruik van een bepaald geneesmiddel. In sommige gevallen kan een dergelijke bevinding de aanleiding zijn voor nader onderzoek.

Een mooi voorbeeld stamt uit 1961, toen de diagnose longembolie werd gesteld bij een jonge vrouw. Deze aandoening is zo zeldzaam op jonge leeftijd dat het gebruik van een oraal anticonceptivum als mogelijke oorzaak werd geopperd. Later werd deze associatie in grootschalig onderzoek bevestigd.

Ten slotte moet de meta-analyse worden genoemd als bijzondere vorm van onderzoek. Bijzonder omdat de resultaten van verschillende onderzoeken de basis vormen

### Kernpunten

- Bij voorspellend onderzoek speelt confounding (vertekening door een derde factor die samenhangt met determinant en uitkomst) geen rol; dit in tegenstelling tot bij causaal onderzoek.
- Bij experimenteel onderzoek bepaalt de onderzoeker welke patiënt aan de te onderzoeken factor wordt blootgesteld en wie niet; bij observationeel onderzoek heeft de onderzoeker geen invloed op de blootstelling.
- In dwarsdoorsnedeonderzoek worden determinant en uitkomst op hetzelfde moment vastgesteld, bij longitudinaal onderzoek wordt de determinant voorafgaand aan de uitkomst gemeten.
- Bij cohortonderzoek wordt de totale groep van personen met een bepaald gemeenschappelijk kenmerk onderzocht. Bij het patiënt-controleonderzoek wordt de frequentie van blootstelling aan een bepaalde determinant gemeten bij patiënten met een bepaalde aandoening en bij een steekproef van personen zonder die aandoening.

voor de analyse in plaats van de gegevens van verschillende individuen; vandaar de naam meta-analyse. Het doel van dit onderzoek is een preciezere schatting te krijgen van een bepaald effect, meestal het effect van een geneesmiddel. Het kan worden beschouwd als een formele en kwantitatieve samenvatting van de resultaten van het tot dan toe beschikbare onderzoek. Meta-analyse wordt nader besproken bij interventie (paragraaf 1.7.3).

### Kernpunt

- De meta-analyse neemt een aparte plaats in bij de verschillende onderzoeksvormen. Dit type onderzoek wordt gebruikt om de gegevens uit meerdere onderzoeken kwantitatief samen te vatten.



#### Intermezzo 1.2

#### Observationeel of gerandomiseerd experimenteel onderzoek naar behandelingseffecten?

Voor het aantonen van een bepaald behandelingseffect wordt het gerandomiseerde experiment in het algemeen als meest valide onderzoeksoptzet beschouwd. De belangrijkste reden hiervoor is dat randomisatie ervoor zorgt dat de toewijzing van een bepaalde behandeling niet afhangt van de prognose; dat wil zeggen dat *confounding by indication* wordt voorkomen. Helaas kleeft er een aantal belangrijke bezwaren aan deze onderzoeksoptzet, in het bijzonder de veelal hoge

kosten en de vaak selecte onderzoeksgroep. Bovendien kan randomisatie ethisch bezwaarlijk zijn als al voorafgaand aan het onderzoek de indruk bestaat dat een van de te onderzoeken behandelingen superieur is.

Het is daarom niet verwonderlijk dat telkens weer de vraag rijst of effecten van interventies niet ook met een observationele opzet bestudeerd kunnen worden. Bij deze opzet zijn de genoemde bezwaren van het gerandomiseerde onderzoek minder uitgesproken of zelfs afwezig. De vraag is dan of de validiteit niet in gevaar komt. Recent is deze discussie weer op gang gekomen naar aanleiding van de bevinding dat observationeel onderzoek duidelijk maakt dat postmenopauzale hormoontherapie de kans op een myocardinfarct verkleint, terwijl uit gerandomiseerd onderzoek blijkt dat deze interventie de kans op een myocardinfarct juist iets groter maakt. Ook het gunstige effect van antioxidanten op het ontstaan van hart- en vaatziekten en kanker, zoals aangetoond in observationeel onderzoek, kon niet worden bevestigd in gerandomiseerd onderzoek.

Toch zijn er ook voorbeelden waarbij observationeel en gerandomiseerd onderzoek exact dezelfde evidence laten zien, zoals voor de effectiviteit van vaccinatie tegen actieve tuberculose. Het hangt van de aard van het onderzoek af of observationeel onderzoek een even valide resultaat kan geven als gerandomiseerd onderzoek. Als er geen duidelijke relatie bestaat tussen de indicatie of contra-indicatie voor een bepaalde interventie en de uitkomstvariabele, kan observationeel onderzoek heel valide zijn. Bijvoorbeeld: in het onderzoek naar de effecten van postmenopauzale hormoontherapie was er geen discrepantie tussen het observationele en het gerandomiseerde onderzoek ten aanzien van de uitkomsten veneuze trombose en borstkanker; dit in tegenstelling tot de uitkomst myocardinfarct. De artsen waren waarschijnlijk terughoudend met hormonen als zij een verhoogd risico op een myocardinfarct vermoedden, maar konden door de onvoorspelbaarheid van veneuze trombose en borstkanker geen rekening houden met het risico op deze aandoeningen. *Confounding by contra-indication* kon zo geen rol spelen bij deze twee uitkomsten (Vandenbroucke, 2004). Overigens is deze verklaring recent verworpen (Vandenbroucke, 2009). In de gerandomiseerde trials bleken de hormonen in de eerste jaren een verhoging van het risico op myocardinfarct te brengen, wat gevolgd werd door een daling. In het observationele onderzoek werden gebruikers vergeleken met niet-gebruikers van hormonen ten tijde van inclusie in het onderzoek. De gebruikers waren al jaren blootgesteld en hadden de periode van verhoogd risico al achter de rug. Zo ontstond een verschil in effectiviteit van postmenopauzale hormoontherapie tussen gerandomiseerd onderzoek en observationeel onderzoek dat niets te maken had met de onderzoeksopzet maar volledig verklaard kon worden door de duur van blootstelling.

## 1.3 Precisie en validiteit

De mate waarin een onderzoek bijdraagt aan de wetenschap hangt in sterke mate af van zijn precisie en validiteit.

### 1.3.1 PRECISIE

De precisie of nauwkeurigheid van een onderzoek kan geïnterpreteerd worden als de mate waarin het onderzoek bij denkbeeldige herhalingen hetzelfde resultaat laat zien. De precisie zegt dus niets over de juistheid van de resultaten, alleen iets over de reproduceerbaarheid. Ze neemt toe met de grootte van de onderzoekspopulatie en neemt af met de grootte van meetfouten. Op basis van de grootte van de onderzoekspopulatie en een schatting van de meetfout uit het onderzoek kan daarom een uitspraak over de precisie worden gedaan, bij voorkeur in de vorm van een 95%-betrouwbaarheidsinterval. Dit interval geeft aan dat in 95 van de 100 denkbeeldige herhalingen van het onderzoek het gevonden 95%-betrouwbaarheidsinterval het werkelijke effect zou bevatten. In plaats van een betrouwbaarheidsinterval wordt de precisie van een onderzoek ook vaak uitgedrukt in de  $p$ -waarde. De  $p$ -waarde geeft aan hoe vaak bij een denkbeeldige herhaling van het onderzoek een waarde voor de effectmaat wordt gevonden die minstens zo groot is als die in het huidige onderzoek, gegeven dat er in werkelijkheid geen effect is, dat wil zeggen dat de nulhypothese van geen effect juist is. De  $p$ -waarde is moeilijk te interpreteren omdat deze vooral afhangt van de grootte van het gevonden effect. Bij het betrouwbaarheidsinterval daarentegen krijgen we direct een indruk van de toevalsvariantie in de grootte van het effect, een maat voor de precisie in de meting van een effect. Daarom wordt in de epidemiologie de voorkeur gegeven aan een betrouwbaarheidsinterval in plaats van aan een  $p$ -waarde.

De vraag blijft hoe we deze maten van precisie betrekken bij de conclusie uit een onderzoek, de statistische inferentie. Hierin bestaan twee scholen. In de traditionele school wordt de nulhypothese verworpen of aanvaard. Volgens conventie wordt bij een  $p$ -waarde kleiner dan 0,05 (de alfa) de nulhypothese ten gunste van de alternatieve hypothese van wel een effect verworpen. Stel dat in een onderzoek het relatieve risico op retinopathie bij patiënten met diabetes, vergeleken met personen zonder diabetes, 2,3 bedraagt met een  $p$ -waarde van 0,04, dan wordt de nulhypothese van geen effect verworpen en wordt geconcludeerd dat er een statistisch significant verband is tussen diabetes en retinopathie. Bij een  $p$ -waarde groter dan of gelijk aan 0,05 wordt geconcludeerd dat dit verband niet kon worden aangetoond. Ook op basis van een betrouwbaarheidsinterval kan men een dergelijke ja-nee-uitspraak doen. Als de nulwaarde – bij een relatief risico is dit 1 – zich buiten het 95%-betrouwbaarheidsinterval bevindt, betekent dit dat er een significant verband is bij een alfa van 0,05. In genoemd voorbeeld zou dan het 95%-be-

trouwbaarheidsinterval 1,4-5,2 kunnen zijn. Omgekeerd zou de conclusie bij een betrouwbaarheidsinterval 0,9-5,7 zijn: geen statistisch significant verband aangetoond. Het is belangrijk te benadrukken dat een dergelijke conclusie geen recht doet aan de informatieve waarde van het betrouwbaarheidsinterval, vooral bij een niet-significant resultaat. Als bijvoorbeeld een relatief risico van 1,1 met een klein 95%-betrouwbaarheidsinterval (0,9-1,3) wordt gevonden, is direct zichtbaar dat waarden boven de 1,3 en onder de 0,9 niet erg plausibel zijn en ook dat het effect niet erg groot is. Een  $p$ -waarde, in dit geval bijvoorbeeld 0,2, is veel minder informatief en leidt er waarschijnlijk slechts toe dat de nulhypothese niet wordt verworpen. Omgekeerd betekent een relatief risico van bijvoorbeeld 9,0 met een 95%-betrouwbaarheidsinterval van 1,6 tot 50 dat de effectschatting weliswaar niet erg precies is, maar wel groot en daardoor (wellicht) relevant. Ook hier is een  $p$ -waarde ( $p < 0,05$ ) veel minder informatief.

