

Statistische aspecten van het geneesmiddelenonderzoek*

DOOR DR. CHR. L. RÜMKE, ARTS**

Veel artsen valt het moeilijk zich een voorstelling te maken van het werk van de medisch statisticus. Sommigen denken dat hij zich in hoofdzaak wijdt aan het turven van bijvoorbeeld de aantallen besmettelijke ziekten en dergelijke en aan het maken van tabellen. Anderen lijken in de mening te verkeren, dat de dagtaak van een statisticus - medisch of niet-medisch - bestaat uit het jongleren met getallen. Dit spel zou kunnen leiden tot meer of minder gerechtvaardigde conclusies over de uitkomsten van onderzoeken. Men krijgt wel eens de indruk, dat velen intuïtief veronderstellen, dat deze conclusies des te betrouwbaarder zijn naarmate het rekenwerk - en het rekentuig waarmee de berekeningen worden uitgevoerd - gecompliceerder zijn. Opvallend is de eerbied waarmee men soms vermeldt, dat een bepaalde conclusie berust op berekeningen met een computer. Hierin klinkt door dat het resultaat daarvan wel veel betrouwbaarder moet zijn dan dat wat met eenvoudiger hulpmiddelen kan worden bereikt.

Van de medisch statisticus wordt bovendien nogal eens verondersteld, dat hij in een ivoren toren van theorie leeft. Daarin zou hij zozeer buiten de praktijk van de geneeskunde staan, dat hij geen idee kan hebben van wat het beoefenen van de geneeskunde in het dagelijks leven betekent.

Wie zich een dergelijk beeld van de statisticus maakt heeft geen goede voorstelling van de werkelijkheid. Vanzelfsprekend houdt hij zich óók bezig met de wiskundige analyse van uitkomsten van waarnemingen. Natuurlijk behoort het woord significant tot zijn dagelijks gebruikt jargon. Het heeft voor hem echter niet de minste zin een uitgebreide statistische analyse van waarnemingsuitkomsten uit te voeren, wanneer de onderzoeken niet op verantwoorde wijze zijn verricht. Verantwoord dan in die zin, dat een antwoord kan worden verkregen op de vraag, die de onderzoekers zich stelden toen zij met hun werk begonnen. *De Jonge* (1964) schrijft: „Soms heeft men kennelijk het idee, dat derderangs waarnemingsmateriaal, aangeraakt door de toverstaf der statistiek, eerste-rangs uitspraken kan opleveren”. Dat kan bepaald niet!

Men kan inderdaad wel eens de indruk krijgen, dat het in de statistiek in hoofdzaak gaat om het maken van tabellen en ingewikkelde berekeningen

*Naar een voordracht, gehouden tijdens het veertiende N.H.G.-congres op 15 november 1969 te Amsterdam

** Afdeling medische statistiek van de faculteit der geneeskunde Vrije Universiteit

en dat deze een doel in zichzelf vormen. Hetgeen opvalt is echter slechts de buitenkant, het technische aspect van de statistiek. Veel minder valt het logische aspect in het oog, dat minstens even belangrijk is.

Het gaat in de eerste plaats om de logica die ten grondslag ligt aan de manier, waarop onderzoeken en experimenten worden opgezet en verricht, en daarna - pas veel later - om de analyse, die slechts kan worden uitgevoerd wanneer het logisch goed blijkt te zitten. Steeds weer gaat het om vragen zoals: *wat* wil men in dit onderzoek precies te weten komen? *Hoe* moet men het onderzoek opzetten? *Hoe* moet men zijn gegevens *verzamen*? *Welke* waarnemingen moet men dan doen en *op welke wijze* moet men deze doen? Hoe moet men handelen om een duidelijk antwoord te krijgen op de gestelde vragen en dubbele bodems in dat antwoord te vermijden?

De vluchtige beschouwer ziet als het werk van de statisticus veelal alleen - en buiten het verband van het geheel - de analyse aan het slot van een onderzoek met alle daaraan verbonden rekenwerk. Meestal is wat daaraan voor de statisticus is voorafgegaan minstens even belangrijk. De hulp die hij in deze fase kan bieden is veel geruislozer en minder spectaculair. Naar mijn mening is dit deel van ons werk echter van minstens even grote betekenis als het rekenwerk.

