

Prof. dr. Kit C.B. Roes is per 1 september 2009 benoemd tot hoogleraar Biostatistiek aan het Universitair Medisch Centrum Utrecht.

Hij studeerde Wiskunde aan de Katholiek Universiteit Nijmegen (nu Radboud Universiteit). Hij is zijn carrière gestart in de industriële statistiek bij Philips en heeft dit voortgezet bij het academische adviesbureau IBIS UvA BV. Op basis hiervan is hij in 1995 gepromoveerd in Amsterdam.

Hij is nu meer dan 15 jaar werkzaam in biostatistiek, zowel in een academische omgeving als in onderzoek en ontwikkeling in de farmaceutische industrie. Binnen Wageningen Universiteit en Research Centrum heeft hij de onderzoeksafdeling Biometrie geleid en in "Biometris" deze afdeling samengebracht met de leerstoelgroep Wiskunde en Statistiek van Wageningen Universiteit. Bij het farmaceutische bedrijf Organon (inmiddels MSD) was hij wereldwijd verantwoordelijk voor biostatistiek in research en development (in Nederland, Amerika en Japan), en tevens voor clinical data management en medical writing.

Sinds september geeft hij met collega's van het voormalige Centrum voor Biostatistiek en de onderzoekslijn Theoretische Epidemiologie en Biostatistiek van het Julius Centrum voor Gezondheidswetenschappen en Eerstelijns Geneeskunde vorm aan onderzoek, advisering en onderwijs in de Biostatistiek. Opzet en analyse van klinische studies, causale inferentie en het raakvlak tussen statistiek en ethiek hebben zijn speciale interesse.



Biostatistiek: Leven in variatie

Oratie dr. Kit C.B. Roes

Biostatistiek: Leven in variatie

Inaugurele rede uitgesproken bij de aanvaarding van het ambt van hoogleraar in de Biostatistiek aan de faculteit der geneeskunde van de Universiteit Utrecht op 15 juni 2010 door dr. Kit C.B. Roes.

Inhoudsopgave

1. Variatie maakt het interessant	4
2. Variatie	5
2.1 Een begin van denken in variatie	5
2.2 De klinische studie	9
3. Variaties voor de toekomst in klinisch onderzoek	13
3.1 De schaal van klinisch onderzoek	13
3.2 Onderzoek naar nieuwe methoden	14
3.3 Interim analyses in klinische studies	15
3.4 Complexe afwegingen voor een DSMC	20
3.5 Een stap verder: Adaptieve klinische studies	23
4. Statistiek en ethiek	25
5. Er meer dan gerandomiseerd klinisch onderzoek:	
Data revoluties	26
6. Biostatistiek en context	29
7. Woorden van dank	32
8. Referenties	36

Mijnheer de Rector Magnificus, leden van de Raad van Bestuur van het UMC Utrecht, zeer gewaardeerde toehoorders,

1. Variatie maakt het interessant

“Een mens moet aan statistiek doen om geld te verdienen, maar je kunt toch veel beter echt interessante wiskunde doen?”. Dat waren de woorden van de (destijds) zeergeleerde dr. Kortram van de Universiteit Nijmegen, toen ik hem meldde dat ik van plan was de afstudeerrichting Mathematische Statistiek te kiezen. Het tekent mijn basisopleiding in de zuivere wiskunde, die ik als essentieel en inspirerend heb ervaren, maar waar ik ook wel aan kan terugdenken als een “esoterisch gebied van exactheid”. In die abstracte wiskundige wereld lijkt alles te kloppen, en het ontlokt gevoelens van schoonheid dat je iets sluitend en beknopt kunt bewijzen. Je wordt niet afgeleid door het gedrag van echte data, laat staan het gedrag van levende wezens. Biostatistiek wordt juist gedreven door de uitdagingen die de toepassing ervan biedt. Dan gaat het om echte problemen, waarover echte data beschikbaar zijn. Het wordt dan wel minder “schoon”; het is niet voor niets dat statistici het onderling niet altijd eens zijn over de conclusies die uit echte data getrokken kunnen worden. Een probleem gegoten in een model is onderdeel van de wiskunde en dan valt veel te bewijzen. Welk model van toepassing is op een gegeven probleem waarvan echte data voorhanden zijn staat wel open voor discussie en debat, en dat steeds vaker tot in de rechtszaal toe. De echte data maken het spannend, en des te spannender als het doel is effecten of bijwerkingen te onderzoeken van nieuwe geneeskundige behandelingen, te speuren naar de

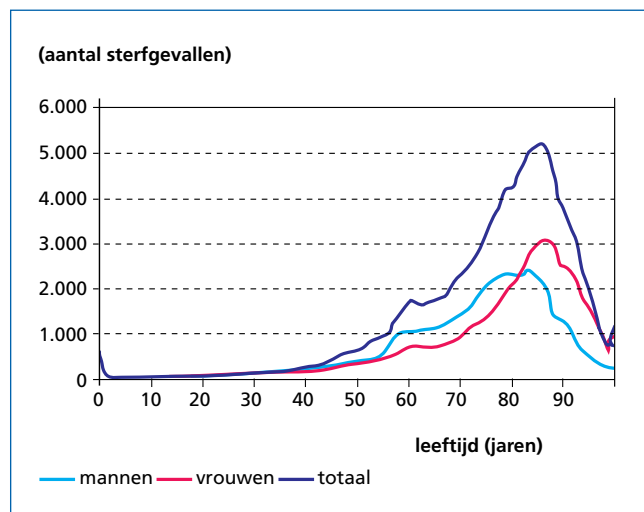
moleculair-genetische basis van belangrijke aandoeningen of breder de mechanismes te ontrafelen die aan biologische processen ten grondslag liggen. De leerstoel die ik vandaag met groot plezier aanvaard staat midden in dergelijke toepassingen van groot maatschappelijk belang. Dat ik u nu zo kan toespreken heeft te maken met mijn eigen voorliefde voor variatie, of afwisseling zo u wilt. Dat zal degenen die mij al wat langer kennen, niet vreemd voorkomen. Ik heb dan ook niet het typische profiel dat u van een hoogleraar biostatistiek zou verwachten. De geselecteerde schets die ik geef is geïnspireerd door de invloed van statistiek die ik in mijn gevarieerde loopbaan heb ervaren en mede heb mogen vormgeven.

2. Variatie

2.1 Een begin van denken in variatie

Het wetenschapsgebied statistiek omvat de opzet van experimenten, en op basis van de gegevens beschrijven, verklaren en voorspellen in een wereld vol variatie. Statistiek vergt denken en wiskundig modeleren in toeval, onzekerheid en kansen. Dat doet een mens van nature niet erg goed. We zien snel patronen: een AEX index die drie achtereenvolgende dagen stijgt wordt uitgebreid van verklaringen voorzien, net zoals de daling (dat heet dan ‘correctie’) die erna op volgt. Maar de handel alleen al, zonder systematische trend in het gedrag, kan voor dergelijke schommelingen zorgen: veel is gewoon ruis. Die neiging tot snelle patroonherkenning is ongetwijfeld evolutionair bepaald. Het was in de natuur zeer nuttig om bij mogelijke patronen van gevaar het zekere voor het onzekere te nemen en snel te handelen. In de complexiteit van

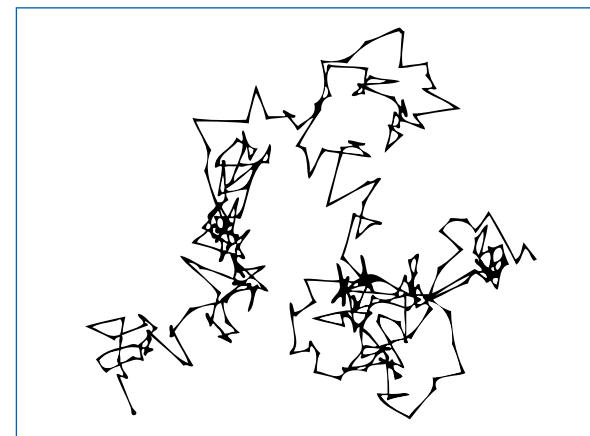
vandaag kan het tot discutabele beslissingen leiden, zoals de zaak van Lucia de Berk heeft laten zien. De volgende twee voorbeelden geven een eenvoudige basis voor het modelleren van variatie. Allereerst de “wetmatigheid” in de variatie van tijdstip van overlijden van mensen. In Shewhart’s¹ woorden: “We come into this world and from that very instant on are surrounded by causes of death seeking our life”.



