

**DE ZOEKTOCHT NAAR
EFFECTIEVE BEHANDELINGEN**

PROF. DR. IR. KEES VAN MONTFORT

**DE ZOEKTOCHT NAAR
EFFECTIEVE BEHANDELINGEN**

Oplage 1000
Omslagfoto Levien Willemse, Rotterdam
Ontwerp Ontwerpwerk, Den Haag
Drukwerk Océ-Nederland B.V., Rotterdam

ISBN 978-90-779-0693-4

© Prof. dr. ir. Kees van Montfort, oratiereeks Erasmus MC
22 december 2011

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd zonder voorafgaande toestemming van de auteur.

Voorzover het maken van kopieën uit deze uitgave is toegestaan op grond van art. 16h t/m 16m Auteurswet 1912 j°. Besluit van 27 november 2002, Stb. 575, dient men de daarvoor wettelijk verschuldigde vergoeding te voldoen aan de Stichting Reprorecht te Hoofddorp (Postbus 3060, 2130 KB).

**DE ZOEKTOCHT NAAR
EFFECTIEVE BEHANDELINGEN**

REDE

In verkorte vorm uitgesproken
ter gelegenheid van het aanvaarden
van het ambt van bijzonder hoogleraar
Biostatistiek in het Klinische Kankeronderzoek
aan het Erasmus MC, faculteit van de
Erasmus Universiteit Rotterdam
op 22 december 2011

door

PROF. DR. IR. KEES VAN MONTFORT

*Mijnheer de Rector Magnificus,
College van Bestuur van de Erasmus Universiteit,
Decaan van de faculteit voor geneeskunde en
gezondheidswetenschappen,
Raad van Bestuur van het Erasmus MC,
Bestuur van de Vereniging Trustfonds Erasmus Universiteit,
Zeer gewaardeerde toehoorders,*

1. Inleiding

In 2006 hield ik een inaugurele rede aan Nyenrode Business Universiteit te Breukelen. De rede werd uitgesproken bij het aanvaarden van het ambt van hoogleraar in de Kwantitatieve Bedrijfskundige Onderzoeksmethoden en had als titel “Meten en weten”. De volgende vraag stond centraal: Stel we hebben meetgegevens over bepaalde verschijnselen. Op welke wijze kunnen dan met behulp van moderne statistische methoden en technieken mogelijke oorzaak-gevolg relaties tussen twee of meer van deze verschijnselen worden vastgesteld? Ofwel: hoe kunnen op grond van verzamelde gegevens eventuele causale verbanden statistisch worden aangetoond? Deze vraag werd door mij aan de hand van bedrijfseconomische gegevens beantwoord.

Vandaag richt ik mij op de medische sector. Ruim een jaar geleden werd ik benoemd als hoofd statistiek bij de unit Trials&Statistiek van het Erasmus Medisch Centrum (Erasmus MC) te Rotterdam. Enige tijd later volgde bij dit centrum tevens een benoeming als bijzonder hoogleraar, met als leerstoel ‘Biostatistiek in het Klinische Kankeronderzoek’. Een belangrijke doelstelling van deze leerstoel is een bijdrage te leveren aan de kwaliteit van het klinische kankeronderzoek en daarmee aan de effectiviteit van oncologische behandelingen. De titel van deze oratie luidt dan ook: “De zoektocht naar effectieve behandelingen”. De centrale vraag is: Op welke wijze kunnen moderne statistische methoden en technieken een bijdrage leveren aan het onderzoek naar effectieve behandelingen voor kankerpatiënten? Het gaat – net als in 2006 – nog steeds over statistiek, maar het toepassingsgebied is totaal anders.

Dat de resultaten van medische wetenschappelijke onderzoeken, en dus ook van de daarop gebaseerde klinische behandelingen, niet altijd een adequate statistische onderbouwing kennen, moge blijken uit het volgende. Enkele jaren geleden baarde onderzoeker John Ioannidis in het Journal of the American Medical Association opzien met zijn bevinding dat van de belangrijkste medische artikelen er na een paar jaar nog maar weinig recht overeind staan (Ioannidis, 2005). Hij stelde op basis van een statistische analyse van 45 zeer frequent geciteerde originele klinische onderzoeken (dus met een hoge impact), dat maar liefst 25 artikelen (56%) werden tegengesproken door latere studies of veel sterkere effecten toonden dan vervolgstudies. Meestal bleek het onderzoek te klein van opzet, de significantie te laag en/of de kans op systematische vertekening te groot. De geclaimde causaliteit wordt dan onvoldoende onderbouwd.

Uiteraard is het ook de vraag of onderzoeker Ioannidis op zijn beurt de juiste conclusies heeft getrokken bij zijn statistische analyses. Een en ander geeft echter wel aan dat bij elk klinisch onderzoek de weg van meten naar weten, en dus ook de weg naar effectieve behandelingen, geen vanzelfsprekend traject is. Het is een hobbelig pad dat steeds met zorgvuldigheid zal moeten worden bewandeld. De toepassing van adequate statistische methoden en technieken is daarbij een noodzakelijke voorwaarde.

