

# ONTWIKKELINGEN IN DE BIostatistiek

Dr. Theo Stijnen

MEDISCHE BIBLIOTHEEK EUR



019600 0025 0625

The logo of Erasmus University, featuring the word "Erasmus" in a stylized, cursive script font.

ERASMUS UNIVERSITEIT ROTTERDAM

## Ontwikkelingen in de biostatistiek

*Rede in verkorte vorm uitgesproken bij  
de aanvaarding van het ambt van  
gewoon hoogleraar in de biostatistiek  
aan de Erasmus Universiteit Rotterdam  
op vrijdag 1 september 2000 door*

Dr. Theo Stijnen

*Mijnheer de Rector Magnificus, zeer gewaardeerde toehoorders,*

### **Inleiding**

De biostatistiek houdt zich bezig met statistische methoden die toegepast worden in het wijde veld van het medische, biologische en landbouwkundige onderzoek. In mijn rede zal ik mij beperken tot het gedeelte van de biostatistiek dat zich bezig houdt met het onderzoek in de geneeskunde. Dit wordt ook wel het vakgebied van de medische statistiek genoemd. In het vervolg zal ik, waar ik biostatistiek zeg, meestal meer specifiek de medische statistiek bedoelen.

Bij het horen van de term medische statistiek zal men snel geneigd zijn te denken aan tellen en inventariseren, het produceren van statistieken zoals het Centraal Bureau voor de Statistiek dat doet. Bijvoorbeeld hoeveel mensen sterven er per jaar in Nederland aan kanker, hoeveel ziekenhuisbedden zijn er, wat is de gemiddelde bedbezetting, etc. Dat is echter niet waar de medische statistiek zich mee bezig houdt. Medische statistiek is een methodologische discipline. Zij houdt zich bezig met methoden van medisch onderzoek, vooral het patiëntgebonden en het epidemiologische onderzoek. Dit betreft het totale traject vanaf het ontwerp van een onderzoek tot en met de analyse van de gegevens die uit het onderzoek voortkomen, alsmede het formuleren van de conclusies. Als zodanig is zij de oudere zuster (of misschien wel de moeder) van nieuwere methodologische disciplines in het medische onderzoeksveld, zoals de klinische epidemiologie en de medische besliskunde.

Aan de ene kant is de medische statistiek een medische discipline, aan de andere kant maakt zij deel uit van het brede vakgebied van de statistiek, dat tientallen internationale wetenschappelijke tijdschriften kent. Een groot aantal daarvan is gericht op de biostatistiek, terwijl de algemene statistische tijdschriften voor een substantieel deel aan biostatistiek gewijd zijn. De onderzoeksmethodologie, die grotendeels statistisch van aard is, heeft in de 20e eeuw een stormachtige ontwikkeling doorgemaakt. Binnen veel wetenschappen zijn aparte disciplines ontstaan gericht op de methoden van onderzoek in het betreffende vakgebied. Men kan denken aan de econometrie binnen de economie, de methodenleer ("methoden en technieken") in de sociale wetenschappen, de biometrie in het landbouwkundig onderzoek, etc. De medische statistiek vervult deze rol binnen de geneeskunde, samen met de klinische epidemiologie en besliskunde, die zelf op hun beurt weer zwaar leunen op statistische methoden.

In deze rede wil ik u eerst een kort globaal overzicht geven van de ontwikkeling van de medische statistiek tot nu toe. Vervolgens zal ik wat uitgebreider stilstaan bij recente ontwikkelingen in de biostatistiek en deze met enige voorbeelden illustreren. Tot slot zal ik kort stilstaan

bij de rol van de medisch statisticus in de praktijk van het medisch onderzoek en het medisch onderwijs.

### **De opkomst van de medische statistiek in de 20e eeuw**

Een zeer belangrijke verandering in het klinisch wetenschappelijk onderzoek in de 20e eeuw was de opkomst van het statistische denken. Was het klinisch onderzoek voorheen vooral gericht op waarnemingen aan het ziekbed van individuele patiënten die bijvoorbeeld konden leiden tot ideeën over pathofysiologie en hypothesen over causaliteit, nu brak het inzicht door dat men ook kan leren over oorzaken van ziekten en effecten van behandelingen door groepen patiënten en controle personen statistisch met elkaar te vergelijken [1]. Dit gaat terug op de 'methode numerique' geïntroduceerd in het Parijs van rond 1830. Na een eeuw lang in de vergetelheid gebleven te zijn, is het numerieke, kwantitatieve of statistische denken in het midden van de 20e eeuw definitief doorgebroken in het klinisch onderzoek. De belangrijkste mijlpaal hierbij was de opkomst van de zog. 'gerandomiseerde gecontroleerde klinische trial'. (Dit is een onderzoeksopzet waarin twee of meer behandelingen op werkzaamheid met elkaar worden vergeleken, waarbij geloot wordt welke behandeling de patiënt krijgt. Idealiter is dit onderzoek dubbelblind, d.w.z. zowel de patiënt als de behandelend arts weten niet welke behandeling de patiënt krijgt.) De naam van de Engelse medisch statisticus Sir Austin Bradford Hill is hiermee onverbreekbaar verbonden. Hij was de motor achter de 'eerste' clinical trial die geheel volgens de nu algemeen aanvaarde regels van kunst werd verricht [2]. Hill schreef in dertiger jaren een serie artikelen in de Lancet over medische statistiek, later gebundeld in het bekende boek [3] dat tientallen jaren lang voor heel veel artsen de eerste kennismaking met de medische statistiek is geweest. In Nederland waren in de zestiger en zeventiger jaren van de 20e eeuw de eerste hoogleraren in de medische statistiek, te weten De Jonge, Rümke en van Strik, de pioniers. Eind jaren vijftig schreef de Jonge zijn tweedelige boek Medische Statistiek [4,5], dat jarenlang in Nederland heeft gefungeerd als de bijbel op dit gebied.

### **Invloed van de statistiek op het medisch-wetenschappelijk onderzoek**

De opkomst van het statistische denken en de impact daarvan op het medisch onderzoek blijkt uit de enorme toename sinds de jaren vijftig van de 20e eeuw in het gebruik van statistische methoden in de medisch-wetenschappelijk literatuur. Het percentage medisch wetenschappelijke artikelen waarin gebruik wordt gemaakt van formele statistische methoden is dramatisch gestegen. Bijvoorbeeld, in de New England Journal of Medicine, een van de meest prestigieuze medisch wetenschappelijke tijdschriften, is dit percentage ongeveer 90%. Het percentage artikelen in dit tijdschrift waaraan een biostatisticus heeft meegewerkt als co-

auteur, is thans meer dan 25%. Een aantal statistische artikelen behoort tot de meest geciteerde in de medische literatuur. Bijvoorbeeld, het artikel van Kaplan en Meier [6] over overlevingscurven en het artikel van Cox [7] over het proportionele hazards model waren gedurende de periode 1985 tot 1999 de nummers een en twee op de lijst van meest geciteerde wetenschappelijke artikelen in de Science Citation Index.

De invloed van de medische statistiek blijkt ook uit het feit dat inmiddels de meeste medische tijdschriften gedetailleerde richtlijnen gemaakt hebben voor het gebruik en de rapportage van statistische methoden. Veel medische tijdschriften laten de aan hun aangeboden manuscripten beoordelen door biostatistici. Opmerkelijk is ook de duidelijk groeiende complexiteit van de statistische methoden die gebruikt worden.

Een veld van onderzoek waar de invloed van het statistische denken bij uitstek aanwezig is, is het onderzoek dat vanuit farmaceutische industrieën gedaan wordt ten behoeve van de registratie van nieuwe geneesmiddelen. Dit soort onderzoek is zeer sterk gereguleerd door de overheid, waarbij de richtlijnen, zoals de Guidelines for Good Clinical Practice [8] en Good Statistical Practice [9] geheel doordrenkt zijn van het streven statistisch zuivere, niet vertekende schattingen van het effect van de geneesmiddelen te krijgen. De participatie van medisch statistici bij dit soort onderzoek is zo goed als wettelijk vereist. In het onderzoek van farmaceutische industrieën spelen biostatistici dan ook een vooraanstaande en zeer gewaardeerde rol.

#### *Ontwikkelingen in de biostatistische methoden*

Parallel aan het toenemend gebruik van statistiek in de medische literatuur hebben de laatste decennia van de vorige eeuw zich gekenmerkt door een stormachtige ontwikkeling van nieuwe biostatistische methoden. Het aantal biostatistische tijdschriften is in de laatste twee decennia krachtig uitgebreid. Een aantal nieuwe tijdschriften is opgericht, bijvoorbeeld *Statistics in Medicine*, dat in 20 jaar groeide van een paar honderd pagina's tot meer dan 3000 per jaar nu. Recentelijk zijn er weer nieuwe bijgekomen, zoals *Statistical Methods in Medical Research*, *JABES* en *Biostatistics*.

Omstreeks 1960 konden alle voor de praktijk relevante medisch statistische methoden worden behandeld in het bekende boek van de Jonge [4,5], terwijl dan nog een groot deel van de inhoud werd besteed aan hoe de methoden m.b.v. pen, papier en een tafelrekenmachine toegepast konden worden. Het boek was gewijd aan de inmiddels als klassiek bekend staande methoden als twee steekproeven toetsen, betrouwbaarheidsintervallen, niet-parametrische toetsen, enkelvoudige regressie, correlatie, variantie-analyse en overlevingstafels. De statistiek bleef voornamelijk beperkt tot het bestuderen van de relatie tussen twee variabelen tegelijkertijd, meestal een uitkomstvariabele en een determinant, bijvoorbeeld een be-

handeling en een variabele waarmee het effect van de behandeling gemeten wordt. Sindsdien heeft de ontwikkeling van nieuwe methoden een enorme vlucht genomen. De jaren zeventig en tachtig werden gekenmerkt door het ontwikkelen van meer geavanceerde methoden, in het bijzonder de multivariabele regressie modellen: methoden waarmee het verband bestudeerd kan worden tussen één uitkomstvariabele en een willekeurig aantal determinanten. De oratie van collega van Houwelingen in 1987 [10] en de afscheidsrede van mijn voorganger van Strik in 1997 [11] waren voor een groot deel gewijd aan de zegeningen die de regressiemodellen met zich meegebracht hebben voor de analyse van medische onderzoeksgegevens. Vele regressiemodellen, voor allerlei verschillende situaties, werden ontwikkeld en op uitgebreide schaal toegepast. De meest bekende zijn het multiple lineaire regressiemodel voor een continue uitkomstvariabele [12], het logistische regressiemodel [13] voor een dichotome uitkomst en het proportionele hazards model [14] voor overlevingsduur uitkomst. Het beschikbaar komen van deze methodologie was een enorme sprong voorwaarts. Veel klinisch onderzoek heeft sterk geprofiteerd van de komst van deze nieuwe mogelijkheden. De opkomst van de epidemiologie in de laatste 30 jaar is mede te danken aan het rigoureuze toepassen van deze methoden.

