

Prof.dr. E.W. Steyerberg

Kansen grijpen



Universiteit
Leiden

Bij ons leer je de wereld kennen

Kansen Grijpen

Oratie uitgesproken door

Prof.dr. E.W. Steyerberg

bij de aanvaarding van het ambt van hoogleraar op het gebied van

Klinische Biostatistiek en Medische Besliskunde

aan de Universiteit Leiden

op vrijdag 23 maart 2018



Universiteit
Leiden

Mijnheer de rector magnificus, hooggeleerde collegae, vrienden en familie en overige toehoorders,

Elke dag word ik vrolijk van de vele interpretaties van het woord 'kans'. Elke dag krijg ik inspiratie van de vele scherpzinnige collega's bij de afdeling Biomedical Data Sciences, tot voor kort de afdeling Medische Statistiek en Bioinformatica van het LUMC. Elke dag ben ik blij dat ik de kans gegrepen heb om naar deze afdeling te komen om nieuwe uitdagingen aan te gaan.

Wat bedoelen we met een kans? In figuurlijke zin kunnen we kansen krijgen en die grijpen; daar wil ik het vandaag met u over hebben, maar vooral ook over de meer statistische aspecten. Bereid u voor op een rollercoaster door epidemiologische, statistische, en besiskundige concepten, en vele open vragen. Bij de statistiek gaat het om het schatten van kansen op basis van empirische gegevens. Statistici voeren als heuse rekenmeesters vaak heel ingewikkelde berekeningen uit. Zo kunnen de kans berekenen op een uitkomst voor een individuele patiënt. Het eenvoudigst is hierbij de frequentistische interpretatie: we begrijpen een kans op het optreden van een onzekere gebeurtenis als de frequentie van het optreden van die gebeurtenis in een lange reeks experimenten. De kans is een getal tussen 0 en 1, of in procenten uitgedrukt tussen 0 en 100%. Als ik bijvoorbeeld de kans op een harten aas wil schatten als ik een willekeurige kaart trek uit een kaartspel met 52 kaarten, dan is de kans $1/52$, oftewel 1,92%. Als ik empirisch wil vaststellen of de kans echt $1/52$ is, moet ik dit trekken van een kaart heel vaak herhalen.

Tegenover de frequentistische interpretatie staat de Bayesiaanse kansopvatting. Hierin is een kans een subjectief begrip, waarmee iemands persoonlijke overtuiging wordt gekwantificeerd, ook op een schaal van 0 tot 100%. De persoonlijke overtuiging is een a priori aanname over de waarschijnlijkheid van een gebeurtenis, die wordt aangepast op basis van informatie die bevat is in empirische waarnemingen. Bij het trekken van een

harten aas is mijn a priori aanname dat de kans $1/52$ is, en kan ik die kans aanpassen op basis van empirische waarnemingen.

Kans of waarschijnlijkheid is een basisbegrip uit de kansrekening en statistiek dat op verschillende wijze geïnterpreteerd kan worden. Belangrijke interpretaties zijn:¹

- Frequentistisch: als relatieve frequentie of frequentiequotiënt
- Bayesiaans: als subjectieve maat van persoonlijke overtuiging

Kansen schatten

Van de theorie naar de praktijk: hoe kunnen we kansen schatten? In het medisch veld zijn kansen op ziekte belangrijk - dit is de kern van het stellen van een diagnose; en kansen voor het beloop van een ziekte - het geven van een prognose. Een diagnose is niet met 100% zekerheid te stellen omdat testen niet perfect zijn. Een prognose is niet met 100% zekerheid te geven omdat we lang niet alle factoren kennen die het beloop van een ziekte bepalen. Ik geef u graag een aantal voorbeelden van diagnostische en prognostische kansen.

Ten eerste is er een behoorlijk aantal risicowijzers op het internet te vinden die kansschattingen geven. Als u in Google zoekt met de term 'risk calculator' komt u makkelijk op een website waar u schattingen kunt krijgen voor uw eigen risico op hart-en vaatziekten.² Dit zijn prognostische kansen, met een tijdshorizon van 10 jaar. De kansberekeningen zijn gebaseerd op systematisch onderzoek onder inwoners van het stadje Framingham in de Amerikaanse staat Massachusetts. Met het invullen van enkele simpele gegevens, de bloeddruk, en de resultaten van een bloedtest op cholesterol krijgen we een kans te zien op het optreden van een hartinfarct of sterfte binnen 10 jaar. Het is een kans; geen zekerheid. En direct rijst dan de vraag: zou deze Amerikaanse kans ook voor mensen in Nederland gelden?

Lynch syndrome prediction model

MLH1, MSH2, MSH6, PMS2, and EPCAM gene mutations

The PREMM₅ model is a clinical prediction algorithm that estimates the cumulative probability of an individual carrying a germline mutation in the *MLH1*, *MSH2*, *MSH6*, *PMS2*, or *EPCAM* genes. Mutations in these genes cause Lynch syndrome, an inherited cancer predisposition syndrome.

In addition to information about the individual being evaluated, the model requires:

- A personal or family history of colorectal cancer, endometrial (uterine) cancer, or other Lynch syndrome-associated cancers
- Types of cancer and ages at diagnosis of first-degree relatives from the affected side of the family (parents, siblings, children)
- Types of cancer and ages at diagnosis of second-degree relatives from the affected side of the family (grandparents, grandchildren, aunts, uncles, nieces, nephews)

1 Patient information

Sex

- ☐ Male
☐ Female

Current age (years)

Has the patient had colorectal cancer?

- ☐ No
☐ Yes

Has the patient had any other Lynch syndrome-associated cancer?

Other Lynch syndrome-associated cancers include ovary, stomach, small intestine, urinary tract/bladder/kidney, bile ducts, brain, pancreas, and sebaceous gland skin tumors.

- ☐ No
☐ Yes

2 Relatives: First-degree

*First-degree relatives include **parents, siblings, children**, only from affected side of family*

How many first-degree relatives have had colorectal cancer?

- ☐ None
☐ One
☐ Two or more

How many first-degree relatives have had endometrial (uterine) cancer?

- ☐ None
☐ One
☐ Two or more

Have any first-degree relatives had other Lynch syndrome-associated cancers?