Figuur 1. Frequentieverdeling sterfte in Nederland 2007

Op individueel niveau is exact en langere tijd vooruit voorspellen van het moment van overlijden een hachelijke, en mijns inziens principieel onmogelijke, zaak – tenzij men er zelf de hand in neemt. Kijken we naar groepen, of dat nu hele populaties zijn of groepen die bepaalde risicofactoren dragen, ziet de frequentieverdeling er stabiel uit – en “constant systeem van oorzaken” (Figuur 1). Dit “constante systeem van oorzaken” maakt voorspellen binnen marges van onzekerheid mogelijk.

Het tweede voorbeeld is gebaseerd op de waarnemingen van de Engelse botanist Brown. In 1827 zag deze onder zijn microscoop kleine deeltjes stuifmeel in suspensie bewegen alsof ze leefden. Ook hier kunnen we exact voorspellen van de volgende positie vergeten, maar de hele beweging is beheerst en dus voorspelbaar binnen grenzen. Elke volgende sprong van het deeltje is blijkbaar onderwerp van hetzelfde “constante systeem van oorzaken”, en heeft geen geheugen voor de plek of vorige sprong. Het systeem van oorzaken is botsing met vele deeltjes. Kijken we bijvoorbeeld naar de verplaatsing in links-rechts richting in 1 stap (Figuur 2), dan zal deze zich (bij benadering) gedragen volgens de bekende normale verdeling: gemiddeld rond 0, en sterk afnemende kans om grotere sprongen te maken.



Figuur 2. De Brownse beweging (cf. Shewhart¹)

Het wiskundige model van deze “Brownse beweging” is fundamenteel gebleken voor het modelleren van *stochastische processen*: variatie van processen in de tijd.

Deze voorbeelden komen niet uit een introductieboek Biostatistiek, integendeel. Ze zijn ontleend aan “Economic Control of Quality of Manufactured Product” van W.A. Shewhart uit 1931¹. Een baanbrekend boek dat de basis vormt voor het hedendaagse, op statistische concepten gebaseerde kwaliteitsmanagement. We kennen die aanpak nu als Zes Sigma (Six Sigma).

W.E. Deming⁴, statisticus en leerling van Shewhart, heeft Shewhart’s ideeën verbreed tot een management aanpak die revolutionaire invloed heeft gehad op de Japanse industrie. Dat was hem gedurende de Tweede Wereldoorlog bij de productie van oorlogsmaterieel in Amerika niet gelukt. Vanaf de jaren 80 van de vorige eeuw heeft de aanpak ook op de Westerse wereld veroverd, in alle sectoren. De kern van de aanpak is het continue streven naar verbetering door het reduceren van variatie, zowel variatie veroorzaakt door toewijsbare oorzaken (“incidenten”) als het door het systeem van “constante oorzaken”. Het modelleren en vervolgens reduceren van variatie in het systeem is gebaseerd op de statistische principes van experimenteren – daar heeft Fisher de basis voor gelegd³. Het succes van de aanpak heeft dus wortels in het statistisch denken en profiteert tot op de dag van vandaag van voortdurende innovatie in statistische methodologie. Maar dat laatste is niet genoeg, anders had Deming in Amerika in de Tweede Oorlog niet gefaald. Succes is ook bepaald door het brengen van het statistisch denken in de juiste context, in dit geval als *de* consistente en door het top management gedragen manier van continu verbeteren. Een succesvol statisticus voldoet dan ook bij voorkeur aan het T-profiel, waarbij de poot de diepte van de eigen vakexpertise is en de dwarsbalk de noodzakelijke vakkennis van het veld waar de statisticus in werkt. Die combinatie van kennis is de basis voor een bijdrage op hoog niveau.

Zie hier de impact van statistiek als het goed wordt ingebed: zonder zou productie van veel geavanceerde producten economisch niet mogelijk zijn. Hetzelfde is het geval bij het leveren van diensten van hoge kwaliteit, op basis van *Lean Six Sigma*. Dat geldt ook voor de zorg: diverse ziekenhuizen in Nederland – waaronder het UMC Groningen – hebben Lean Six Sigma succesvol omarmd als hun verbeteringsaanpak⁸.



Figuur 3. Het T-profiel

2.2 De klinische studie

In klinisch onderzoek is goed begrip van variatie letterlijk van levensbelang. Nieuwe geneesmiddelen worden getest en toegelaten op de markt op basis van resultaten van klinische studies. Maar dat geeft geen absolute garanties: resultaten van klinische studies hebben ook te maken met variatie en onzekerheden. Middelen die op het moment van registratie effectief, waardevol en voldoende veilig werden bevonden, blijken na een aantal jaren beschikbaarheid onacceptabele veilig-

heidsrisico's te geven. De cardiovasculaire risico's van Vioxx (rofecoxib), een COX-2 inhibitor gebruikt bij reumatische pijnklachten, dwongen Merck het product van de markt te halen. Min of meer dezelfde vragen dringen zich op bij andere recente problemen met de veiligheid van geneesmiddelen: de associatie van het gebruik van anti-depressiva (SSRIs) met vergrote neiging tot het plegen van zelfmoord en cardiovasculaire risico's bij het gebruik van Avandia (rosiglitazone), een middel voor de behandeling voor diabetes.

Laten we voor het begrip van deze onzekerheden naar het principe van de klinische studie kijken. Stel, u wilt de effectiviteit van een nieuwe behandeling bepalen, bijvoorbeeld tibolone voor de preventie van osteoporose bij post-menopausale vrouwen (de LIFT studie⁹). Het doel is de effectiviteit te onderzoeken op het voorkomen van wervelfracturen bij een follow-up van 3 jaar. Onderzoeken is vergelijken, dus in een gerandomiseerde klinische studie richt u dit als volgt in. Een groep patiënten wordt in de loop van de tijd voor de studie gerekruteerd. In dit geval is het plan de helft van de patiënten te behandelen met tibolone en de andere helft met placebo, een nepmiddel. De toedeling van patiënten aan behandeling wordt door het lot bepaald: gerandomiseerd. In zijn eenvoudigste vorm is dit equivalent met bij elke nieuwe patiënte die zich aandient een muntje opgooien, bij kop wordt het tibolone en bij munt placebo. Het is ook nog zo ingericht dat de behandelende arts/onderzoeker niet weet of het kop of munt is geworden en de patiënte ook niet. Dat is ook niet te zien aan de vervolgens toegediende behandeling: het is dubbelblind. Ik ga op dit punt niet in op de ethische aspecten van randomiseren en de ongemakkelijke gedachte voor de patiënte dat het lot en niet medisch inzicht de behandeling bepaalt die zij ontvangt. De gerekruteerde patiënten zijn

afkomstig van zorgvuldig geselecteerde klinieken en onderzoekers, die bereid zijn geweest mee te doen. Ze vormen dus doorgaans niet een representatieve steekproef uit de (totale) patiëntenpopulatie die behandeld wordt. Wat ons dan rest voor een betrouwbare en valide conclusie is de randomisatie binnen de gerekruteerde groep. Daarmee zorgen we ervoor dat de twee behandelde groepen *gemiddeld* vergelijkbaar zullen zijn wat betreft de karakteristieken die kunnen samenhangen met het beloop van hun ziekte – in dit geval het ontstaan van osteoporotische wervelfracturen. Een aangetroffen verschil in resultaat tussen de twee groepen kunnen we dan toewijzen aan het verschil in behandeling (Tabel 1). Centraal in de interpretatie van resultaten staan het geschatte effect, de p-waarde en het betrouwbaarheidsinterval.