Bij de meer recent ontwikkelde statistische inferentie volgens Bayes wordt de nulhypothese niet verworpen of aanvaard. In plaats daarvan wordt de statistische informatie uit het onderzoek gebruikt om de geloofwaardigheid van een hypothese zoals die vóór het onderzoek bestond, de a-priori-geloofwaardigheid, te veranderen in een achteraf (a-posteriori)-geloofwaardigheid. Informeel speelt dit proces zich veelvuldig af bij de meningsvorming in de geneeskunde, en ook daarbuiten. Stel dat een arts ervan overtuigd is dat antibiotica bij de behandeling van een maagzweer niet effectief zijn, dan zal door één onderzoek dat met statistische significantie de effectiviteit van deze behandeling wel aantoonde, die overtuiging niet of nauwelijks worden bijgesteld. Als deze arts echter de week erop kennis neemt van nog een onderzoek waarin weer een significant effect van antibiotica bij maagzweren wordt aangetoond, zal hij meer geneigd zijn zijn mening te herzien. De voortdurende verandering van het geloof in een bepaalde hypothese op basis van steeds nieuwe gegevens is de essentie van de Bayesiaanse inferentie. Voor een nadere beschouwing verwijzen wij naar Hofman et al. (1996).

#### Kernpunt

- Precisie is de mate waarin het onderzoek bij denkbeeldige herhalingen hetzelfde resultaat heeft, en wordt uitgedrukt met een  $p$ -waarde of, beter, een betrouwbaarheidsinterval.

### 1.3.2 VALIDITEIT

Bij de beoordeling van de validiteit van een onderzoek wordt onderscheid gemaakt tussen de interne en externe validiteit.

De externe validiteit, beter aangeduid als de *generaliseerbaarheid*, is de mate waarin de onderzoeksresultaten

van toepassing zijn op personen met duidelijk andere karakteristieken dan de personen die tot de onderzoekspopulatie behoorden. Stel bijvoorbeeld dat de effectiviteit van een nieuw antibioticum voor de behandeling van pneumonie werd vastgesteld bij mannen, dan is het zeer aannemelijk dat dit middel ook effectief is bij vrouwen. Als het onderzoek echter de diagnostische waarde van het inspanningselektrocardiogram voor coronairsclerose zou betreffen, is de generaliseerbaarheid van mannen naar vrouwen minder vanzelfsprekend gezien het grote geslachtsverschil in ernst en voorkomen van hart- en vaatziekten. Generaliseerbaarheid kan overigens niet gegarandeerd worden door representatieve steekproeven; ze berust in wezen op een subjectief oordeel waarbij ook informatie uit ander(soortig) onderzoek moet worden betrokken.

De interne validiteit is de mate waarin – afgezien van toevalsfouten – het gevonden effect gelijk is aan het werkelijke effect bij personen met overeenkomstige kenmerken als die van de onderzoekspopulatie. Vertekening, ook wel aangeduid als *bias*, is datgene wat de interne validiteit aantast. Er wordt in het algemeen onderscheid gemaakt tussen drie soorten bias: selectiebias, informatiebias en confounding bias.

Selectiebias treedt op als de inclusie van personen met de te bestuderen uitkomst niet onafhankelijk gebeurt van de determinant, of andersom. Als in een onderzoek naar de diagnostische waarde van ST-depressie op het ecg voor de diagnose coronairsclerose patiënten op een hartafdeling worden onderzocht, zal er vertekening van het resultaat optreden als deze patiënten juist omwille van deze ecg-afwijking op de hartafdeling liggen.

Informatiebias is het gevolg van onvergelykbare waarneming van de determinant of de uitkomst tussen de te vergelijken groepen. Stel dat in een clinical trial de nieuwste ACE-remmer wordt onderzocht op effectiviteit bij de behandeling van hypertensie. Als degene die de bloeddruk meet, op de hoogte is van het feit dat een deelnemer het nieuwe middel krijgt of een placebo, is het niet onaanvaardbaar dat bij de deelnemers aan de ACE-remmergroep de bloeddrukmeter systematisch anders wordt afgelezen dan bij personen uit de placebogroep. Informatiebias wordt onder andere tegengegaan door blinding van de onderzoeker en de onderzochte personen voor de aard van de determinant; in een trial: de aard van de interventie.

Een derde vorm van bias betreft *confounding*, letterlijk ‘verwarring’. Er is sprake van confounding als een derde factor die samenhangt met de onderzochte determinant, los van het te onderzoeken causale pad, de uitkomst beïnvloedt. In dat geval zijn de te contrasteren groepen niet meer vergelijkbaar wat betreft de kans op de uitkomst. Indien men bijvoorbeeld onderzoek heeft gedaan naar het serumcholesterol als risicofactor voor een tweede myocardinfarct, kan bij analyse blijken dat de groep personen met een hoog cholesterolgehalte veel meer rokers bevat dan de referentiegroep met een normaal tot laag chole-

terolgehalte. Omdat de twee groepen louter op grond van de proportie rokers prognostisch onvergelykbaar zijn, kan een verschil in prognose niet meer uitsluitend aan het verschil in cholesterolgehalte worden toegeschreven. Men spreekt van confounding door roken, dat wil zeggen: roken (de *confounder*) belemmert een valide vergelijking van de twee groepen.

Terwijl de mogelijkheid van selectiebias en informatiebias moet worden voorkomen in de onderzoeksopzet, zijn er bij confounding bias meer opties. Confounding kan allereerst worden voorkomen door het onderzoek te beperken tot één categorie van de confounder, dat wil zeggen restrictie, bijvoorbeeld door alleen 50- tot 60-jarigen in het onderzoek op te nemen, of alleen niet-rokers. Verder kan bij follow-uponderzoek confounding worden voorkomen door de onderzoekspopulatie zo te selecteren dat de confounder gelijkelijk wordt verdeeld over de blootgestelde en de referentiegroep. Dit proces wordt *matching* genoemd. Verder kan als onderdeel van de data-analyse door statistische aanpassing gecorrigeerd worden voor een of meer confounders. Deze methoden vereisen alle dat bekend is wat de mogelijke confounders zijn en dat ze goed gemeten kunnen worden. Dit voorbehoud geldt niet als men randomiseert, dat wil zeggen dat men het lot laat bepalen wie de behandeling krijgt en wie niet. Op deze manier worden 'automatisch' alle factoren die de uitkomst kunnen beïnvloeden in gelijke mate verdeeld over de te contrasteren groepen.

Zoals gezegd is confounding alleen aan de orde als men geïnteresseerd is in de causale relatie tussen één bepaalde factor en de uitkomst zoals in etiologisch onderzoek, dus onderzoek naar de oorzaken van het ontstaan of beloop van ziekten. Bij voorspellend onderzoek ligt dit anders; er is louter interesse in de mate waarin variabelen bijdragen aan een juiste schatting van de uitkomst. Zo is men bij diagnose alleen geïnteresseerd in hoeverre bepaalde tests gezamenlijk onderscheid maken tussen ziek en niet ziek, en niet hoe één bepaalde testuitslag oorzakelijk gerelateerd is aan de aanwezigheid van ziekte.

#### Kernpunten

- Interne validiteit is de mate waarin de onderzoeksresultaten met de, onwaarneembare, werkelijkheid overeenstemmen. Ze wordt bedreigd door selectiebias, informatiebias en confounding bias.
- Externe validiteit of liever generaliseerbaarheid is de mate waarin onderzoeksresultaten van toepassing zijn op personen die wezenlijk anders zijn dan die in de onderzoekspopulatie.

Ten slotte willen wij erop wijzen dat er bij de beoordeling van de waarde van onderzoek een duidelijke hiërarchie van criteria is: de interne validiteit staat voorop, dan volgt precisie en ten slotte generaliseerbaarheid. Wanneer bij-

voorbeeld een onderzoek zeer precies is maar niet intern valide, heeft het geen enkele waarde voor de wetenschap. Omgekeerd kan een zeer valide maar niet zo precies onderzoek wel degelijk van waarde zijn, bijvoorbeeld door zijn bijdrage aan een meta-analyse (zie paragraaf 1.7.3). De generaliseerbaarheid komt pas aan de orde wanneer het onderzoek intern valide en voldoende precies is.

#### Kernpunten

- Er is een hiërarchie in criteria voor de waardering van onderzoek: interne validiteit, precisie en ten slotte generaliseerbaarheid.
- Het verdient aanbeveling om bij de beoordeling van een wetenschappelijk onderzoek deze drie criteria op een systematische manier langs te lopen.

## 1.4 Risico

Het begrip risico wordt in deze paragraaf in engere zin gebruikt, namelijk als de kans op het ontstaan van ziekte. In het klinisch-epidemiologisch jargon wordt risico ook vaak gebruikt als synoniem van kans, zoals het risico van een bepaalde complicatie of andere uitkomst bij prognostisch of interventieonderzoek. Dit betekent dat de techniek van het meten van risico zoals hieronder wordt uiteengezet ook van toepassing is bij onderzoek van prognose en interventie.

Zoals gezegd is het risico van ziekte op te vatten als de kans of de waarschijnlijkheid dat ziekte optreedt. Om de kans op ziekte vast te stellen, moet men de frequentie van het optreden van ziekte kennen. Voor onderzoek van risico en prognose is daarbij de frequentie van nieuwe ziekten (*incidentie*) en niet zozeer van reeds bestaande ziekten (*prevalentie*) van belang. Uit de incidentie kan namelijk de kans op ziekte en daarmee het risico van ziekte worden vastgesteld. Het risico van ziekte kan op twee manieren worden gemeten: direct uit de *cumulatieve incidentie*, al of niet met gebruikmaking van de levenstafel, en indirect uit het *incidentiecijfer*.

### 1.4.1 METEN VAN RISICO

De belangrijkste frequentie maat voor ziekte in de klinische epidemiologie is de *cumulatieve incidentie*. Onder cumulatieve incidentie wordt verstaan het aantal nieuwe ziekten in een bepaalde tijdsperiode, gedeeld door het totale aantal mensen dat ziektevrij (en daarmee *at risk*) was aan het begin van de tijdsperiode. Schematisch kan de cumulatieve incidentie (CI) in de periode  $t_0$  tot  $t_1$  als volgt worden weergegeven:

$$CI = \frac{\text{nieuwe patiënten in periode } t_1 - t_0}{\text{aantal mensen zonder ziekte bij } t_0}$$

De cumulatieve incidentie is een proportie of fractie (bijvoorbeeld een percentage). Het is een kans die het risico van ziekte direct aangeeft en die kan variëren van 0 tot 1.

Hoewel in de geneeskunde bij voorkeur wordt gewerkt met de cumulatieve incidentie omdat het risico van ziekte of complicatie het patiëntbeleid bepaalt, moet soms noodgedwongen gebruik worden gemaakt van een andere maat: het incidentiecijfer (IC). Ook hierbij gaat het om het aantal nieuwe patiënten, maar nu gedeeld door het aantal persoonsjaren van follow-up:

$$IC = \frac{\text{nieuwe patiënten}}{\text{aantal persoonsjaren van follow-up}}$$

Onder persoonsjaren wordt verstaan het product van het aantal mensen en de periode dat deze mensen worden geobserveerd. Als honderd mensen gedurende tien jaar worden geobserveerd, levert dit duizend persoonsjaren op. Het incidentiecijfer kan variëren van 0 tot +oneindig. Het incidentiecijfer van enkele belangrijke ziekten in Nederland wordt gegeven in tabel 1.2.