Mijn leerstoel richt zich op de betekenis van biostatistiek voor klinisch kankeronderzoek. Voor een groot deel van mijn tijd ben ik betrokken bij de statistische begeleiding van zogeheten 'HOVON-studies'. Ik schets nu eerst de achtergrond en opzet van deze studies. Vervolgens wordt ingegaan op de rol van statistiek in de klinische trials die in het kader van de HOVON-studies worden uitgevoerd. Daarbij wordt aangegeven op welke punten verbetering in de statistische methodologie van klinische trials wenselijk is. Aansluitend schenk ik aandacht aan statistische analyses van high throughput data met betrekking tot het DNA van kankerpatiënten. Ik eindig met enkele opmerkingen over de interne en externe samenwerking bij klinisch kankeronderzoek en over mijn bijdrage aan het biostatistische onderwijs binnen het Erasmus MC.

2. HOVON-studies

Een belangrijk gedeelte van mijn werkzaamheden bestaat uit de statistische begeleiding van HOVON-studies. De afkorting HOVON staat voor de Stichting Hemato-Oncologie Voor Volwassenen Nederland. HOVON is in 1985 door de afdelingen Hematologie van de acht Nederlandse academische ziekenhuizen opgericht om de kwaliteit van zorg voor hemato-oncologische patiënten in Nederland en België te verbeteren. Deze kwaliteitsverbetering van patiëntenzorg wordt vooral nagestreefd door patiëntgebonden studies te verrichten waarin nieuwe medicaties en nieuwe behandelstrategieën worden geëvalueerd.

HOVON heeft de uitvoering van haar multi-center studies neergelegd bij het Hovon Data Centrum (HDC). De afgelopen jaren is het HDC uitgegroeid tot een professionele organisatie die een groot aantal multi-center fase II- en multi-center fase III-studies ontwikkelt en uitvoert. Aanvankelijk werden de studies in Nederland en België uitgevoerd, maar tegenwoordig wordt er steeds meer in intergroup studies met buitenlandse groepen samengewerkt. Denk aan de samenwerking met SAKK (Zwitserland), GMMG (Duitsland), GELA (Frankrijk), MRC en BNLI (Verenigd Koninkrijk), EMN (Italië) en de Nordic Lymphoma Group (Scandinavische landen). Het data- en statistische management vindt doorgaans bij het HDC in Rotterdam plaats.

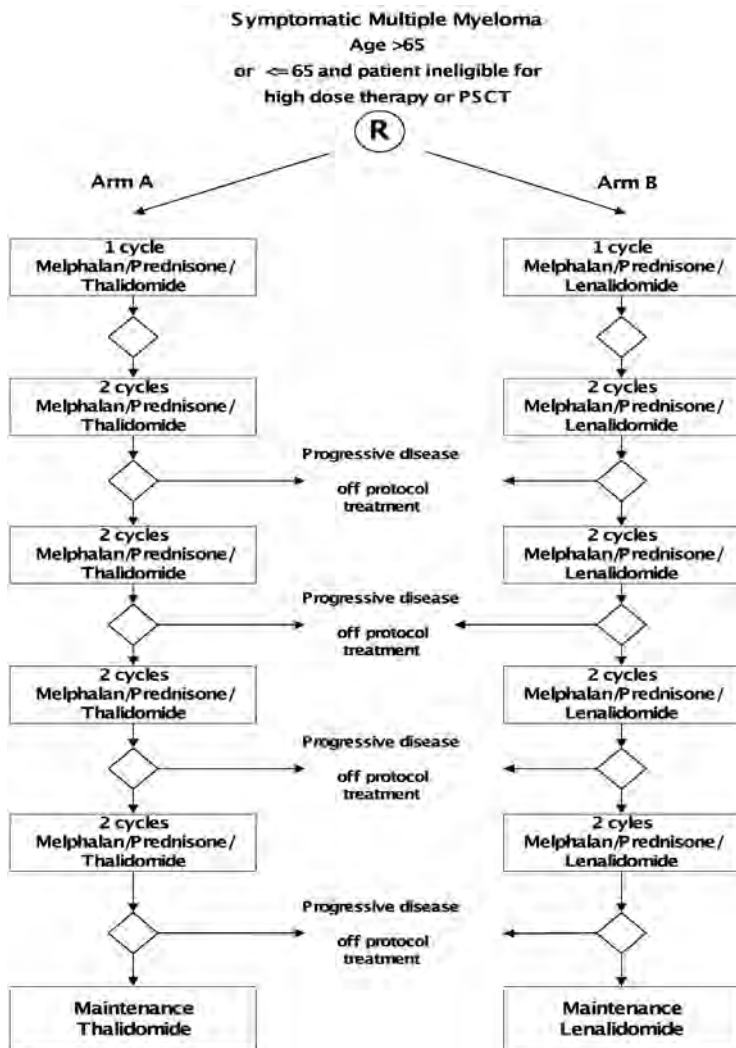
Inmiddels zijn er meer dan 100 HOVON-studies gestart. Over het overgrote deel van deze studies zijn kwaliteitspublicaties in de internationale literatuur uitgebracht. De productie van “peer reviewed” artikelen is in 1990 op gang gekomen. Sindsdien zijn er ruim 100 gepubliceerd. HOVON en het HDC zijn zo uitgegroeid tot een bekend “merk” in de internationale netwerken van hemato-oncologie.