Other Lynch syndrome-associated cancers include ovary, stomach, small intestine, urinary tract/bladder/kidney, bile ducts, brain, pancreas, and sebaceous gland skin tumors.

- ☐ No
☐ Yes

3 Relatives: Second-degree

*Second-degree relatives are **grandparents, grandchildren, aunts, uncles, nieces, nephews**, only from affected side of family*

Schermafdruck van het PREMM model (PREMM: “PREdiction Model for gene Mutations”). Dit model helpt om te voorspellen of er een genetische mutatie aanwezig is in een patiënt met kanker. Als er een mutatie gevonden wordt, kan verder getest worden in familieleden. Indien bij hen een mutatie gevonden wordt, is intensief vervolgen geboden aangezien zij een sterk verhoogd risico op kanker hebben. Het PREMM model is uitgebreid geëvalueerd, steeds verder verbeterd, en wordt nu aanbevolen voor klinische toepassing door Amerikaanse professionele verenigingen, zoals het “National Comprehensive Cancer Network”, “American College of Gastroenterology”, en de “U.S. Multi-Society Task Force on Colorectal Cancer”.¹⁴

Laat ik u ook een aantal voorbeelden geven uit mijn eigen onderzoekspraktijk van de afgelopen 30 jaar. Ik heb door de jaren heen steeds als wetenschappelijke missie gehad om kansen te schatten, zodat we beter inzien wat het nut is van een test of behandeling voor een individuele patiënt.

Restmassa's opereren na chemotherapie voor testis carcinoom?

Ten eerste een voorbeeld van een prognostisch model, waar ik als 21-jarige student ooit aan begon te werken met twee Leidse begeleiders (dr. Jo Hermans van de Medische Statistiek en dr. Jan Keizer, de helaas te jong overleden klinisch oncoloog). Het was een heroïsche, en achteraf gezien misschien zelfs onverantwoorde, poging om de prognose te schatten van patiënten die geopereerd waren na chemotherapie voor een uitgezaaide testistumor. Er waren slechts 86 mannen in deze studie, waarvan er 11 overleden.³ Dan is het makkelijk te begrijpen dat het moeilijk is om een betrouwbare schatting te geven van de prognose in het algemeen. En nog moeilijker om aan te geven wie er een duidelijk slechtere prognose had, en wie een duidelijk betere prognose.

In latere studies combineerden we de gegevens van 6 studiegroepen, om zodoende tot adequate aantallen te komen voor statistische analyses.⁴ We maakten gedetailleerde logistische regressiemodellen om aan te geven wie er wel en wie niet geopereerd zou moeten worden na chemotherapie.

Operatierisico bij een verwijde slagader (aneurysma)

Het probleem van relatief kleine aantallen kwam ook naar voren bij mijn afstudeerproject bij de afdeling Medische Statistiek, onder begeleiding van met name dr., later prof., Ronald Brand. In een groep van 246 patiënten die geopereerd waren in het AZL overleden er slechts 18 kort na operatie. Dit kleine aantal maakt het betrouwbaar grijpen van kansen onmogelijk. Samen met prof. Hans van Houwelingen vonden we een aanpak om informatie uit de literatuur over soortgelijke patiënten te gebruiken, en zo beter de kansen te kunnen schatten en aan te kunnen geven wie er een duidelijk slechtere prognose had, en wie een duidelijk betere prognose.^{5,6} Patiënten met een sterk

verhoogd risico zouden kunnen afzien van de operatie, of eerst risicoverlagende interventies kunnen ondergaan.

Voorspellen van uitkomsten na schedelhersenletsel

Met grotere aantallen werkt het beter; we kunnen bijvoorbeeld de gegevens van verschillende studies proberen samen te analyseren in een zogenaamde meta-analyse. Dit gebeurde in het 'IMPACT' project onder leiding van prof. Andrew Maas.⁷ In dit project combineerden we de gegevens van 11 studies naar patiënten met ernstig schedelhersenletsel. Gemiddeld stierf ongeveer een kwart van de patiënten kort na het ernstige trauma. We besteedden veel tijd aan 'data management' en met name het uniformeren van definities tussen de verschillende studies. We konden vervolgens prognostisch modellen opstellen om de kans te schatten om 6 maanden na het oplopen van het trauma nog in leven te zijn, waarbij we de gegevens gebruikten van bijna 10.000 patiënten.⁸ Het resulterende IMPACT model is samen met het CRASH model, dat ook gebaseerd is op ongeveer 10.000 patiënten, een doorbraak geweest voor het schatten van de prognose na hersenletsel.⁹

Wie testen voor een erfelijke aandoening?

En ook voor diagnostiek werken grote aantallen beter. In Boston werkte ik met gastro-enterologen die wilden weten wie ze moesten testen op een erfelijke aanleg voor darmkanker. Hierbij speelt de familiegeschiedenis een belangrijke rol. In 2003 hadden we de beschikking over de gegevens van ongeveer 1.000 patiënten met darmkanker. Bij ongeveer 15% werd een erfelijke oorzaak voor de kanker vastgesteld; het Lynch syndroom. Twee jaar later hadden een nieuwe serie van nog eens 1.000 patiënten, zodat we onze modellen konden valideren op deze meer recente patiënten.¹⁰ Daarna werden er nog veel meer patiënten getest, zodat we meer dan 20.000 patiënten konden gebruiken voor de meest recente analyses.^{11,12,13} Zo kunnen we inmiddels betrouwbare schattingen geven over wie een laag en wie een hoog risico heeft op een erfelijke oorzaak als onderliggende verklaring voor de darmkanker. Dit model wordt gepresenteerd op een veel bezochte website van het Dana Farber Cancer Center.¹⁴

PROSTAATWIJZER

Prostaatwijzer

Over ons

Taal

Begin opnieuw

Wat is de PSA waarde in ng/mL?

Deze waarde moet tussen 0,4 en 50 liggen.

3

Is er een eerder negatief biopt?

JA

NEE

Is er in de afgelopen 6 maanden een rectaal toucher verricht?

JA

NEE

Bereken

Wat is de PSA waarde in ng/mL?