Helaas krijgen wij veel onderzoeken pas onder ogen, wanneer deze al zijn uitgevoerd. Het kan dan gebeuren dat bij het telefonisch maken van een afspraak voor een consult wordt gezegd: „Het zal u slechts weinig tijd kosten. Het is de bedoeling dat u alleen maar even kijkt of het significant is”. In sommige gevallen blijft ons dan geen andere mogelijkheid dan - zoals Green het formuleerde - „performing a kind of postmortem upon the numerical remains of a badly planned study”. Overeenkomstige problemen levert het beoordelen van artikelen voor tijdschriften. Het wordt dan vaak een soort wetenschappelijke „accountancy”: mag de onderzoeker werkelijk *uit dit materiaal* de conclusie trekken, die hij wenst te trekken? Blijkt uit deze cijfers nu werkelijk, dat het gebruik van het nieuwe geneesmiddel M over het geheel genomen tot betere resultaten voert bij een bepaalde aandoening dan het tot dusver gangbare middel C? Verdient de toegepaste nieuwe operatietechniek werkelijk de voorkeur boven de tot dusver gebruikelijke?

Om ieder misverstand over het werk van de sta-

tisticus te vermijden is het goed te bedenken, dat hij geen oordeel uitspreekt over de juistheid van de conclusie *ten opzichte van de werkelijkheid*. Wanneer hij zegt: „Dit onderzoek bewijst niet, dat middel M een aanwinst vormt voor de therapie”, dan zegt hij niet: „Dit middel *is* geen aanwinst voor de therapie”. Bij het formuleren van zijn oordeel moet hij bijvoorbeeld steeds rekening ermee houden, dat dit later in een goed opgezet onderzoek, wel degelijk het geval zou kunnen blijken te zijn.

Huisartsen kunnen uiterst belangrijke bijdragen leveren bij het geneesmiddelenonderzoek. Voor bepaalde soorten onderzoekingen zijn ze zelfs de enigen, die deze kunnen verrichten. De vraag wordt gesteld, op welke punten een statisticus hen daarbij kan helpen. Het antwoord zou kortweg kunnen luiden: op alle punten, waarbij dit ook mogelijk is voor alle andere artsen. Daaraan kan dan meteen worden toegevoegd, dat de moeilijkheden van huisartsen bij het opvolgen van de door de statisticus te geven adviezen niet groter behoeven te zijn dan die van andere artsen. Zij zitten in het algemeen niet krapper in hun tijd; evenmin is het gevoel voor logica bij hen minder ontwikkeld dan bij hun collega's die op andere terreinen werken.

Een statisticus kan de huisarts helpen aan het eind van een onderzoek wanneer het erom gaat te bepalen, welke conclusies uit de waarnemingen kunnen worden getrokken. Daarbij kan het bijvoorbeeld gaan om het berekenen van gemiddelden, standaardafwijkingen, betrouwbaarheidsintervallen, correlatiecoëfficiënten en dergelijke, of om het uitvoeren van toetsen. (Sommigen gebruiken hiervoor de nogal ongelukkig gekozen uitdrukking significantieonderzoek.) Op dit soort hulp zullen wij thans niet nader ingaan. Wij willen hier onze aandacht in hoofdzaak richten op de hulp, die een statisticus aan huisartsen vóór het begin van een geneesmiddelenonderzoek kan bieden bij het opzetten daarvan. Wij volstaan met een korte vermelding van verschillende punten, waarop deze hulp betrekking kan hebben; over elk punt zou in principe een veel bredere beschouwing kunnen worden gegeven.

In een bepaald geneesmiddelenonderzoek zullen in het algemeen de effecten van twee geneesmiddelen worden vergeleken bij de behandeling van patiënten met een bepaalde aandoening. Wanneer verschillende (huis-)artsen aan een dergelijk onderzoek deelnemen moet worden getracht zoveel mogelijk eenheid te bereiken bij het stellen van de diagnose van de beschouwde aandoening door alle deelnemende artsen. Een statisticus kan behulpzaam zijn bij de pogingen deze eenheid te bewerkstelligen. Men zal duidelijk moeten afspreken naar welke anamnestiche gegevens in het kader van het onderzoek door iedere deelnemer zal moeten worden gevraagd. Er zal - bij voorkeur schriftelijk - moeten worden vastgelegd, bij de aanwezigheid van welke symptomen men zal besluiten, dat de be-

schouwde aandoening aanwezig is. De wijze van omschrijven van deze symptomen is van het grootste belang, opdat zoveel mogelijk eenheid in diagnostiek ontstaat. Met enkele voorbeelden kan worden gedemonstreerd, hoe verschillende artsen bij een onderzoek van vergelijkbare groepen mensen tot sterk uiteenlopende antwoorden kunnen komen.