Het incidentiecijfer is een veel gecompliceerdere frequentie maat dan de cumulatieve incidentie. Het kan niet eenvoudig worden geïnterpreteerd en het is vooral geen proportie of percentage en geeft dus niet direct een kans of risico aan. Het is echter mogelijk het incidentiecijfer om te rekenen in een cumulatieve incidentie. Bij een redelijk laag incidentiecijfer en een korte follow-up geldt als vuistregel het volgende verband:  $CI^{\Delta t} = IC \times \Delta t$ . Zo is de vijfjaars cumulatieve incidentie van dementie voor een 70-jarige Nederlandse man af te leiden uit het incidentiecijfer voor de leeftijdsgroep 70 tot 74 jaar (1,98/1000 manjaren). Dit geeft een schatting van het risico om binnen vijf jaar dementie te worden van  $CI_5 = 0,00198 \times 5 = 1\%$ . Een uitvoeriger bespreking met meer algemene formules voor de berekening van het risico uit het incidentiecijfer kan elders worden gevonden (Vandenbroucke en Hofman, 1997).

#### Kernpunten

- Een risico is een kans op een – meestal ongewenste – gebeurtenis binnen een bepaalde periode.
- Voor onderzoek van risico en prognose is de frequentie van nieuwe zieken (incidentie) en niet zozeer van reeds bestaande zieken (prevalentie) van belang.
- Er worden twee soorten incidentie onderscheiden: de cumulatieve incidentie en het incidentiecijfer. In de klinische epidemiologie is de cumulatieve incidentie het meest relevant omdat het direct het risico weergeeft.

#### 1.4.2 VORMEN VAN RISICO

Tot nu toe is het meten van het risico van een individu om ziekte te krijgen besproken. Men kan twee vormen van

**Tabel 1.2** Incidentiecijfer (per 1000 persoonsjaren) van enkele ziekten op basis van Nederlandse zorgregistraties naar geslacht.

	mannen	vrouwen
acuut hartinfarct	2,3	1,3
beroerte	2,1	2,3
longkanker	0,85	0,28
borstkanker	-	1,3
prostaatkanker	0,85	-
aids	0,01	0,002
influenza	20,4	20,1
diabetes	3,5	3,8
heupartrose	0,9	1,6
heupfractuur	0,5	1,5
depressie	5,4	10,0
dementie	0,8	1,4
astma	6,2	6,0

Bron: Volksgezondheid Toekomst Verkenning, Nationaal Kompas Volksgezondheid. Bilthoven: RIVM; www.nationaalkompas.nl, versie 2.7, 19 mei 2004.

risico onderscheiden: het absolute risico en het relatieve risico.

Het *absolute risico* is gelijk aan de kans op ziekte, complicatie van ziekte of overlijden. Een absoluut risico moet altijd voor een tijdsperiode worden gegeven. Dit kan een concreet omschreven periode zijn, zoals het tienjaarsrisico op een hartinfarct bij een 40-jarige man, maar ook een abstracte tijdsperiode, zoals het *life-time* risico op leukemie, waarmee wordt bedoeld de kans dat iemand in zijn leven leukemie krijgt. Over lange periodes wordt de cumulatieve incidentie en daarmee het risico van sterfte 1. Absolute risico's komen in de geneeskunde in vele gedaanten voor (tabel 1.3). De *vijf- of tienjaarsoverleving* is een cumulatieve incidentie. Onder *letaliteit* of *case fatality* wordt meestal verstaan het risico van overlijden bij patiënten die lijden aan een bepaalde ziekte; het is dan een cumulatieve incidentie van sterfte. Het begrip *attack rate* wordt veel gebruikt voor infectieziekten en is een cumulatieve incidentie over een korte periode. In de medische praktijk wil men het absolute risico zo specifiek mogelijk weten. Zo hebben we weinig aan het kennen van de kans op een hartinfarct van een Nederlander. Men wil die kans weten voor mannen en vrouwen, voor verschillende leeftijden, voor mensen met of zonder hypertensie, met of zonder een positieve familieanamnese voor een hartinfarct, dus het risico gespecificeerd volgens de determinanten van ziekte. Deze determinanten worden vaak risico-indicators genoemd omdat het risico

**Tabel 1.3 Enkele voorbeelden van absoluut risico bij verschillende vraagstellingen.**

vraagstelling	voorbeeld	technische term	risico
risico	kans op myocardinfarct binnen tien jaar bij 40-jarige man	10-jaars cumulatieve incidentie (CI)	2%
	kans op myocardinfarct binnen tien jaar bij 60-jarige man	10-jaars-CI	10%
	kans op opname wegens dementie binnen vijf jaar voor 70-jarige man	5-jaars-CI	1%
	kans op longkanker voor 75e jaar, Nederlandse man	75-jaars-CI	3%
	kans op longkanker voor 75e jaar, Nederlandse vrouw	75-jaars-CI	0,5%
prognose	life-time risico van leukemie	CI over gehele leven	0,6%
	kans op overlijden binnen een jaar na myocardinfarct	letaliteit, case fatality, 1-jaars-CI	6%
interventie	kans op in leven zijn vijf jaar na diagnose ovariumkanker	5-jaarsoverleving, 5-jaars-CI	40%
	kans op in leven zijn vijf jaar na coronaire bypasschirurgie	5-jaarsoverleving, 5-jaars-CI	90%

van ziekte varieert met categorieën van de determinanten. Wanneer de risico-indicator een causale relatie heeft met de ziekte, wordt gesproken over een *risicofactor*. Zo zijn leeftijd en geslacht risico-indicators voor een hartinfarct, terwijl bloeddruk en serumlipiden risicofactoren zijn; leeftijd is een risico-indicator voor longkanker en rookgewoonten een risicofactor. (Sommige auteurs gebruiken de term risicofactor echter zonder causale implicatie.)

#### Kernpunten

- Het absolute risico is gelijk aan de kans op ziekte, complicatie van ziekte of overlijden in een bepaalde periode.
- Een van de bekendste verschijningsvormen van het absolute risico in de klinische epidemiologie is de vijf- of tienjaarsoverleving.

Er zijn twee relatieve maten van risico: het *risicoverschil* (RV) en de *risicoratio*, vrijwel altijd aangeduid als *relatief risico* (RR).

Het relatieve risico geeft het aantal keren dat het absolute risico hoger is in één categorie van de determinant (de indexgroep) tegenover een andere categorie (de referentiegroep). Zo is het relatieve risico op longkanker bij roken ongeveer 10. Dat betekent dat het absolute risico van longkanker onder rokers ( $R_i$ ) tien keer zo hoog is als onder niet-rokers ( $R_o$ ). Het relatieve risico kan worden berekend door de absolute risico's van indexgroep ( $R_i$ ) en referentiegroep ( $R_o$ ) op elkaar te delen:  $RR = R_i/R_o$ . Uit patiënt-controleonderzoek kan men meestal geen absolute risico's berekenen, maar wel de verhouding van de blootstellingsfrequenties bij patiënten en controles. Deze verhouding wordt de *oddsratio* (OR) genoemd en geldt als benadering van het relatieve risico. Overigens moet de betekenis van het RR altijd worden beoordeeld in combinatie met het absolute risico. Een hoog RR bij een zeer laag absoluut risico is nog steeds

een laag absoluut risico, en dus vaak van weinig betekenis.

Het *risicoverschil* is een maat die het verschil tussen twee absolute risico's aangeeft. Als  $R_i$  het risico in de indexgroep en  $R_o$  het risico in de referentiegroep is, dan geldt  $RV = R_i - R_o$ . Deze maat kan worden gebruikt om na te gaan hoeveel het absolute risico van een patiënt hoger is dan dat van een referentiepatiënt. De reciproke van het risicoverschil, dat wil zeggen  $1/RV$ , kan worden gebruikt om het aantal personen te berekenen dat moet worden behandeld om één *event* (complicatie, dood) te voorkomen (het *number needed to treat* of NNT).

Een voorbeeld is ontleend aan een trial bij patiënten met milde hypertensie die werd uitgevoerd door de Engelse Medical Research Council. In deze MRC-trial werden matig hypertensieve patiënten behandeld met een antihypertensivum (bètablokker of diureticum) of placebo. Bij de vergelijking van de effectiviteit van het antihypertensivum met de placebo bleek dat in de medicatiegroep (de indexgroep) de tienjaars cumulatieve incidentie van het cerebrovasculaire accident 1,4% was ( $R_i = 0,014$ ), tegen 2,6% ( $R_o = 0,026$ ) in de placebogroep. Het risicoverschil bedroeg dus 1,2%. Dat betekent dat men gedurende tien jaar  $1/0,012 = 83$  patiënten met een matig hoge bloeddruk moet behandelen om één CVA te voorkomen. Uit hetzelfde onderzoek bleek dat men 333 patiënten met milde hypertensie tien jaar moest behandelen om één myocardinfarct te voorkomen.

Tot nu toe hebben we de relatie tussen één determinant en de kans op ziekte besproken. Voor de klinische praktijk is het van veel grotere waarde de gezamenlijke invloed van meerdere determinanten op het absolute risico van een uitkomst te kennen. Er is daarbij geen causale vraag en voorspellend onderzoek is aangewezen. De gezamenlijke invloed van determinanten kan kwantitatief worden beschreven met een *risicofunctie*. De risicofunctie is in feite

een verzameling van zogenoemde gewichten die aangeven hoeveel de afzonderlijke determinanten bijdragen aan het absolute risico en hoe deze gecombineerd moeten worden. De gewichten worden doorgaans afgeleid van de coëfficiënten uit een regressiemodel. Vaak heeft de risicofunctie de vorm van een scoresysteem dat bij een individuele patiënt wordt toegepast door voor de aanwezigheid van elke determinant een aantal punten (het gewicht) toe te kennen en deze op te tellen. Het totale aantal punten kan dan worden vertaald in een absoluut risico.

Een goed voorbeeld is een groot Nederlands follow-uponderzoek waarin een risicofunctie voor de tienjaarsmortaliteit werd afgeleid (Hoes et al., 1993). Hiermee kan de clinicus aan de hand van een aantal risico-indicatoren en risicofactoren op eenvoudige wijze berekenen wat het risico voor een patiënt is om binnen tien jaar te overlijden. Voor bijvoorbeeld een 60-jarige man met een lengte van 1,74 meter, een gewicht van 87 kilo, een systolische bloeddruk van 165 mmHg waarvoor hij behandeld wordt, een hartfrequentie van 85 slagen/min, en een cholesterolgehalte van 6,3 mmol/l, die 30 sigaretten per dag rookt en geen diabetes of hartinfarct heeft gehad, is dit risico volgens deze functie gelijk aan 35%.