#### *De rol van de computer*

De ontwikkeling van de statistische methodologie en de grote toename van het gebruik daarvan in medisch onderzoek waren niet denkbaar geweest zonder de opkomst van de computer en de daarmee gepaard gaande ontwikkeling van statistische programmatuur. Sinds de laatste 10 jaar heeft elke medische onderzoeker de beschikking over uitgebreide en zeer gebruikersvriendelijke statistische pakketten in de personal computer op het bureau. Het praktisch uitvoeren van ingewikkelde statistische methoden is daarom in de praktijk geen enkel probleem meer; daarvoor heeft men geen biostatisticus meer nodig. Het is echter een misverstand te denken dat daarmee de biostatisticus in de praktijk overbodig is geworden, in tegendeel. Het grotere aantal beschikbare technieken en de toenemende complexiteit ervan maakt het moeilijker om de meest geschikte te vinden. Wanneer men een statistisch programma pakket start op de PC, dan wordt men geconfronteerd met een woud aan mogelijkheden. Van SPSS en SAS, de meest gebruikte en bekende statistische programmapakketten, bevatten de "manuals", waarin de beschikbare technieken worden beschreven, duizenden bladzijden. De biostatisticus zal veelal onmisbaar blijven om de onderzoeker te begeleiden bij het vinden van de optimale methode en vervolgens te helpen bij de interpretatie van de uitkomsten.

### **Moderne ontwikkelingen in de biostatistiek**

Uit het voorgaande is duidelijk dat de medische statistiek in de afgelopen halve eeuw een krachtige ontwikkeling heeft doorgemaakt en dat de ontwikkelde methoden op uitgebreide schaal worden toegepast door medische onderzoekers. Men zou zich kunnen afvragen of de statistiek niet zo langzamerhand af is nu er zo'n groot aantal methoden beschikbaar is onder handbereik van iedere medische onderzoeker. Dat is echter nog bij lange na niet het geval. In het medisch onderzoek worden steeds meer gegevens verzameld, met een steeds ingewikkelder structuur, hetgeen vraagt om verdergaande ontwikkeling van methoden. De biostatistiek blijft een bruisend vakgebied met volop nieuwe ontwikkelingen. Enkele belangrijke recente en huidige ontwikkelingen wil ik de revue laten passeren.

### **Modellen voor afhankelijke waarnemingen**

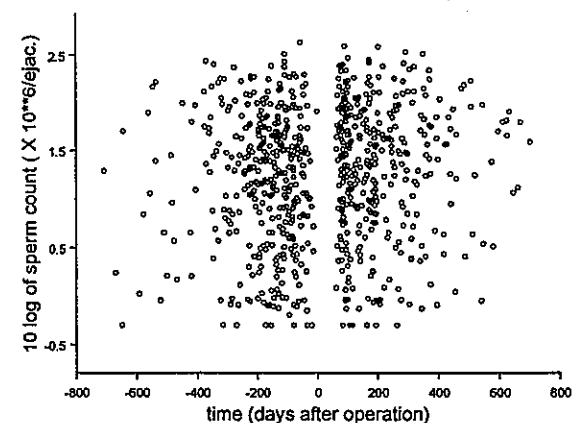
Ik noemde al dat er een sterke tendens waarneembaar is naar steeds ingewikkelder gegevensstructuren. Het belangrijkste kenmerk hiervan is het optreden van allerlei afhankelijkheden tussen variabelen, afhankelijkheden die niet genegeerd kunnen worden in de statistische analyses, en die dus op een adequate manier in de statistische modellering verdisconteerd moeten worden. De modellering van afhankelijke data is één van de belangrijkste thema's van het hedendaagse biostatistische onderzoek. Als voorbeelden wil ik nu een aantal verschillende situaties de revue laten passeren waarin afhankelijkheid een centrale rol speelt.

### **Herhaalde waarnemingen.**

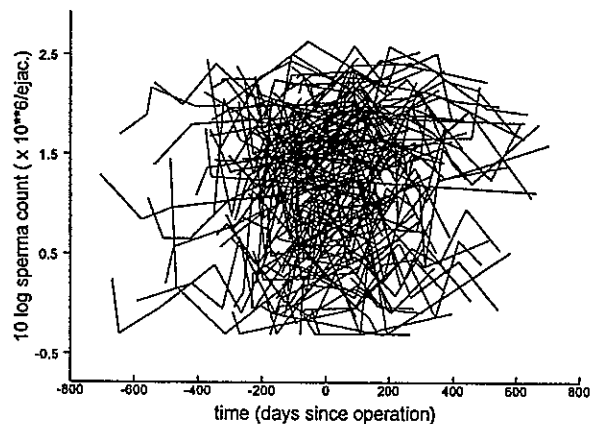
In medisch onderzoek is men vaak geïnteresseerd in de relatie tussen een aantal factoren of determinanten enerzijds en een bepaalde uitkomst anderzijds. In de praktijk van het huidige medische onderzoek wordt de uitkomstvariabele steeds vaker meer dan één keer gemeten, met het doel om het verloop in de tijd van de uitkomstvariabele te bestuderen. Stel bijvoorbeeld dat men in een onderzoek naar oorzaken van hoge bloeddruk de bloeddruk elk half jaar gemeten heeft gedurende een zekere periode. Wanneer men de invloed van determinanten op het verloop van de bloeddruk wil bestuderen, zou men dit naïef kunnen aanpakken met een regressie methode waarbij elke bloeddrukmeting als een afzonderlijke uitkomst wordt beschouwd. Dit is echter niet correct, omdat standaard regressiemodellen uitgaan van onafhankelijke uitkomsten. Aan deze veronderstelling is nu niet voldaan, omdat bloeddrukwaarden van een zelfde individu meer op elkaar zullen lijken dan die van verschillende individuen. Er is dus correlatie, ofwel afhankelijkheid, tussen de uitkomsten van een zelfde persoon. Het statistische model moet hiermee rekening houden. Geen rekening houden met de-

ze afhankelijkheid kan in het slechtste geval leiden tot onjuiste conclusies, en in het gunstigste geval tot een sub-optimale analyse, waarbij minder uit de gegevens wordt gehaald dan er eigenlijk in zit.

Een voorbeeld van een situatie met herhaalde waarnemingen ontleen ik aan een onderzoek van Pierik et al. [15] dat enkele jaren geleden is uitgevoerd bij de afdeling Andrologie van het Academisch Ziekenhuis Rotterdam. In figuur 1 ziet u de gegevens van 139 verminderd vruchtbare mannen weergegeven die geopereerd werden aan een varicocele (dat is een abnormale uitzetting van de aders van de zaadstreng in het scrotum). Op een aantal tijdstippen werd een zaadmonster afgenomen en het aantal zaadcellen bepaald. Op de verticale as in de figuur is (op logaritmische schaal) het aantal spermacellen weergegeven. Op de horizontale as staat het tijdstip van de meting in dagen voor of na de operatie. Tijdstip 0 correspondeert dus met de datum van de operatie. In figuur 2 zijn dezelfde gegevens weergegeven, maar nu zijn de opeenvolgende metingen van een zelfde patiënt met elkaar verbonden. Er is hier sprake van zeer ongebalanceerde data: het aantal metingen varieert zeer sterk tussen de patiënten (van 1 tot 8), en de tijdstippen waarop de metingen verricht zijn vertonen ook een grote variatie.



Figuur 1. Aantal spermacellen x 106 per ejaculaat, op 10log schaal, uitgezet tegen de tijd van de monsterafname.

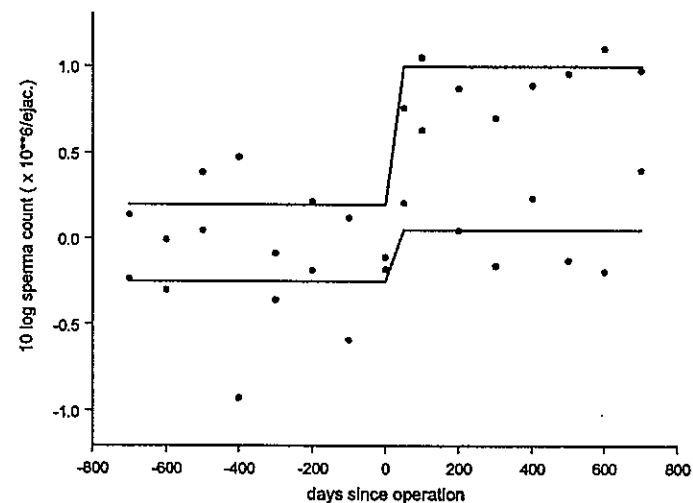


Figuur 2. Gegevens als in figuur 1, waarbij de opeenvolgende metingen van een zelfde patiënt met een lijnstuk verbonden zijn.

De centrale vraag in het onderzoek was of de zaadproductie na de operatie gemiddeld hoger is dan voor de operatie, en vervolgens hoe dit verschil afhangt van allerlei factoren zoals leeftijd, diagnose, het niveau van zaadproductie voor de operatie en de ernst van de varicocele. Dit is een gegevensstructuur die 10 jaar geleden lastig te analyseren geweest zou zijn. Tegenwoordig kunnen deze data echter adequaat geanalyseerd worden met behulp van het zog. algemene lineaire gemengde model (zie bijv. [16]). Dit model is een generalisatie van het gewone lineaire regressie model, waaraan een extra component is toegevoegd die de correlatie tussen de waarnemingen van een zelfde individu modelleert. Het model heet gemengd omdat het twee componenten bevat: een die de gemiddelde uitkomst beschrijft en een die de correlatie structuur tussen de uitkomsten beschrijft. Het is een zeer flexibel en algemeen model, met talloze toepassingen. Voor deze methode is het niet langer vereist om gebalanceerde data te hebben, zoals in de klassieke variantie-analyse. Patiënten mogen op verschillende tijdstippen gemeten worden, het aantal metingen per patiënt mag variëren en missende waarnemingen vormen onder zekere voorwaarden ook geen probleem. Recentelijk is in enkele veel gebruikte statistische pakketten, onder meer in SAS en S-Plus een module voor het algemene lineaire gemengde model ingebouwd, zodat tegenwoordig de toepassing hiervan op grote schaal in de praktijk mogelijk is.

De gegevens in het voorbeeld werden geanalyseerd gebruik makend van een random coëfficiënten regressie model. Dit is een bijzonder geval van het genoemde algemene lineaire ge-

mengde model. Het idee is dat de gemiddelde zaadproductie van een individuele patiënt wordt verondersteld constant te zijn zowel voor als na de operatie, met een mogelijke sprong op het moment van de operatie. De metingen van een patiënt variëren random om het eigen niveau heen. In figuur 3 ziet u geïllustreerd hoe dit er voor individuele patiënten uit zou kunnen zien.

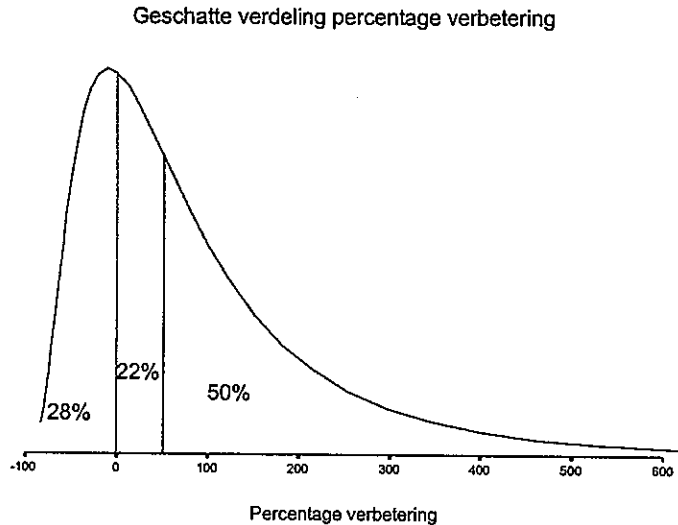


Figuur 3. Model illustratie.