Cochrane en medewerkers verrichtten een onderzoek bij vier gelijkwaardige groepen van elk ongeveer 250 mijnwerkers. In het kader daarvan werd door vier verschillende artsen een korte anamnese opgenomen. Op de vraag: „Hoest u wel eens?”, werden in de vier groepen aanzienlijk verschillende percentages positieve antwoorden verkregen: deze liepen uiteen van 23 tot 40. Een nog groter verschil werd gezien tussen de percentages positieve antwoorden op de vraag, of de patiënt opgaf. Deze liepen van 13 tot 42. Het moet onwaarschijnlijk worden geacht, dat deze verschillen werden veroorzaakt door verschillen tussen de vier groepen onderzochten. Zij waren namelijk door loting uit 1.000 mijnwerkers gevormd. Het uiteenlopen van de percentages moet worden toegeschreven aan de verschillende manieren van vragen en het op verschillende manieren interpreteren van het gegeven antwoord, eventueel na het vragen van verdere inlichtingen aan de onderzochte, voordat ten slotte ja of neen werd genoteerd.

In het genoemde onderzoek bleek, dat er weinig verschil tussen de percentages positieve antwoorden bestond wanneer aan dezelfde mijnwerkers werd gevraagd, of zij wel eens pleuritis hadden gehad. De antwoorden waren voor de onderzoekers niet voor tweeërlei uitleg vatbaar. Het waren - in tegenstelling tot het voorafgaande - harde feiten. De percentages lopen hier slechts uiteen van 8 tot 13. Met harde feiten had men echter niet meer te maken bij de antwoorden op de vraag, of men wel eens last van de maag had. Hier liepen de mogelijkheden uiteen van enige last na een goede maaltijd tot een volledige maagzweer. Het oordeel van de onderzoeker moest weer worden verwerkt bij het noteren van het antwoord. Het percentage positieve antwoorden liep van 10 tot 27. *Fletcher* en zijn medewerkers hebben later laten zien, dat bij het standaardiseren van de vragen een groot stuk van de subjectiviteit bij het noteren van de antwoorden tot verdwijning kan worden gebracht. Wanneer slechts werd gevraagd: „Hoest u regelmatig, ja of neen?”, en dit ja of neen werd zonder meer genoteerd, bleken bij drie groepen van honderd postbeambten nauwelijks verschillen te worden gevonden (20 tot 25%).

Acht geroutineerde longartsen verrichtten een eenvoudig fysisch onderzoek bij twintig patiënten met emfyseem (*Fletcher*, 1952) en vergeleken hun resultaten. Terwijl één van de onderzoekers bij slechts vijf van de twintig onderzochten noteerde, dat hij een vatvormige thorax had, deed een ander dit zestien maal. Deze laatste onderzoeker nam nimmer hyperresonantie waar; een ander deed dit

zeventien keer, en meldde viermaal dat dit in bijzonder sterke mate het geval was. De aanwezigheid van kyfose werd tussen vier en twaalf keer genoteerd. Het gebruik van accessoire spieren bij de ademhaling nam één onderzoeker nimmer waar. Twee anderen deden dit in vijftien van de twintig gevallen. Uit dit soort gegevens kan de conclusie worden getrokken, dat een duidelijke omschrijving wanneer men een symptoom aanwezig kan achten, dringend noodzakelijk kan zijn, wanneer verschillende klinisch chemische laboratoria 35 hemoglobinebepalingen verricht volgens de Sicca-methode. De uitkomsten liepen uiteen van 13,3 tot 16,8%.

In het kader van een door het Rijksinstituut voor de Volksgezondheid georganiseerd onderzoek (Holtz; De Jonge, 1963) werden in elk van 35 verschillende klinisch chemische laboratoria 35 hemoglobinebepalingen verricht volgens de Sicca-methode. De uitkomsten liepen uiteen van 13,3 tot 16,8%.

Bij 119 spectrofotometrische bepalingen was het verschil niet kleiner. Bij beide groepen vielen de uitkomsten bovendien over het geheel genomen te laag uit. Dit soort waarnemingen maakt duidelijk, dat het de moeite waard is een onderzoek, waarin bijvoorbeeld hemoglobinebepalingen moeten plaatsvinden, zodanig op te zetten, dat de systematische verschillen tussen de door de verschillende onderzoekers met verschillende apparaten verkregen uitkomsten zoveel mogelijk worden geëlimineerd.