Eindpunt	Tibolone (N=2249) Aantal	Placebo (N=2257) Aantal	Relatieve hazard (95% betrouwbaarheidsinterval)	p-waarde
Nieuwe wervelfracturen	70	126	0.55 (0.41 - 0.74)	<0.001

Tabel 1. Resultaten effectiviteit LIFT studie⁹

Het geschatte effect, relatieve hazard van 0.55, duidt op een bijna halvering van het risico op nieuwe wervelfracturen, ten opzichte van placebo. Dat is een schatting, maar hoe zeker zijn we nu van onze zaak? De p-waarde drukt de kans uit op het waarnemen van een dergelijk groot verschil (effect), als er in werkelijkheid geen verschil is – de kans dat dit een vals positieve bevinding is. Als deze kans klein is, dan zal de nulhypothese van geen verschil wel niet kloppen – en vinden we dat een verschil is aangetoond. Dit gehele concept hebben

we aan Fisher te danken³. De hele wereld hanteert de norm 5% als kleine kans (p-waarde), overigens zonder enige wetenschappelijke basis en zo had Fisher het niet bedoeld. De p-waarde is hier minder dan 1 promille, we zijn dus behoorlijk zeker van onze zaak. Het 95% betrouwbaarheidsinterval geeft een range van ongeveer 40% tot 75%. Een overtuigend effect. Maar niet 100% zeker, dat bestaat niet gegeven de onvermijdelijke variatie.

Een drempel van 5% maakt het wel erg zwart/wit: haal je de drempel ben je binnen (ook voor je publicatie), haal je hem niet – pech gehad. Al probeert men soms de suggestie te geven een waarde te kiezen die past bij de uitgangssituatie, de 5% ‘terreur’ blijft hardnekkig – ondanks de gesuggereerde toegenomen nadruk op betrouwbaarheidsintervallen. Voor de Amerikaans Food and Drug Administration is dit zelfs in een wettelijk kader vastgelegd. Gij zult twee studies hebben die deze drempel op het primaire eindpunt gehaald hebben.

Er is debat mogelijk over de interpretatie dat de groepen onderling “*gemiddeld vergelijkbaar*” zijn als gevolg van randomisatie. In een precies wiskundig raamwerk kan u uitgelegd worden (of zelfs uitgelegd zijn) dit te zien als: “gemiddeld over het heel vaak herhalen van hetzelfde experiment in dezelfde omstandigheden”. Door Senn¹¹ is dit de “marathon experimenter” genoemd, en al houd ik van hardlopen – hier kan ik moeilijk mee uit de voeten. Het is praktisch weinig bruikbaar. Zeker klinische studies zijn vaak unieke gebeurtenissen. Als de groepen in de studie niet al te klein zijn, kunt u wat mij betreft “gemiddeld” ook bij benadering als gemiddeld binnen de studie opvatten. Dan drukt dus de 95% van het betrouwbaarheidsinterval de betrouwbaarheid uit, gegeven de diverse variantie componenten in de studie (in een

zeker te postuleren model)¹¹. En al te kleine klinische studies moet u volgens mij helemaal niet doen – dus dan zit dit wel goed.

Het gebruik van het instrument van randomisatie gaat verder dan onderlinge vergelijkbaarheid. Het resultaat van de vergelijking van behandelingen kan causaal geïnterpreteerd worden: de behandeling is de oorzaak van het gevonden verschil, naar alle waarschijnlijkheid – er blijft immers onzekerheid. Dat is uitermate krachtig; in niet gerandomiseerd onderzoek is de vraag van causaliteit doorgaans niet sluitend te beantwoorden op grond van statistische analyse alleen.

En al mag een gerandomiseerde klinische studie een “grof instrument” lijken, het geeft een hoge standaard van “bewijs”, het maakt medische interventies evidence based. Ik wens me weleens dat andere terreinen – het basis en voortgezet onderwijs bijvoorbeeld – vergelijkbare standaarden zouden toepassen op effectiviteit van nieuwe interventies. Dat zou ongetwijfeld veel kommer en kwel schelen.

3. Variaties voor de toekomst in klinisch onderzoek

3.1 De schaal van klinisch onderzoek

Vandaag de dag is de hoeveelheid klinische studies onvoorstelbaar groot. Wat statistieken (in de klassieke zin, zoals u ook in de sport gewend bent): Op www.clinicaltrials.gov stonden op 28 april 2010 88.856 studies met locaties in 172 landen geregistreerd, waarvan 33.355 studies nog rekruteerden. Het is al met al een veilige schatting dat op enig moment miljoenen mensen wereldwijd participeren in

klinische studies. Deze schaal en het ingrijpende karakter van een klinische studie voor patiënten leggen een grote verantwoordelijkheid bij alle disciplines die verantwoordelijk zijn voor opzet, uitvoering, analyse, rapportage, publicatie en gebruik van de resultaten van klinische studies. Onderzoek naar nieuwe statistische methoden voor de opzet van klinische studies wordt sterk gedreven – en uitgevoerd – door de farmaceutische bedrijven. Zij hebben dan ook de meerderheid van biostatistici in dienst – en dat wereldwijd. De bijdrage aan en sturing van dit onderzoek vanuit de academische wereld moet (veel) groter worden, en wel om de volgende redenen. De integrale verantwoordelijkheid van academische instellingen voor het klinisch onderzoek, naar geneesmiddelen en andere interventies, is enorm toegenomen en zal blijven toenemen in de komende decennia. Ten tweede zal meer statistisch onderzoek naar opzet en analyse van klinisch onderzoek door academische groepen een verdere verrijking betekenen. Onderzoeksvraagstellingen en oplossingen worden ook bepaald door het perspectief dat men heeft. En dat perspectief is voor farmaceutische bedrijven, registratieautoriteiten en academische instellingen niet gelijk.

3.2 Onderzoek naar nieuwe methoden

Verbetering van statistische methoden voor opzet en analyse van onderzoek leidt, vereenvoudigd bekeken, tot enerzijds een juister, zuiverder interpretatie van de gegevens en anderzijds tot een kleinere onzekerheid. Het eerste punt draagt bij aan een beter inzicht in de biologische mechanismen en de feitelijke effecten. Het tweede punt, reduceren van de onzekerheid, is een belangrijke basis voor het kunnen reduceren van het aantal benodigde proefdieren in dieren-

experimenteel onderzoek en het aantal proefpersonen in klinisch onderzoek. Daarmee worden ook de kosten en doorlooptijd van onderzoek gereduceerd. Nieuwe behandelingen komen sneller ter beschikking. Vaak leidt betrokkenheid van een statisticus ook tot verbetering van onderzoek op andere punten: verscherping van de vraagstelling, verbetering van de uitvoerbaarheid en het terugdringen van onnodige gegevensverzameling.

Mijn belangrijkste drijfveer voor verdere ontwikkeling van de statistiek is deze impact op kwaliteit van onderzoek, processen en kwaliteit van leven. Die impact wordt bepaald door de innovativiteit van de methoden, maar ook door de wijze waarop (bio)statistici methodologie en statistisch denken weten in te bedden in de werkwijze van vele disciplines.