HOVON en het HDC werken in een internationale context en hebben een uitstekende internationale reputatie. De HOVON-onderzoekers werken vanuit ziekenhuizen in Nederland, België, Noorwegen, Denemarken, Duitsland, Zwitserland, Italië, Frankrijk en Estland. De nieuwe HOVON-studies worden vanuit internationaal samengestelde studiewerkgroepen bedacht en door het HDC verder ontwikkeld. In de internationale studiewerkgroepen komen de hematologen en statistici samen, die vooral geïnteresseerd zijn in een bepaald soort ziektebeeld. De resultaten van de HOVON-studies worden doorgaans op internationale wetenschappelijke congressen gepresenteerd en in internationale wetenschappelijke tijdschriften met een hoge impact gepubliceerd.

Bij al deze internationale wetenschappelijke activiteiten is er sprake van een sterk internationaal netwerk dat de afgelopen jaren gestaag is gegroeid. Hierbij zijn het trial management, het centraal data management en de statistische analyses van het HDC van doorslaggevend belang.

3. Klinische trials en power berekeningen

Onderstaand schema geeft een beknopte weergave van de design van een HOVON klinische trial (de HOVON 87), waar ik zelf als statisticus bij betrokken ben.



Figuur 1

De structuur van het design is relatief eenvoudig (zie ook Hordijk, 2011). Voor de ziekte multiple myeloom zijn er twee behandelingen beschikbaar: de standaard behandeling met Thalidomide en de experimentele behandeling met Lenalidomide. De gerandomiseerde patiënten krijgen één van beide behandelingen. Eén van de keuzen, die vooraf genomen moet worden, betreft de vaststelling van de beoordelingscriteria, de zogenoemde eindpunten. De keuze van het eindpunt of de eindpunten hangt sterk af van het te bestuderen ziektebeeld. Veel gebruikte eindpunten in het oncologische onderzoek zijn de zogenoemde event free survival en de overall survival.

Stel dat de overall survival (d.w.z. de tijdsduur van de start van de behandeling tot het moment van overlijden) het primaire eindpunt (d.w.z. het beoordelingscriterium) is. Vooraf moet dan worden ingeschat hoe groot de sample size (d.w.z. het aantal patiënten dat deelneemt aan de trial) minimaal moet zijn. Dit noemen we de power berekening van de studie. Geen medisch ethische commissie geeft toestemming voor het starten van een wetenschappelijk onderzoek als er geen sample size berekening wordt aangeleverd. Het is zelfs zo dat de resultaten van een experimenteel onderzoek niet meer gepubliceerd kunnen worden in veel vooraanstaande medische tijdschriften als er geen sample size berekening in de rapportage wordt opgenomen.

Bij het berekenen van de sample size wordt doorgaans een standaardaanpak gevolgd. Daarvoor is, gegeven de gangbare betrouwbaarheidsdrempel en power van respectievelijk 0.05 en 0.80, de volgende informatie nodig:

- * Het gemiddelde aantal events per tijdseenheid in de controle groep (hoe hoger het aantal events per tijdseenheid, des te lager het aantal benodigde patiënten);
- * Het gemiddelde aantal events per tijdseenheid in de experimentele groep (hoe hoger het aantal events per tijdseenheid, des te lager het aantal benodigde patiënten; hoe groter het verschil tussen het aantal events per tijdseenheid in de controle en de experimentele arm, des te lager het aantal benodigde patiënten);
- * Het aantal patiënten dat jaarlijks wordt geïncludeerd in de trial;
- * De resterende looptijd van de trial nadat de laatste patiënt is geïncludeerd (hoe langer deze looptijd, des te lager het aantal benodigde patiënten).

Het eerste en het derde gegeven zijn van tevoren goed in te schatten. Het vierde gegeven kan door de principal investigators worden gekozen. Voor het tweede gegeven, te weten het gemiddelde aantal events per tijdseenheid in de experimentele groep, kan hoogstens een inschatting worden gemaakt. Deze inschatting heeft grote gevolgen voor de berekening van de sample size (en de kosten van de klinische trial).

Het voorgaande kan met behulp van een eenvoudig rekenvoorbeeld over de HOVON 87-studie worden toegelicht. Informatie uit de wetenschappelijke literatuur levert de volgende cijfers op:

- * Het gemiddelde aantal events per maand in de controle groep = $1/28$;
- * Het gemiddelde aantal events per maand in de experimentele groep = $1/42$;
- * Het gemiddelde aantal patiënten dat jaarlijks wordt geïncludeerd in de trial = 155.
=> de berekende sample size is 452 patiënten (bij een follow-up tijd van 12 maanden).

Gebaseerd op resultaten uit recentere wetenschappelijke literatuur kwam men tot andere cijfers:

- * Het gemiddelde aantal events per maand in de controle groep = $1/20$;
- * Het gemiddelde aantal events per maand in de experimentele groep = $1/28$;
- * Het gemiddelde aantal patiënten dat jaarlijks wordt geïncludeerd in de trial = 155;
=> de berekende sample size is 668 patiënten (bij een follow-up tijd van 12 maanden).

Dit rekenvoorbeeld toont aan dat de uitkomst van een power berekening iets arbitrairs kan hebben en sterk afhangt van de gemaakte aannames. In het tweede geval zijn er voor de uitvoering van de klinische trial veel meer patiënten nodig.