#### Kernpunten

- Er zijn twee relatieve maten van risico: het risicoverschil en het relatieve risico. Vaak wordt het relatieve risico weergegeven als een oddsratio.
- Bij het weergeven van een behandelingseffect heeft het risicoverschil de voorkeur. De reciproke is het aantal personen dat moet worden behandeld om één event te voorkomen (number needed to treat of NNT, zie ook paragraaf 1.7).
- De gezamenlijke invloed van een aantal determinanten op het risico van een ziekte of ziektecomplicatie kan kwantitatief worden weergegeven door een risicofunctie.



#### Intermezzo 1.3

##### De doctrine van evidence-based medicine

Het concept van de evidence-based medicine werd ontwikkeld aan de McMasters Medical School in Canada onder leiding van de internist David Sackett. Het werd in 1992 als nieuw paradigma voor de geneeskunde gepresenteerd in de *JAMA* (Evidence-Based Medicine Working Group, 1992). Het idee is ontstaan vanuit de bewustwording dat voor een arts pathofysiologisch redeneren, individuele klinische ervaring en intuïtie weliswaar onmisbaar zijn, maar niet altijd tot

het juiste medisch handelen leiden. Daarvoor is empirische onderbouwing door systematisch, degelijk en veelal praktijkgericht onderzoek nodig.

De Nederlandse definitie van evidence-based medicine luidt: 'Het gewetensvol, expliciet en oordeelkundig gebruik van het huidige beste bewijsmateriaal om beslissingen te nemen voor individuele patiënten.' In de praktijk betekent dit de integratie van klinische ervaring, pathofysiologisch redeneren en de beste evidence uit systematisch onderzoek. De evidence-based-medicine-beweging stelt daarbij dat de voorkeuren, wensen en verwachtingen van de patiënt een grote rol moeten spelen. De beweging heeft een grote vlucht genomen en men kan zeggen dat het bedrijven van evidence-based medicine tot norm verheven is. Er is een uitstekende Nederlandse inleiding in evidence-based medicine beschikbaar, waarin behalve een degelijke behandeling van de principes ook veel praktische informatie is te vinden (Offringa et al., 2008).

Behalve lof heeft de evidence-based-medicine-doctrine sinds haar oprichting ook een aantal bezwaren en kritieken opgeroepen. Veel daarvan blijken op een foutief begrip van het concept te berusten. Zo wordt bijvoorbeeld beweerd dat de evidence-based-medicine-beweging aanzet tot 'kookboekgeneeskunde' en geen rekening houdt met de wensen van de individuele patiënt. Deze kritiek is ongegrond omdat evidence-based medicine alleen betekent dat er gebruik moet worden gemaakt van het beste bewijsmateriaal, niet dat elke patiënt identiek zou moeten worden behandeld. Andere problemen die genoemd worden, zijn niet uniek voor evidence-based medicine, maar gelden voor alle manieren van geneeskundig handelen. Er wordt bijvoorbeeld gesteld dat er moeilijkheden ontstaan bij het toepassen van de gevonden evidence op een individuele patiënt omdat die altijd min of meer verschilt van de onderzochte populaties. Dit is een probleem van generalisatie en heeft niets met evidence-based werken op zichzelf te maken. Ook pathofysiologische theorieën en klinische ervaring zullen nooit perfect van toepassing zijn op één bepaalde patiënt.

Door de nadruk op het belang van evidence-based werken in de huidige medische opleiding ligt het voor de hand dat deze vorm van geneeskunde steeds vanzelfsprekender zal zijn. Het bedrijven van evidence-based medicine vraagt echter wel een aantal vaardigheden van de arts die in de opleiding nog maar weinig aandacht krijgen. Het gaat dan vooral om het methodologisch beoordelen en samenvatten van onderzoek. Gelukkig is dit werk voor veelvoorkomende problemen vaak al gedaan en zijn de resultaten gepubliceerd in systematische reviews. Een organisatie die zich daarop speciaal heeft toegelegd, is de Cochrane Collaboration. Dit internationale samenwerkingsverband stelt zich ten doel te helpen bij het nemen van wetenschappelijk onderbouwde beslissingen over interventies in de gezondheidszorg. De Cochrane Database of Systematic Reviews (onderdeel van de

Cochrane Library) is een bijzonder nuttig product van de Cochrane Collaboration, waarin vier keer per jaar systematische literatuuroverzichten van hoge kwaliteit worden aangeboden.

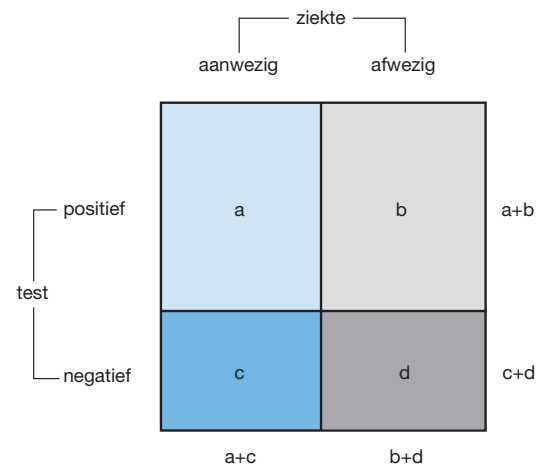
Voor de practicus zijn er ook richtlijnen beschikbaar met evidence-based, landelijk geldende, vakinhoudelijke aanbevelingen voor optimale zorg. De voornaamste kracht van deze richtlijnen is dat toepassing ervan zal leiden tot een kwalitatief betere maar zeker ook efficiëntere zorg. Een grotere efficiëntie kan bijvoorbeeld bereikt worden doordat opvolging van de richtlijn voorkomt dat er onnodige diagnostische tests worden gedaan. Een inherente zwakte is dat een richtlijn algemeen is geformuleerd. Bij de toepassing ervan zal de hulpverlener altijd rekening moeten houden met de individuele eigenschappen en omstandigheden van de patiënt, zoals zijn wensen, draagkracht en sociale omstandigheden. De algemeenheid van een richtlijn betekent ook dat deze niet altijd overal implementeerbaar is, bijvoorbeeld als gevolg van lokale organisatorische omstandigheden.

## 1.5 Diagnose

Een centrale activiteit in de geneeskunde is het stellen van een diagnose. De onderliggende vraag of een patiënt een ziekte al of niet heeft, kan worden uitgedrukt in waarschijnlijkheidstermen. De ziekte is immers niet direct constateerbaar, maar laat zich door haar tekenen met beperkte zekerheid kennen. Een diagnose is dan: de kans dat een patiënt een ziekte heeft, gegeven bepaalde bevindingen. Deze bevindingen kunnen zijn: demografische gegevens zoals leeftijd en geslacht, symptomen of uitslagen van een diagnostische test. Diagnostisch onderzoek is voorspellend onderzoek waarbij als gevolg confounding geen rol speelt.

Een *diagnostische test* voor een bepaalde ziekte levert gegevens die kunnen worden weergegeven zoals in figuur 1.1. Daarbij is een hypothetisch voorbeeld gegeven van het inspannings-ecg als diagnosticum voor angiografisch vastgestelde coronairsclerose. De *sensitiviteit* van een diagnostische test is de proportie zieke mensen met een positieve testbevinding. In de figuur is dat  $a/(a+c)$ . De *specificiteit* is de proportie mensen zonder de onderzochte ziekte met een negatieve testbevinding. In de figuur is dat  $d/(b+d)$ . De *fout-negatieven* worden aangegeven met  $c/(a+c)$ , en de *fout-positieven* met  $b/(b+d)$ .

De sensitiviteit geeft een antwoord op de vraag: Wat is de kans dat een testbevinding positief is, gegeven de aanwezigheid van ziekte? De klinische vraag die wordt gesteld, is precies andersom, namelijk: Wat is de kans dat een patiënt de ziekte heeft, gegeven het feit dat de testbevinding positief is? Deze kans wordt aangegeven met de *predictieve waarde* van een test. In figuur 1.1 is de predictieve waarde van een *positieve* testbevinding, de positieve pre-



$$\text{sensitiviteit: } \frac{a}{a+c}$$

$$\text{specificiteit: } \frac{d}{b+d}$$

$$\text{fout-negatieven: } \frac{c}{a+c}$$

$$\text{fout-positieven: } \frac{b}{b+d}$$

$$\text{positief predictieve waarde: } \frac{a}{a+b}$$

$$\text{negatief predictieve waarde: } \frac{d}{c+d}$$

.....  
**Figuur 1.1** Analyse van diagnostische gegevens.

dictieve waarde, gelijk aan  $a/(a+b)$ . De *negatieve* predictieve waarde is gegeven met  $d/(c+d)$ .

### Kernpunt

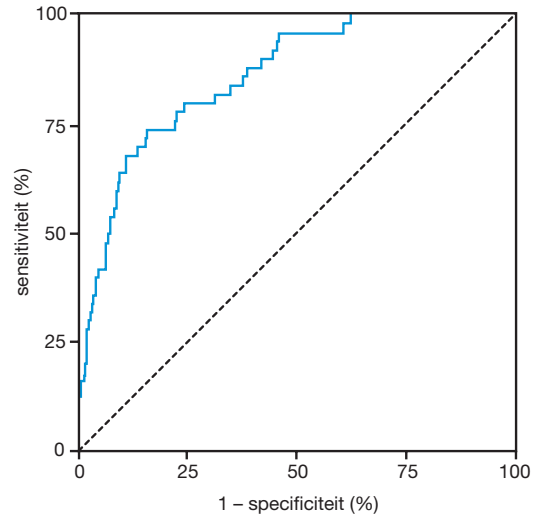
- Voor de klinische praktijk zijn de predictieve waarden van groter belang dan de sensitiviteit en de specificiteit van een diagnostische test.

In de klinische praktijk wordt de waarschijnlijkheid van een bepaalde ziekte nooit op basis van een enkele test, maar altijd op basis van verscheidene tests en bevindingen bepaald. Zo zal de kans op de aanwezigheid van coronairsclerose worden geschat op basis van leeftijd, geslacht, bloeddruk, cholesterolgehalte van het bloed en het ecg. Het combineren van diagnostische informatie tot een kans op ziekte is niet eenvoudig en gebeurt in de klinische praktijk meestal informeel op basis van ervaring. Het belangrijkste probleem vormt het gegeven dat verschillende testuitslagen meestal niet los van elkaar staan. Stel dat men bij de diagnostiek van coronaire hartziekte het lichaamsgewicht al gemeten heeft, dan zal de diagnostische waarde van een bloeddrukmeting kleiner

zijn dan wanneer men geen weet had van het lichaamsgewicht. Dit is het gevolg van de nauwe relatie tussen gewicht en bloeddruk; een deel van de diagnostische waarde van de bloeddruk zit al in het lichaamsgewicht. De vraag bij diagnostiek is daarom wat de onafhankelijke bijdragen zijn van de verschillende tests. Hierbij wordt meestal uitgegaan van demografische variabelen en anamnese, en wordt vervolgens onderzocht wat de toegevoegde waarde is van steeds meer belastende tests (voor patiënt en budget). Bij de patiënt bij wie men een coronaire hartziekte vermoedt, zou men bijvoorbeeld achtereenvolgens kunnen onderzoeken wat de toegevoegde waarde is van lichamelijk onderzoek, bloedonderzoek en het inspannings-ecg.