3.3 Interim analyses in klinische studies

Het boeiende van statistiek is dat nieuwe methoden en technieken ontwikkeld voor het ene toepassingsgebied ook voor ander toepassingen relevant blijken. Interim analyses en sequentiële analyses van klinische studies zijn daar een aansprekend voorbeeld van. Die hebben een startpunt in de industriële statistiek. Statistische methoden voor het beheersen van productieprocessen kunnen efficiënter gemaakt worden door gebruik te maken van het feit dat gegevens in de loop van de tijd worden waargenomen: het gebeurt *sequentieel*. Dit stelt de onderzoeker (of operator, of...) in staat op elk tijdstip alle gecumuleerde gegevens tot dan toe te gebruiken. Een verschil dat consequent bestaat in de loop van de tijd wordt op deze manier sneller opgepikt, dan wanneer een vaste periode wordt afgewacht. Dit geeft dus direct

efficiency winst. De pionier voor het ontwikkelen van sequentiële statistische analyse is Wald¹², die overigens hevig werd bekritiseerd door Fisher^{13,14}, omdat het naar Fisher's inzicht geen goed experimenteel design betrof. George Barnard¹⁶ heeft verdere theoretische onderbouwing gegeven, gebaseerd op wiskundige modellering van stochastische processen – de eerder geïntroduceerde Brownse beweging. Peter Armitage (geboren in 1924), wordt erkend als een van de grondleggers van de hedendaagse medische statistiek. Hij was gedurende de Tweede Wereldoorlog werkzaam voor de afdeling Wapen Inkoop van het Engelse Ministerie van Bevoorrading onder de leiding van diezelfde George Barnard. Hij werkte daar onder andere aan problemen op het vlak van kwaliteitsbeheersing en steekproef inspectie. Zoals gezegd was dit het type probleem waar sequentiële methoden bij uitstek geschikt zijn en werden toegepast. Na de oorlog aanvaardde Armitage een positie bij de Engelse Medical Research Council, in een periode waar gerandomiseerde klinische studies op steeds grotere schaal werden toegepast. Ook klinische studies ontrollen zich in de tijd: patiënten worden achtereenvolgens gerekruteerd en voor kortere of langere tijd gevolgd. Doorgaans is een klinische studie een kwestie van jaren. Het is dan ook niet verbazend dat Peter Armitage¹⁰ voor deze studies sequentiële methoden heeft ontwikkeld. Hij was daar zijn tijd – in 1960! – ver mee vooruit.

Het belang van het tussentijds kunnen analyseren van grote klinische studies lijkt evident. Als er al eerder duidelijk bewijs is dat de behandeling effectief is, is het verantwoord te stoppen en de behandeling ter beschikking te stellen. Er kunnen zich grote veiligheidsproblemen aftekenen, die doorgaan niet ethisch maken. Of er ontstaan in dezelfde tijd wezenlijke andere inzichten over de effectiviteit en veiligheid op

basis van andere studies. Om tussentijds beslissingen te kunnen nemen, zijn *interim analyses* noodzakelijk. In principe wordt bij een interim analyse voor effectiviteit dezelfde vergelijking tussen de behandelingen gemaakt als aan het einde van de studie, maar dan op basis van de tussentijds beschikbare data. Ook dat kunnen we statistisch toetsen, bij een zeker significantie niveau; maar het toeval, de variatie compliceert de zaak. Herhaald analyseren, herhaald toetsen heeft consequenties voor het uiteindelijke significantie niveau in de studie, en dus de feitelijke betrouwbaarheid van de uitkomst. Het 'casino van het leven' heeft hier zijn effect. Als bij elke interim analyse getoetst wordt tegen het bekende 5% significantie niveau, loopt de kans dat er bij een van die analyses ten onrechte een significant verschil gevonden wordt – wanneer er eigenlijk geen sprake is van een verschil – snel op met het aantal interim analyses (Tabel 2). De oplossing is het significantieniveau per interim analyse aan te passen, zodat het overall significantie niveau behouden blijft. Dus tussentijds strenger zijn: het verschil tussen behandelingen moet statistisch beduidend overtuigender zijn om vroegtijdig te stoppen. De theorie hiervoor is, in geval van 1 duidelijk eindpunt, goed ontwikkeld en maakt flexibele opzetten mogelijk^{17,18}. Het significantie niveau, ook wel α , kan verdeeld worden over de loop van de studie – en totaal op 5% blijven komen. In dergelijke opzetten kunnen interim analyses leiden tot stoppen vanwege bewezen effectiviteit, maar desgewenst ook stoppen vanwege futiliteit: een (te) kleine waarschijnlijkheid dat de studie – indien gecontinueerd – aan het eind nog een positief resultaat laat zien.

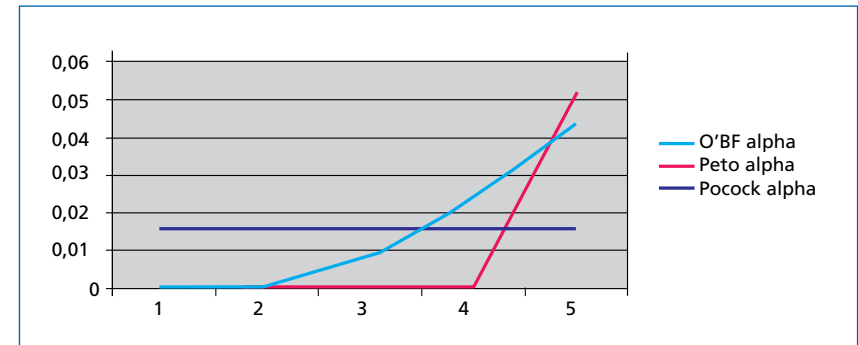
Totaal aantal interim "looks" (gelijkverdeeld)	Kans op verschil aantonen onder nulhypothese van geen verschil
1	5%
2	8%
3	11%
4	13%
5	14%
10	19%

Tabel 2. Consequenties van naïeve interim analyses

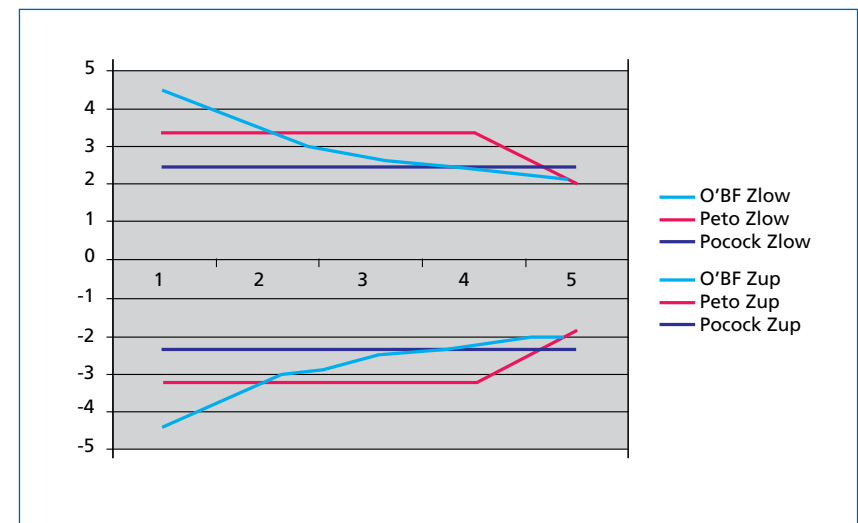
Voorbeelden van populaire regels om de α te verdelen over de loop van de studie staan in Figuur 4.