Dit zijn enkele voorbeelden van problemen rond de opzet van een onderzoek waarin men een statisticus kan betrekken. Hij kan helpen het probleem te signaleren en hij kan trachten mede te werken aan het oplossen daarvan. Ook met behulp van de statistiek zal men er nooit in slagen de beschreven en soortgelijke verschillen geheel tot verdwijnen te brengen. De statistiek kan echter wel een bijdrage leveren bij de poging de invloed van deze factoren op de uitkomst van het onderzoek zo gering mogelijk te maken. Tegen elke prijs moet men trachten te voorkomen, dat deze verschijnselen de oorzaak worden van foutieve conclusies.

Andere punten waarop onderzoekers en statisticus kunnen samenwerken zullen hier slechts kort worden aangeduid. De statisticus kan helpen bij het kiezen van zo objectief mogelijk te bepalen grootheden als maat voor het therapeutisch succes van de te vergelijken geneesmiddelen. Deze grootheden dienen zo relevant mogelijk te zijn voor het succes van de behandeling. Het lijkt haast overbodig op te merken, dat het succes van bijvoorbeeld een behandeling van coronairinfarcten in de eerste plaats wordt afgemeten aan de frequentie van nieuwe infarcten en aan de duur van de overleving ervan door de patiënt. Een verlaging van het ge-

halte aan onverzadigde vetzuren bij dergelijke patiënten kan een nuttige zaak zijn. Aan een dergelijke verlaging wordt echter het succes van een therapie van coronairinfarcten niet gemeten. (Yerushalmy sprak in dit verband van een „substitution game”). Helaas komt men nogal eens rapporten tegen, waarin de schijn wordt gewekt, dat dit wél het geval is.

Een statisticus kan hulp bieden bij het uitvoeren van een loting bij het bepalen welke patiënten behandeling A en welke behandeling B zullen krijgen. In tegenstelling tot wat velen menen, behoeft het aselekt verdelen van de behandelingen over de patiënten aan de behandelende artsen nagenoeg geen tijd te kosten. Bij de aselectering kan er voor worden gezorgd, dat iedere huisarts die deelneemt aan het onderzoek, in principe een ongeveer gelijk aantal A's en B's in zijn onderzoek krijgt; al is het slechts één van elk.

Een statisticus kan hulp bieden bij het vaststellen van het aantal patiënten dat in het onderzoek zal worden betrokken. Meestal zal hij meer patiënten willen laten onderzoeken dan haalbaar is. Er is dan een discussie mogelijk over de gevolgen, wanneer een compromis met de realiteit moet worden gesloten. In sommige gevallen kan de statisticus de onderzoeker er voor behoeden een onderzoek te beginnen, dat door het geringe aantal patiënten nooit tot een conclusie kan leiden, die enige betekenis kan hebben voor de praktijk.

Samenvattend kan worden gezegd, dat de statistiek op verschillende punten een bijdrage kan leveren bij het uitvoeren van een geneesmiddelenonderzoek. Dit geldt voor een onderzoek dat door huisartsen wordt uitgevoerd evenzeer als bij geneesmiddelenonderzoek door anderen. Deze bijdrage van de statistiek is niet van mysterieuze, bijna esoterische aard; ze kan wel degelijk betrekking hebben op zeer praktische, medisch direct belangrijke aspecten van het onderzoek. Het is wenselijk, dat de statisticus in de vroegste fase van het maken van de plannen voor een dergelijk onderzoek daarin wordt betrokken; niet in een eenmalig consult, maar als partner in een werkgroep.

- Cochrane, A. L., P. J. Chapman en P. D. Oldham (1951) *Lancet* I, 1007.
- Fletcher, C. M. (1952) *Proc. roy. Soc. Med.* 45, 577.
- Fletcher, C. M. Criteria for diagnosis and assessment in clinical trials. In: A. Bradford Hill, *Controlled Clinical Trials*. Blackwell, Oxford, 1960.
- Green, F. H. K. (1954) *Lancet* II, 1085.
- Holtz, A. H. (1960) *Verslagen en Mededelingen betreffende de Volksgezondheid*. I, 211.
- Jonge, H. de, *Inleiding tot de Medische Statistiek I*. Tweede druk, Nederlands Instituut voor Praeventieve Geneeskunde, 1963, blz. 41.
- Jonge, H. de (1964) *Statist. Neerl.* 18, 417.
- Yerushelmy, J. On inferring causality from observed associations. In: Ingelfinger and others. *Controversy in Internal Medicine*. W. B. Saunders, Philadelphia, 1966.