Prostaat-specifiek antigeen (PSA) is een enzym dat uitsluitend in de prostaat wordt geproduceerd. Er is een simpele bloedtest waarmee de PSA waarde kan worden gemeten. Dit kan helpen bij het ontdekken van prostaatkanker in een vroeg stadium. De kans op prostaatkanker wordt groter naarmate de PSA waarde stijgt. Toch hoeft een hoge PSA waarde niet altijd te betekenen dat een man prostaatkanker heeft. Een verhoogde PSA waarde kan ook worden veroorzaakt door een goedaardige vergroting van de prostaat (BPH), een urineweginfectie of een ontsteking van de prostaat.

Het risico op het detecteren van prostaatkanker bij biopsie

De kans op een positieve biopsie bepaalt de volgende stap.

Kans kleiner dan 12.5%:

Geen biopsie.

Kans tussen de 12,5% en 20,0%:

Overweeg biopsie, afhankelijk van comorbiditeit en indien het risico op agressieve prostaatkanker groter is dan 4%.

Kans groter dan 20,0%:

Prostaatbiopsie.

UITSLAG

17%

Het risico op het detecteren van prostaatkanker bij biopsie

6%

Risico op hooggradig of gevorderd prostaatkanker

Begin opnieuw

Risico op hooggradig of gevorderd prostaatkanker

Er is sprake van een klinische relevante prostaatkanker als het tumorstadium groter is dan T2b of als de Gleason score gevonden bij prostaatbiopsie groter of gelijk is aan 7.

6

Schermafdrucken van prostaatwijzer #3 voor het schatten van de kans op het detecteren van (hooggradig of gevorderd) prostaatkanker voor een hypothetische patiënt met een PSA waarde van 3 ng/ml.¹⁶

PROF.DR. E.W. STEYERBERG

Welke mannen hebben een gevaarlijke prostaatkanker?

Een laatste voorbeeld: Screening van prostaatkanker is een controversieel onderwerp, waarbij het denken in kansen en risico's essentieel is. Het lukt niet goed om alleen op basis van het prostaat specifiek antigeen (PSA) mannen te vinden met een hoog risico op een gevaarlijke kanker. Met alleen PSA als screeningtest vinden we vrij veel afwijkingen die we prostaatkanker noemen, maar die nooit een probleem zouden gaan vormen voor de betreffende mannen (overdiagnose). Er worden dan veel mannen behandeld die daar geen baat bij hebben (overbehandeling). Het probleem van overdiagnose en overbehandeling kan worden beperkt door betere voorspellers te gebruiken. Met prof. Monique Roobol deed ik jarenlang onderzoek naar het gebruik van betere markers om die mannen te identificeren die werkelijk baat hebben bij screening. Ook willen we mannen identificeren voor wie behandeling van de kanker kan worden uitgesteld of zelfs afgesteld ('Active Surveillance').¹⁵ We ontwikkelden een serie modellen voor diverse, inmiddels goedbezochte, websites.¹⁶

Samenvattend:

- Kansen schatten is van belang voor diagnostiek en prognose
- Schatten gaat beter met grotere aantallen
- Schatten gaat beter met sterkere voorspellers

Kansen met en zonder behandeling

Wat hebben we aan een geschatte kans? Waarom zouden we een kans willen grijpen? Ten eerste willen we vaak graag begrijpen welke factoren de kansen bepalen: welke factoren hangen samen met een kans op ziekte of het beloop van ziekte? Liever nog gaan we verder dan associaties vaststellen, namelijk het claimen van een causale relatie. Dat geeft ons pas echt inzicht in ziekteprocessen en houdt de belofte in van aanknopingspunten voor betere behandelingen. Het onderzoeken van causale relaties is een mijnenveld waar allerlei biases onze interpretatie van het empirisch zichtbare beeld van de biologie verstoren.

Als we ons alleen op de kansschattingen richten, zien we dat de meer kwantitatieve benadering steeds meer opgang doet zowel rond diagnose als prognose, zoals geïllustreerd met de eerdere voorbeelden en de risicowijzers die steeds meer op internet te vinden zijn. De klassieke redenering voor diagnostiek: eerst de ziekte weten, want dan kunnen we de juiste behandeling kiezen. Dit is een hoeksteen van de geneeskunde. Voor prognose is de redenering: "als de prognose slecht is moeten we er iets aan doen". De winst van behandeling hangt af van de prognose: bij een slechtere prognose is er in absolute zin meer behandelwinst te behalen, bij een goede prognose minder.¹⁷ Voor zowel diagnose als prognose is het daarom belangrijk om kansen te schatten. Voor het effect van behandelingen constateer ik dat we van veel interventies niet precies weten hoe ze precies werken. Het is al heel mooi als we weten dat ze werken op groepsniveau.

Gerandomiseerde studies

Solide bewijs op groepsniveau komt veelal het best uit een gerandomiseerde studie. Hierin worden twee groepen deelnemers gevormd die door het lot ingedeeld worden voor de ene of de andere behandeling. Er is dan geen systematische toewijzing door een arts, die de ene patiënt misschien liever de ene dan de andere behandeling had gegeven. Door de toevallige indeling ontstaat vergelijkbaarheid van de twee groepen. Er zijn veel moeilijkheden bij dit soort studies, meest praktisch

van aard. Veel gerandomiseerde studies zijn op zichzelf te klein om overtuigende resultaten te laten zien. Ik wil deze gelegenheid gebruiken om een publieke bekentenis te doen: ik beken u hier dat ik een track record heb van een behoorlijk aantal mislukte gerandomiseerde studies. Ik heb teruggezocht bij hoeveel gerandomiseerde studies ik betrokken was gedurende de afgelopen 30 jaar; van de 23 hebben er 8 overtuigende, positieve resultaten opgeleverd over een behandeling. Ik kan de gerandomiseerde studies indelen in ‘teleurstellingen’: studies die we enthousiast begonnen, in de hoop een doorbraak op het spoor te zijn; ‘successen’, de 8 studies die een overtuigend resultaat lieten zien; en een tussencategorie, waar de resultaten per se niet overtuigend waren maar wel bij hebben gedragen aan de voortschrijdend wetenschappelijk inzicht.