De sensitiviteit en specificiteit zijn geschikt om de waarde van één test in isolement te kwantificeren en in de literatuur wordt hiervan vaak verslag gedaan. Het gebruik in de kliniek van deze geïsoleerd onderzochte parameters is echter problematisch. Omdat de uitslagen van verscheidene tests niet losstaan van elkaar, kunnen ze niet valide gecombineerd worden tot een kans op ziekte. Toch is het mogelijk op formele wijze de kans op ziekte op basis van verscheidene tests te schatten, waarbij rekening wordt gehouden met de onderlinge afhankelijkheid van de uitslagen en waarbij de toegevoegde diagnostische waarde van een bepaalde test kan worden onderzocht. Dit gebeurt net als bij een risicofunctie door het gebruik van een regressiemodel. Het eindproduct van diagnostisch onderzoek is dan ook een zogenoemde *diagnostische functie*. De diagnostische functie geeft de kans op aanwezigheid van ziekte op basis van de onafhankelijke bijdragen van verschillende determinanten; in de diagnostiek de testbevindingen. Deze voorspelde kans kan gedichotomiseerd worden tot een diagnostische test, waarna de bekende testparameters sensitiviteit, specificiteit en de predictieve waarden berekend kunnen worden. Ook kan men een *receiver operating characteristic (ROC)-curve* maken (figuur 1.2). Dit is een curve van combinaties van de sensitiviteit en 1 – specificiteit voor verschillende afkappunten van een diagnostische test waarvan de uitslag een continue variabele is; bij een diagnostische functie de voorspelde kans. Naarmate de oppervlakte onder deze curve groter is dan 0,5, heeft de diagnostische functie een beter voorspellend vermogen.

Een diagnostische functie zal vooral accuraat zijn bij patiënten die behalve wat betreft het soort klinisch beeld ook wat betreft ernst lijken op die waarbij de functie werd ontwikkeld. Omdat de gemiddelde ernst in hoge mate afhankelijk is van het echelon waar de patiënt zich meldt, zal een diagnostische functie die in een universitair centrum werd ontwikkeld, hoogstwaarschijnlijk slecht functioneren in een huisartspraktijk en vice versa.



Figuur 1.2 Receiver operating characteristic (ROC)-curve.

Als voorbeeld noemen we de diagnostische functie die in 1990 werd ontwikkeld voor patiënten met een acute respiratoire aandoening, die zich melden op een afdeling voor spoedeisende hulp (Heckerling et al., 1990). Bij deze patiënten werd de kans op de aanwezigheid van pulmonale infiltraten gerelateerd aan een eenvoudige score. Voor elk van de volgende bevindingen werd een punt toegekend en de totale score was eenvoudigweg de som van de punten: koorts, hartfrequentie hoger dan 100/min, rhonchi, verminderd ademgeruis en de afwezigheid van astma. De diagnostische functie werd vervolgens op twee andere afdelingen voor spoedeisende hulp gevalideerd. Bij een voorafkans van 12% bleek de score 0 overeen te komen met een kans op pulmonale infiltraten van 1%, en een score van 5 met een kans van 75%.

#### Kernpunten

- Diagnose is de kans op een bepaalde ziekte gegeven bepaalde bevindingen, inclusief leeftijd en geslacht.
- Diagnostisch onderzoek is voorspellend onderzoek; bijgevolg speelt confounding geen rol.
- De vraag bij diagnostisch onderzoek is hoe groot de toegevoegd voorspellende waarde van een bepaalde bevinding is.
- Voor de rapportage van een diagnostisch onderzoek is eveneens een richtlijn beschikbaar: de STARD (STAndards for the Reporting of Diagnostic accuracy studies) statement ([www.stard-statement.org](http://www.stard-statement.org)).

## 1.6 Prognose

De individuele prognose van een patiënt moet, evenals iemands risico en diagnose, worden uitgedrukt in waarschijnlijkheidstermen. Prognose is dan: de kans op een bepaalde ziekte-uitkomst (bijvoorbeeld dood, invaliditeit, volledige genezing, complicaties) na een bepaalde periode, gegeven de diagnose en bepaalde prognostische factoren. Het is het uitdrukken van het beloop van de ziekte in een individuele kans. Het kan daarbij gaan om het natuurlijke, dat wil zeggen onbeïnvloede, beloop of het beloop van de ziekte bij een bepaalde behandeling. In deze zin kan een therapeutische maatregel worden opgevat als een prognostische factor. Om de prognose te bestuderen, is het van belang aandacht te besteden aan drie aspecten: de ziekte, de uitkomst en de prognostische factoren die het verband tussen ziekte en uitkomst bepalen. Deze drie aspecten worden achtereenvolgens kort aangeduid.

Wat de ziekte betreft, is het van belang te weten hoe de patiënten van wie we de prognose willen vaststellen, bij ons zijn gekomen, wat het verwijzingspatroon is geweest en hoe de toegang is tot het diagnostische proces. Dit bepaalt of er een bepaalde selectie en daarmee mogelijke selectiebias is opgetreden en daarnaast geeft het inzicht in de generaliseerbaarheid van de gegevens. De uitkomst van een ziekte kan op verschillende manieren worden gedefinieerd. Het kan gaan om genezing of dood, om functieverlies of restverschijnselen, of om complicaties door andere ziekten. Hoe de uitkomst ook wordt gedefinieerd, het is van het grootste belang dat de criteria voor deze definitie duidelijk en objectief zijn en bij voorkeur eenvoudig, maar bovenal klinisch relevant. Voor het onderzoek naar de prognose is het belangrijk dat de uitkomst bij alle onderzochte patiënten wordt nagegaan (complete follow-up) en dat de uitkomst zonder vertekening wordt vastgesteld. Dit betekent dat informatiebias moet worden voorkomen. Een van de middelen hiertoe is, zoals gezegd, een *geblindeerde* vaststelling van de uitkomst, waarop bij de bespreking van de clinical trial nader wordt ingegaan. Voor het schatten van de prognose op basis van verschillende indicators kan, analoog aan de risicofunctie en de diagnostische functie, een *prognostische functie* worden ontwikkeld. Prognostisch onderzoek moet worden opgevat als causaal onderzoek als het doel is een oorzakelijke relatie tussen determinant en uitkomst aan te tonen, met als ultiem doel de frequentie of ernst van de uitkomst via de determinant te modificeren. Een voorbeeld is depressie na een hartinfarct. Als men in gedachte heeft de depressie te gaan behandelen met als doel de prognose van het hartinfarct te verbeteren, moet causaal onderzoek worden gedaan. In dat geval moet voor verschillende confounders (onder andere roken) worden gecorrigeerd. Als daarentegen het doel is depressie te gebruiken bij risicostratificatie na een hartinfarct, volstaat voorspellend onderzoek.

Een voorbeeld van een prognostische functie uit de acute geneeskunde is de Glasgow Coma Scale, die een schatting geeft van de prognose na een hoofdtrauma (Jennet et al., 1976). Een patiënt die de ogen spontaan opent, krijgt 4 punten, iemand die dat alleen bij een akoestische prikkel doet 3 punten, alleen bij een pijnprikkel 2 punten en anders 1 punt. Op gelijke wijze worden punten toegekend voor de motorische en verbale respons. Het totale aantal punten is dan een maat voor de prognose van de patiënt.

Tegenwoordig worden veel prognostische functies gepubliceerd en de vraag rijst of deze alle toepasbaar zijn in de klinische praktijk. Om dit na te gaan, is het in elk geval belangrijk te weten of de prognostische functie voldoende werd gevalideerd, dat wil zeggen of is aangetoond dat hij ook accuraat is in andere patiëntenpopulaties dan die van het onderzoek. Dit geldt overigens evenzeer voor risicofuncties en diagnostische functies.

### Kernpunten

- Prognose is de kans op een bepaalde ziekte-uitkomst na een bepaalde periode, gegeven de ziekte en prognostische indicators.
- Prognostisch onderzoek is causaal onderzoek als het doel is de ziekte-uitkomst te verklaren, en het is voorspellend onderzoek als het doel louter risicostratificatie is.



### Intermezzo 1.4 Equipoise en de farmaceutische industrie

Een van de voorwaarden om te mogen randomiseren is dat men werkelijk niet weet wat de beste behandeling is bij een bepaalde aandoening. Dit wordt het onzekerheids- of equipoise-beginsel genoemd. Zodra er al serieuze aanwijzingen zijn dat behandeling A superieur is ten opzichte van behandeling B, is het onethisch te randomiseren naar behandeling B. Gerandomiseerde trials die door de industrie worden gesponsord, laten vaker een positief resultaat zien voor het nieuwe, experimentele middel. Een van de mogelijke verklaringen hiervoor is dat in dit onderzoek het equipoise-principe vaker dan gemiddeld wordt geschonden. De controlegroep zou een a priori slechtere behandeling krijgen en derhalve is de kans op een positieve bevinding voor het nieuwe middel groter dan 50%.

Deze hypothese is onderzocht door Djulbegovic et al. (2000). Het uitgangspunt was dat indien het equipoise-principe wordt gerespecteerd, *achteraf* 50% van alle trials een voorkeur voor de nieuwe experimentele behandeling moet laten zien. De onderzoekers verzamelden 136 gepubliceerde trials betreffende de behan-

deling van het multipole myeloom en bepaalden bij elk onderzoek of het gesponsord was door een farmaceutische industrie en of het onderzoek superioriteit van de nieuwe experimentele behandeling liet zien. Het resultaat was dat bij de trials die door non-profitorganisaties waren gefinancierd, inderdaad in ongeveer de helft van de gevallen de experimentele behandeling superieur werd bevonden; dat wil zeggen dat in dit onderzoek het onzekerheidsbeginsel werd gerespecteerd. De trials die al of niet gedeeltelijk door een farmaceutische industrie waren gesponsord, lieten in 74% van de gevallen een gunstiger resultaat voor het nieuwe experimentele middel zien. De onderzoekers concludeerden dat schending van het onzekerheidsprincipe een mogelijke verklaring vormt voor bias in de resultaten van onderzoek dat wordt gesponsord door farmaceutische industrieën.