Het verwachte uiteindelijke aantal patiënten dat nodig is voor een studie met interims zal kleiner zijn, vergeleken met een studie zonder interim analyses. Deze designs zijn dus inderdaad gemiddeld efficiënter dan klassieke designs

Armitage baseerde zich voornamelijk op frequente, bijna continue evaluatie (dus na elk paar patiënten). Pocock¹⁹ heeft met de *group sequential* benadering de aansluiting verzorgd met wat er vaak feitelijk gebeurt in een klinische studie. Een *Data and Safety Monitoring Committee (DSMC)* evalueert de data in intervallen en zal dus de interim analyses groepsgewijs willen doen.



Figuur 4a. Polulaire - spending regels en bijhorende grenzen



Figuur 4b. Polulaire α - spending regels en bijhorende grenzen

3.4 Complexe afwegingen voor een DSMC

Uit het voorgaande zou men kunnen afleiden dat het plannen van interim analyses, en het uitvoeren van hun uitkomst relatief eenvoudig kan worden geïmplementeerd. Men beschrijft de criteria zorgvuldig in het studie protocol, en de statisticus voert de analyses op de geplande tijd uit. De stopregel(s) bepalen vervolgens de actie: doorgaan of stoppen. Een computer kan “de was doen”. In werkelijkheid is het niet zo simpel. Dat heeft te maken met de totaliteit van data en informatie die een DSMC in haar overwegingen mee moet nemen, zoals het volgende

		Progressie naar ARC / AIDS / Sterfte		Sterfte		p-waarde	
Meeting	Subject-jaren	Onmiddelijk	Uitgesteld	Onmiddelijk	Uitgesteld	Progressie	Sterfte
1	-	-	-	-	-	-	-
2	?	1	4	0	0	0.22	-
3	443	7	13	0	0	0.19	-
4	-	-	-	-	-	-	-
5	920	24	34	0	0	0.19	-
6	1711	56	74	14	7	0.12	0.13
7	2181	76	104	25	14	0.037	0.081
8	3042	114	148	29	24	0.036	0.50
9	-	-	-	-	-	-	-
10	4258	202	227	59	45	0.23	0.11
11	5328	263	284	95	76	0.37	0.15

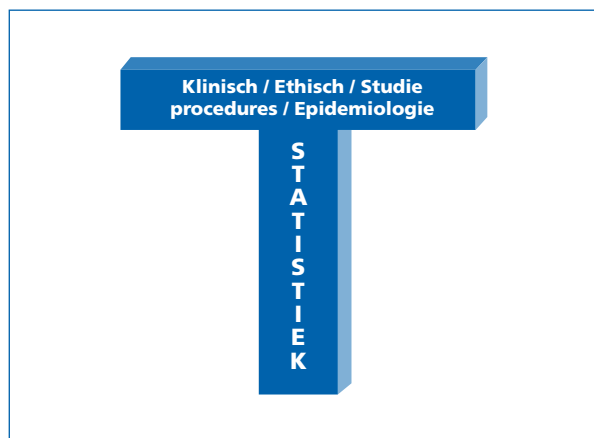
Tabel 3. Data beschikbaar voor de Concorde DSMC²¹

voorbeeld laat zien. De Concorde studie²¹ betrof een gerandomiseerde klinische studie onder mensen die weliswaar HIV besmet waren, maar nog symptoombvrij. In deze studie werd voor die groep onmiddellijke behandeling met zidovudine (AZT) vergeleken met uitgestelde behandeling. Uitgesteld hield in dat deze groep eerst placebo ontving, tot het moment van optreden van symptomen: AIDS, AIDS gerelateerd complex (ARC), of voortdurende lage CD4 tellingen. Voor de Concorde studie zagen de belangrijkste interim resultaten gedurende de studie er als volgt uit (Tabel 3). De eerste, veel vaker voorkomende, moeilijkheid is de suggestie van een zich aftekende trends, maar in tegengestelde richting: onmiddellijke behandeling van symptoombvrije HIV patiënten lijkt een voordeel op te leveren wat betreft uitstellen van progressie, maar er is een suggestie van nadeel met betrekking tot sterfte (wit gearceerd).

De DSMC heeft in dit geval de moed gehad door te gaan, waardoor we de betekenis van variatie, “nature’s guessing game” kunnen zien: een trend die zich lijkt af te tekenen, zet zich niet vanzelfsprekend door. U kunt zich voorstellen dat de DSMC het besluit tot doorgaan niet licht genomen heeft, ongeacht de stopregels. Vele additionele analyses zijn uitgevoerd. In die analyses passeren alle aspecten de revue die analyse en interpretatie van resultaten van klinische studies kunnen bemoeilijken, maar dan in een “snelkookpan”: de invloed van ontbrekende gegevens en incomplete follow-up (die samen kan hangen met beloop of behandeling), mogelijke invloed van ongelijke randomisatie omdat patiënten aantallen nog relatief klein zijn, en het tegelijkertijd evalueren van verschillende eindpunten die onderling kunnen samenhangen.

Over de werkwijze van DSMCs is sinds begin jaren negentig van de vorige eeuw veel gepubliceerd en gediscussieerd^{22,23}. De afgelopen decennia hebben DSMCs wereldwijd ingang gevonden, en de statistische theorie van sequentiële klinische studies op effectieve wijze geïmplementeerd. Toonaangevende statistici hebben – naast het leveren van het statistische instrumentarium – een grote invloed gehad op de belangrijkste aspecten: samenstelling, openheid en geslotenheid van bijeenkomsten, onafhankelijkheid, evaluatie en rapportage. Een succesvolle DSMC statisticus heeft dus bij uitstek een “goed gevuld” T-profiel (Figuur 5). Met de complexiteit van de afwegingen is het ook duidelijk dat een DSMC niet veel heeft aan een statisticus die zich keurig “aan de (stop-)regels houdt”.

De statistische uitdagingen voor het analyseren en interpreteren van klinische studies komen juist bij data en safety monitoring bij elkaar: omgaan met ontbrekende gegevens en drop-out, samengestelde eind-



Figuur 5. T-profiel van de DSMC statisticus

punten, evalueren van meerdere, mogelijke samenhangende eindpunten en betere kwantitatieve afweging van effectiviteit en veiligheid (benefit/risk).

De ontwikkeling van statistische methoden voor data en safety monitoring zijn op deze vlakken niet af, en zal voor ons een belangrijk onderwerp van onderzoek zijn. Met het onderzoek van Ingeborg van der Tweel is er in het UMC Utrecht al een duidelijke basis. Op dit moment participeren wij (met op dit moment drie statistici) in ongeveer 15 DSMCs. Belangrijk werk, met naar mijn overtuiging positieve invloed op de kwaliteit van het uitgevoerde onderzoek. Ook een inspiratiebron en garantie dat ons onderzoek op dit gebied aansluit bij de praktijk.

