

Epiloog

J.A. KNOTTNERUS
A. VOLOVICS

De artikelenserie die de afgelopen twee jaar onder het motto 'Huisarts, wetenschap en statistiek' in dit tijdschrift is verschenen, wordt met deze epiloog beëindigd. 'Voltooid' zou geen passend woord zijn, want daarvoor is het gebied te groot en te dynamisch. De doelstellingen zoals die in het openingsartikel zijn genoemd, lijken echter in voldoende mate bereikt, en de geïnteresseerde lezer kan voor nadere studie de aangehaalde literatuur raadplegen. In deze laatste bijdrage wordt nog eens expliciet ingegaan op de relatie tussen de statistische benadering en het medisch handelen.

Knottnerus JA, Volovics A. Epiloog [Huisarts, wetenschap en statistiek]. Huisarts Wet 1989; 32(10): 389-91.

Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht.

Prof. dr. J.A. Knottnerus, hoogleraar huisartsgeneeskunde, vakgroep Huisartsgeneeskunde; A. Volovics, vakgroep Medische Informatica en Statistiek.

Correspondentie: Prof. dr. J.A. Knottnerus.

Inleiding

Twee jaar geleden spraken wij de hoop uit dat de serie 'Huisarts, wetenschap en statistiek' ertoe zou bijdragen dat de geïnteresseerde huisarts:

- 1 inzicht zou krijgen in de uitgangspunten van veel gehanteerde methoden en technieken in de medische statistiek, voor zover van belang voor het begrijpen van de meeste onderzoeksverslagen en het voorbereiden van een onderzoeksopzet;
- 2 aan de statistische benadering een zekere steun zou hebben bij het handelen in concrete situaties;
- 3 naast de mogelijkheden ook de beperkingen van de statistiek zou leren kennen;
- 4 op zinvolle wijze gerichte consultatie en ondersteuning zou kunnen vragen aan een professionele statisticus.¹

Een ieder kan voor zich nagaan in hoeverre de serie – waarvan de afleveringen nog eens zijn opgesomd op pagina 390 – aan de verwachtingen heeft voldaan. Zelf denken we dat de doelen 1, 3 en 4 goede kansen hebben gekregen. Ten aanzien van punt 2, dat de overgang tussen de statistische beschouwing en de individuele medische verantwoordelijkheid betreft, geldt dat op het eerste gezicht misschien wat minder. Toch hopen wij dat ook dit doel in een aantal opzichten is bereikt: een kritischer blik bij het beoordelen van wetenschappelijk 'bewijsmateriaal' dat consequenties kan hebben voor het medisch handelen; een beter inzicht in de onvermijdelijkheid van variabiliteit en variatie van metingen en medische beoordelingen, en een toegenomen vermogen om hiermee om te gaan; een beter oordeel over de zeggingskracht van diagnostische gegevens en de zeeffunctie van de huisarts. En tenslotte: de wens om meer bewust en systematisch bij te dragen aan de opbouw van de gezamenlijke *body of knowledge* op basis van de dagelijkse zorgervaringen, met als doel deze kennis (over groepen) weer terug te vertalen naar die (individuele) zorg.

Wetenschap en zorg

Er is niettemin alle reden voor een blijven-de wederzijdse bescheidenheid in de ver-

houding tussen wetenschap en individuele zorg. Een groot deel van het complex van relaties tussen beide verloopt impliciet, en in zekere zin intuïtief. De verantwoordelijkheden zijn verbonden, maar ook verschillend, en alleen al het feit dat eenduidige, wetenschappelijk 'harde' richtlijnen voor de praktijk maar in zeer beperkte mate beschikbaar zijn, onderstreept dat 'science' en 'art' tot elkaar zijn veroordeeld. De vertaling van algemene wetmatigheden met betrekking tot groepen naar veel meer gevarieerde individuele situaties (waarbij sprake kan zijn van bijzondere aspecten en preferenties) is het domein, de kunst, van de dokter – en wellicht vooral van de huisdokter.