## 1.7 Interventie

Het doel van een interventie is meestal het zodanig beïnvloeden van het natuurlijke beloop van ziekten dat de kans op een bepaalde ziekte-uitkomst (dood, invaliditeit, complicaties) kleiner wordt. Een ander doel van interventie is de preventie van een bepaalde uitkomst bij een persoon die nog niet ziek is maar wel een verhoogd risico heeft. In de klinische epidemiologie zijn er twee aspecten van interventie die speciaal aan de orde komen: de effectiviteit en de veiligheid. De effectiviteit van een interventie, ook wel het bedoelde effect genoemd, kan uitstekend worden bestudeerd in experimenteel onderzoek. In het therapeutische experiment, vaak clinical trial genoemd, bepaalt de onderzoeker, bij voorkeur at random, welke patiënt een interventie krijgt en welke niet. De veiligheid van interventies, en daarmee de bijwerkingen of onbedoelde effecten, kan om redenen die verder worden besproken moeilijk in experimenteel onderzoek worden bestudeerd. Voor het vaststellen van bijwerkingen van een interventie nemen we dan ook onze toevlucht tot niet-experimenteel (observatieel) onderzoek.

In de paragrafen hierna ligt de nadruk op onderzoek van geneesmiddelen, maar de geldigheid betreft evenzeer onderzoek naar andere (preventieve) interventies.

### 1.7.1 EFFECTIVITEIT

In een clinical trial wordt het effect van een behandeling in kwantitatieve termen bepaald. Het effect kan worden uitgedrukt in vermindering van de frequentie van optreden van ziekte, vermindering van de tijdsduur, ernst of complicaties van de ziekte, of vermindering van het niveau van risicofactoren voor ziekte. In het laatste geval is de aanname dat een reductie in een bepaalde risicofactor voor de ziekte gepaard gaat met een reductie in de frequentie van ziekte. Het grote praktische voordeel is dat het optreden van ziekte niet hoeft te worden afgewacht.

De clinical trials met cholesterolverlagende geneesmiddelen zoals simvastatine en pravastatine bleken behalve een cholesterolverlagend effect later ook een gunstig effect op het risico van coronaire hartziekten te hebben.

Dat verandering in de risicofactor niet altijd verandering in de ziektefrequentie hoeft te betekenen, wordt geïllustreerd door de bevindingen van een grote trial met natriumfluoride voor de behandeling van osteoporose. Uit deze trial bleek dat natriumfluoride weliswaar een aanzienlijke stijging van de botmineraaldichtheid teweegbrengt, maar dat dit niet gepaard gaat met een reductie van de kans op fractures, integendeel, de incidentie van fractures was in de fluoridegroep zelfs hoger.

Bij het vaststellen van het effect van een behandeling hebben we te maken met drie elementen die eerder uitvoerig zijn besproken door De Lang en Lubsen (1987). Het eerste is het natuurlijke beloop van de ziekte, dat van patiënt tot patiënt sterk kan variëren. In de tweede plaats is er een invloed van meetfouten bij de vaststelling van het effect. Zowel de patiënt als degene die het effect moet vaststellen – vaak de arts – is hierin subjectief. Deze subjectiviteit kan de vaststelling van een effect van een behandeling beïnvloeden. In de derde plaats is het effect van een behandeling afhankelijk van het farmacologische effect van een geneesmiddel. In een clinical trial wil men dit farmacologische effect geïsoleerd vaststellen, dus los van het natuurlijke beloop en de subjectieve waarnemingen van patiënt en arts. In een clinical trial gebeurt dit door een interventiegroep die het medicament krijgt, te vergelijken met een referentiegroep (vaak ook placebogroep genoemd). Om het effect van verschillend natuurlijk beloop tussen interventie- en referentiegroep uit te schakelen, wordt gerandomiseerd, dat wil zeggen: de patiënten in een clinical trial worden willekeurig aan de interventie- of referentiegroep toegewezen. Om de subjectiviteit in waarneming van patiënt en arts uit te sluiten, wordt de ziekte-uitkomst vastgesteld zonder dat de waarnemer op de hoogte is van de groep waarin de patiënt zit. Als zowel de waarnemer-arts als de patiënt hiervan niet op de hoogte is, wordt van een dubbelblind onderzoek gesproken. Dit dubbelblinde karakter is goed te bereiken door de referentiegroep een placebo te geven. Vandaar dat in het medisch jargon vaak wordt gesproken over een ‘randomized, placebo-controlled, double-blind clinical trial’. In een dergelijke trial is het mogelijk het verschil in farmacologisch effect tussen de interventie- en de referentiegroep vast te stellen. Hieronder wordt wat dieper ingegaan op de redenen voor randomisatie en blinde vaststelling van de uitkomst.

Het valide meten van het effect van een interventie wordt bedreigd door *regressie naar het gemiddelde* en prognostische onvergelykbaarheid tussen interventie- en referentiegroep, voornamelijk door *confounding by indication*. Door de

patiënten at random toe te wijzen aan een interventie- of controlegroep kunnen de effecten van beide fenomenen worden bestreden.

*Regressie naar het gemiddelde.* Dit fenomeen treedt op indien bij mensen die geselecteerd worden op een hoge waarde voor een variabele, die variabele enige tijd later opnieuw wordt gemeten. Bij hermeting zullen personen met aanvankelijk een hogere waarde nu gemiddeld een lagere waarde hebben. Zo zullen bij mensen die bij de eerste meting hypertensie hebben, er enkelen zijn die bij hermeting spontaan regressie, dat is daling van de bloeddruk, vertonen. Dit verschijnsel is het gevolg van de individuele variatie in de bloeddruk. Patiënten die geselecteerd zijn op een hoge bloeddruk bij de eerste meting, hebben door die variatie bij de tweede meting gemiddeld een grotere kans een lagere waarde te hebben. Indien patiënten op een lage meetwaarde worden geselecteerd, zal bij de nieuwe meting gemiddeld een hogere waarde worden vastgesteld. De regressie treedt in dat geval op in de vorm van een stijging. Regressie naar het gemiddelde is een belangrijk motief voor experimenteel onderzoek, omdat het kan worden uitgesloten door het regressie-effect vast te stellen in de referentiegroep. Zo is in een hypertensietrial het werkelijke effect van een antihypertensief medicament op de bloeddruk vast te stellen uit het verschil in bloeddruk na behandeling tussen interventie- en controlegroep.

Randomisatie kan sterk bijdragen aan de vergelijkbaarheid aan het begin van het onderzoek van de interventie- en de referentiegroep. Door een aselechte indeling van patiënten zal aan het begin van een clinical trial de prognose van beide groepen in principe gelijk zijn. Als namelijk de arts die het experiment uitvoert niet verplicht wordt te randomiseren, zal hij de neiging hebben de ernstiger zieke patiënten het te onderzoeken geneesmiddel te geven en de minder ernstig zieke patiënten de placebo. Deze onvergelijkbaarheid van prognose wordt *confounding by indication* genoemd en is een belangrijk motief voor gerandomiseerd onderzoek.

Zo is de ernst van een myocardinfarct, bijvoorbeeld uitgedrukt als de linkerventrikelfunctie, een belangrijke indicatie voor het al of niet instellen van behandeling met bijvoorbeeld een ACE-remmer. Daarbij is de ernst van het infarct een belangrijke prognostische factor voor een herinfarct of voor overlijden. De directe vergelijking van patiënten met een infarct en medicatie (in dit geval de ACE-remmer) en patiënten met een infarct zonder medicatie kan daarmee een zeer verstoorde vaststelling van het effect van het medicament opleveren.

Een ander voorbeeld is het feit dat de prognose van behandelde hypertensiepatiënten gemiddeld slechter is dan die van onbehandelde hypertensiepatiënten. Dit is niet zozeer het gevolg van het feit dat de anti-

hypertensieve medicatie ineffectief of zelfs schadelijk is, maar wel van het feit dat patiënten die voor hypertensie worden behandeld als regel een veel ernstigere hypertensie en ook op andere wijze een slechtere prognose hebben dan onbehandelde hypertensiepatiënten.

### Kernpunten

- Randomisatie kan sterk bijdragen aan de vergelijkbaarheid aan het begin van het onderzoek van de interventie- en de referentiegroep. Door een aselechte indeling van patiënten zal aan het begin van een clinical trial de prognose van beide groepen in principe gelijk zijn.
- Het doel van de blinde vaststelling van de ziekte-uitkomst is de vermindering van informatiebias die kan optreden ten gevolge van de subjectieve waarneming door arts of patiënt van het effect van een behandeling.

Randomisatie is evenwel geen garantie voor vergelijkbaarheid. Enkele belangrijke debatten in de medische literatuur zijn terug te voeren op het feit dat er ondanks randomisatie verschillen bestonden in prognostische factoren tussen interventie- en placebogroep.

In de jaren zeventig van de vorige eeuw werd in het University Group Diabetes Program (een gerandomiseerde trial) de frequentie van cardiovasculaire complicaties van diabetes vergeleken tussen patiënten die met het orale bloedglucoseverlagende middel tolbutamide (een sulfonylureumderivaat) werden behandeld en patiënten die insuline kregen. In de tolbutamidegroep werd een hogere cardiovasculaire mortaliteit geconstateerd dan in de insulinegroep. Het debat ontstond doordat ondanks de randomisatie bleek dat in de tolbutamidegroep relatief meer personen met een voorgeschiedenis van hart- en vaatziekten zaten en statistische aanpassing het wezen van de randomisatie zou aantasten. (Een recent Brits onderzoek toonde aan dat behandeling met sulfonylureumderivaten niet beter of slechter was dan behandeling met insuline.)

Het doel van de blinde vaststelling van de ziekte-uitkomst is, zoals gezegd, de vermindering van informatiebias die kan optreden ten gevolge van de subjectieve waarneming door arts of patiënt van het effect van een behandeling. Een belangrijk principe hierbij is dat symmetrie in de vaststelling van het effect moet worden nagestreefd. Dit kan het best worden bereikt door een placebogroep in te stellen, waardoor het voor de patiënt en de waarnemer onmogelijk is onderscheid tussen interventie- en referentiegroep

te maken. In dat geval is de symmetrie in dataverzameling gegarandeerd.

In de *opzet* van de clinical trial zijn er verschillende mogelijkheden, die hier slechts kort worden genoemd. Het klassieke design is dat van de parallelle groepen, waarbij de interventie- en referentiegroep parallel in de tijd met elkaar worden vergeleken. Dit is het meest voorkomende design en wordt hier niet verder besproken. Een alternatief is het *cross-overdesign*, waarbij de interventie- en placebomedicatie in aselecte volgorde aan dezelfde patiënt worden toegediend. Deze opzet kan alleen worden toegepast om de effectiviteit van geneesmiddelen voor de reductie van de ernst of frequentie van symptomen van chronische aandoeningen te onderzoeken. Een van de problemen is het zogenoemde *carry-overeffect* dat bij veel geneesmiddelen niet is uit te sluiten. Dit betekent dat het middel dat in de eerste periode werd gegeven, nog doorwerkt in de tweede periode. Om deze reden wordt vaak een *wash-outperiode* tussen beide behandelingen voorgeschreven, waarin de patiënt geen behandeling krijgt. Het voordeel van *cross-overonderzoek* is de veel grotere precisie, doordat de wetenschappelijke vergelijking binnen één patiëntengroep in plaats van tussen twee patiëntengroepen wordt gemaakt. Daardoor kunnen deze trials aanzienlijk kleiner zijn.