3.5 Een stap verder: Adaptieve klinische studies

Het onderzoek zal zich ook richten op een van de “hot topics” in het ontwerp van klinische studies: “adaptive design”. Dat dit “hot” is wordt sterk gedreven door de farmaceutische industrie. De productiviteit van ontwikkeling van nieuwe geneesmiddelen laat al lange tijd te wensen over. Kort gezegd is het aantal werkelijke nieuwe geneesmiddelen veel te klein, gegeven de grote investeringen, en dus veel te kostbaar. Bovendien duurt het veel te lang. Een van de mogelijkheden tot verbetering wordt gezocht in de manier van opzetten en uitvoeren van klinische studies naar nieuwe geneesmiddelen. In zogenaamde “adaptive designs” is het op basis van interim analyses mogelijk aspecten van de studie aan te passen – zonder de validiteit van de resultaten aan te tasten. De aanpassingen waar u aan moet denken zijn onder andere: het aanpassen van de groepsgrootte voor het resterende

stuk van de studie, kiezen van een ander primair eindpunt, het discontinueren van een of meer van de onderzochte groepen of het beperken van de oorspronkelijke populatie tot een specifieke hoog-risico groep. Ik heb dit wel de “heilige graal” genoemd voor de klinisch onderzoeker. Een klinische studie kan gestart worden, lopende de rit op diverse punten worden aangepast en levert aan het eind toch valide resultaten. Tevens is het verwachte aantal patiënten dat behandeld wordt met een inferieur middel kleiner, voor bepaalde designs. Dit ethische aspect wordt zwaar aangezet door de voorstanders, ik ga daar nog op in. Ondanks dat de basistheorie (zowel frequentistisch als Bayesiaans, zie^{24,25,26}) elegant is, zijn er veel haken en ogen. Dergelijke studies zijn logistiek beduidend complexer om uit te voeren. Het is in veel gevallen allerm minst duidelijk hoe de uiteindelijke resultaten geïnterpreteerd moeten worden vanuit het perspectief van toekomstig te behandelen patiënten. En daar ging het toch om?. Dus zorg is geboden bij deze complexe designs. De recente draft guidance van de Amerikaanse Food and Drug Administration onderscheidt toepasselijk twee categorieën adaptive designs: (1) designs die we goed begrijpen en (2) designs die we niet zo goed begrijpen. De eerste categorie valt bijna geheel in de categorie group sequential designs die ik hiervoor introduceerde. Kortom, we begrijpen de echte adaptive designs nog niet zo goed – en nader onderzoek is nodig. We missen bovendien de praktische ervaring: er zijn nog maar zeer weinig werkelijk adaptieve studies afgerond waarop toelating van nieuwe geneesmiddelen is gebaseerd. We zitten dus in een fase van “experimenteren met experimenten”, die juist omdat het onderzoek met mensen betreft met grote zorgvuldigheid doorlopen moet worden. In ons onderzoek zal nadruk liggen op de mogelijkheden van adaptive

designs om klinische en wetenschappelijke vragen te beantwoorden, waarvoor traditionele designs niet geschikt zijn. Er moeten wetenschappelijk overtuigende redenen zijn om te kiezen voor een dergelijk design en de extra complicaties in uitvoering en interpretatie voor lief te nemen.

4. Statistiek en ethiek

Op een aantal punten heb ik al even aan de ethiek geraakt. Ethiek en integriteit treden nadrukkelijker op de voorgrond naarmate het wetenschappelijk proces van onderzoek en de daarmee samenhangende risico's en onzekerheden transparanter worden voor de maatschappij als geheel. En die transparantie neemt in de huidige open informatie maatschappij zeer snel toe, en dat moet ook. Dat is evident voor onderzoek met mensen, al worstelen we vanuit de wetenschap regelmatig met de mate en vorm van transparantie over die risico's en onzekerheden. Zoals gezegd, van nature hebben we moeite variatie en kansen te bevatten en dat is een grote drempel in de communicatie. Ethiek maakt het ook persoonlijk, met haar fundament in normen en waarden. Het wordt door betrokkenen te vaak verengd tot wat volgens regelgeving en wetten mag. De normen en waarden van de onderzoekers zelf zijn een even noodzakelijk kompas. Dit kompas is onmisbaar in een omgeving waarin onderzoek in complexe netwerken van publieke en private partijen plaatsvindt. Juist statistici hebben en voelen doorgaans in onderzoek vanuit ethisch perspectief een vakinhoudelijke grote verantwoordelijkheid. Door hun inbreng in de opzet van een studie of experiment en de aantallen patiënten of proefdieren, bepalen zij in sterk mate de “last” die geaccepteerd wordt om een

bepaalde onderzoeksvraag te beantwoorden. Het kan dus best voorkomen dat een statisticus in discussie met de klinisch onderzoeker woorden in de mond neemt als “ik vind deze studie met dit kleine aantal patiënten ethisch niet verantwoord.” En nu wordt zelfs een hele klasse van adaptive designs door statistici aanbevolen met een expliciet ethisch argument. Ik denk dat dit niet zo simpel ligt, en dat er nog een wetenschappelijk inhoudelijke brug te slaan is tussen statistiek en ethiek. Dat vraagt van statistici dieper inzicht in ethiek en bijbehorende perspectieven (zie bijv. Senn¹¹), en van ethici dieper begrip van het denken in en modelleren van variatie en onzekerheden. Voorwaar niet eenvoudig. Ik tref het in het Julius Centrum; ik deel de interesse hiervoor met mijn collega hooggeleerde Hans van Delden en de onlangs gepromoveerde ethica dr. Rieke van der Graaf. Dit interessante vraagstuk staat dus op de onderzoeksagenda.

5. Er meer dan gerandomiseerd klinisch onderzoek:

Data revoluties

Miljoenen deelnemers aan klinische studies leveren al een grote hoeveelheid onderzoeksgegevens op, maar er zijn twee ontwikkelingen al enige tijd aan de gang die in omvang veel grotere volumes aan data genereren.

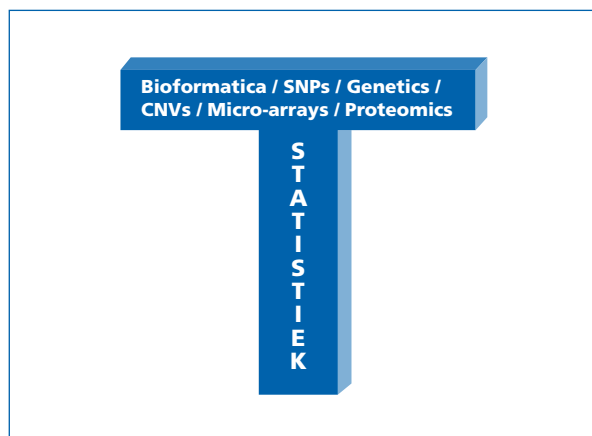
De eerste stille revolutie heeft een zeker “big brother” gehalte. De mogelijkheden om gegevens over gezondheid, behandeling, uitkomsten van operaties en leefstijl te registreren en te koppelen met behoud van privacy nemen snel toe. U bent allen potentieel donor van

onderzoeksgegevens, alleen zal men vanwege de privacy niet kunnen en niet willen ontdekken dat u dat ook daadwerkelijk was. Daar wordt de grootste zorg aan besteed en vraagt op zich al innovatie in processen en IT systemen. Dit alles levert een schatkamer – of misschien is op dit moment rommelkamer nog een beter woord – aan gegevens voor onderzoek. Deze gegevens zijn van groot belang; lang niet elke relevante vraag over effectiviteit en risico’s van behandelingen of preventie kan met een gerandomiseerd klinisch onderzoek onderzocht worden. Vanwege de potentiële onderlinge samenhang van mogelijke voorspellers en zogenaamde “verstremeling” of confounding, is analyse en interpretatie complexer. Innovatie in statistische methoden op dit gebied, ook wel aangegeven met causale inferentie, zijn in volle gang – zowel vanuit wiskundig vernieuwende aanpakken als de epidemiologische toepassingen. Wij zijn betrokken in meerdere samenwerkingsprojecten, onder andere met farmaceutische wetenschappen, waarin dit een belangrijke rol zal spelen.