Een interessant punt is dat de standaard software voor de power-berekening veronderstelt dat bijvoorbeeld de overall survival een exponentiële kansverdeling heeft. Dit is vaak niet het geval, hetgeen met verschillende statistische toetsen kan worden aangetoond. Hierdoor zal de berekende sample size niet juist zijn. De komende jaren zal ik mij onder andere gaan bezighouden met de zogeheten verdelingsvrije power-berekeningen, waarbij de aanname van exponentieel verdeelde eindpunten niet meer nodig is.

Een tweede tekortkoming van de beschikbare standaard software is de aanname van een constante patiënteninclusie. Aan deze aanname wordt bij multi-center studies zelden voldaan. Er wordt met één ziekenhuis begonnen en stap voor stap breidt het aantal participerende ziekenhuizen zich uit. Er is dus sprake van een toenemend aantal geïncludeerde patiënten per tijdseenheid. Zonder in detail te treden, leidt het gebruik van standaard software met de aanname van een constante patiënteninclusie tot een stevige onderschatting van het benodigde aantal patiënten. De komende tijd zal ik werken aan de ontwikkeling van methoden die het verloop van de patiënteninstroom meenemen in de berekeningen.

Tenslotte valt op te merken dat door een andere keuze voor de eindpunten het benodigde aantal patiënten in de trial kan worden verminderd. Bij het vervangen van het eindpunt 'overall survival' door bijvoorbeeld de eindpunten 'overall survival na één jaar' en 'overall survival na drie jaar' zal de trial-omvang drastisch kunnen afnemen. De onderzoekers hebben dan veel minder datapunten nodig. De vraag is alleen of redacties van wetenschappelijke tijdschriften aan dit idee zullen kunnen wennen en traditionele aanpakken zullen kunnen loslaten.

4. Klinische trials en interim analyses

De bevindingen van een interim-analyse worden vaak gerapporteerd aan een Data Safety Monitoring Board (DSMB), ook wel aangeduid als een Data Safety Monitoring Committee (DSMC). Een dergelijke toezichthoudende instantie in het kankeronderzoek bestaat doorgaans uit drie personen: twee hematologen en één statisticus. Zij hebben geen enkele formele relatie met de ziekenhuizen en organisaties die betrokken zijn bij de desbetreffende studie.

Om tussentijds beslissingen te kunnen nemen, zijn interim analyses noodzakelijk. Dit kunnen beslissingen zijn met betrekking tot de optimale dosering en de veiligheid van de experimentele behandeling. Hiervoor moet bij een interim analyse een vergelijking tussen de behandelingen worden gemaakt op basis van de tussentijds beschikbare data. De beperkte beschikbaarheid van data heeft consequenties voor het uiteindelijke significantieniveau in de analyses en de feitelijke onbetrouwbaarheid van de uitkomsten. Stel dat er geen sprake is van een verschil tussen de resultaten van de controle behandeling en de resultaten van de experimentele behandeling (op het primaire eindpunt van de interim analyse). Als er bij een interim analyse wordt getoetst tegen het algemeen gebruikelijke 5% significantieniveau, neemt de kans dat er bij een analyse ten onrechte een significant verschil wordt gevonden toe bij een afnemend aantal geïncludeerde patiënten. Als er wel sprake is van een verschil tussen de resultaten van de controle behandeling en de resultaten van de experimentele behandeling, treedt er een tegenovergesteld verschijnsel op. De power, dat wil zeggen de kans dat er een significant verschil tussen de resultaten van de beide behandelingen wordt aangegeven gegeven het feit dat er feitelijk een verschil is, neemt dan af bij een afnemend aantal geïncludeerde patiënten.

De conclusie van dit alles is dat als er wel een verschil tussen de resultaten van de beide behandelingen is, bij een interim analyse een onderschatting van dit verschil zal plaatsvinden. Als er geen verschil tussen beide behandelingen is, zal het verschil tussen beide behandelingen juist worden overschat.

Een mogelijke oplossing hiervoor is om het significantieniveau van de interim analyse aan te passen. Dit betekent dat we bij een interim analyse strenger zijn dan bij een eindanalyse. Het verschil tussen de resultaten van behandelingen moet statistisch overtuigender zijn om ten onrechte een statistisch verschil tussen de controle en de experimentele arm aan te geven

Een andere vraag die bij interim analyses kan opkomen is of in een vroegtijdig stadium van de studie al beoordeeld kan worden of de doelstelling van de studie uiteindelijk gehaald kan worden. Bijvoorbeeld: zal aan het einde van de studie de overall survival bij de experimentele behandeling daadwerkelijk beter blijken te zijn dan bij de standaard behandeling? Hiervoor kan men bijvoorbeeld gebruik maken

van futility analyses. Hierbij vindt met behulp van de gegevens die bij de interim analyse beschikbaar zijn, een berekening plaats van de kans dat de doelstelling van de studie uiteindelijk wordt gehaald. Deze statistische technieken worden in het kankeronderzoek helaas nog weinig toegepast.

Kleine patiëntenaantallen leiden bij interim-analyses tot statistische problemen. Eén van de methoden ter verbetering is een andere manier van opzetten en uitvoeren van de klinische trials. Bij het zogenaamde 'adaptive design' kan de opzet van de trial tussentijds worden aangepast op basis van de interim-resultaten. De aanpassingen waaraan gedacht kan worden, zijn onder andere:

- * Het aanpassen van de groepsgrootte voor de resterende studie;
- * Het aanpassen van de inclusiecriteria voor de participerende patiënten;
- * Het kiezen van andere eindpunt(en).