Bijvoorbeeld patiënt 1 heeft een relatief hoog niveau voor de operatie en maakt een grote sprong na de operatie. Deze patiënt lijkt veel voordeel van de operatie te hebben. Patiënt 2 heeft een relatief lage productie voor de operatie, en maakt een veel kleinere sprong. Deze patiënt vertoont een veel kleiner effect.

De belangrijkste resultaten van de analyse waren als volgt. De schatting van het gemiddelde verschil (op de <sup>10</sup>log schaal) tussen na en voor de operatie was 0.18, met een 95% betrouwbaarheidsinterval (0.11, 0.25). Dit komt overeen met een mediane verbetering van 51%, met een 95% betrouwbaarheidsinterval lopend van 29% tot 78%. Er blijkt verder een grote variabiliteit te zijn in de mate van verbetering tussen de patiënten. De standaarddeviatie tussen patiënten in de grootte van het verschil op de log schaal is naar schatting 0.31 (95% be-

trouwbaarheidsinterval 0.06 tot 0.18). In figuur 4 is geïllustreerd wat dit betekent op de oorspronkelijke schaal.



Figuur 4. Geschatte verdeling van het percentage verbetering in het gemiddeld niveau van het aantal spermacellen per ejaculaat na de operatie ten opzicht van het niveau voor de operatie.

Naar schatting verbetert de zaadproductie van helft van de patiënten met 51% of meer. Ongeveer 22% van de patiënten gaat er tussen de 0 en 51% op vooruit. Naar schatting 28% vertoont een verslechtering.

De hier getoonde analyse is een voorbeeld van de eenvoudigste analyse die we zouden kunnen doen. In verdergaande analyses zouden we gemakkelijk bijvoorbeeld de zaadproductie van de tijd kunnen laten afhangen, zowel voor als na de operatie, en kunnen we invloed van allerlei factoren, zoals leeftijd en de ernst van de varicocele erin betrekken.

Tot zover dit voorbeeld betreffende de analyse van herhaald gemeten uitkomsten.

#### Genetische data

De statistiek heeft de laatste 10 á 15 jaar grote vorderingen gemaakt in het ontwikkelen van geschikte methoden voor de analyse van gegevens met afhankelijkheid. Deze afhankelijk-

heid komt men niet alleen tegen in de situatie waarin de uitkomstvariabele per persoon vaker gemeten wordt, maar ook in allerlei andere situaties. Bijvoorbeeld ook in het genetische onderzoek. Daar heeft men de uitkomst, bijvoorbeeld het al of niet krijgen van een bepaalde ziekte, bij families waargenomen. Ook hier stuit men op afhankelijkheid: de uitkomsten van familieleden lijken meer op elkaar dan die van personen die geen familie van elkaar zijn. Dit soort familie gegevens geeft dikwijls aanleiding tot een zeer ingewikkelde gegevensstructuur met allerlei complicaties. Bijvoorbeeld gecensureerde gegevens, problemen die samenhangen met de manier waarop de families geselecteerd zijn (het zog. ascertainment probleem), en incomplete gegevens (sommige familieleden willen zich niet laten onderzoeken of zijn al overleden) etc. In onze groep doet dr. Jeanine Houwing-Duistermaat, in samenwerking met de groep genetische epidemiologie, onderzoek naar statistische methoden voor dit soort genetische gegevens.

Nu genetische data aan de orde komen wil ik in dit verband ook noemen de massale hoeveelheden genetische gegevens die in de nabije toekomst verzameld zullen worden. Het Menselijk Genoom Project is inmiddels zo goed als afgerond en dankzij allerlei nieuwe technieken zullen er zeer grote hoeveelheden genetische informatie per persoon voor grote onderzoekspopulaties beschikbaar komen. Een nieuwe discipline, de bio-informatica, zal zich hiermee bezig gaan houden. Het is in dit verband merkwaardig dat men, geconfronteerd met uitdagend problemen betreffende analyse van data met ingewikkelde nieuwe structuren, zich in eerste instantie voornamelijk lijkt te wenden tot de computerwetenschap voor oplossingen, zonder dat men veel oog lijkt te hebben voor de actuele en potentiële bijdrage van de statistiek, die toch de oudste en een zeer succesvolle "informatie wetenschap" is [17]. De problemen zijn voor een groot deel statistisch van aard en het analyseren van deze 'post genome' data vormt een van de grootste uitdagingen voor statistici in de nabije toekomst.

#### Cluster randomised trials

Een heel andere situatie waarin men met afhankelijke gegevens te maken krijgt is de steeds meer populair wordende cluster randomised trial [18]. Het randomiseren van individuele patiënten is niet geschikt voor het evalueren van het effect van sommige interventies, bijvoorbeeld organisatorische veranderingen in huisartsenpraktijken, zoals het invoeren van een geprotocolleerde behandeling volgens een bepaalde consensus. Dan is het beter om niet de individuen maar de praktijken te randomiseren in een groep waarin de organisatorische verandering wordt doorgevoerd en een controle groep van praktijken waarin dat niet wordt gedaan. Een gevolg van deze onderzoeksopzet is dat uitkomsten van patiënten binnen eenzelfde cluster, de huistartspraktijk in het voorbeeld, afhankelijk van elkaar zijn. Immers, patiënten binnen een zelfde cluster zullen uitkomsten hebben die meer op elkaar lijken dan de

uitkomsten van patiënten van verschillende clusters. De meest eenvoudige analyse, die men nog regelmatig in de literatuur kan aantreffen, is eenvoudig de clustering te negeren. Dit is echter niet correct. De nauwkeurigheid van de schatting van het behandelingseffect wordt dan overdreven en de P-waarde wordt te klein, hetgeen gemakkelijk tot onjuiste conclusies kan leiden. Een eenvoudige en correcte aanpak is per cluster de gemiddelde uitkomst te nemen, en vervolgens standaard methoden op deze clustergemiddelden toe te passen. Een nadeel hiervan is echter dat dit statistisch niet de meest efficiënte manier is en dus sub-optimaal. Een correcte en tegelijkertijd efficiënte statistische analyse modelleert de uitkomsten op patiënt niveau en houdt vervolgens rekening met de correlatie binnen clusters [19]. Hiervoor is een klasse methoden ontwikkeld die bekend staat als 'random effecten' modellen, waarin de afhankelijkheid gemodelleerd wordt door elke cluster zijn eigen niveau te geven en een plausibele verdeling voor deze niveaus aan te nemen. In feite zijn deze modellen weer speciale gevallen van het al eerder genoemde algemene lineaire gemengde model. Een voordeel van deze aanpak boven die gebaseerd op clustergemiddelden is bovendien de mogelijkheid om de variatie in de uitkomsten op patiënt niveau te bestuderen en gebruik te maken van karakteristieken die op patiënt niveau gemeten zijn, zoals leeftijd, geslacht en ernst van de ziekte. Overigens, iets wat nog al eens vergeten wordt, ook in het stadium van het ontwerp van het onderzoek moet men al rekening houden met de afhankelijkheid binnen de clusters met betrekking tot de keuze van het aantal en de grootte van de clusters. Immers, tengevolge van de afhankelijkheid geeft een cluster randomised trial minder informatie dan een trial met hetzelfde aantal patiënten waarin op patiënt niveau gerandomiseerd wordt. Hoe hoger de correlatie binnen een cluster, des te groter is het verlies aan onderscheidingsvermogen. Bij de keuze van de steekproefomvang dient men hier terdege rekening mee te houden.

#### *Multi-level data*

De structuur van de data wordt nog ingewikkelder als in een cluster randomised trial de uitkomst herhaald gemeten wordt binnen een patiënt. Men kan ook denken aan onderwijskundig onderzoek waarbij scholen worden gerandomiseerd, en de prestaties van leerlingen gemeten worden. De leerlingen clusteren dan in klassen, en de klassen clusteren binnen scholen. Zo kan onderzoek vele lagen van variatie bevatten. Voor dit soort gegevens zijn de zogenaamde multi-level modellen ontwikkeld, een generalisatie van de eerder genoemde random effecten modellen voor cluster randomised trials [20].

#### *Meta-analyse*

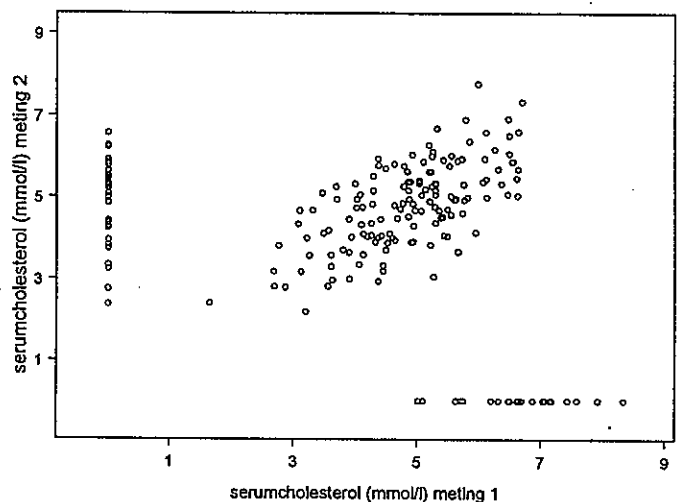
Het laatste decennium zijn meta-analyses zeer populair geworden. Dit zijn systematische overzichten van alle artikelen die betrekking hebben op een bepaalde medische vraagstelling, bijvoorbeeld de effectiviteit van een bepaalde behandeling. Wanneer men bij zo'n overzicht gebruik maakt van formele statistische methoden, spreekt men vaak van meta-analyse. In dit verband denkt u misschien niet onmiddellijk aan afhankelijke waarnemingen. Echter ook meta-analyse data kunnen opgevat worden als data met een bepaalde vorm van afhankelijkheid. Immers, de uitkomsten van patiënten binnen een zelfde trial zullen meer op elkaar lijken dan uitkomsten van patiënten van verschillende trials. Een moderne aanpak van meta-analyse data is dan ook gebruik te maken van een bepaalde vorm van het al eerder genoemde algemene gemengde lineaire model [21] als een natuurlijk kader voor de statistische analyse. In mijn biostatistiek groep is drs. Lidia Arends bezig met onderzoek naar toepassing van deze modellen in allerlei verschillende meta-analyse situaties [22].

#### **Methoden voor missende gegevens**

Ik kom nu toe aan een ander belangrijk onderzoeksthema in de biostatistiek: ontbrekende gegevens. In het laatste decennium is veel vooruitgang geboekt op dit gebied [23]. In elk medisch onderzoek, hoezeer we ook ons best doen om dat te voorkomen, zullen er gegevens ontbreken, gegevens die we hadden willen hebben maar die we om een of andere reden niet gekregen hebben. Als de gegevens niet geheel willekeurig (random) missend zijn, maar selectief missend, kan dat leiden tot ernstige vertekening van de onderzoeksresultaten. Bijvoorbeeld, stel we willen de inkomensverdeling van medisch specialisten onderzoeken door middel van een enquête. Wanneer de specialisten met de hoogste inkomens minder vaak responderen dan die met een relatief laag inkomen, zal de inkomensverdeling in het onderzoek lager uitvallen dan in werkelijkheid.