Het tweede spoor betreft de kennisvermeerdering die wordt gedreven door nieuwe technologische ontwikkelingen. Beeldvormende en moleculaire technieken genereren grote hoeveelheden gegevens. Daarmee kan inzicht verkregen worden in ketens van processen, van onze genen, via eiwitten en alle tussenliggende niveaus tot ontwikkeling van ziekten. Naast inzicht levert dit ook betere mogelijkheden tot voorspellen en aanpassen van behandeling op individuele karakteristieken van de patiënt en zijn ziekte. De gegenereerde data levert een fundamenteel statistisch probleem: Het aantal experimentele eenheden is in deze gevallen vaak extreem veel kleiner dan de hoeveelheid variabelen. 1000 patiënten waarbij 100000 genetische of beeldvormende merkers

bepaald zijn is inmiddels geen uitzondering. Deze merkers vertonen bovendien complexe afhankelijkheden in de data – zelfs als er geen directe causale samenhang is – bepaald door evolutie en overerving. Dit is het onderzoeksgebied waar al minstens een decennium relatief de meeste academisch biostatistici op werkzaam zijn. Diverse nieuwe hoogleraren (bio) statistiek in Nederland hebben dit thema in hun oratie al flink aan bod laten komen. Wat dat betreft gaat het goed met het vak in Nederland; alleen al dit academisch jaar houden 6 nieuwe hoogleraren statistiek hun rede, waarvan 2 deze week.

Met de komst van dr. Rene Eijkemans naar Utrecht zijn we ook hier in staat een bijdrage te leveren. Die wordt gedreven door concrete samenwerking met klinische disciplines – en in moderne UMC Utrecht termen: speerpunten. Onze onderzoeks aandacht zal sterk gericht zijn op de keuze, kwaliteit en validiteit van statistische modellen. Het gebied wordt belaagd door veel vals positieve bevindingen, waarop soms zelf



Figuur 6. T-profiel van de "omics" statisticus

klinische studies worden gestart die vervolgens stopgezet moeten worden als de "misser" aan het licht komt.

Het verbeteren van de kwaliteit van modelkeuze en –validatie heeft het potentieel verspilling van geld en menselijke belasting tegen te gaan. De uitdaging voor de statisticus zit ook hier deels in het T-profiel. De techniek en inhoudelijke kennis ontwikkelen zeer snel en voor een relevante bijdrage moet een statisticus dit "bijbenen". Dit is een belangrijke reden waarom we binnen de statistiek steeds verdergaande specialisatie zien ontstaan.

6. Biostatistiek en context

Statistiek staat midden in de toepassingen en er is dus meer dan onderzoek alleen. Het samenwerken met andere onderzoekers is een kerntaak, zeker in de huidige omgeving waarin data en onderzoeksvraagstellingen complexer worden. Dat levert niet altijd meteen statistische vernieuwing op, maar dat zie ik dan ook niet als doel op zich.

We hebben een grote rol in onderwijs. Ik ben uitgesproken voorstander van het opleiden van velen, er kunnen wat mij betreft niet genoeg mensen zijn met enig verstand van statistiek. En dat is niet overal het geval, getuige de volgende anekdote. Enige tijd terug vertelde Bill O'Reilly (US vertegenwoordiger) op het Amerikaanse Fox News hoe vreselijk Amsterdam is: iedereen gebruikt drugs, het is een paradijs voor criminelen, het is er levensgevaarlijk, enzovoorts. Twee Amsterdammers maakten als antwoord hierop een filmpje met

wat feiten over Amsterdam, zie [the Truth About Amsterdam](#). Eén van die feiten was: 40,3% van de Amerikanen heeft cannabis geprobeerd, tegenover 22,6% van de Nederlanders. Het antwoord van O'Reilly: "The way they use statistics in The Netherlands is different. Plus, it's a much smaller country."

Er ligt wel een uitdaging op onderwijsvlak. De meeste in de zaal die ooit statistiek onderwijs hebben gehad, zullen niet als eerste associatie hebben dat het "leuk", "boeiend" of "interessant" was. We zijn al aan de slag om hier verbetering in te brengen. Belangrijke richtingen zijn ten eerste meer nadruk leggen op concepten dan formules. Mijn kinderen hebben geen idee wat DOS of machinetaal is en kunnen maar beperkt programmeren, maar slagen erin geavanceerde zaken te doen op hun computer. Dat moet met statistiek ook kunnen. Een tweede punt, dat ikzelf in bedrijfsomgeving al heel lang toepas, is "just-in-time" in plaats van "just-in-case". Het heeft in mijn ogen geen zin studenten met statistiek te overladen, als ze nog maar beperkt zicht hebben op het vak dat ze feitelijk studeren. Ze kunnen dit dan nog moeilijk plaatsen. Dus meer accent tegen het einde van de studie. Dat accent moet er dan wel zijn. Juist vanuit Geneeskunde Master studenten krijgen we de – op zich opmerkelijke – klacht dat ze te weinig statistiek krijgen!

Statistici worden natuurlijk ook zelf opgeleid, maar nog niet in belangrijke mate in Nederland. En er is een schreeuwend tekort. Recent is een prima initiatief gestart voor een zelfstandige Master opleiding Statistiek voor de Sociale en Levenswetenschappen als samenwerking tussen diverse universiteiten met een basis in Leiden. Een opleiding waarin de "poot" van mijn T-profiel goed verzorgd wordt. Utrecht

beschikt over een uitstekende Master opleiding Epidemiologie. Om opleiding van statistici in Nederland verder te versterken, hebben we het initiatief genomen een specialisatie Medische Statistiek aan deze Master toe te voegen. Daarmee wordt een stevige statistische poot toegevoegd aan studenten die daarvoor geschikt zijn en al een goede "dwarsbalk" van de T meebrengen. Ik denk dat dit een profiel medisch statistici zal opleveren dat uitstekend aansluit bij de vraag vanuit het klinische en epidemiologisch onderzoek.

Ik prijs me gelukkig met de start die ik heb kunnen maken in Utrecht. Er zijn interessante samenwerkingsprojecten op alle onderwerpen die ik besproken heb; sommige heb ik al genoemd. De diversiteit, de variatie, is groot. Het Escher project is voor mij een belangrijke brug. Dit project moet binnen Top Instituut Pharma de verbinding brengen tussen innovaties in het proces van geneesmiddelenonderzoek zelf en de regulering er van. Daar is voor mij directe samenwerking met het College ter Beoordeling van Geneesmiddelen bijgekomen. Het is interessant methodologische ontwikkeling te kunnen bekijken "van de andere kant van de tafel" in vergelijking met mijn rol bij Organon/Schering-Plough. Binnen Utrecht worden dwarsverbanden makkelijk gelegd en onderhouden, variërend van de diverse divisies in het UMC Utrecht tot de sterke statistische afdeling bij sociale wetenschappen en de afdelingen diergeneeskunde en wiskunde.

Het lijkt me duidelijk dat ik dit niet allemaal alleen doe. Mijn groep komt voort uit het Centrum voor Biostatistiek, dat dit jaar 20 jaar zou hebben bestaan. Met deze nieuwe start streven we ernaar op basis van een stevig netwerk een platform voor samenwerking en

ontwikkeling van de biostatistiek te kunnen vormgeven – waar alle samenwerkende partijen welkom zijn en van kunnen profiteren. Het begin is gemaakt.

7. *Woorden van dank*

Mijnheer de rector, geachte toehoorders, ik sluit graag af met woorden van dank. De Raad van Bestuur van het UMC Utrecht en het College van Bestuur van de Universiteit Utrecht wil ik bedanken voor het in mij gestelde vertrouwen de stimulering van de Biostatistiek in mijn handen te geven.

Ik heb het voorrecht gehad van velen te leren. Bij het Center for Quantitative Methods (CQM), destijds onderdeel van Philips, is de stevige basis gelegd. Jos de Kroon was met het CQM zijn tijd ver vooruit, en zijn strategische inzichten voor statistiek als vakgebied zijn voor mij nog steeds leidend.

Mijn promotor, Hooggeleerde Ronald Does heeft me op het wetenschappelijke pad gekregen, op zijn geheel unieke wijze. Zonder ons gezamenlijk pad had ik hier zeker niet gestaan. Does & Roes is voor een aantal bedrijven een tijdje een begrip geweest, ik sluit niet uit dat dit weer kan gaan gebeuren.