Teleurstellingen waren er onder andere bij de volgende studies:

- We deden twee trials bij kinderen met koorts en een eerdere koortsstuipt. Hieruit bleek dat ibuprofen koorts beter verlaagt dan paracetamol,¹⁸ maar in een placebo-gecontroleerde vergelijking traden niet minder nieuwe koortsstuipten door ibuprofen.¹⁹
- Veel moeite is geïnvesteerd in de ESES trial bij patiënten met kritieke ischemie van de benen. Spinale elektrische stimulatie hielp niet tegen het risico op amputatie.²⁰
- Een kleine studie liet zien dat een speciale klep niet werkte tegen reflux in de slokdarm.²¹
- Op de ‘intensive care’ helpt glutamine in parenterale voeding niet voor betere uitkomsten in jonge aan hun darm geopereerde kinderen.²²
- Betere besluitvorming kunnen we proberen te ondersteunen met geautomatiseerde systemen. Het lukte in twee trials naar kinderen met koorts op de spoedeisende hulp niet om positieve effect op de zorg te laten zien.^{23,24} We zagen wel enige positieve effecten in ORS gebruik in een trial bij kinderen met een acute darmondsteking.²⁵
- Bij kinderen met cerebrale parese verhogen botuline injecties de kansen op beter functioneren niet vergeleken met alleen intensieve fysiotherapie.²⁶

In de tussencategorie vallen een aantal trials:

- Bij patiënten met een instabiele heupfractuur bleek er weinig verschil tussen een nieuwe pin en de ‘gamma nail’.²⁷
- Palliatie van slokdarmkanker kan goed verzorgd worden door lokale bestraling van de tumor, of door het plaatsen van een stent. De SIREC trial toonde aan dat stents alleen aantrekkelijk waren in plaats van bestraling bij een slechte prognose.^{28,29} De analyse naar prognose was ook relevant bij de vergelijking van stents met chirurgie bij maagkanker in de SUSTENT trial.³⁰
- Een andere stent studie liet zien dat bij verstopping van de galwegen plastic of metalen stents gebruikt kunnen worden zonder overtuigende voor- of nadelen.³¹ Ook vonden we weinig verschillen tussen 3 typen stent voor plaatsing in de slokdarm (Ultraflex stent vs Polyflex stent vs Niti-S stent).³²

Tenslotte de ‘succes’ categorie; 8 gerandomiseerde trials met overtuigende resultaten.

- NT-proBNP is een marker die kan helpen om onderscheid te maken tussen cardiale en pulmonale problemen bij patiënten die zich op de spoedeisende hulp presenteren met kortademigheid. In een gerandomiseerde trial toonden we duidelijke voordelen aan van het snel bepalen van deze diagnostische marker.³³
- De follow-up van patiënten met kanker bleek in 2 trials goed overgenomen te kunnen worden door verpleegkundigen.^{34,35} We vonden betere of vergelijkbare uitkomsten voor de patiënten en hun mantelzorgers.
- Drie chirurgische trials waren succesvol. Bij liesbreukoperatie bleek de TEP procedure het best in de LEVEL trial.³⁶ Littenbreuken na electieve abdominale chirurgie kunnen worden voorkomen door kleinere steken te gebruiken (STITCH trial),³⁷ en een chirurgisch matje (PRIMA trial).³⁸
- Een groot succes is de CROSS trial bij slokdarmkanker patiënten, waar we aantoonde dat een slim gekozen chemotherapie regime weinig bijwerkingen heeft in combinatie met bestraling, en gevolgd door chirurgie een veel

betere kans op lange termijn overleving geeft dan chirurgie alleen.^{39,40}

- Ook een groot succes is de MRCLEAN trial bij patiënten met een herseninfarct. Voor deze patiënten leidt intra-arteriële therapie door veel betere uitkomsten dan standaard behandeling.⁴¹

Deze laatste categorie bevat wetenschappelijk gezien echte doorbraken, die de patiëntenzorg direct ten goede komen. Patiënten met een beroerte hebben nu een betere uitkomst met de intra-arteriële therapie, en patiënten met slokdarmkanker een betere overleving met een combinatie van chemotherapie, bestraling, en operatie.

Bij de mislukte trials kunnen we teruggrijpen op het dogma: meer is beter. Waren de studies simpelweg te klein om op zichzelf tot betrouwbare conclusies te kunnen leiden? We kunnen soms een analyse uitvoeren waarin we resultaten van verschillende studies combineren: een meta-analyse. De noodzaak tot combinatie van bewijs uit verschillende trials komt voort uit het feit dat de verschillen in uitkomsten tussen patiënten met verschillende therapie veelal veel kleiner zijn dan de verschillen in uitkomsten tussen patiënten met verschillende prognostische factoren, zoals leeftijd en geslacht, ziektegeschiedenis en andere kenmerken. De relatief kleine effecten van behandeling zijn alleen goed vast te stellen met grote aantallen.

Samenvattend:

- Effecten van behandeling stellen we vast op groepsniveau
- Sommige gerandomiseerde studies zijn succesvol: ze laten betere uitkomsten voor patiënten zien en hebben een directe invloed op de medische praktijk
- Gerandomiseerde studies zijn vaak te klein om duidelijk bewijs van effect van een behandeling aan te tonen
- Verschillen tussen patiënten met verschillende therapie zijn veelal veel kleiner dan de verschillen tussen patiënten met verschillende prognostische factoren

Grotere aantallen: *Big Data*

In deze rede kan ik niet voorbijgaan aan een belangrijk opkomend thema: *Big Data*. Een definitie is niet makkelijk te vinden. Ik beperk me hier tot *Big Data* als een grote set gegevens: observationele data die niet voortkomen uit een systematische studie om onderzoek te doen. *Big Data* gaat verder dan observationele studies zoals we die al veel langer kennen, bijvoorbeeld de Nederlandse kankerregistratie, of de meer recent opgezette DICA database van het “Dutch Institute for Clinical Auditing”. Kan *Big Data* dienen als bron van informatie voor kansen op een ziekte, het beloop van een ziekte, en voor het effect van behandeling?

Succesverhalen rond *Big Data* zijn er zeker: ik gebruik de Google zoekmachine dagelijks om informatie op het internet te vinden. Ik gebruik ook veel “Google Scholar”, de zoekmachine die met name wetenschappelijke informatie zoekt, onder het mooie motto: “Staan op de schouders van reuzen”. Nog steeds heel nuttig vind ik de PubMed zoekmachine, die al sinds 1997 miljoenen medische artikelen toegankelijk maakt voor iedereen die op zoek is naar medische vakliteratuur.