Hoe gaat men in de praktijk van het medisch onderzoek om met missende gegevens? Laten we een voorbeeld bekijken. Stel we hebben bij een steekproef van 200 patiënten uit een bepaalde populatie de serumcholesterolconcentratie gemeten. Na een jaar werd de meting bij dezelfde groep herhaald. In figuur 5 ziet u de waargenomen gegevens.





Figuur 5. Spreidingsdiagram van serumcholesterol meting 1 tegen meting 2. Punten op de horizontale resp. verticale as geven de individuen aan waarvoor de tweede resp. de eerste meting missend is.

Op de horizontale as staat de eerste meting en op de verticale as de tweede. Van sommige individuen ontbreekt de eerste waarneming, maar hebben we wel de tweede. Die zijn weergegeven als punten op de verticale as. Voor sommige individuen hebben we alleen de eerste, die zijn weergegeven als punten op de horizontale as. 15% van de gegevens is missend in dit voorbeeld, een in de praktijk zeker geen ongebruikelijk hoog percentage. Dit zijn m.b.v. de computer gesimuleerde data, d.w.z. ik heb de data zelf gemaakt, en zelf een aantal missende waarden aangebracht. Eerst heb ik geheel willekeurig 18% van de eerste metingen missend gemaakt, en vervolgens zijn missende waarden op de tweede meting gecreëerd met een kans die toenam naarmate de eerste meting hoger was. Merk op dat dit in het geheel niet onrealistisch is: het is goed denkbaar dat juist patiënten met een zeer hoge, of een zeer lage cholesterolwaarde, een grotere kans hebben om uit te vallen bij het onderzoek.

#### Traditionele aanpak

Hoe gaat men in de praktijk met ontbrekende gegevens om? Ik noem de in de praktijk meest gevolgde strategieën.

#### Analyse van de complete patiënten

De eenvoudigste en in de praktijk meest gevolgde strategie is dat men zich beperkt tot de patiënten met complete data. Een nadeel van deze aanpak is dat dit kan leiden tot een drastische reductie van de hoeveelheid beschikbare gegevens, in de praktijk gemakkelijk tot een reductie van 50% of meer, en daardoor tot een verlies van statistische betrouwbaarheid. In ons voorbeeld houden we 140 van de 200 patiënten over, een reductie van 30%. Een belangrijker nadeel is dat de resultaten van de analyse alleen valide zijn, d.w.z. zonder systematische vertekening (bias), als verondersteld mag worden dat de 140 overblijvende patiënten nog steeds een geheel willekeurige (en daarmee naar verwachting een representatieve) steekproef vormen uit de populatie. Dit is een strenge voorwaarde die de 'volledig random missend' voorwaarde wordt genoemd. Laten we in ons voorbeeld kijken wat het resultaat van deze analyse is. In de volgende tabel staan de gemiddelden van meting 1 en 2 en het gemiddelde verschil.

Serumcholesterol (mmol/l)	Complete patiënten analyse (n=140)	Volledige data (n=200)
Gemiddelde meting 1	4.85	5.12
Gemiddelde meting 2	4.81	4.99
Gemiddelde verschil	0.04	0.13

Omdat het hier gesimuleerde data betreft, kennen we ook de volledige data en kunnen daarmee vergelijken. We zien dat het gemiddelde op meting 1, op meting 2, en het gemiddelde verschil alle te laag zijn. Deze bias wordt niet veroorzaakt door de groep patiënten die missend waren op eerste meting, deze groep was immers geheel willekeurig, voldoende aan de voorwaarde van volledig at random missende gegevens. De bias wordt veroorzaakt door de groep die missend is op de tweede meting. Deze patiënten scoren hoog op beide metingen en veroorzaken een negatieve vertekening in het gemiddelde van meting 1 en 2. Deze biases vallen niet tegen elkaar weg in het verschil, zodat ook dit is vertekend. Het verschil is niet statistisch significant ( $P = 0.57$ ), terwijl dit voor de volledige data wel zo is ( $P = 0.03$ ). We kunnen concluderen dat in dit voorbeeld de standaard analyse gebaseerd op de complete patiënten niet adequaat is.

#### Analyse van de beschikbare data

Een andere vaak in de praktijk gevolgde aanpak is voor elke schatting alle patiënten te gebruiken die beschikbaar zijn. Een voordeel van deze aanpak is dat minder gegevens verloren gaan dan bij de analyse van de complete patiënten. Daarom is de statistische nauwkeurigheid meestal groter. Maar consequent doorvoeren van deze aanpak is in de praktijk heel lastig. Een belangrijker nadeel is dat ook deze aanpak alleen valide is als voldaan is aan de strenge eis van volledig at random missende gegevens. Laten we in de volgende tabel bekijken hoe deze analyse uitvalt in ons voorbeeld.

Serumcholesterol (mmol/l)	Analyse beschikbare data (n=140)	Volledige data (n=200)
Gemiddelde meting 1	5.11 (n=164)	5.12
Gemiddelde meting 2	4.81 (n=176)	4.99
Gemiddelde verschil	0.28 (n=164;176)	0.13

We zien dat het gemiddelde van meting 1 niet is vertekend. Dat komt natuurlijk doordat de 36 ontbrekende waarnemingen geheel willekeurig tot stand waren gekomen. De tweede meting is ernstig gebiased. Ook het verschil is vertekend, veel erger nog dan in de complete patiënten analyse. We concluderen dat ook de 'beschikbare gegevens' aanpak in dit voorbeeld zeer inadequaaf uitpakt.

#### Andere methoden

Voor de volledigheid noem ik de enkelvoudige imputatie methoden. In plaats van patiënten uit de data analyse weg te laten probeert men dan de ontbrekende gegevens in te vullen (dit heet "imputeren") op basis van bepaalde veronderstellingen. Dit kan op allerlei manieren. Het gemakkelijkst is gewoon het gemiddelde van de niet ontbrekende waarnemingen in te vullen. In de praktijk wordt dit vooral toegepast in regressie analyses om de waarden van de voorspellende variabelen te completeren. In het algemeen moet deze methode sterk worden ontzaden. Zelfs als aan de strenge voorwaarde van volledig at random ontbrekend voldaan is, kunnen de resultaten ernstig vertekend zijn. Bovendien zijn de resulterende betrouwbaarheidsintervallen altijd te smal, zodat men een grotere zekerheid omtrent de resultaten van het onderzoek suggereert dan op grond van de data gerechtvaardigd is. Hierdoor kan men tevens ten onrechte significante verschillen tussen schattingen vinden.

Een andere methode die vaak wordt toegepast in regressie-analyses is het gebruik van zogenaamde indicatorvariabelen. Recentelijk is het inzicht gegroeid dat ook deze methode tot ernstige bias kan leiden en moet worden afgeraden [24].

#### Moderne modelmatige aanpak.

In het laatste decennium is naarstig onderzoek gedaan naar statistische analyse methoden die geldig blijven onder minder strenge voorwaarden dan het volledig at random ontbreken van de waarnemingen, terwijl tegelijkertijd alle geobserveerde gegevens worden gebruikt. Er zijn inmiddels verscheidene methoden beschikbaar gekomen die geschikt zijn voor vele in de praktijk voorkomende situaties. Een vaak geschikte methode is de likelihood methode. Een andere, sterk in opkomst zijnde methode, is de multiple imputatie methode. Wij illustreren beide methoden aan de hand van ons voorbeeld.

#### Likelihood methode

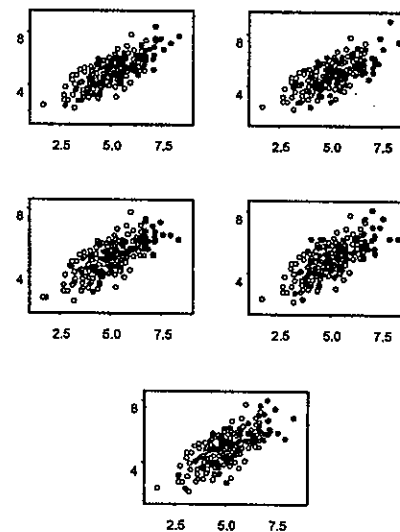
Theoretisch kan bewezen worden dat deze methode geldig blijft onder een minder strenge voorwaarde die missend at random wordt genoemd. Dit betekent dat de ontbrekende waarden wel in zekere mate selectief mogen zijn: de kans dat een gegeven ontbreekt mag wel van de data afhangen, maar alleen van geobserveerde data, en niet van niet geobserveerde data. In ons voorbeeld waren de missende waarden wel degelijk selectief missing. De patiënten met een missende tweede waarde scoren gemiddeld hoog. Maar aan de missend at random voorwaarde is in ons voorbeeld wel voldaan: het missend zijn op meting 2 hing alleen af van de score op meting 1, en niet van de ontbrekende tweede meting zelf. Dit stelt ons in principe in staat om de ontbrekende waarnemingen min of meer te reconstrueren uit de wel geobserveerde data. De essentie van de likelihood methode is dat gebruik wordt gemaakt van de samenhang tussen de variabelen. In ons voorbeeld 'ziet' de methode als het ware dat de patiënten met een ontbrekende meting 2 relatief hoog scoren op meting 1. De methode anticipeert er dan op dat de ontbrekende tweede waarnemingen wel relatief hoog geweest zullen moeten zijn vanwege de correlatie tussen de twee variabelen, en het gemiddelde van de geobserveerde van meting 2 wordt dan automatisch naar boven gecorrigeerd. Er komt steeds meer software beschikbaar waarmee in allerlei praktisch situaties de likelihood analyse gemakkelijk kan worden uitgevoerd. In mijn voorbeeld heb ik gebruik gemaakt van de procedure Proc Mixed uit het SAS pakket [16]. De resultaten zijn weergegeven in de volgende tabel.

Serumcholesterol (mmol/l)	Likelihood analyse	Volledige data (n=200)
Gemiddelde 1e meting	5.107	5.12
Gemiddelde 2e meting	4.976	4.99
Gemiddelde verschil	0.131	0.13

De resultaten benaderen opvallend goed de schattingen op grond van de volledige data. De nauwkeurigheid is wel iets minder (een standaardfout van 0.07 voor het verschil, i.p.v. 0.06 voor de volledige data) en de P-waarde voor het verschil is iets groter (0.07 i.p.v. 0.04); dat weerspiegelt netjes het feit dat de likelihood analyse op minder data gebaseerd is.

#### Meervoudige imputatie

Deze methode, die evenals de likelihood methode gebruik maakt van alle waarnemingen en geldig is als de ontbrekende waarnemingen aan de minder strenge 'missend at random' voorwaarde voldoen, bestaat uit twee stappen [25]. De eerste stap is het completeren van de dataset door de ontbrekende waarnemingen in te vullen, de tweede stap is de eigenlijke data analyse. In feite worden er meerdere gecompleteerde datasets gemaakt, bijvoorbeeld 5, waarbij de verschillen tussen deze gecompleteerde datasets de onzekerheid representeren ten aanzien van de ontbrekende gegevens. De imputaties worden gebaseerd op statistische modellering van de data, iets waarop ik in dit bestek niet kan ingaan. Essentieel is weer dat gebruik gemaakt wordt van de relaties tussen de variabelen. De analyse van de gecompleteerde datasets kan gedaan worden met standaard methoden, op elke gecompleteerde dataset afzonderlijk. Vervolgens worden de resultaten daarvan gecombineerd, een soort middeling. Ik heb deze methode toegepast op ons voorbeeld. In de volgende figuur ziet u de



Figuur 6. Vijf geïmputeerde data sets. De geïmputeerde punten zijn vetgedrukt weergegeven.