Strikt genomen wordt de eerste mogelijkheid voor het aanpassen van het trial design af en toe toegepast bij HOVON-studies. Los daarvan denk ik dat er in de toekomst meer gewerkt zal worden met adaptive designs. Het is immers een aantrekkelijk idee dat een klinische trial kan worden gestart en dat het ontwerp vervolgens lopende de rit op basis van de nieuwe beschikbaar gekomen data kan worden aangepast.

5. Klinische trials en eindpunten

Bij de afweging of een experimentele behandeling wel of niet beter is dan de standaardbehandeling maken we gebruik van één of meerdere beoordelingscriteria, de zogenaamde eindpunten. Als we de scores bij deze eindpunten voor de verschillende behandelingen met elkaar vergelijken, kunnen we onder meer tegen de volgende statistische problemen aanlopen:

- * klinische relevantie p-waarden,
- * causaliteit,
- * imputatietechnieken,
- * competing risks.

Klinische relevantie p-waarden

De p-waarde (oftewel de probability waarde of de overschrijdingskans) geeft, gesimplificeerd gezegd, aan de kans dat er geen verschil is tussen de resultaten van de beide behandelingen. De beslissing om wel of geen verschil te veronderstellen, wordt genomen aan de hand van een vooraf gekozen grenswaarde, die doorgaans gelijk aan 0.05 wordt gesteld. Als de p-waarde dus kleiner dan 0.05 is, moeten we concluderen dat de nieuwe behandeling beter is. We zeggen dan dat het verschil tussen de twee behandelingen statistisch significant is. Dit is een alles of niets conclusie: of de nieuwe behandeling is effectiever dan de standaard behandeling of beide behandelingen zijn even effectief. Bovendien geldt: als de p-waarde kleiner dan 0.05 blijkt te zijn, is het een stuk waarschijnlijker dat de onderzoeker zijn onderzoeksresultaat in een wetenschappelijk tijdschrift met hoge impact factor kan publiceren.

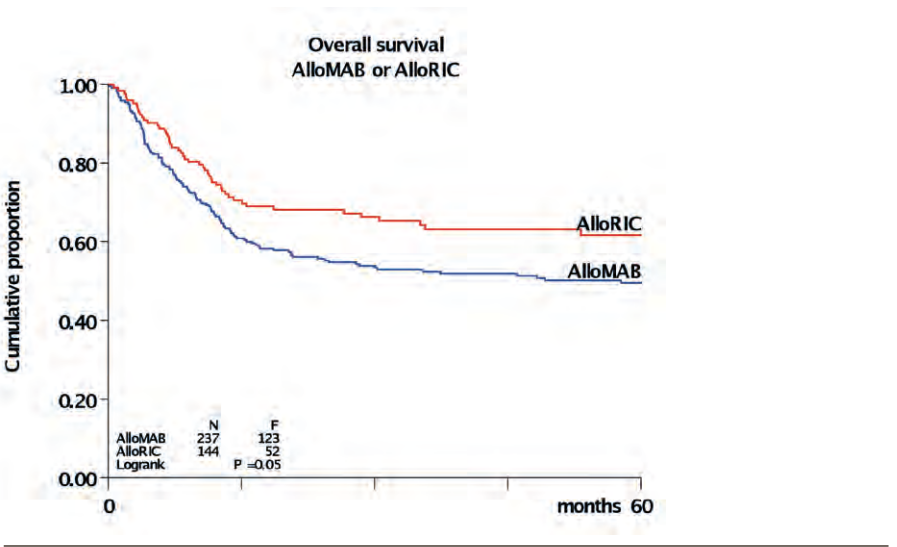
In het algemeen zijn statistici het er over eens dat de grens van 0.05 wel erg arbitrair is. Desondanks is voor de Amerikaanse Food and Drug Administration wettelijk geregeld dat een nieuw medicijn wordt 'erkend' als de drempelwaarde van 0.05 op het primaire eindpunt bij twee verschillende studies is gehaald.

Overigens dient opgemerkt te worden dat een p-waarde niets zegt over de grootte van het verschil tussen twee behandelingen. Ook zegt een statistisch significant verschil (d.w.z. een p-waarde kleiner dan 0.05) nog niet dat het verschil klinisch significant is (zie Ziliak & McCloskey, 2008). Of het statistisch significante verschil tussen de resultaten van twee behandelingen ook echt klinisch relevant is, moet strikt genomen vooral worden vastgesteld door de klinische onderzoeker op basis van zijn medische expertise en ervaring. De statisticus zal de klinische onderzoeker moeten voorzien van de juiste informatie om die afweging goed te kunnen laten verlopen. Dit betreft zowel informatie over de statistische significantie als informatie over de grootte van het effect van afzonderlijke behandelingen. Er zal in de toekomst meer gekeken moeten worden of statistische verschillen tussen behandelingen ook klinisch relevant zijn.