Wilding et al. (1997) onderzochten de effectiviteit van een inhalatie- $\beta_2$ -mimeticum bij 101 astmapatiënten die hiervoor reeds corticosteroiden gebruikten. Patiënten kregen twee keer gedurende een halfjaar in willekeurige volgorde eerst het inhalatie- $\beta_2$ -mimeticum of eerst de placebo, met daartussen een wash-outperiode van een maand. Gemiddeld bleek dat de patiënten gedurende het gebruik van dit nieuwe middel minder astma-aanvallen hadden en minder corticosteroiden gebruikten dan tijdens placebogebruik.

De zogenoemde *n=1-trial* is een recent gepresenteerde onderzoeksvorm die in zekere zin een formalisering is van eeuwenlange geneeskundige praktijk. In dergelijk onderzoek wordt verschillende malen in willekeurige volgorde aan één patiënt verschillende medicamenten of één medicament in verschillende doses toegediend om daarmee op objectieve wijze het meest werkzame medicament of de meest werkzame dosis bij de betrokken patiënt vast te stellen. Deze vorm wordt ook vaak binnen grotere clinical trials toegepast om de medicatie van de afzonderlijke patiënten te titreren.

Naast de effectiviteit is de laatste decennia ook de kosteneffectiviteit van een interventie een belangrijk onderwerp van onderzoek. Het budget voor de gezondheidszorg is eindig en beleidsmakers hebben evidence nodig van de doelmatigheid van een nieuwe interventie, zij het diagnostisch, therapeutisch, preventief of anderszins. De doel-

Magee en Redman (1996) beschreven een *n=1-trial* bij een vrouw met ernstige hyperemesis gravidarum die slecht bleek te reageren op de reguliere therapie. Omdat er verder geen behandelingsopties waren, werd de vrouw *informed consent* gevraagd en werd in willekeurige volgorde gedurende periodes van vijf dagen prednisolon of placebo voorgeschreven, telkens gescheiden door een tweedaagse wash-outperiode. Helaas konden de onderzoekers geen gunstig effect van prednisolon op de mate van misselijkheid aantonen.

matigheid of ook wel efficiëntie van een interventie geeft aan hoe de verhouding is tussen de gezondheidswinst en de daarvoor in te zetten middelen: de kosten-batenverhouding. Voor een enkele interventie is deze verhouding niet zo interessant. Het gaat om de vergelijking van de kosten-batenverhouding tussen behandelingsalternatieven. Daarom worden in kosteneffectiviteitsonderzoek altijd meerdere alternatieven vergeleken; meestal een nieuwe interventie ten opzichte van de gangbare zorg. Met deze gegevens wordt het voor beleidsmakers mogelijk om met meenemen van alle relevante informatie een afgewogen besluit te kunnen nemen.

Niet alleen van beleidsmakers maar ook van artsen wordt verwacht dat zij de kosteneffectiviteit van een interventie laten meewegen bij hun besluitvorming. Dit blijkt uit de Nederlandse artseneed, waarin onder andere wordt gesteld: 'Ik zal de beschikbaarheid en de toegankelijkheid van de gezondheidszorg bevorderen.' De eed verplicht hiermee de arts doelmatig te werken, omdat ondoelmatig, niet-kosteneffectief handelen onverantwoordbaar is met het bevorderen van de beschikbaarheid en de toegankelijkheid van de zorg. Overigens wordt kosteneffectiviteitsonderzoek ook wel aangeduid als economische evaluatie, doelmatigheidsonderzoek, *medical technology assessment* (MTA) of *health technology assessment* (HTA).

Veel vooraanstaande tijdschriften eisen dat onderzoekers voor het publiceren van een randomised controlled trial zich houden aan de CONSORT (Consolidated Standards of Reporting Trials) statement ([www.consort-statement.org](http://www.consort-statement.org)). Hierin staat in een checklist voorgeschreven welke punten in een artikel over een gerandomiseerde clinical trial aan de orde moeten komen. De checklist is zeer uitgebreid en kan behalve noodzakelijk voor rapportage ook van nut zijn in alle andere fasen van het onderzoek: van onderzoeksontwerp tot en met data-analyse.

#### Kernpunten

- Pragmatische versus verklarende (ook wel pathofysiologische) trial: terwijl men bij een verklarende trial alleen geïnteresseerd is in de effecten (efficacy) van één specifiek deel van de interventie, zoals een bepaalde chemisch stof, moet de pragmatische

trial de vraag beantwoorden wat in de praktijk de beste behandelingsstrategie is. De placebocontroleerde dubbelblinde trial is daarom van het verklarende type. Een trial waarin bijvoorbeeld een chirurgische behandeling wordt vergeleken met een medicamenteuze behandeling is pragmatisch. De interesse is in het geheel van effecten van de chirurgische ingreep, inclusief die van onder andere narcose, incisie, placebo-effecten en bloedverlies, vergeleken met het geheel van effecten van medicamenteuze therapie: effectiveness. Belangrijkste consequentie is dat pragmatische trials open zijn, dat wil zeggen niet dubbelblind. Verder wordt in een pragmatische trial een nieuwe interventie meestal vergeleken met een bestaande interventie, niet met placebo zoals bij een verklarende trial. Ten slotte worden in een pragmatische trial bij voorkeur patiënten geïnccludeerd zoals die zich in de dagelijkse praktijk voordoen. Bij verklarende trials wordt vaak sterke selectie toegepast.

- Number needed to treat (NNT): het aantal patiënten dat behandeld moet worden om één uitkomst te voorkomen. Het is numeriek gelijk aan de reciproke van het risicoverschil:  $1/RV$ . Als het risico van een heupfractuur bij een bepaalde behandeling voor osteoporose 5% is en dit risico bij placebobehandeling gelijk is aan 10%, is het NNT gelijk aan  $1/0,05 = 20$ . Het NNT is, evenals het RV, afhankelijk van de follow-upduur.
- Naast het meten van de effectiviteit van een bepaalde interventie is het bepalen van de kosteneffectiviteit een belangrijke opdracht van de klinische epidemiologie.

### 1.7.2 VEILIGHEID

Niet-experimenteel onderzoek (observationeel onderzoek) is van beperkte waarde bij het bestuderen van het effect van behandeling. Zoals hiervoor beschreven, is dat vooral het gevolg van fenomenen zoals regressie naar het gemiddelde en confounding by indication. De *onbedoelde effecten of bijwerkingen* van geneesmiddelen kunnen heel goed in niet-experimenteel onderzoek worden bestudeerd. Dat is in het bijzonder het geval omdat het probleem van confounding by indication niet speelt. De ziekte, dat is de indicatie waarvoor het geneesmiddel dat leidt tot de bijwerking wordt gegeven, is immers in het algemeen niet gerelateerd aan die bijwerking. Hieronder wordt in het kort een opsomming gegeven van de verschillende onderzoeksmogelijkheden om bijwerkingen te bestuderen (zie ook tabel 1.1).

Het *case report of de patiëntenserie* is van oudsher een zeer belangrijke bron van kennis over bijwerkingen (Vandenbroucke en Hofman, 1997).

*Cohortonderzoek of follow-uponderzoek* kan goed worden gebruikt om bijwerkingen na te gaan. Hierbij wordt

Als voorbeelden kunnen worden gegeven: agranulocytose na analgeticagebruik, retroperitoneale fibrose na gebruik van de bètablokker practolol, en meer in het verleden retrolentale fibrose na zuurstoftoediening bij premature baby's en het optreden van vulvacarcinoom bij dochters van moeders die tijdens de zwangerschap di-ethylstilbestrol hadden gebruikt. Een recent voorbeeld is het vermeende hogere risico op acuut myocardinfarct bij sildenafilgebruik. Dit werd beschreven in *The Lancet* naar aanleiding van één 65-jarige patiënt (Feenstra et al., 1998). Zoals te verwachten ontstond discussie of de patiënt die werd beschreven niet al voorafgaand aan het gebruik van sildenafil een verhoogd risico op myocardinfarct had.

een indexgroep van gebruikers van het geneesmiddel vergeleken met een referentiegroep van mensen die het geneesmiddel niet gebruiken. Een belangrijk principe bij de selectie van de referentiegroep is dat het basisrisico van een bijwerking in de referentiegroep even groot moet zijn als in de indexgroep. Met andere woorden: indien de indexgroep geen geneesmiddel zou hebben gebruikt, zouden er evenveel events die als bijwerkingen worden aangeduid moeten zijn opgetreden als in de referentiegroep. Dit onderzoekdesign wordt toegepast in zogenaamde *postmarketingsurveillance* van geneesmiddelen.

*Patiënt-controleonderzoek* kan uitstekend worden toegepast om zeldzame bijwerkingen vast te stellen. De patiëntengroep wordt dan gevormd door mensen met de ziekte die het gevolg is van de bijwerking (bijvoorbeeld agranulocytose), en een referentie- of controlegroep wordt gevormd door een groep mensen die deze ziekte ten gevolge van de bijwerking niet heeft – het kan overigens zeer wel een andere ziekte zijn. In een patiënt-controleonderzoek wordt dan de frequentie van geneesmiddelengebruik in de patiëntengroep vergeleken met die in de referentiegroep. Zo wordt het gebruik van analgetica in de agranulocytosegroep vergeleken met het gebruik van analgetica in de referentiegroep.

Een voorbeeld waarin deze opzet in een Nederlands onderzoek is gebruikt, is het onderzoek naar plotse hartdood als mogelijke bijwerking bij patiënten die in verband met hypertensie met niet-kaliumsparende diuretica worden behandeld. In dit patiënt-controleonderzoek werden patiënten met hypertensie die plotseling overleden vergeleken met vergelijkbare hypertensiepatiënten die niet overleden. De frequentie van het gebruik van diuretica, en vooral van niet-kaliumsparende diuretica, werd vastgesteld. Zo kon worden nagegaan of er een oversterfte was bij patiënten die niet-kaliumsparende diuretica kregen voorgeschreven.