De grote aantrekkingskracht van *Big Data* ligt in het woord **Big**. Bij alle voorbeelden die ik u tot nu toe gegeven heb speelde de grootte van de studie een rol; het is de eerste zorg van de statisticus dat toeval overal is, en ons makkelijk kan misleiden. Is *Big Data* de panacee? Of brengt *Big Data* andere misleiding dan misleiding door kleine aantallen?

Samenvattend:

- *Big Data* is aantrekkelijk voor het verkleinen van de rol van toeval
- *Big Data* houdt een belofte in meer te leren over prognose en effectieve behandeling

Andere onzekerheid: bias (vertekening)

Grote aantallen zorgen voor minder invloed van toeval. Dit verschuift onze aandacht naar bronnen van systematische vertekening ('bias'). Observationele data sets zijn vaak groot, en dat is op zich gunstig, maar brengen grote uitdagingen met zich mee als het gaat om het leren over oorzakelijke (causale) relaties van risico factoren voor ziekte, of het effect van een behandeling op het beloop van een ziekte. Klassieke biases gaan over informatie, de selectie van patiënten, en vertekende factoren ('confounding'). Deze biases treden alle in sterkere mate op bij *Big Data* dan bij meer klassieke epidemiologische studies.

Als voorbeeld van bias rond behandel-effecten noem ik een recente studie naar de beste behandeling voor patiënten met multipale sclerose.⁴² De onderzoekers analyseerden een grote registratie waarin patiënten gevolgd werden terwijl ze behandeld werden met 1 van 6 bekende middelen. De onderzoekers wilden weten welk middel het best werkt voor welke patiënt. Dat is ambitieus, maar niet ongebruikelijk, gemotiveerd door de term "Precision Medicine"⁴³, een variant op termen als "Personalized Medicine" en "Stratified Medicine".

Bij een dergelijke studie spelen een aantal grote problemen. De gegevens werden niet systematisch verzameld: met allerlei verschillende meetinstrumenten; in verschillende ziekenhuizen in verschillende landen. Als een geneesmiddel niet goed werkte werd overgegaan naar een ander middel, en eventueel naar een derde als het tweede middel ook niet goed werkte. Dit veroorzaakt een complexe selectie van patiënten, die moeilijk volledig in statistische modellen te vangen is. Tenslotte gebruikten de onderzoekers geavanceerde statistiek, maar op een suboptimale manier. Het eindproduct was een webpagina waarop de prognose geschat kon worden op basis van individuele gegevens zoals leeftijd, geslacht, en ernst van de ziekte, in combinatie met het specifieke medicijn voor behandeling. Gezien de vele biases in deze studie is deze prognose schatting onbetrouwbaar, en vermoedelijk zelfs misleidend.⁴⁴

Precision Medicine (NIH definition):

"an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person." This approach will allow doctors and researchers to predict more accurately which treatment and prevention strategies for a particular disease will work in which groups of people. It is in contrast to a one-size-fits-all approach, in which disease treatment and prevention strategies are developed for the average person, with less consideration for the differences between individuals.

Samenvattend:

- Bias is een bedreiging voor de validiteit van vele epidemiologische studies, en met name analyses met Big Data
- Het schatten van op de persoon gefocuste behandel-effecten ("Precision Medicine") uit Big Data is vaak omgeven met zo veel bias dat we uitkomen op misleiding

Nieuwe analyses: “Machine Learning”

In het voorbeeld dat ik net besprak werd gebruik gemaakt van niet-standaard statistiek voor het analyseren van uitkomsten van behandelingen. Dit is een ontwikkeling die we systematisch zien: *Big Data* analyseren we met *Machine Learning* technieken. Hierbij wordt gebruikt gemaakt van technieken met fancy namen, zoals ‘random forests’, ‘support vector machines’, en ‘deep learning’. Deze technieken worden vaak gepresenteerd als tegengesteld aan klassieke technieken zoals regressie analyse. Regressie is maar ouderwets. Toch zijn er duidelijke relaties tussen de verschillende technieken aan te wijzen, en zijn er moderne varianten van klassieke regressietechnieken (“ridge” of “penalized regression”, LASSO^{45,46}) die ook als *Machine Learning* op te vatten zijn.⁴⁷

Machine Learning beroept zich op een modern dogma: hypothese vrij leren, laat de data voor zich spreken. Dat zou in principe kunnen werken indien we met hele grote data sets van hoge kwaliteit aan de slag gaan. Er zijn successen met ‘deep learning’ geboekt bij het interpreteren van beelden die door een radioloog of patholoog worden beoordeeld voor de diagnostiek van allerlei ziekten. Bij medische vragen is het echter veelal verstandig om enige externe kennis te gebruiken over hoe ziekteprocessen werken en welke modelstructuur redelijk is om mee te starten. In verschillende studies met relatief eenvoudige data structuren zagen we geen enkel voordeel van modernere technieken om kansen mee te schatten vergeleken met meer klassieke technieken.⁴⁸ We zagen wel heel veel overoptimisme; misleiding ligt alweer op de loer.

Samenvattend:

- Machine learning is populair voor de analyse van Big Data
- Machine learning lijkt in veel opzichten op klassieke statistische technieken, en heeft geen voordeel in situaties met relatief eenvoudige data structuren (waar klassieke technieken al goed werken)

De dilemma’s voor *Precision Medicine*

In het bovenstaande heb ik u meegenomen langs enkele aspecten bij het grijpen van kansen voor diagnose, prognose, en behandel-effecten; van het dogma ‘meer is beter’, naar de risico’s van bias en misleidingen met moderne terminologie als *Big Data* en *Machine Learning*. Hoe komen we dan tot de beste schattingen voor het nut van een behandeling voor een individuele patiënt?