Causaliteit

Causaliteit, dat wil zeggen het bestaan van een oorzaak-gevolg relatie, lijkt geen issue te zijn bij klinische trials. Als op het primaire eindpunt bij de ene behandeling beter wordt gescoord dan bij de andere behandeling, zal dit wel het gevolg zijn van de verschillende behandelingen. Deze in klinische onderzoeken vaak toegepaste redenering gaat echter alleen op als er vooraf een goede randomisatie, dat wil zeggen een aselechte toewijzing van patiënten aan de twee verschillende behandelingen, heeft plaatsgevonden. De patiënten moeten met andere woorden op een zodanige wijze aan de twee behandelgroepen worden toegewezen dat de patiënten in beide behandelgroepen gemiddeld genomen dezelfde eigenschappen hebben. Denk hierbij aan leeftijd, ernst van de ziekte, geslacht, etc. Als men bij randomisatie echter een bepaalde relevante patiënteigenschap over het hoofd ziet, kan dit tot verkeerde conclusies leiden.

In onderstaande grafiek zijn de overall survival curven voor twee verschillende behandelingen voor de ziekte Acute Myeloid Leukemia (AML) weergegeven. Op basis van de p-waarden bij de logrank test (p-waarde =0,05) moeten we concluderen, dat AlloRIC een betere Overall Survival oplevert dan AlloMAB. Als we echter corrigeren voor de variabelen "Risk Group (very poor/poor/intermediate/good)" en "t-cell depletion (yes/no)" kan bovenstaande conclusie in twijfel worden getrokken, hetgeen blijkt uit een p-waarde voor de logrank test van 0,30. Blijkbaar werd het verschil in performance voor de twee behandelingen niet verklaard door de behandelingen op zich.



Figuur 2

Imputatietechnieken

Regelmatig is er bij klinische studies sprake van ontbrekende gegevens. Stel, we willen bij patiënten op drie verschillende tijdstippen metingen verrichten. Het is dan zeer wel mogelijk dat meerdere patiënten na één of twee meetmomenten afhaken, waardoor voor hen niet op alle drie de meetmomenten de observatiewaarden beschikbaar zijn.

We kunnen op verschillende manieren met deze ontbrekende observaties omgaan. Een gangbare oplossing is om de desbetreffende patiënten uit de database te verwijderen. Een betere manier is het imputeren van de ontbrekende gegevens. Dit betekent dat we een getal moeten invullen op de plek in de dataset waar nog een getal ontbreekt. Een eenvoudige aanpak is om voor een missende observatie het ongewogen gemiddelde over de gegevens die we wel hebben in te vullen. Er zijn echter vele genuanceerdere imputatie technieken beschikbaar (zie onder andere het boek van Little en Rubin uit 2002, en diverse publicaties van Stef van Buuren), die voor een specifieke situatie geschikt kunnen zijn. Het verdient mijn inziens aanbeveling om bij ontbrekende observaties niet onmiddellijk de eenvoudigste oplossing te kiezen: namelijk het buiten beschouwing laten van de desbetreffende patiënt of het werken met ongewogen gemiddelden.

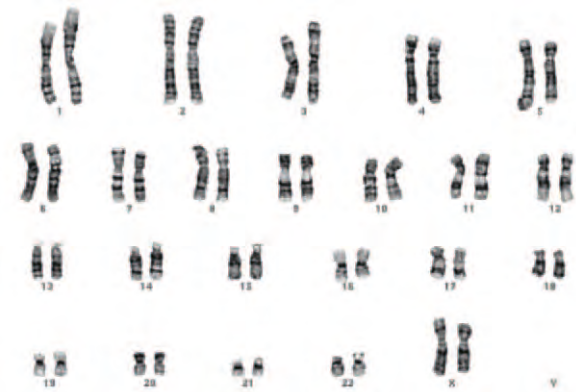
Competing risk

Competing risk is ook een onderwerp dat bij evaluatie van de eindpunten van belang is. Enige tijd geleden was ik betrokken bij een studie waarin vrouwen die ovariumkanker hadden gehad werden gevolgd. De vrouwen werden op twee verschillende manieren begeleid. Het primaire eindpunt was het optreden van borstkanker. Op basis van de survival curven (eigenlijk Kaplan-Meier curven) voor beide groepen konden we concluderen dat in één van beide groepen borstkanker minder snel optrad. Bij deze studie was er echter sprake van een competing risk (d.w.z. concurrerend risico), namelijk het overlijden van een patiënt zonder dat er borstkanker had opgetreden. Als we rekening houden met dit concurrerend risico krijgen we twee nieuwe survival curven (wederom eigenlijk Kaplan-Meier curven), die nauwelijks van elkaar verschillen. Op basis van deze nieuwe survival curven moeten we concluderen dat vrouwen in beide groepen evenveel kans op borstkanker hebben. Moraal van het verhaal: denk van tevoren goed na over de exacte definities van de eindpunten en houdt rekening met competing risks.

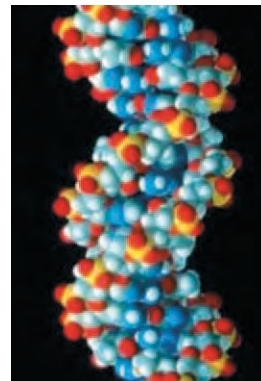
6. High throughput data

In het voorafgaande heb ik een en ander verteld over de statistiek die bij klinische trials binnen het kankeronderzoek van belang is. Biostatistiek speelt ook een belangrijke rol bij de analyse van grote gegevensbestanden die worden verkregen door nieuwe technologische ontwikkelingen. Beeldvormende en moleculaire technieken genereren grote hoeveelheden gegevens. Met behulp van fluorescentie, analyses van snp's, GWA scans en analyses van gen-expressies kan inzicht worden verkregen in de ketens van onze genen en de ontwikkelingen van ziekten. Dit kan de mogelijkheid geven om te voorspellen welke behandeling voor welke patiënt het meest geschikt is op basis van individuele karakteristieken van de patiënt en de ziekte.