Voor elk van de 5 datasets heb ik de gemiddelden van meting 1 en 2 en het gemiddelde verschil uitgerekend en vervolgens gemiddeld, zie de volgende tabel. Het resultaat benadert opmerkelijk goed de volledige data, en is bijna identiek aan de likelihood analyse.

Serumcholesterol (mmol/l)	Multiple imputatie methode	Volledige data (n=200)
Gemiddelde 1e meting	5.105	5.12
Gemiddelde 2e meting	4.976	4.99
Gemiddelde verschil	0.128	0.13

Voor het toepassen van de likelihood methode en/of multiple imputatie komt steeds meer software beschikbaar zowel in speciale pakketten, bijvoorbeeld SOLAS [26], als langzamerhand ook in standaard statistische pakketten zoals SAS en S-Plus. In veel praktijksituaties kunnen deze methoden inmiddels toegepast worden, maar nog lang niet in alle. Veel onderzoek moet op dit terrein nog gebeuren worden, veel ervaring moet nog worden opgedaan, en zeer veel software nog ontwikkeld. Het toepassen van deze methoden in de praktijk is ge-

specialiseerd statistisch werk, dat vooralsnog in de meeste gevallen beter kan worden overgelaten aan professionele statistici.

### 3. Bayesiaanse methoden

Een derde terrein van onderzoek waarop ik de aandacht wil vestigen is het gebied van Bayesiaanse methoden. Er zijn twee concurrerende filosofische scholen in de statistiek: de frequentistische en de Bayesiaanse. De statistici die frequentistische methoden toepassen zijn nog steeds verreweg in de meerderheid, en vrijwel alle statistische analyses die u in de medische literatuur aantreft zijn frequentistisch van aard. De Bayesianen zijn minder in getal en wat meer geconcentreerd op mathematisch statistische afdelingen van universiteiten. De toenemende rekenkracht en beschikbaarheid van computers en programmatuur brengt echter verandering in deze situatie [26]. In de laatste 10 á 15 jaar zijn er doorbraken geweest op het gebied van de numerieke methoden die het toepassen van Bayesiaanse statistiek in de praktijk een stuk gemakkelijker hebben gemaakt [27].

Bayesiaanse methoden zijn gebaseerd op het uitgangspunt dat onbekende parameters een verdeling hebben. Stel bijvoorbeeld dat we de prevalentie van diabetes type II in Rotterdam bij ouderen van boven de 60 jaar willen onderzoeken. We nemen een aselechte steekproef van 1000 individuen uit deze populatie en vinden bijvoorbeeld 25 mensen met diabetes. De standaard frequentistische statistiek geeft dan een schatting van 2.5%, met een 95% betrouwbaarheidsinterval dat loopt van 1.53% tot 3.47%. In deze vorm van statistiek wordt alle kennis die we hebben over de prevalentie van diabetes in andere populaties geheel genegeerd. Stel bijvoorbeeld dat wij weten dat de prevalentie van diabetes in deze leeftijdsgroep in de verschillende regio's in Nederland ongeveer varieert tussen de 1% en 3%. De Bayesiaanse statistiek maakt dan van deze kennis gebruik middels het opstellen van een zogenaamde apriori verdeling, die de voorafkennis weerspiegelt over de mogelijke waarden van de parameter in kwestie. Deze apriori verdeling specificeert hoe waarschijnlijk de mogelijke waarden van de onbekende parameter zijn, vooraf aan het verzamelen van de data. Als apriori verdeling zou in dit voorbeeld een normale verdeling met gemiddelde 2% en een standaardafwijking 0.5% genomen kunnen worden. Wanneer de data eenmaal verzameld zijn, wordt met behulp hiervan de posterior verdeling afgeleid. In deze verdeling wordt de informatie uit de data met de voorafkennis geïntegreerd en deze verdeling weerspiegelt zodoende onze kennis over de onbekende parameter na afloop van het onderzoek.

In ons voorbeeld is de Bayesiaanse schatting van de Rotterdamse prevalentie 2.41% met een 95% "betrouwbaarheidsinterval" dat loopt van 1.54% tot 3.28%. Dit Bayesiaanse betrouwbaarheidsinterval heeft een simpele interpretatie: de kans is 95% dat de Rotterdamse diabetes prevalentie tussen de 1.54% en 3.28% ligt, een interpretatie die bij het gewone, fre-

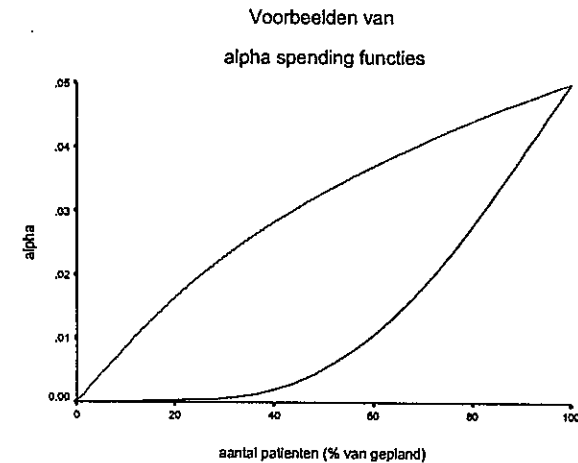
quentistische, betrouwbaarheidsinterval niet is toegestaan. Het Bayesiaanse betrouwbaarheidsinterval is in twee opzichten verschillend van het gewone. Het is het een beetje verschoven in de richting van het centrum van de apriori verdeling, dus naar het gemiddelde van alle regio's, en verder is het korter dan het conventionele, omdat van meer kennis gebruik wordt gemaakt.

De Bayesiaanse benadering is aantrekkelijk, elegant, en sluit beter aan bij de denkwijze van onderzoekers. U zult zich afvragen waarom de Bayesiaanse statistiek dan zo op de achtergrond is gebleven. Het is een misverstand te denken dat biostatistici en masse de frequentistische statistiek tot dogma verheven hebben en principieel tegen Bayesiaanse statistiek zijn. De reden is tweërlei. Ten eerste is het opstellen van de apriori verdeling een grote moeilijkheid. De specificatie hiervan is voor een groot deel subjectief en beïnvloedt de conclusies van het onderzoek. Hier kan goeddeels aan tegemoet gekomen worden door de door de apriori verdeling weinig informatief te specificeren, maar aan de andere kant is dat eigenlijk weer tegen de achterliggende filosofie. De andere reden is veel meer praktisch van aard. De Bayesiaanse benadering in reële toepassingen kan gemakkelijk leiden tot onhanteerbare numerieke problemen, hetgeen toepassing van deze methoden in de praktijk zeer heeft belemmerd. In het laatste decennium zijn echter grote vorderingen gemaakt met het oplossen van deze problemen, de zogenaamde MCMC methoden [28]. Deze nieuwe methoden, samen met de komst van krachtige computers, maken tegenwoordig het toepassen van Bayesiaanse methoden in de praktijk gemakkelijker mogelijk. Het is nu zelfs zo dat allerlei ingewikkelde problemen soms makkelijker in het Bayesiaanse framework kunnen worden opgelost dan in het frequentistische en daarom aantrekkelijk zijn. Deze ontwikkelingen zullen in de nabije toekomst een krachtige impuls geven aan het toepassen van Bayesiaanse methoden in de praktijk van het medisch onderzoek. U zult hier zeker meer van gaan zien in de medische literatuur.

### (Groeps)sequentiele methoden (interim analyses)

Een andere recente ontwikkeling waar ik uw aandacht voor wil vragen zijn de sequentiële methoden voor experimenteel klinisch onderzoek. In een standaard klinisch onderzoek naar het verschil in effectiviteit tussen twee behandelingen vindt de data analyse, d.w.z. het berekenen van de schatting van het verschil in effect, het bijbehorende betrouwbaarheidsinterval en de P-waarde, pas plaats wanneer de uitkomsten van alle patiënten die in het onderzoek zouden worden betrokken zijn waargenomen. Steeds vaker worden echter zogenaamde interim analyses gedaan: tussentijdse data analyses wanneer nog niet alle patiënten uitkomsten zijn waargenomen. Wanneer interim analyses de consequentie kunnen hebben dat de trial tussentijds gestopt wordt, bijvoorbeeld als blijkt dat er op dat moment al een duidelijk verschil

in effect is, spreekt men vaak van formele interim analyses. De redenen voor formele interim analyses zijn vooral ethisch van aard. Als halverwege de trial al duidelijk had kunnen zijn welke behandeling beter is, dan is het ethisch onaanvaardbaar dat patiënten nog worden blootgesteld aan een behandeling waarvan men had kunnen weten dat hij inferieur is. Vaak realiseren onderzoekers zich niet dat formele interim analyses specifieke statistische complicaties met zich meebrengen. Stel dat interim analyses worden uitgevoerd na 25, 50 en na 75% van het geplande aantal patiënten in de trial. Steeds toetst men de nulhypothese dat er geen verschil is tussen de behandelingen. Een naïeve aanpak zou zijn het onderzoek af te breken zodra het behandelingseffect statistisch significant is, zeg bij 5% significantie niveau. Het probleem hiervan is dat elke keer dat een statistische toets wordt uitgevoerd, de fout positieve kans op een significant verschil 5% is. Als men meerdere keren tussentijds toetst, dan cumuleert deze kans. Bij drie interim analyses is de kans dat ten onrechte tot een verschil in effect geconcludeerd wordt, opgelopen tot ongeveer 13%, een stuk hoger dan de 5% die in praktijk aanvaardbaar wordt geacht. Wanneer men vaker interim analyses doet, kan deze kans nog veel hoger oplopen. Ook het betrouwbaarheidsinterval klopt niet meer. De gebruikelijke methoden kunnen in de situatie van Interim analyses dus niet zonder meer worden toegepast. Er zijn aanpassingen nodig die er voor zorgen dat de onterechte significantie kans in de hand wordt gehouden, netjes op het niveau van 5%. Dat is echter nog niet zo heel eenvoudig. Een belangrijke tak van onderzoek in de biostatistiek gedurende de laatste 10 á 15 jaar is het ontwikkelen van onderzoeksopzetten en analyse methoden voor dit soort zog. (groeps)sequentiële klinische trials. Men begon met designs voor klinische trials met een vast aantal, tevoren geplande, interim analyses op gelijke afstanden van elkaar. Hoewel een verbetering, was dit voor de praktijk nog steeds een heel rigide design. Een belangrijke volgende stap in de ontwikkelingen was de introductie van de zog.  $\alpha$ -spending functie. Daarbij staat  $\alpha$  voor de onbetrouwbaarheid, d.w.z. de fout positieve kans op een significant verschil in effect na afloop van de trial. In de praktijk fixeert men deze vrijwel altijd op de bekende 5%. Het idee is dat iedere keer als er een interim analyse gedaan wordt, er een stukje van de totale  $\alpha$  gebruikt mag worden. Het is aan de onderzoeker om, via een vrij te kiezen  $\alpha$ -spending functie, te specificeren hoe in de loop van de trial de totale  $\alpha$  verbruikt wordt. In de volgende figuur ziet u twee voorbeelden van  $\alpha$ -spending functies.