In Wageningen, bij het latere Biometris, heb ik het voorrecht gehad in een omgeving met uitstekende wetenschappers door te groeien. Als leidinggevende kun je het treffen en in je team talenten hebben die verdere groei stimuleren. Mijn toenmalige teamleden Hooggeleerden

Hans Heesterbeek en Ritsert Jansen zijn van dat kaliber gebleken, ik dank ze voor de samenwerking in die periode.

Het klinisch geneesmiddelenonderzoek bij Organon heeft me voor mijn huidige opdracht wellicht het meest gevormd. Ik ben veel collega's van alle disciplines dank verschuldigd. Het zijn er oprecht teveel om op te noemen. Ik prijs me daarom gelukkig dat de samenwerking tot op de dag van vandaag doorloopt.

De samenwerking met Hooggeleerde Rick Grobbee is de directe aanleiding voor het feit dat ik hier sta, en daar ben ik dankbaar voor. Ik heb daarmee een geweldige kans gekregen.

De directe collega's in de nieuwe omgeving, mijn groep, hooggeleerde Carl Moons en de collega's van het Julius Centrum, hebben mij warm onthaald. Op de steun van hooggeleerden Yolanda van der Graaf en Arno Hoes kan ik bouwen.

Mijn moeder, mijn ouders, dank ik voor het geduld dat ze vroeger met me hebben gehad en hun vertrouwen dat het wel goed zou komen.

Okko, Jelger en Niek omdat ze me scherp en jong houden en hard op weg zijn in veel dingen beter te worden dan ik ben.

En natuurlijk Sabine (mijn bron van variatie), voor haar is dit laatste citaat.

“You haven't told me yet,” said Lady Nuttal, “what it is your fiancé does for a living.”

“He’s a statistician,” replied Lamia, with an annoying sense of being on the defensive.

Lady Nuttal was obviously taken aback. It had not occurred to her that statisticians entered into normal social relationships. The species, she would have surmised, was perpetuated in some collateral manner, like mules.

“But Aunt Sara, it’s a very interesting profession,” said Lamia warmly.

“I don’t doubt it,” said her aunt, who obviously doubted it very much.

“To express anything important in mere figures is so plainly impossible there must be endless scope for well-paid advice on how to do it. But don’t you think that life with a statistician would be rather, shall we say, humdrum?”

Lamia was silent. She felt reluctant to discuss the surprising depth of emotional possibility which she had discovered below Edward’s numerical veneer.

“It’s not the figures themselves,” she said finally, “it’s what you do with them that matters.”

K.A.C. Manderville, *The Undoing of Lamia Gurdleneck*

Ik heb gezegd.

8. Referenties

1. Shewhart W.A. (1931). *Economic Control of Quality of Manufactured Product*. New York: D. van Nostrand Company Inc. (1980, 50th Anniversary Commemorative Reissue by ASQC).
2. Fisher R.A. (1958). *Statistical Methods for Research Workers*. Edinburgh: 13th Ed.
3. Fisher R.A. (1947). *The Design of Experiments*. Edinburgh: Oliver & Boyd: 4th Ed.
4. Deming W.E. (1986). *Out of the crisis*. Cambridge, MA: MIT, Center for Advanced Engineering Study.
5. Roes K.C.B. and R.J.M.M. Does (1995). Shewhart-type charts in nonstandard situations (with discussion). *Technometrics*: 37, 15-40.
6. Roes C.B. (1995). *Shewhart-type charts in Statistical Process Control*. Amsterdam, Thesis Publishers.
7. Does R.J.M.M., Roes C.B. and Trip A. (1999). *Statistical Process Control in Industry*. Dordrecht, Kluwer Academic Publishers.
8. Trip A., Does, R.J.M.M. (2008). Lean Six Sigma blijkt het ideale middel om betere zorg te krijgen tegen lagere kosten.. *Kwaliteit in bedrijf*. Juni 2008: 6-9.
9. Cummings S.R. et al. for the LIFT Trial Investigators (2008). The Effect of Tibolone in Older Postmenopausal Women *New England Journal of Medicine*, 359, 7: 697-708.
10. Armitage P. (1975). *Sequential Medical Trials*. 2nd Ed.(1st Ed. 1960). Oxford: Blackwell Scientific Publications.
11. Senn S. (2003). *Dicing with Death. Chance, Risk and Health*. Cambridge University Press.
12. Wald, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons.
13. Fisher RA (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society. Series B (Methodological)* 17, 69-78. (criticism of statistical theories of J. Neyman and A. Wald).
14. Neyman J. (1956). Note on an Article by Sir Ronald Fisher. *Journal of the Royal Statistical Society. Series B (Methodological)* 18 (2): pp. 288–294 (reply to Fisher 1955).
15. Page, E. S. (1954). Continuous Inspection Scheme. *Biometrika*, 41 (1/2): 100–115.
16. Barnard, G.A. (1959). Control charts and stochastic processes. *Journal of the Royal Statistical Society B (Methodological)*, 21: 239–271.
17. Whitehead J. (1992) *The Design and Analysis of Sequential Clinical Trials*. 2nd Ed. London: Ellis Horwood.
18. Jennison C., Turnbull B.W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall.
19. Pocock S.J. (1982). Interim Analyses for randomized Clinical Trials: The Group Sequential Approach. *Biometrics*, 38:153-162.
20. DeMets D.L., Lan G.K.K. (1984). An overview of sequential methods and their application in clinical trials. *Communication in Statistics – Theory and Methods*, 13 (19): 2315-2338.
21. Armitage P. on behalf of the Concorde and Alpha Data and Safety Monitoring Committee (1999). *Data and Safety Monitoring in the Concorde Alpha Trials*. *Controlled Clinical Trials* 20: 207-228.
22. Ellenberg SS, Fleming, TR and DeMets DL. *Data Monitoring Committees in Clinical Trials: A Practical Perspective*. John Wiley & Sons, Chichester, 2002.
23. Sydes M.R., Spiegelhalter D.J., Altman D.G., Babiker A.B., Parmar M.K.B., and DAMOCLES Group (2004). Systematic qualitative review of the literature on data monitoring committees for randomized controlled trials. *Clinical Trials*: 1: 60-79.
24. Bauer P., Köhne K. (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics*, 50: 1029-1041.

25. Dragalin V. (2006). Adaptive Designs: Terminology and Classification. *DIA Journal*, 40:425-435.
26. Berry D. (2001). Adaptive trials and Bayesian statistics in drug development. *Biopharm Rep.*, 9:1-11.
27. Bassler D., Briel M., Montori V.M. et al. (2010). Stopping Randomized Trials Early for Benefit and Estimation of Treatment Effects: Systematic Review and Meta-regression Analysis. *JAMA*, 303 (12): 1180-1187.
28. Stallard N, Whitehead J, Cleall S. (2005). Decision-making in a phase II clinical trial: a new approach combining Bayesian and frequentist concepts. *Pharmaceutical Statistics*, 4: 119-128.
29. Solomon, P.J. (2009). Some statistics in bioinformatics: The fifth Armitage Lecture. *Statistics in Medicine*, doi::10.1002/sim.3668.

Colofon

Uitgave

© Universitair Medisch Centrum Utrecht/Universiteit Utrecht,
december 2010

Foto cover

Jaqueline Bosman

Vormgeving

Atelier van GOG, Amsterdam

Druk

Rijser Grafische Communicatie

UMC Utrecht

Bezoekadres:
Heidelberglaan 100
3584 CX Utrecht

Postadres:
Postbus 85500
3508 GA Utrecht

www.umcutrecht.nl