De dokter zal zeker geïnteresseerd zijn in het al dan niet statistisch significant zijn van mogelijk richtinggevende, klinisch relevante onderzoeksbevindingen. Tegelijkertijd dient hij, mede op grond van zijn medisch inzicht en gezond verstand, statistische conclusies met de nodige relativering en soms scepsis te bezien. Een dergelijke relativering kan te maken hebben met een kritische beschouwing van de aan de statistische analyse ten grondslag liggende aannames (zoals een gepostuleerde normale verdeling, het gekozen significantieniveau, onafhankelijkheid van verschillende invloeden, of het aselekt zijn van steekproeven), de aantallen (veel te klein of juist veel te groot²) of met een beoordeling van de studie-opzet (in verband met eventuele vertekening). Maar evenzeer of zelfs in de eerste plaats gaat het om de vraag in hoeverre het onderzoeksresultaat plausibel of zinvol is in het licht van wat reeds bekend was, of met betrekking tot de dagelijkse praktijk.

Wat betreft de verhouding tussen wetenschap en zorg is in deze serie ook enkele malen benadrukt dat zij van elkaars aanpak kunnen leren. Zo is de wetenschapper die de effectiviteit van diagnostische en therapeutische interventies onderzoekt, erbij gebaat in een vroeg stadium feedback te krijgen over de haalbaarheid van de te onderzoeken interventies in de medische praktijk. En de praktizerende arts kan de kwaliteit van zijn handelen verbeteren door bijvoorbeeld meer aandacht te besteden

aan de objectiviteit van zijn waarnemingen, onder gebruikmaking van de methoden die onderzoekers daarvoor hebben ontwikkeld. Een fraaie, op de individuele zorg gerichte toepassing hiervan is de zogenaamde 'n=1 trial', die enkele jaren geleden werd geïntroduceerd.³⁻⁵

Het was niet de bedoeling in deze serie volledig te zijn. Onder andere is afgezien van het geven van een overzicht van en inzicht in allerlei statistische technieken en formules. Het ging om de centrale concepten, en die zijn voor veel technieken dezelfde. Toch was het af en toe noodzakelijk om enkele formules te presenteren, teneinde de concepten concreet voelbaar te maken en ze op exemplarische wijze 'in jargon' te steken. Aldus kon ook beter duidelijk gemaakt worden dat de statistische benadering en de medische, inhoudskundige benadering complementair zijn: zo bleek in de aflevering over statistisch toetsen dat een reële schatting van een benodigde steekproefomvang met behulp van statistische formules alleen te maken is, als de medicus aangeeft welke verschillen klinisch relevant – dus het ontdekken waard – zijn, en hoe vaak bepaalde fenomenen te verwachten zijn, dan wel hoe ze verdeeld zijn.²

Voor een zinvolle toepassing van de statistiek moet de dokter dus met zijn billen bloot: de statistiek dwingt hem tot nadenken over de medisch inhoudelijke essenties. Daartoe behoren ook vragen als: wat zijn bij een bepaalde vraagstelling nu wel en niet *confounding factors* (verstoring van een juiste waarneming) en modifierende variabelen (een groter differentiatie van de waar te nemen werkelijkheid), wat is de relevante onderzoekspopulatie? En: wat is medisch-inhoudelijk het meest relevante criterium of 'eindpunt' bij de beantwoording van de vraagstelling en hoe is dit zo nauwkeurig mogelijk te meten? Deze vragen liggen uiteindelijk veeleer op medisch dat op statistisch-methodologisch gebied.

De aannemelijkheidsschaal

Wanneer is in de medische wetenschap nu eigenlijk iets 'bewezen' en wat is daarbij de rol van de statistiek? Aan deze kernvraag zijn we in deze serie hoogstens indirect

Huisarts, wetenschap en statistiek

- 87-349 Medisch-statistische besprekingen voor de huisarts
- 87-351 Het beschrijven en samenvatten van gegevens van groepen
- 87-391 Associatiematen voor etiologische en prognostische verbanden
- 88-018 Correlatie en regressie
- 88-100 Statistisch toetsen
- 88-135 Over validiteit, vergelijkbaarheid en versturende variabelen
- 88-207 Het analyseren van de invloed van meer dan één variabele
- 88-301 Gecontroleerd interventie-onderzoek op basis van randomisatie
- 89-056 Overeenstemming tussen beoordelaars
- 89-338 Het onderscheidend vermogen van diagnostische tests
- 89-383 Het beschrijven van latente dimensies en complexe relaties
- 89-389 Epiloog