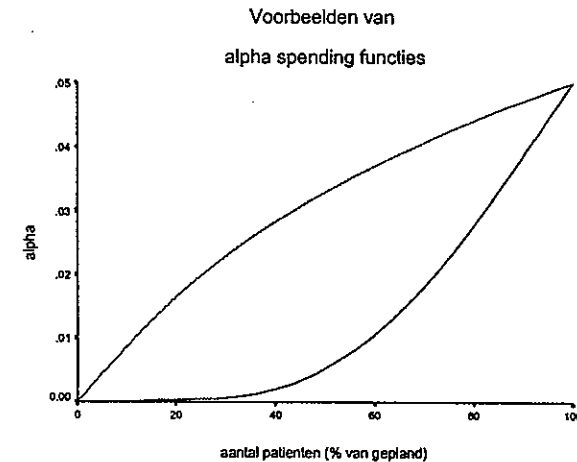


Figuur 7. Twee voorbeelden van  $\alpha$ -spending functies.

Op de horizontale as ziet u het percentage patiënten waarvan de uitkomst op een bepaald moment bekend is. Op de verticale as is het gedeelte van de totaal beschikbare  $\alpha$  uitgezet dat tot op dat moment gebruikt mag worden. De onderste curve is een voorbeeld van een  $\alpha$ -spending functie die heel conservatief is in de eerste helft van de trial. De kans om te stoppen is in het begin heel gering. In het andere voorbeeld wordt relatief veel van de beschikbare  $\alpha$  al in het begin gebruikt en is de kans om vroeg te stoppen relatief groot.

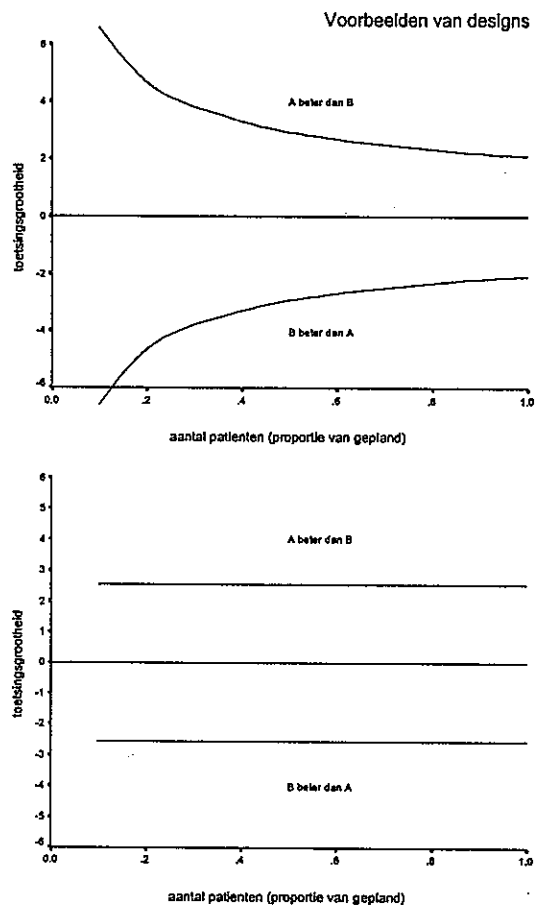
De  $\alpha$ -spending functie aanpak is heel flexibel. Het is niet meer nodig van te voren vast te stellen na welke aantallen patiënten de interim analyses verricht zullen worden. Deze mag de onderzoeker geheel vrij kiezen gedurende de loop van de trial. Wanneer op een bepaald moment tot een interim analyse besloten wordt, kan op grond van de  $\alpha$ -spending functie, rekening houdend met het gedeelte van  $\alpha$  dat bij eventuele eerdere interim analyses al gebruikt is, een stopregel worden bepaald die zegt hoe klein de P-waarde moet zijn om het onderzoek te mogen afbreken met de conclusie dat een verschil in behandelingseffect is aangetoond. Enkele schematische voorbeelden van designs ziet u in de volgende figuur.

In effect is, spreekt men vaak van formele interim analyses. De redenen voor formele interim analyses zijn vooral ethisch van aard. Als halverwege de trial al duidelijk had kunnen zijn welke behandeling beter is, dan is het ethisch onaanvaardbaar dat patiënten nog worden blootgesteld aan een behandeling waarvan men had kunnen weten dat hij inferieur is. Vaak realiseren onderzoekers zich niet dat formele interim analyses specifieke statistische complicaties met zich meebrengen. Stel dat interim analyses worden uitgevoerd na 25, 50 en na 75% van het geplande aantal patiënten in de trial. Steeds toetst men de nulhypothese dat er geen verschil is tussen de behandelingen. Een naïeve aanpak zou zijn het onderzoek af te breken zodra het behandelingseffect statistisch significant is, zeg bij 5% significantie niveau. Het probleem hiervan is dat elke keer dat een statistische toets wordt uitgevoerd, de fout positieve kans op een significant verschil 5% is. Als men meerdere keren tussentijds toetst, dan cumuleert deze kans. Bij drie interim analyses is de kans dat ten onrechte tot een verschil in effect geconcludeerd wordt, opgelopen tot ongeveer 13%, een stuk hoger dan de 5% die in praktijk aanvaardbaar wordt geacht. Wanneer men vaker interim analyses doet, kan deze kans nog veel hoger oplopen. Ook het betrouwbaarheidsinterval klopt niet meer. De gebruikelijke methoden kunnen in de situatie van interim analyses dus niet zonder meer worden toegepast. Er zijn aanpassingen nodig die er voor zorgen dat de onterechte significantie kans in de hand wordt gehouden, netjes op het niveau van 5%. Dat is echter nog niet zo heel eenvoudig. Een belangrijke tak van onderzoek in de biostatistiek gedurende de laatste 10 á 15 jaar is het ontwikkelen van onderzoeksopzetten en analyse methoden voor dit soort zog. (groeps)sequentieële klinische trials. Men begon met designs voor klinische trials met een vast aantal, tevoren geplande, interim analyses op gelijke afstanden van elkaar. Hoewel een verbetering, was dit voor de praktijk nog steeds een heel rigide design. Een belangrijke volgende stap in de ontwikkelingen was de introductie van de zog.  $\alpha$ -spending functie. Daarbij staat  $\alpha$  voor de onbetrouwbaarheid, d.w.z. de fout positieve kans op een significant verschil in effect na afloop van de trial. In de praktijk fixeert men deze vrijwel altijd op de bekende 5%. Het idee is dat iedere keer als er een interim analyse gedaan wordt, er een stukje van de totale  $\alpha$  gebruikt mag worden. Het is aan de onderzoeker om, via een vrij te kiezen  $\alpha$ -spending functie, te specificeren hoe in de loop van de trial de totale  $\alpha$  verbruikt wordt. In de volgende figuur ziet u twee voorbeelden van  $\alpha$ -spending functies.



Figuur 7. Twee voorbeelden van  $\alpha$ -spending functies.

Op de horizontale as ziet u het percentage patiënten waarvan de uitkomst op een bepaald moment bekend is. Op de verticale as is het gedeelte van de totaal beschikbare  $\alpha$  uitgezet dat tot op dat moment gebruikt mag worden. De onderste curve is een voorbeeld van een  $\alpha$ -spending functie die heel conservatief is in de eerste helft van de trial. De kans om te stoppen is in het begin heel gering. In het andere voorbeeld wordt relatief veel van de beschikbare  $\alpha$  al in het begin gebruikt en is de kans om vroeg te stoppen relatief groot. De  $\alpha$ -spending functie aanpak is heel flexibel. Het is niet meer nodig van tevoren vast te stellen na welke aantallen patiënten de interim analyses verricht zullen worden. Deze mag de onderzoeker geheel vrij kiezen gedurende de loop van de trial. Wanneer op een bepaald moment tot een interim analyse besloten wordt, kan op grond van de  $\alpha$ -spending functie, rekening houdend met het gedeelte van  $\alpha$  dat bij eventuele eerdere interim analyses al gebruikt is, een stopregel worden bepaald die zegt hoe klein de P-waarde moet zijn om het onderzoek te mogen afbreken met de conclusie dat een verschil in behandelingseffect is aangetoond. Enkele schematische voorbeelden van designs ziet u in de volgende figuur.

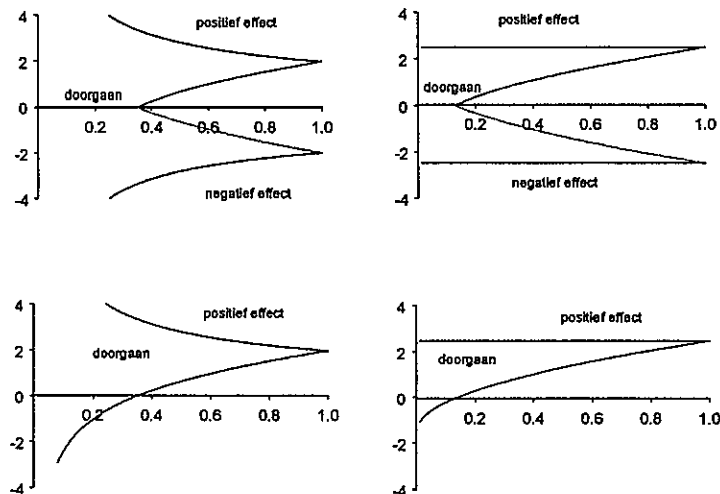


Figuur 8. Twee voorbeelden van sequentiële designs.

Deze designs geven de stopgrenzen weer voor het geval er 10 interim analyses gedaan worden, steeds na gelijke aantallen patiënten. Op de horizontale as is de proportie weergegeven van het totaal aantal patiënten waarvan ten tijde van de interim analyse de uitkomst al is waargenomen. Verticaal zijn de stopgrenzen weergegeven in termen van een standaardnormaal verdeelde toetsingsgrootheid. Als deze bij een interim analyse boven de bovengrens uitkomt, dan wordt de trial gestopt met de conclusie dat behandeling A beter is dan B. Als de toetsingsgrootheid onder de onderste grens uitkomt, dan stopt de trial met conclusie B is beter dan A. Zolang de toetsingsgrootheid tussen de grenzen blijft, wordt de trial voortgezet. Zit

hij aan het eind nog steeds tussen de grenzen, dan is de conclusie dat geen verschil tussen de behandelingen kon worden aangetoond. Merk op dat de stopgrenzen ruim buiten de  $\pm 1.96$  liggen, de grenzen die gebruikt zouden worden als men naïef de gebruikelijke grenswaarde voor significantie  $P=0.05$  zou gebruiken. Het verschil tussen de twee designs is dat de kans om vroeg te stoppen voor de eerste veel kleiner is dan voor de tweede. Voor het tweede voorbeeld is de grenswaarde steeds gelijk aan 2.56, overeenkomend met een P-waarde van 0.01.