Ten eerste willen we goed weten welke behandelingen werken, al is het slechts op een gemiddeld niveau. “Evidence-based medicine” is een beweging die zich heeft ingezet voor het systematisch gebruiken van bewijs in de geneeskunde, met een sterke focus op gerandomiseerde studies en het combineren van bewijs uit verschillende studies.¹⁷

Analyse van gerandomiseerde studies: behandel-effect

De meta-analyse van gerandomiseerde studies de hoeksteen van de wetenschappelijke activiteiten op het gebied van “Evidence-based medicine”, m.n. door de Cochrane Collaboration.⁴⁹ Toen ik bij mijn epidemiologische scholing voor het eerst over meta-analyse hoorde, leek mij het nut van een dergelijke analyse een “no-brainer”. Ik geloofde namelijk heilig in het dogma: meer is beter. Een schatting gebaseerd op meer studies is daarom beter dan een schatting van een behandel-effect gebaseerd op minder studies. Ik herkende wel dat er soms systematische verschillen kunnen zijn tussen studies, waardoor er net andere onderliggende behandel-effecten kunnen bestaan per studie. Een elegante oplossing is dan om een zogenaamd “random effects” model te gebruiken. Hierin wordt toevallige variatie binnen studies onderscheiden van systematische variatie in het behandel-effect tussen studies. Pas meer recent lig ik echt wakker van random effect meta-analyse: wat schatten we nu eigenlijk als er systematische verschillen zijn tussen studies? Heeft het wel zin om een meta-analyse uit te voeren als er heterogeniteit is? Hoe kunnen we een effect schatting van heterogene studies gebruiken voor een individuele patiënt?

Analyse van gerandomiseerde studies: subgroepen

In plaats van over studies heen, kunnen we binnen gerandomiseerde studies meer in detail proberen te leren van het effect van behandeling. Het behandel-effect zal niet voor alle patiënten hetzelfde zijn. Daarom is het gebruikelijk om analyses uit te voeren in subgroepen, bijvoorbeeld op basis van geslacht: is het relatieve effect voor mannen anders dan voor vrouwen? Recent analyseerden we de betrouwbaarheid van dergelijke subgroep analyses. Voor analyses naar geslacht was de situatie dramatisch: er werden nauwelijks meer statistisch significante verschillen in relatieve effecten gevonden dan op basis van toeval verwacht zou worden als er in feite geen verschillen waren.⁵⁰ Dit ondermijnt mijn vertrouwen in subgroep analyses als methode om meer individuele effecten van behandeling te schatten.⁵¹

Een verklaring is dat de verschillen in effecten relatief klein zijn, en dat er daarom veel grotere studies nodig zijn om subgroep effecten betrouwbaar vast te stellen.⁵² De huidige praktijk is echter dat subgroepen worden onderzocht in gerandomiseerde studies die al te klein zijn om een gemiddeld effect van behandeling betrouwbaar vast te stellen. Dit is een recept voor misleiding.

Analyse van gerandomiseerde studies: complexere modellering

Tenslotte kunnen we binnen gerandomiseerde studies meer complexe statistische modellen gebruiken dan subgroep analyses. Subgroep analyses kijken naar 1 factor, terwijl we weten dat de prognose van patiënten door een combinatie van vele factoren bepaald wordt. Het eenvoudigste model noemde ik al eerder: we modelleren de prognose met een set kenmerken, en schatten daarbij 1 relatief effect van behandeling. Dit leidt dan tot heel verschillende kansen van succes door behandeling per profiel van kenmerken. Een effect van behandeling dat *relatief* constant is, leidt tot een effect dat in *absolute* zin heel verschillend is als de kansen op uitkomsten verschillen tussen patiënten.⁵³ Dergelijke prognostische heterogeniteit is voldoende voor verschillen in absolute behandel-effecten; heterogeniteit in relatieve behandel-effecten is geen vereiste, in tegenstelling tot hetgeen vele aanhangers van *Precision Medicine* denken. Zij zullen complexere varianten toejuichen, die we inderdaad ook

in gerandomiseerde studies kunnen onderzoeken, namelijk het laten variëren van de relatieve behandel-effecten met patiëntkenmerken.^{54,55,56} Hierbij geldt dezelfde beperking als bij subgroep analyse: een gerandomiseerde studie is vaak te klein voor betrouwbare schatting van al deze effecten.⁵⁷

Analyse van observationele data: behandel-effect

Gerandomiseerde studies zijn duur en mislukken vaak. Het lijkt daarom aantrekkelijk om uit bestaande observationele data te proberen te leren wat het effect van behandeling is. Ten eerste het gemiddelde effect. Zoals ik al besprak worden we hier geteisterd door allerlei vormen van vertekening; en het is maar de vraag in welke mate we hier voor kunnen corrigeren. Een veelbelovende richting is om te proberen te leren van verschillen tussen zorgaanbieders. Als in het ene ziekenhuis altijd één variant van behandeling wordt gekozen, en in een ander ziekenhuis systematisch een andere behandeling, kunnen we de uitkomsten goed vergelijken. Een voorbeeld is de behandeling van bepaalde hersentumoren (laaggradige gliomen). Twee Noorse centra hadden een heel ander beleid, namelijk vroeg opereren tegenover eerst afwachten. De kansen op overleving waren beter in het meer agressieve ziekenhuis; een belangrijke bevinding.⁵⁸ Echter, indien de keuze voor een behandeling afhangt van patiëntkenmerken is het maar de vraag in hoeverre een eerlijke vergelijking mogelijk is; bias ligt weer op de loer, en geen enkele analysetechniek kan echt de oplossing bieden.

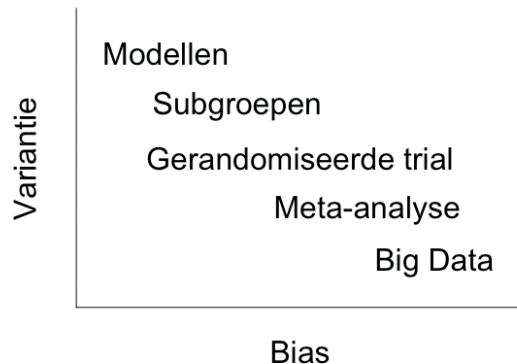
Analyse van observationele data: prognostische modellen

Routinematig verzamelde data kunnen wel heel geschikt zijn om predictiemodellen mee te maken. De data komen uit minder geselecteerde groepen, en de aantallen zijn meestal groter dan in gerandomiseerde studies. Een specifieke kans ligt in het bestuderen van verschillen tussen settings, bijvoorbeeld tussen ziekenhuizen of tussen landen. Idealiter is een prognostisch model goed generaliseerbaar, dat wil zeggen: valide in verschillende omstandigheden. Dit valt in de praktijk vaak tegen helaas. Zo zagen we bij de prognose van schedelhersenletsels behoorlijke verschillen tussen de gemiddelde uitkomst in verschillende studies.^{8,48} Deze

verschillen konden we niet verklaren door verschillen in samenstelling van de patiëntengroepen. Het is dan eigenlijk onmogelijk om te spreken van “de prognose” van een patiënt; de prognose hangt af van allerlei factoren die we niet kennen.