De eerste twee cijfers verwijzen naar het jaar van verschijnen, de volgende drie cijfers naar het paginanummer.

toegekomen: Een bespreking hiervan zou echter een apart hoofdstuk vergen, met een flink stuk wetenschaps-filosofische inbreng en onze serie was hiervoor niet bedoeld. Toch willen we hierover nog enkele opmerkingen maken.

Bij klinisch wetenschappelijk onderzoek is er zelden sprake van een onomstotelijk bewijs in de mathematische 'alles of niets'-betekenis. Indien dit wel het geval is, gaat het vaak om een vanuit de klinische ervaring zo evidente zaak, dat door velen niet eens onderzoek verricht zou zijn. Het gaat in de regel meer om 'bewijsmateriaal' of 'aanwijzingen' en de interpretatie daarvan, dan om 'bewijs'. De aannemelijkheid van bepaalde hypothesen of opvattingen wordt dan, in het licht van de onderzoeksresultaten, groter of kleiner, of blijft gelijk. Men wordt bij wijze van spreken in een bepaalde mate verschoven over een aannemelijkheidsschaal die loopt van 0 procent tot 100 procent.

Voor het individu is dit nogal een subjectief proces; in de kring van medische beroepsbeoefenaren zal er gewerkt worden in

de richting van consensus, ofwel een gemeenschappelijke positie op de aannemelijkheidsschaal. Daarbij is er in principe altijd ruimte voor dissidente interpretaties, al maken die soms een verdachte indruk. Zo zijn er nog steeds artsen die ontkennen dat er een causaal verband is tussen roken en longkanker. In hun ogen zou de bij herhaling gevonden sterke relatie tussen beide het gevolg kunnen zijn van *confounding*: er is bijvoorbeeld geopperd dat een bepaalde, tot nu toe nog niet geïdentificeerde, constitutie zou predisponeren tot zowel het krijgen van longkanker als het gaan roken. Het is zelfs nog niet zo lang geleden dat de tabaksindustrie een folder verspreidde waarin, uitgaande van een gevonden verband tussen stress en kanker, gesuggereerd werd dat voor rokers het stoppen met roken stress-verhogend en daardoor carcinogeen zou zijn.

Gelukkig zullen de meesten onder ons onderscheid kunnen maken tussen gereede wetenschappelijke twijfel en dit soort (ook methodologische) spitsvondigheden die door vooringenomenheid zijn geïnspireerd.

Men kan objectieve gegevens subjectief verwerken, in het kader van een geheel van voorkennis, maar ook preoccupaties en vooroordelen. Er is, net als in de diagnostiek, een 'a priori' situatie van objectieve gegevens en persoonlijke kennis en opvattingen. De a priori situatie heeft een sterke invloed op de betekenis en de interpretatie van nieuwe bevindingen. Sterker nog, zij kan ook bepalend zijn voor de vraag of men überhaupt een onderzoek gaat doen. Men kan al zo zeker van zijn zaak zijn, in positieve of negatieve zin (in de buurt van 100 procent of 0 procent op de aannemelijkheidsschaal, bijvoorbeeld omdat er al veel onderzoek of waarnemingen in een bepaalde richting wijzen), dat men een (extra) onderzoek overbodig vindt. Het resultaat daarvan, wat het ook zal zijn, zal voor de meeste vakgenoten de inzichten niet veranderen; bij een onverwacht resultaat zou men eerder twijfelen aan de kwaliteit van het onderzoek dan aan de eigen visie.