In bovenstaande opzet is het alleen mogelijk bij interim analyses te stoppen als een verschil tussen de behandelingen is aangetoond. Een nieuwe ontwikkeling is gebruik te maken van een design dat ook de mogelijkheid toelaat vroeger te stoppen met de conclusie dat geen verschil in effectiviteit tussen de behandelingen is aangetoond. Voor een dergelijk design moet naast een  $\alpha$ -spending functie ook een  $\beta$ -spending functie gespecificeerd worden.  $\beta$  staat voor de kans op de fout negatieve conclusie dat er geen verschil is in effect, terwijl er in werkelijkheid wel een verschil is.  $\beta$  is niets anders dan het complement van het onderscheidingsvermogen, de power, van de trial. In de praktijk worden patiëntenaantallen voor trials zodanig gekozen dat het onderscheidingsvermogen bij een bepaalde grootte van het verschil in effectiviteit 80% of 90% is, overeenkomend met  $\beta = 0.20$  of  $\beta = 0.10$ . De  $\beta$ -spending functie, weer vrij te kiezen door de onderzoeker, zegt hoe in de loop van de trial de totaal beschikbare kans op een fout-negatieve conclusie wordt opgesnoept. Doordat de keuze van  $\alpha$ - en de  $\beta$ -spending vrij is, zijn er talloze verschillende designs mogelijk waaruit de onderzoeker een geschikte keus kan maken. Een viertal voorbeelden zijn in de volgende figuur weergegeven. Concreet kunt u hierbij bijvoorbeeld denken aan een gerandomiseerde klinische trial waarin een cholesterolverlagend middel wordt vergeleken met een placebo met betrekking tot de kans op een hartinfarct gedurende een behandelingsperiode van 5 jaar. In alle getoonde designs is de kans  $\alpha$  om ten onrechte te concluderen tot verschil in effect gelijk aan 0.05 en het onderscheidingsvermogen  $1-\beta$  is 80% tegen een halvering van de kans op een hartinfarct.



Figuur 9. Enkele voorbeelden van sequentiële designs.

Op de horizontale as staat weer het aantal patiënten dat op het moment van de interim analyse beschikbaar is, als proportie van het geplande maximale aantal. Verticaal staat de bij de interim analyse berekende toetsingsgrootte. In de bovenste twee designs geldt dat wanneer de toetsingsgrootte boven de bovenste of beneden de onderste uitkomt, de trial wordt gestopt met de conclusie dat de cholesterolverlagende behandeling een positief resp. een negatief effect heeft. Wanneer de waarde van de toetsingsgrootte tussen de twee binnenste curven uitvalt, dan wordt de trial gestopt met als conclusie dat geen verschil tussen beide behandelingen is aangetoond. In de andere gevallen wordt de trial voortgezet. In de onderste twee designs wordt de trial gestopt met een positieve conclusie zodra men de bovenste curve kruist, en met de conclusie dat geen effect van de behandeling is aangetoond als de onderste curve gekruist wordt.

De bovenste twee designs in de figuur zijn symmetrisch: het onderscheidingsvermogen tegen een positief effect van de cholesterolverlager is even groot als tegen een even groot negatief effect. In ons voorbeeld is dit niet erg logisch, omdat het niet interessant zal zijn om aan te tonen dat de behandeling slechter is dan placebo. Asymmetrische designs, zoals de onderste twee liggen in dit geval meer voor de hand. In deze designs zal men, als de behandeling geen of een ongunstig effect heeft, met grote kans spoedig de onderste curve kruisen en

stoppen met de conclusie dat geen positief effect van de behandeling kan worden aangetoond. Dit design heeft grote voordelen: in het geval dat de behandeling niet het gewenste positieve effect heeft, worden er niet onnodig veel patiënten blootgesteld aan een inferieure behandeling. Naast dit ethische voordeel zijn er ook economische voordelen. In de situatie dat men twee effectieve behandelingen in een trial met elkaar vergelijkt om te onderzoeken welke de beste is, ligt juist een symmetrisch design zoals de twee bovenste voor de hand. Enkele karakteristieken van de verschillende designs zijn weergegeven in de volgende tabel (tweezijdige  $\alpha = 0.05$ ; eenzijdige  $\alpha = 0.025$ ; power  $1 - \beta = 0.80$ ; kans op hartinfarct in placebo groep naar schatting 4%).

Design	Aantal patiënten			
	I	II	III	IV
	2-sided	2-sided	1-sided	1-sided
Maximum	2871	4497	2890	4725
Verwachte aantallen				
Kansen gelijk	1659	1532	1246	1070
Kans 50% lager onder behandeling	1834	1680	1820	1652

De karakteristieken van de designs verschillen behoorlijk. Het maximale benodigd aantal patiënten loopt uiteen van ongeveer 2900 tot ruim 4700. Het verwachte aantal benodigde patiënten als de cholesterolverlager geen effect heeft varieert tussen 1070 en 1659. De aantallen benodigde patiënten onder halvering van de kans op een hartinfarct door de cholesterol verlagende behandeling blijken niet zoveel te verschillen. Niet weergegeven in de tabel zijn de verwachte patiëntenaantallen als de cholesterolverlager onverhoopt een negatief effect zou hebben. Het aantal patiënten zal dan in de onderste twee designs dramatisch lager uitvallen.

Voor een trial met een vaste steekproefgrootte, maar de zelfde  $\alpha$  en  $\beta$ , zou het benodigde aantal bijna 2500 zijn. Gemiddeld gesproken zijn de aantallen patiënten voor de vier getoonde sequentiële designs dus een stuk lager. Er is slechts een kleine kans dat het aantal boven de 2500 uitvalt.

U begrijpt dat er in principe talloze designs mogelijk zijn. Door de eigenschappen van de verschillende designs met elkaar te vergelijken kan de onderzoeker een keus maken. Hiervoor kan gebruik gemaakt worden van de inmiddels beschikbaar gekomen commerciële software pakketten [29, 30].

Een kanttekening is hierbij op zijn plaats. Ook in de statistiek geldt: niet voor niets. Tegenover het gebruik van dit soort sequentiële studieopzetten staat de prijs van veel ingewikkelder sta-



tistiek. Standaard statistische methoden, geïmplementeerd in de bekende programmapakketten, kunnen niet meer gebruikt worden. Speciale methoden zijn vereist, die correct rekening houden met de gebruikte studieopzet. Vooralsnog zijn de beschikbare analyse mogelijkheden zowel in de theorie als in de praktijk kleiner dan voor trials met een vaste steekproefgrootte. Veel is al wel beschikbaar in de speciale software pakketten, maar veel is ook nog in ontwikkeling. Het is dan ook verstandig om interim analyses vooralsnog vooral te beperken tot de situatie waarin het eindpunt ernstig is (bijvoorbeeld overlijden of een hartinfarct), en het ethische argument voor interim analyses echt belangrijk is, of tot grootscheepse trials waarin ook het economische argument doorslaggevend kan zijn.

In dit korte bestek is het slechts mogelijk u een globale indruk te geven van de methodologie van groepssequentiële onderzoeken, de mogelijkheden en de onmogelijkheden ervan, en de voor- en de nadelen. Recentelijk zijn er twee boeken verschenen die een goed overzicht geven van de huidige 'state of the art' [31,32]. Ik volsta verder met op te merken dat de methodologie op dit gebied voldoende volwassen is geworden om op veel uitgebreider schaal in de praktijk toegepast te worden dan nu het geval is.

In het voorafgaande heb ik u enige belangrijke moderne ontwikkelingen in de biostatistiek geschetst. Ik zou er nog vele aan toe kunnen voegen. De tijd schiet te kort hierop in te gaan. Ik hoop u echter een goede indruk gegeven te hebben van de moderne ontwikkelingen in de medische statistiek. Het is een springlevend vak dat ook in de toekomst een belangrijke en onmisbare rol zal blijven spelen in het medisch onderzoek.

#### ***De medisch statisticus in de praktijk***

Biostatistici zijn in de praktijk natuurlijk niet alleen bezig met nieuwe methodologie te ontwikkelen. In feite besteden zij in de praktijk daar meestal slechts een relatief klein deel van hun tijd aan, de een wat meer en de ander wat minder. Zoals aangekondigd wil ik ook, zij het kort, stilstaan bij de rol van medisch statistici in de praktijk. Bij het bespreken daarvan beperk ik mij tot statistici die werkzaam in medische faculteiten, met de nadruk op wat wij zelf in Rotterdam doen.

#### ***Methodologische ondersteuning***

Gemiddeld gesproken besteden biostatistici verreweg het grootste deel van hun tijd aan methodologische begeleiding van medisch onderzoek. Ik beschouw dit als onze primaire taak.

Dit betreft alle stadia van onderzoek, vanaf de opzet tot en met de data-analyse en het schrijven van de publicaties. De manier waarop dit plaats vindt is zeer divers. Dit varieert tussen een kort contact, een eenmalig advies ten aanzien van bijvoorbeeld de steekproefomvang of de statistische analyse, tot een diepgaande jarenlange participatie in een onderzoeksproject of een reeks onderzoeksprojecten van eenzelfde onderzoeksgroep. Tegenwoordig heeft elke medische onderzoeker direct de beschikking over uitgebreide en gebruiksvriendelijke statistische programma's. Dit leidt echter geenszins tot verminderde behoefte aan biostatistici. Zoals al eerder opgemerkt leidt de steeds groeiende statistische methodologie en de toenemende complexiteit van de gegevens eerder tot meer dan minder behoefte aan statistische inbreng. Ook in de activiteiten van mijn groep ligt de nadruk op de statistische begeleiding van het onderzoek dat in het Erasmus *Universitair* Medisch Centrum Rotterdam, het EMCR, wordt uitgevoerd. Er zullen in het academisch ziekenhuis en faculteit nauwelijks afdelingen te vinden zijn waarmee wij niet in meer of mindere mate contacten onderhouden. Over het algemeen is deze samenwerking niet geformaliseerd. Mijn groep zal in de toekomst streven naar meer formele samenwerkingsverbanden, waarin de biostatisticus zich voor een vast gedeelte van haar of zijn tijd kan concentreren op het onderzoek van eenzelfde onderzoeksgroep of afdeling. Dit leidt tot een diepere betrokkenheid en meer inzicht in het betreffende onderzoeksterrein en derhalve tot vergroting van de kwaliteit en effectiviteit van zijn of haar inbreng.

In 1993 was onze groep, samen met de Klinische Besliskunde en de Klinische Epidemiologie, medeoprichter van het Centrum voor Patiëntgebonden Onderzoek, het CPO. Later werd het CPO uitgebreid met deskundigen op het gebied van kosten-effectiviteitsanalyse van het Instituut Medical Technology Assessment en met deskundigen op het gebied van klinimetrie van het Instituut Medische Psychologie en Psychotherapie. Het CPO adviseert en participeert in de opzet en uitvoering van klinisch onderzoek in het gehele EMCR en staat open voor alle afdelingen van daarvan. Het CPO heeft zich inmiddels een gewaardeerde plaats verworven, hoewel de bekendheid ervan onder medische onderzoekers op de werkvloer nog wel eens te wensen overlaat.

#### ***Onderwijs***

Een andere, even belangrijke, taak voor medisch statistici werkzaam bij medische faculteiten is vanzelfsprekend het onderwijs, in de eerste plaats het onderwijs aan medische studenten. Om de medisch wetenschappelijke literatuur te kunnen lezen en om onderzoek te kunnen doen is een behoorlijke statistische basiskennis onmisbaar. Medische studenten krijgen niet meer dan het absolute minimum dat hiervoor nodig is in hun reguliere programma aangereikt. Vergeleken met studenten in andere vakken, bijvoorbeeld biologie, gezondheidswetenschap-

pen of sociale wetenschappen, is dit relatief weinig. Dit is een gevolg van het feit dat de studie geneeskunde vooral een beroepsgerichte opleiding is. In Rotterdam zal de komst van het nieuwe curriculum "Arts 2007" naar verwachting een nieuwe impuls geven. Een onderwijsblok gewijd aan de methodologie geplaatst vlak voor het keuze-onderzoek en daarmee sterk verbonden, geeft de kans om een krachtige bijdrage te leveren aan de profilering van de Rotterdamse artsenopleiding als een opleiding die sterk onderzoeksgericht is.