De gegeneerde data leveren statistische problemen op. Het aantal patiënten is doorgaans extreem veel kleiner dan de hoeveelheid gemeten variabelen. 100 Patiënten met 10.000 genetische of beeldvormende markers zijn statistisch moeilijk te analyseren. Deze markers kunnen bovendien complexe afhankelijkheden vertonen.



Figuur 3



Figuur 4

Bovenstaande figuur 3 vertoont de 23 chromosomen paren van een mens. Een chromosoom is opgebouwd uit allerlei verschillende genen en eiwitten. Naar schatting zijn er ongeveer 26.000 genen mogelijk.

Met behulp van de statistiek kan gekeken worden of bepaalde genen met een bepaald ziektebeeld corresponderen. Bij dit soort onderzoek worden steeds grotere databestanden aangelegd. Bij zeer grote databestanden is het kleinste verschil echter al statistisch significant. Dit bemoeilijkt de interpretatie van de statistische analyses behoorlijk. De vraag is natuurlijk of de statistische significantie ook klinisch relevant

is. Ten gevolg van het feit dat bij zeer grote databestanden het kleinste verschil al statistisch significant wordt, zijn er inmiddels talloze wetenschappelijke publicaties met klinisch onbruikbare conclusies en aanbevelingen verschenen.

Afgelopen decennia is er veel tijd, energie en geld in het verzamelen en analyseren van genen gestoken. De “grote doorbraak” op het gebied van het kankeronderzoek heeft echter nog niet plaatsgevonden. De komende jaren zal ik mij eveneens inzetten om een positieve bijdrage aan dit onderzoek te leveren. Enkele aandachtspunten hierbij zijn:

- * In kankercellen kan het DNA op talloze manieren veranderen, afhankelijk van het type en het stadium van de tumor. Dit kan men met microarrays voor SNP's onderzoeken;
- * Bovendien geven genexpressies informatie over de aard en het stadium van de tumoren. Dit kan helpen bij het kiezen van een behandeling of het voorspellen van overlevingskansen;
- * Door de correlaties in genexpressies bij verschillende patiënten en/of op verschillende tijdstippen te analyseren, kan men meer te weten komen over hoe genen elkaar beïnvloeden (pathways);
- * Aan de histonen in de genen kunnen zich moleculen hechten, die de vorming van RNA belemmeren (mythylering);
- * De huidige opzet van het onderzoek spoort effecten van individuele genen op. Mogelijk is het zinvoller om naar gecombineerde effecten van genen te kijken.

Bij marktonderzoek kan eveneens sprake zijn van analyses van omvangrijke databestanden. Denk bijvoorbeeld aan het analyseren van kassagegevens van een winkelketen. Mogelijk zijn statistische en/of econometrische technieken uit het marktonderzoek bruikbaar bij statistische analyses van genen binnen het kankeronderzoek.

7. Samenwerking

We zijn bij de unit Trials & Statistiek van het Erasmus MC op zoek naar effectieve behandelingen voor kankerpatiënten. Bij dit klinische onderzoek zijn verschillende partijen betrokken. Het HOVON Data Centrum (HDC) dat het trial management, het centraal data management, de monitoring, de safety desk activiteiten en de statistische ondersteuning voor de HOVON studies verricht, krijgt het leeuwendeel van haar financiering indirect van de farmaceutische industrie. De farmaceutische industrie is dus een belangrijke samenwerkingspartner in het kankeronderzoek.

Het zal duidelijk zijn dat de farmaceutische industrie bij de start van een HOVON-studie commerciële mogelijkheden kan zien. Hierbij valt bijvoorbeeld te denken aan het aantonen van de werkzaamheid van een bepaald medicijn voor een specifiek ziektebeeld. Het spreekt voor zich dat in zo'n geval de regio van de HOVON-studies volledig bij het HOVON/HDC ligt en dat er geen enkele belangenverstrengeling bij HDC-medewerkers en klinisch onderzoekers wordt toegestaan.

Het is moeilijk om studies van de grond te krijgen waarbij er op voorhand geen potentieel voordeel voor een farmaceut valt te verwachten. Studies die naar verwachting onvoldoende financieel voordeel voor een farmaceut zullen opleveren, worden vooral uitgevoerd met behulp van KWF-geld. Hier zou de overheid, die relatief weinig geld in oncologisch onderzoek steekt, eigenlijk moeten bijspringen.

Trial management, centraal datamanagement en statistiek gaan bij de unit Trials & Statistiek hand in hand. Vanuit de statistiek is het verstandig om bij de datacollectie betrokken te zijn en intensief met centraal datamanagement en trial management samen te werken. Zo kan de statisticus een beter inzicht verkrijgen in de beschikbare data en de eventuele onvolkomenheden in deze data. Ook kan dit leerzaam zijn voor de opzet van een volgende nieuwe studie.