Iets anders is dat een onderzoek – weer net als in de diagnostiek – een fout-positief resultaat kan hebben. Er wordt dan bijvoorbeeld na statistische toetsing ten on-

rechte een significant verschil (met een p -waarde $< .05$) gevonden. Dit kan per definitie gebeuren in 1 op de 20 gevallen waarin er in werkelijkheid geen verband is. De kans hierop neemt toe als men een groot aantal statistische toetsingen verricht, ook als er geen sprake is van a priori enigszins aannemelijke hypothesen. Daartegenover staat dat men bedacht moet zijn op een fout-negatief resultaat indien een tevoren zeer aannemelijke hypothese niet bevestigd wordt. Men moet daarom onder meer kritisch kijken naar de 'statistische power' van het onderzoek.²

Een onderzoek lijkt al met al, evenals een diagnostische test, het meeste rendement op te leveren als bron van nieuwe informatie, als de argumenten in een bepaalde richting nog niet erg sterk zijn.

Hypothesenvorming en vernieuwing

Als men zich bij de vorming en toetsing van hypothesen alleen baseert op hetgeen verwacht wordt op grond van reeds bekende gegevens en inzichten, staat men niet open voor volstrekt nieuwe bevindingen. Waar men niet kijkt, kan men niet zien. Corrigeren van ten onrechte aanvaarde inzichten is dan niet mogelijk, en men raakt eerder verstrikt in cirkelprocessen dan dat men de wetenschappelijke vooruitgang dient.

De druk in conserverende richting is nogal sterk, omdat veel onderzoekers die hun naam gevestigd hebben, niet te veel

risico willen nemen, hun eigen spoor niet gauw verlaten, en liever een voorspelbare positieve dan een onverhoopt negatieve bevinding doen. Bovendien is een onderzoek aantrekkelijker voor subsidie-verlenende instanties, als men de kans op een positief resultaat hoog inschat, en als de onderzoeker een goede staat van dienst heeft ten aanzien van het bevestigen van hypothesen.

Het is daarom nodig dat bij de beoordeling van een onderzoek ook gelet wordt op het mogelijk innoverende karakter ervan, en om die reden is in de criterialijst van de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) de kans op een doorbraak een belangrijk aandachtspunt. Daarnaast zal men onverwachte bevindingen als uitgangspunt nemen voor nader onderzoek, eventueel ook via exploratie van voor een ander doel verzamelde gegevens. Dat voorkomt enerzijds dat mogelijke toevalsbevindingen direct tot belangrijk resultaat worden verheven, terwijl anderzijds het kind niet met het badwater wordt weggegooid.

De bijdrage van de medicus practicus aan het ontwikkelen van originele en mogelijk innoverende onderzoekshypothesen is onmisbaar. Om dat te onderstrepen willen we eindigen met de slotzin van ons openingsartikel:

Benadrukt zij 'dat huisartsen die onderzoek willen initiëren, niet eerst statisticus hoeven te worden. Hun bijdrage ligt in de

eerste plaats in het vormen van onderzoekshypothesen op grond van al dan niet gemeenschappelijke ervaringen. Was het niet de eerste indruk van enkele artsen rond de eeuwwisseling over de schadelijkheid van roken, die van primair belang is geweest? Daardoor ontstonden de hypothesen die door epidemiologen en statistici onderzocht konden worden.'

Dit vormen van relevante hypothesen kan ondersteund worden door huisartsgeneeskundige conferenties, goede casuïstiekbesprekingen en systematisch overleg met (huisarts-)onderzoekers. In deze zin is er ruimte voor meer professionalisering van dit deel van het wetenschappelijk onderzoek, onder leiding van de huisarts.

Literatuur

- ¹ Knottnerus JA, Volovics A. Medisch-statistische besprekingen voor de huisarts. Huisarts Wet 1987; 30: 349-50.
- ² Knottnerus JA, Volovics A. Statistisch toetsen. Huisarts Wet 1988; 31: 100-5.
- ³ Guyatt G, Sackett D, Taylor W, et al. Determining optimal therapy - randomized trials in individual patients. N Engl J Med 1986; 314: 889-92.
- ⁴ Mcleod RS, Taylor DW, Cohen Z, Cullen JB. Single patient randomized. Clinical trial. Lancet 1986; i: 726-8.
- ⁵ Ingelfinger JA, Mosteller F. Biostatistics in clinical medicine. 2nd ed. New York: Macmillan, 1987. ■