Een opvallende ontwikkeling in de laatste 10 jaar is het steeds belangrijker worden van post-academische opleidingen en cursussen. In alle medische faculteiten heeft de biostatistiek een vaste plaats in het cursorische gedeelte van de opleiding van Assistenten In Opleiding. In een aantal medische faculteiten zijn postacademische onderwijsprogramma's ontstaan waarin de biostatistiek een substantiële plaats in neemt. In de onderzoeksschool waarin wij participeren, het Netherlands Institute of Health Sciences, is gedurende het laatste decennium een stelsel van internationale Master of Science onderzoekersopleidingen gegroeid. Een substantieel deel van het programma is gewijd aan biostatistiek en wordt door onze groep verzorgd.

#### *Onderzoek*

Een derde belangrijke taak van medisch statistici is het leveren van een bijdrage aan de verdergaande ontwikkeling van biostatistische methoden. Dit kan op verschillende niveaus gebeuren. Het kan zijn het introduceren van een methode die standaard is in andere wetenschapsgebieden maar nieuw in de medische discipline. Het kan zijn het toepassen van recent gepubliceerde maar nog niet praktisch toegepaste methoden die nieuw zijn voor de gemeenschap van biostatistici en het proberen deze toe te passen in medische projecten waar men in betrokken is. Het derde niveau is het ontwikkelen van nieuwe modellen en theorie voor situaties waarvoor nog geen goede oplossingen zijn.

Ook aan het biostatistische onderzoek leveren wij als Rotterdamse biostatistiek groep onze bijdrage en streven er naar dit verder uit te bouwen. Bij voorkeur zullen wij dit doen in nauwe samenwerking met de epidemiologische onderzoeksgroepen uit ons eigen instituut en met klinische onderzoeksgroepen uit het EMCR. In dit verband noem ik de reeds lopende samenwerkingen met de groepen genetische-, klinische-, hart- en vaatziekten-, en neuro-epidemiologie, en het samenwerkingsverband met de afdeling Maag- Darm- en Leverziekten.

#### *Overige taken*

Naast de al besproken statistische advisering, onderwijs en onderzoek zijn er vele andere, minder omvangrijke maar daarom niet minder belangrijke, taken van biostatistici, zoals het

reviewen van manuscripten voor medische tijdschriften en het beoordelen van onderzoeksvoorstellen als lid van allerlei beoordelingscommissies, bijvoorbeeld medisch-ethische toetsingscommissies (METC's). Nu de recent in werking getreden wet op het medisch onderzoek met mensen voorschrijft dat er in iedere METC een methodoloog moet zitten, wordt er een toenemend beroep hiervoor gedaan op biostatistici. Dit soort activiteiten kan heel tijdrovend zijn. Hoewel belangrijke taken in het kader van het algemeen belang, zullen biostatistici deze vaak noodgedwongen beperken omdat in de regel compensatie ontbreekt.

#### *Geachte toehoorders,*

Biostatistici, dat brengt de aard van het vak met zich mee, doen hun werk in medische faculteiten en academische ziekenhuizen tamelijk onopvallend en op de achtergrond. Zij verzorgen de public relations over het algemeen tamelijk slecht. Dat past misschien bij hun aard: in een editorial in de Lancet van een aantal jaren geleden worden zij beschreven als een "modest and shy" type mensen. Uit deze rede moge duidelijk zijn dat daar geen enkele reden voor is. De "onzichtbaarheid" van biostatistici staat niet in verhouding tot de impact die zij hebben op het medisch onderzoek en, vooral het postacademische, onderwijs. Een groot deel van mijn rede heb ik besteed aan het schetsen van een aantal voor het medisch wetenschappelijk onderzoek zeer relevante hedendaagse ontwikkelingen in de biostatistiek. Er is geen reden om aan te nemen dat deze ontwikkelingen zich in de toekomst niet krachtig zullen doorzetten. Er is wel alle reden om te verwachten dat de medische statistiek ook in de toekomst een even essentiële bijdrage aan de voortgang van het medisch onderzoek zal leveren als zij in het verleden gedaan heeft.

#### *Dankwoord*

Aan het einde van mijn oratie gekomen wil ik nog enkele persoonlijke woorden spreken. Het College van Bestuur van de Erasmus Universiteit en het bestuur van de Faculteit der Geneeskunde en Gezondheidswetenschappen wil ik bedanken voor het in mij gestelde vertrouwen. Ik zal mij maximaal inspannen om de mij toevertrouwde taak naar beste kunnen uit te voeren.

#### *Hooggeleerde van Strik, beste Roel,*

Meer dan 10 jaar heb ik in grote harmonie met je mogen samenwerken. Dat heb ik altijd zeer gewaardeerd. Eerder in deze rede heb ik je een van de pioniers van de medische statistiek in Nederland genoemd. Ik hoop mij een waardig opvolger van jou te zullen betonen.

*Hooggeleerde van Houwelingen, beste Hans*

Destijds bij het Instituut voor Mathematische Statistiek in Utrecht, inmiddels al 25 jaar geleden, was jij mijn leermeester die mij leidde bij mijn eerste schreden op het pad van de statistiek. Hoewel niet formeel, was ik eigenlijk jouw eerste promovendus. Ook later heb ik veel geleerd van je scherpe inzicht in de statistische methodologie. Wij hebben altijd intensieve contacten onderhouden zowel in de privé sfeer als in professionele sfeer. Dit heb ik altijd zeer gewaardeerd en dat zal ongetwijfeld in de toekomst zo blijven.

*Collegae van het Instituut Epidemiologie en Biostatistiek.*

In de eerste plaats Hooggeleerde Hofman, beste Bert. Ik heb bewondering voor de visie waarmee je leiding geeft aan het instituut. Tot die visie behoorde ook de overtuiging dat een leerstoel biostatistiek onmisbaar was. De epidemiologie en de biostatistiek zijn binnen ons instituut geïntegreerd tot een succesvol geheel. Dat is mede aan jou te danken.

*Mijn directe collegae van de biostatistiek groep:* Lidia, Bettina, Wim, Jeanine, Paul, Simone en Maria. Hoewel wij allen zeer druk en soms individualistisch bezig zijn, vormen wij toch een hechte groep met een prettige en hartelijke sfeer. Het is een plezier om met jullie samen te werken en het enthousiasme voor ons mooie vak te delen.

*De leiders van de andere onderzoekslijnen van het Instituut Epidemiologie en Biostatistiek:* De contacten tussen jullie groepen en de biostatistiek worden steeds intensiever. Ik stel dat zeer op prijs en hoop dat onze samenwerking in de toekomst verder uitgebouwd zal worden.

*Lieve Coby, Frank, Klaas-Jan, Erik en Christiaan:*

Biostatistiek is een mooi vak, en het werk dreigt mij soms meer op te slokken dan goed is. Toch valt het belang van mijn werk geheel in het niet bij het hebben van een harmonieus en gelukkig leven met elkaar. En dat geven jullie mij.

Ik heb gezegd.

**Referenties**

1. Vandenbroucke JP. De opkomst van de medische statistiek en de epidemiologie in het klinisch wetenschappelijk onderzoek van de afgelopen eeuw. (1999) Nederlands Tijdschrift voor Geneeskunde 1999;143:2625-8
2. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. British Medical Journal 1948;ii:425-429.
3. Hill AB. *A short textbook of medical statistics*. (1984) 11th edn, Hodder and Stoughton, London.
4. H. de Jonge. Inleiding tot de Medische Statistiek. Deel 1: Fundamentele begrippen en technieken, verdelingsvrije methoden. (1963) Wolters-Noordhoff NV, Groningen.
5. H. de Jonge. Inleiding tot de Medische Statistiek. Deel 2: Klassieke methoden. (1964) Wolters-Noordhoff NV, Groningen.
6. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. Journal of the American Statistical Association 1958; 53:457-181.
7. Cox DR. Regression analysis and life tables (with disussion). Journal of the Royal Statistical Society B 1972; 34:187-220.
8. Good Clinical Practice. ICH Harmonised Tripartite Guideline (1996)
9. Statistical Principles for Clinical Trials. ICH Harmonised Tripartite Guideline (1998)
10. Houwelingen JC van. (1987) Het Statistisch Instrumentarium. (Oratie)
11. Strik R van. (1996) Lof der Biostatistiek. (Afscheidscollege)
12. Kleinbaum DG, Kupper LL, Muller KE. Applied Regression Analysis and Other Multivariate Methods (1988) PWS-Kent Publishing Company, Boston.
13. Hosmer DW, Lemeshow S. Applied Logistic Regression (1989). John Wiley & Sons, New York.

14. Klein JP, Moeschberger ML. (1997). *Survival Analysis. Techniques for Censored and Truncated Data*. Springer-Verlag, New York.
15. Pierik FH, Vreeburg JTM, Stijnen T, Roijen JH van, Dohle GR, Lameris JS, Timmers T, Weber RFA. Improvement of sperm count and motility after ligation of varicoceles detected with colour Doppler ultrasonography. *International Journal of Andrology* 1998; 21: 256-260.
16. Verbeke G, Molenberghs G (editors). *Linear Mixed models in Practice. A SAS Oriented Approach*. (1997) Springer Verlag, New York.
17. Balding D. *Bioinformatics*. RSS News 2000; Volume 27, No 10.
18. Houwelingen JC van. Dwalingen in de methodologie. III. Randomisatie op het niveau van behandelaars. *Nederlands Tijdschrift voor Geneeskunde* 1998;142:1662-5.
19. Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. *British Medical Journal* 1998; 317:1171-1172.
20. Goldstein H. *Multilevel Statistical Models. Second Edition* (1995). Arnold, London.
21. Houwelingen JC van, Arends LR, Stijnen Th. Advanced methods in meta-analysis: multivariate approach and meta-regression. Submitted.
22. Arends LR, Hoes AW, Lubsen J, Grobbee DE, Stijnen T. Baseline risk as predictor of treatment benefit; three clinical meta-re-analyses. *Statistics in Medicine* 2000;19:
23. Stijnen Th, Arends LR. Dwalingen in de methodologie. XVI. Wat te doen met ontbrekende waarnemingen? *Nederlands Tijdschrift voor Geneeskunde* 1999; 143:1996-2000.
24. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American Journal of Epidemiology* 1995 ;142:1255-64.
25. Schafer JL. *Analysis of Incomplete Multivariate Data* (1997). Chapman & Hall, London.
26. SOLAS for Missing Data Analysis (1997) Statistical Solutions Ltd, Cork.

27. J Martin Bland, Douglas G Altman. Bayesians and frequentists. *British Medical Journal* 1998;317:1151-1160.
28. Gilks WR, Richardson S, Spiegelhalter DJ. (Eds) (1996) *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
29. EAST. A Software Package for the Design and Interim Monitoring of Group Sequential Trials. Cytel Software Corporation
30. PEST 4.0 Software Package for the Planning and Evaluation of Sequential Trials (2000). Medical and Pharmaceutical Statistics Research unit, University of Reading.
31. Whitehead J. *The Design and Analysis of Sequential Clinical Trials. Second Edition* (1992), Ellis Horwood, London.
32. Jennison C, Turnbull BW. *Group Sequential methods with Applications tot Clinical Trials* (2000) Chapman & Hall/CRC.