Het aan het begin van dit jaar opgerichte Centrum voor Quantitatieve Methoden (CQM) is een belangrijk initiatief om de samenwerking tussen statistici binnen het Erasmus MC te bevorderen. Gelukkig hebben we tot dusverre nog geen klachten ontvangen van het gelijknamige CQM, dat al ruim 25 jaar als afsplitsing van het gerenommeerde Natuurkundig Laboratorium van Philips statistische dienstverlening verricht. Ons eigen Rotterdamse CQM heeft ten doel dat de verschillende statistiegroepen en statistiekgebruikers binnen het Erasmus MC, met als trekker prof. Emmanuel Lesaffre, hun krachten gaan bundelen. Met name de CQM-werkgroep Clinical Trials is interessant voor het kankeronderzoek.

8. Onderwijs

Mijn onderwijsactiviteiten zullen na verloop van tijd gaan plaatsvinden binnen één of meer masteropleidingen van onze faculteit en binnen het onderwijsaanbod voor PhD-studenten. Daarnaast zal ik enkele PhD-studenten bij hun promotieonderzoeken gaan begeleiden. Hierbij zal ik gebruik maken van de ervaring die ik in de afgelopen jaren aan de VU Amsterdam en bij Nyenrode Business Universiteit heb opgedaan als begeleider van promovendi.

Doel van deze onderwijsactiviteiten zal zijn om de deelnemers op te leiden tot het zelfstandig kunnen modelleren van medische en biologische problemen en tot het zelfstandig kunnen uitvoeren van hieraan gekoppelde statistische analyses. Het kunnen herkennen van de eventuele eigen beperkingen op het terrein van statistische kennis en vaardigheden lijkt mij eveneens een zinvolle doelstelling.

Ik zal niet betrokken zijn bij het bacheloronderwijs binnen de Faculteit der Geneeskunde en Gezondheidswetenschappen. Dat een wetenschappelijke bacheloropleiding voldoende ruimte voor statistiekonderwijs moet inruimen, lijkt mij evident. Een van de hoofddoelen van een wetenschappelijke bacheloropleiding is immers dat studenten in de eindfase van de opleiding zelfstandig onderzoek kunnen verrichten. Zonder statistische kennis en vaardigheden is dat niet mogelijk.

De alom gehuldigde opvatting dat statistiekonderwijs saai zou zijn, wil ik met klem tegenspreken. Tegenwoordig zijn er vele studieboeken beschikbaar met goede praktische voorbeelden en vraagstukken, zodat de link met de toepassingspraktijk van statistiek al snel duidelijk wordt. Bij mijn verhuizing van de VU Amsterdam naar het Erasmus MC heb ik ruim 6 meter aan presentexemplaren van statistiekboeken aan studenten uitgedeeld of in mijn boekenkast op mijn VU-werkkamer achtergelaten. Er zijn, kortom, genoeg goede statistiekboeken beschikbaar. Verder werd bij mijn vorige werkgevers het statistiekonderwijs door de studenten doorgaans duidelijk bovengemiddeld geëvalueerd. Voor mij is statistiek nog steeds een boeiend vakgebied.

9. Dankwoord

Op deze plaats maak ik graag gebruik van de mogelijkheid om enkele mensen te bedanken.

Bij het proces van een hoogleraarbenoeming moeten doorgaans één of meer personen zich sterk maken voor de kandidaat. In dit verband wil ik prof. Emmanuel Lesaffre en prof. Bob Lowenberg, net als de andere leden van de benoemingsadviescommissie, bedanken voor hun steun en vertrouwen.

Verder bedank ik mijn collega's van de unit Trials & Statistiek van het Erasmus MC. Hun open houding heeft ertoe bijgedragen dat ik me relatief snel in de wereld van de statistiek van het klinische kankeronderzoek ben gaan thuis voelen. In dit verband zou ik met name Wim van Putten, Ronnie van der Holt en Marleen van Waarden willen noemen.

Ook wil ik het bestuur van mijn thema, Thema Daniel, en het bestuur van het HOVON bedanken. Hierbij denk ik met name aan prof. Pieter Sonneveld. Zij scheppen gunstige randvoorwaarden voor mijn werkzaamheden en leveren een wezenlijke bijdrage aan een plezierig en productief werkklimaat.

Ten slotte wil ik mijn vrouw en kinderen, mijn ouders en mijn naaste familie bedanken voor hun voortdurende betrokkenheid en morele steun. Ook voor hen is dit een bijzonder feestelijke gebeurtenis.

Ik heb gezegd.



Referenties

Hordijk, M. (2011). Klinische trial krijgt concurrentie. *Medisch Contact*, 66, 35, 2045-2047.

Ioannidis, J.P.A. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association*, Vol. 294, No. 2, 20 pages.

Little, R.J.A. & Rubin, D.B. (2002). *Statistical analysis with missing data*. Wiley, New York.

Van Montfort, K. (2006). *Meten en Weten: statistische analyses van causale relaties*, oratie Nyenrode Business Universiteit, 10 maart 2006, 27 pages.

Ziliak, S. & McCloskey, D. (2008). *The cult of significance*. University of Michigan Press, Michigan.

*Deze publicatie betreft een oratie aan
de Erasmus Universiteit Rotterdam*

ISBN 978-90-779-0693-4