**Kernpunt**

- De onbedoelde effecten of bijwerkingen van geneesmiddelen kunnen meestal goed in niet-experimenteel onderzoek worden bestudeerd, omdat het probleem van confounding by indication niet zo speelt. Dit geldt alleen voor die bijwerkingen die onvoorspelbaar zijn. Als een bijwerking logischerwijs voortvloeit uit de werking van het middel, zoals orthostatische hypotensie bij het gebruik van antihypertensiva, kan confounding by indication wel een probleem vormen.

**1.7.3 META-ANALYSE**

Vaak is het nodig op basis van de resultaten van verschillende onderzoeken de typische grootte van een bepaald effect te schatten, bijvoorbeeld voor het opstellen van een richtlijn. Dat is niet simpel. Onder andere door de beperkte precisie, vooral van de kleinere onderzoeken, lopen de resultaten vaak sterk uiteen, waardoor het typische effect niet direct zichtbaar is. De meta-analyse is bedoeld om de resultaten van verschillende onderzoeken op formele wijze kwantitatief samen te vatten en daarmee een precieze schatting van het effect te maken. Vaak gaat het om de schatting van het typische effect van een interventie, maar niet altijd. Het kan ook gaan om de waarde van een diagnostische procedure of het risico op ziekte.

Een voorbeeld van een meta-analyse van een vraag van risico is het onderzoek naar de relatie tussen de botmineraaldichtheid en de kans op fracturen (Marshall et al., 1996). In dit onderzoek werden de originele gegevens van elf observationele onderzoeken verzameld, zodat men de uiteindelijke analyses met 90.000 persoonsjaren en meer dan 2000 fracturen kon doen. De kans op een fractuur bleek met gemiddeld 50% toe te nemen per standaarddeviatieafname van de botmineraaldichtheid.

Na zorgvuldige selectie van de onderzoeken waarvan men de resultaten wil samenvatten, moet de effectparameter worden gekozen. Bij een dichotome uitkomst zou dat een risicoverschil kunnen zijn en bij een continue uitkomst het gemiddelde verschil. Met behulp van verschillende technieken is het mogelijk vervolgens een samenvattend effect te schatten, waarbij rekening wordt gehouden met de verschillende groottes van de onderzoeken. Het samenvatten van de verschillende effecten uit verscheidene onderzoeken in één effect veronderstelt dat er in werkelijkheid maar één typisch onderliggend effect is en dat verschillen tussen onderzoeken uitsluitend berusten op toevalsvariantie. Deze aanname van homogeniteit is waarschijnlijk lang niet altijd juist. Zo is het bijvoorbeeld zeer

waarschijnlijk dat het effect van een interventie varieert met de ernst van de aandoening of met het behandelingscentrum. Het is in een meta-analyse dan ook absoluut noodzakelijk dergelijke verschillen goed te onderzoeken. Dit kan door middel van zogenoemde *forest plots*, waarin de effectparameters met betrouwbaarheidsintervallen uit de verschillende onderzoeken worden afgebeeld. Er zijn ook statistische toetsen voor homogeniteit beschikbaar. Deze moeten echter voorzichtig geïnterpreteerd worden en men moet zeker niet bij een niet-significant resultaat concluderen dat er homogeniteit bestaat. Indien men concludeert dat er sprake is van heterogeniteit, kan met de verschillen tussen onderzoeken rekening worden gehouden met zogenoemde *random-effects*-modellen. Ook kan men aparte schattingen voor subgroepen maken of afzien van het berekenen van een enkele effectparameter. Meta-analysen worden vaak gedaan in het kader van een systematisch literatuuroverzicht, waarin men volgens protocol relevant onderzoek identificeert en kwalificeert en waarvan men vervolgens de resultaten samenvat in een meta-analyse. Zoals gezegd publiceert de Cochrane Collaboration systematische literatuuroverzichten met meta-analysen over tal van onderwerpen.

Een belangrijk probleem bij een meta-analyse ontstaat doordat statistisch significante trials eerder in de medische literatuur terechtkomen dan de niet-significante. Aangezien trials waarin een groot effect wordt gevonden eerder statistisch significant zijn, wordt in een meta-analyse van gepubliceerde trials het therapeutische effect in onbekende mate overschat. Deze vertekening wordt publicatiebias genoemd.

Een duidelijk voorbeeld van publicatiebias werd aan het licht gebracht door Turner et al. (2008). Zij analyseerden de resultaten van 74 onderzoeken naar de effectiviteit van selectieve serotonineheropnameremmers (SSRI's) bij depressie die tussen 1987 en 2004 bij de Amerikaanse Food and Drug Administration (FDA) aangemeld werden. Van de 38 onderzoeken die volgens de FDA 'positief' waren, dat wil zeggen dat ze een gunstig effect op depressie hadden, werden er 37 (99%) gepubliceerd. Van de 36 onderzoeken met een volgens de FDA neutraal (twijfelachtig) of negatief resultaat werden er drie (8%) gepubliceerd, 22 (61%) niet gepubliceerd en elf (31%) met een te gunstig effect gepubliceerd. Of de publicatiebias ontstond op het niveau van de onderzoekers of sponsors die negatieve bevindingen achterhielden, of op het niveau van editors en reviewers kon niet worden achterhaald.

Een betere schatting zou dus op basis van alle gepubliceerde en niet-gepubliceerde onderzoeken moeten gebeuren. De gegevens uit de niet-gepubliceerde onderzoeken zijn echter lastig te achterhalen en mogelijk kwalitatief

inferieur omdat ze nog niet aan *peer review* werden onderworpen. In de praktijk wordt de eerste selectie van onderzoeken voor meta-analyse uitgevoerd met behulp van computerprogramma's (bijvoorbeeld MEDLINE via PubMed, EMBASE) die in bestanden van de gepubliceerde literatuur kunnen zoeken. Hoewel in mindere mate dan bij het klassieke literatuuroverzicht, speelt ook bij de meta-analyse de subjectiviteit van de onderzoeker een grote rol. Uiteindelijk bestaat er een aanzienlijke vrijheid ten aanzien van de selectie van onderzoek, de analyse van de gegevens en uiteraard ook de conclusies. De auteur van de meta-analyse heeft daarom, ook gezien het gewicht dat doorgaans aan het resultaat wordt toegekend, een grote verantwoordelijkheid. Tegelijkertijd moet ook de lezer zich bewust zijn van de subjectieve kanten van een meta-analyse alvorens hij vergaande conclusies trekt.

Voor het schrijven van artikelen over een meta-analyse is een richtlijn verschenen, getiteld Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), die net als alle andere richtlijnen gratis beschikbaar is gemaakt via internet ([www.prisma-statement.org](http://www.prisma-statement.org)).

#### Kernpunten

- Het valide meten van het effect van een interventie wordt bedreigd door regressie naar het gemiddelde, door prognostische onvergelykbaarheid tussen interventie- en referentiegroep, vooral door confounding by indication, en door informatiebias door asymmetrische vaststelling van het effect.
- Deze fenomenen kunnen worden bestreden door randomisatie en blinding.
- Onvoorspelbare effecten van geneesmiddelen kunnen vaak valide met niet-experimenteel onderzoek worden bestudeerd.
- Naast de onderzoeksopzet is de data-analyse van groot belang voor de precisie, validiteit en toepasbaarheid van resultaten uit klinisch-epidemiologisch onderzoek.

#### Literatuur

1. Djulbegovic B, Lacey M, Cantor A, et al. The uncertainty principle and industry-sponsored research. *Lancet*. 2000;356:635-8.
2. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*. 1992;268:2420-5.
3. Feenstra J, Drie-Pierik RJHM van, Laclé CF, Stricker BHCh. Acute myocardial infarction associated with sildenafil. *Lancet*. 1998;352:957-8.
4. Heckerling PS, Tape TG, Wigston RS, et al. Clinical prediction rule for pulmonary infiltrates. *Ann Intern Med*. 1990;113:664-70.
5. Hoes AW, Grobbee DE, Valkenburg HA, et al. Cardiovascular risk and all-cause mortality; a 12 year follow-up study in the Netherlands. *Eur J Epidemiol*. 1993;9:285-92.
6. Hofman A, Grobbee DE, Lubsen J. *Klinische epidemiologie*. Utrecht: Bunge, 1996.
7. Jennett B, Teasdale G, Braakman R, et al. Predicting outcome in individual patients after severe head injury. *Lancet*. 1976;1:1031-4.
8. Lang R de, Lubsen J. *Klinisch geneesmiddelenonderzoek*. Utrecht: Bunge, 1987.
9. Magee LA, Redman CWG. An N-of-1 trial for treatment of hyperemesis gravidarum. *Br J Obstet Gynaecol*. 1996;103:478-80.
10. Marshall D, Johnell O, Wedel H. Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. *BMJ*. 1996;312:1254-9.
11. Offringa M, Assendelft WJJ, Scholten RJPM. *Inleiding in evidence-based medicine*, 3e druk. Houten: Bohn Stafleu van Loghum, 2008.
12. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358:252-60.
13. Vandenbroucke JP, Hofman A. *Grondslagen der epidemiologie*, 5e druk. Utrecht: Bunge, 1997.
14. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet*. 2004;363:1728-31.
15. Vandenbroucke JP. The HRT controversy: observational studies and RCTs fall in line. *Lancet*. 2009;373:1233-5.
16. Wilding P, Clark M, Coon JT, et al. Effect of long-term treatment with salmeterol on asthma control: a double blind, randomised crossover study. *BMJ*. 1997;314:1441-6.

#### Links naar enkele relevante websites:

17. [www.cochrane.nl](http://www.cochrane.nl)  
De website van de Nederlandse afdeling van de Cochrane Collaboration (zie o.a. intermezzo 1,3).
18. [www.sph.emory.edu/~dkleinb/activepi.htm](http://www.sph.emory.edu/~dkleinb/activepi.htm)  
Homepage van David Kleinbaum over zijn commercieel verkrijgbare computerprogramma ActivEpi, een uitstekende interactieve multimedia cursus over de grondslagen van de epidemiologie.
19. [www.consort-statement.org](http://www.consort-statement.org)  
Het CONSORT-statement is bedoeld voor het verbeteren en standaardiseren van de rapportage van randomised controlled trials en bestaat uit richtlijnen voor het beschrijven van methode, resultaten en discussie van wetenschappelijke artikelen over dit type onderzoek.
20. [www.prisma-statement.org](http://www.prisma-statement.org)  
Idem, maar dan speciaal met betrekking tot systematische literatuuroverzichten en meta-analysen.
21. [www.strobe-statement.org](http://www.strobe-statement.org)  
Idem, maar dan met betrekking tot observationeel onderzoek.
22. [www.stard-statement.org](http://www.stard-statement.org)  
Idem, maar dan voor diagnostisch onderzoek.
23. [www.nationaalkompas.nl](http://www.nationaalkompas.nl)  
Het Nationaal Kompas van het Rijksinstituut voor Volksgezondheid en Milieu biedt informatie over gezondheid, ziekte, risicofactoren, zorg en preventie in Nederland.