Precision Medicine: wat is het effect voor een individuele patiënt?

Als we niet kunnen spreken van ‘het behandel-effect’ op basis van verschillende gerandomiseerde studies, en ook zien dat we niet kunnen spreken van ‘de prognose’ van een patiënt, hoe kunnen we dan iets zinnigs zeggen over het effect van behandeling voor een individuele patiënt? Systematische verschillen tussen studies maken het ons lastig om die kans te grijpen.

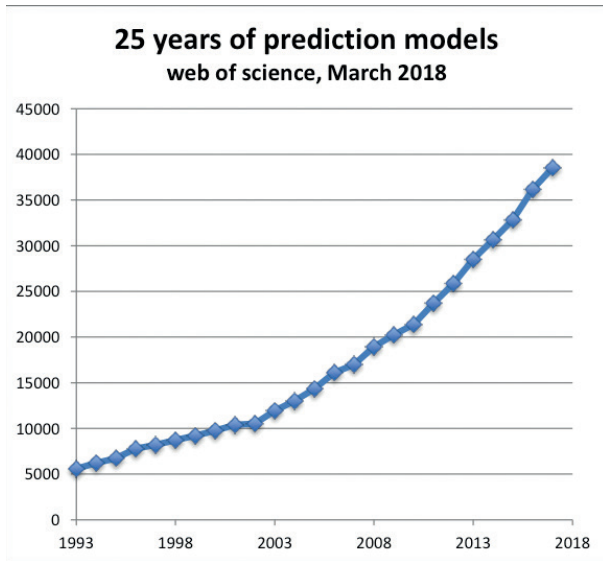


Schematische weergave van de afweging van variantie tegenover bias (vertekening) bij het schatten van een behandel-effect voor een individuele patiënt. Een gerandomiseerde trial heeft naar verwachting geen bias op gemiddeld niveau. Subgroep analyses geven in potentie minder bias, maar een hogere variantie. Nog minder bias zou bereikt kunnen worden met complexere modellering, waarbij de rol van toeval echter nog groter wordt. Minder variantie treedt op bij de analyse van meerdere trials in een meta-analyse, of met *Big Data*, maar deze analyses gaan ten koste van een hoger risico op bias.

Onderzoeksagenda

Nu we de dilemma's helder hebben is het tijd voor een onderzoeksagenda.

- De komende jaren zal de beschikbaarheid van steeds meer en diversere data toenemen. Helder houden voor welke onderzoeksvragen deze data wel en niet gebruikt kunnen worden is volgens mij een centrale missie van de afdeling *Biomedical Data Sciences* van het LUMC. In welke situaties kunnen we Big Data goed gebruiken?
- De toename van biologische kennis zal leiden tot het beschikbaar komen van steeds meer ‘markers’ die geassocieerd zijn met het optreden van ziekte en het beloop van ziekte. Deze ontwikkelingen vragen om het steeds verder verfijnen van statistische methoden. Een aspect hierbij is om *Machine Learning* en klassieke statistiek niet tegenover elkaar te plaatsen maar in elkaars verlengde. Hoe kunnen *Machine Learning* en statistische perspectieven elkaar versterken?
- Mijn oratie in 2007 aan de Erasmus Universiteit Rotterdam had als titel: “Kansrijk beslissen”. Hierin was mijn centrale stelling dat predictiemodellen essentieel zijn voor het nemen van betere beslissingen: “Goed beslissen vraagt om goed voorspellen”. Tot mijn grote vreugde zien we een enorme toename in het aantal en in de kwaliteit van gepubliceerde predictiemodellen. Werden er in 1995 nog ongeveer 7.000 predictie modellen gepubliceerd, in 2005 was dat verdubbeld tot 14.000, en weer 10 jaar later, in 2015 werden er 32.000 gepubliceerd. Er is veel methodologisch onderzoek gedaan, dat onder andere met de “TRIPOD” richtlijn (“Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis”) meer onder de aandacht van toegepast onderzoekers komt.⁵⁹ Ik ben daarom positief gestemd over het ontwikkelen en valideren van predictiemodellen, maar verder methodologisch en toegepast onderzoek is nodig.⁶⁰ Met name is een grote vraag wanneer een predictiemodel wel of niet valide is voor een bepaalde setting? Beter begrip van verschillen tussen studies is gewenst.⁶¹



De toename in jaarlijks gepubliceerde predictiemodellen tussen 1993 en 2018 zoals vastgelegd in de database van Web of Science.

- Vandaag heb ik geprobeerd met u stil te staan bij een grotere ambitie dan het schatten van de prognose, namelijk het schatten van een geïndividualiseerd behandel-effect, in lijn met de ambitie van *Precision Medicine*.⁴³ Hoever kunnen we daar mee komen? Moeten we vasthouden aan een robuuste aanpak, namelijk de combinatie van een predictiemodel voor absolute kansen met het relatieve behandel-effect uit een gerandomiseerde trial; of zijn er omstandigheden waarin we echt verder kunnen gaan? Hoe kunnen observationele data ons hierbij helpen? Welke statistische of *Machine Learning* methoden zijn hierbij nuttig? Als we modellen voor geïndividualiseerde winst door behandeling hebben, kunnen we deze dan aanbieden in een geautomatiseerde omgeving, zodat het zelflerende systemen worden? Onder welke voorwaarden en op welke manier kunnen dit soort voorspelmodellen de klinische praktijk werkelijk

gaan verbeteren? Ik verheug mij erop om in samenwerking met vele anderen met deze en gerelateerde vragen aan de slag te gaan de komende jaren; zo gaan we op weg van *Big Data* naar *Big Science*.⁶²

Samenvattend:

- Waarvoor kunnen we *Big Data* wel, en waarvoor niet goed gebruiken?
- Welke statistische methoden kunnen de toenemende biologische kennis gebruiken voor betere voorspellingen van individuele behandel-effecten?
- Wat bepaalt of een predictiemodel wel of niet valide is voor een bepaalde setting?
- Hoe kunnen we goed invulling geven aan *Precision Medicine*, waarbij we het geïndividualiseerde behandel-effect willen schatten?

Academische omgeving

De afdeling Biomedical Data Sciences biedt prachtige kansen om deze onderzoeksagenda vorm te geven. De kennis opgebouwd in het gebied van Moleculaire Epidemiologie kan gebruikt gaan worden voor betere voorspellingen en inzicht in behandelmechanismen. De groep van prof. Eline Slagboom, met haar uitmuntende collega's waaronder prof. Ingrid Meulenbelt en dr. Bas Heijmans, doet hier onderzoek op topniveau. Allerlei aspecten rond dataverzameling en datakwaliteit staan centraal in de groep "Advanced Data Management", die door prof. Ronald Brand tot grote bloei is gebracht, en nu voortvarend door Karin van der Pal wordt geleid. De Medische Statistiek groep heeft internationale experts in haar midden, specifiek de hoogleraren Hein Putter, Jelle Goeman, Saskia le Cessie en Jacco Wallinga, met wie het een genot is om samen te werken en van te leren. De Medische Besliskunde groep, met Anne Stiggelbout als hoogleraar, is voor mij de brug van biostatistiek naar de medische praktijk. "Shared decision making", implementatie, "e-health", kosteneffectiviteit en kwaliteit van zorg zijn allen gerelateerd aan het centrale thema van geïndividualiseerde behandeling.

Daarbij prijs ik mij gelukkig met de uitstekende contacten met de afdeling Klinische Epidemiologie, waarbij bijvoorbeeld Saskia le Cessie en onze rising star Rolf Groenwold een dubbelaanstelling hebben bij de afdelingen Biomedical Data Sciences en Klinische Epidemiologie. Samen gaan we zo spoedig mogelijk een Centrum voor Kwantitatieve Geneeskunde vormen, dat een nog meer vooraanstaande internationale rol zal spelen op het gebied van het ontwikkelen en toepassen van epidemiologische en statistische methodologie. Hierbij wil ik graag de brug naar de Medical Delta maken; er zijn al samenwerkingen met de TU Delft rond computationele biologie (prof. Marcel Reinders) en met het Erasmus MC (de afdelingen Maatschappelijke Gezondheidszorg en Biostatistiek, met name met prof. Dimitris Rizopoulos). Binnen en buiten Nederland verheug ik mij op het vormgeven van verdere samenwerking rond methodologisch georiënteerde projecten.

Samenvattend

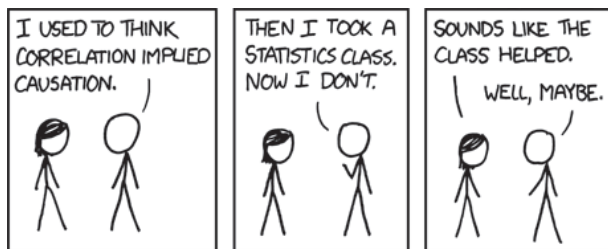
- Het LUMC biedt een prachtige omgeving om onderzoek naar individuele behandel effecten vorm te geven, met een sterke afdeling Biomedical Data Sciences
- Voor nog meer succes is een bundeling van krachten wenselijk op lokaal niveau (een Centrum voor Kwantitatieve Geneeskunde met de afdeling Klinische Epidemiologie), nationaal (in Medical Delta context en breder) en internationaal.

Onderwijs

Uit het bovenstaande volgt een aantal implicaties voor het onderwijs. We moeten studenten een kwantitatieve denkwijze bijbrengen. Dit is voor de meeste mensen onnatuurlijk, en daarom moeilijk. Mensen denken nu eenmaal makkelijker in simpele heuristieken, in regeltjes, dan in expliciete kansen.⁶³ Toch vraagt beter beslissen om een meer expliciete afweging van de voor- en nadelen van een keuze.

Voor onderzoekers en klinici is een gedegen kennis van kernbegrippen uit de biostatistiek essentieel, te beginnen bij bias, onzekerheid, en vooruit, ook iets over de p-waarde; en voorzichtigheid rond causale interpretaties van statistische associaties. Verder moet het belang van goede data kwaliteit steeds weer benadrukt worden. Mijn afdeling is zeer actief in het onderwijs op deze thema's, en blijft dat van harte doen.

16



Samenvattend

- Onderwijs moet gericht zijn op een kwantitatieve denkwijze zodat betere beslissingen genomen kunnen worden
- Onderwijs moet op zijn minst kernbegrippen uit de biostatistiek bijbrengen

Dankwoord

Tot slot wil ik allen bedanken die hebben bijgedragen aan mijn wetenschappelijke ontwikkeling tot nu toe, en die dat hopelijk ook nog geruime tijd blijven doen. Ik kan slechts een enkeling hier specifiek noemen.

Uiteraard denk ik aan collega's van de afdeling Maatschappelijke Gezondheidszorg van het Erasmus MC, waarmee ik ruim 25 jaar gewerkt heb. Ik bewaar uitstekende herinneringen aan de eerste jaren daar, waarin ik veel met René Eijkemans (inmiddels hoogleraar in het UMCU) optrok, en het werken met vele promovendi en collega klinisch onderzoekers. Ik ben dank verschuldigd aan teveel personen in Rotterdam om hier allen op te noemen. Toch noem ik specifiek mijn directe collega's bij het Centrum voor Medische Besliskunde, waaronder mijn meer dan uitstekende opvolger in Rotterdam, Hester Lingsma, en David van Klaveren, die mij gevolgd heeft naar Leiden.

Uiteraard bedank ik bij deze ook mijn leermeester van het eerste uur, Dik Habbema. Zijn humorvolle werkwijze blijft een inspirerend voorbeeld. Ik denk ook aan de vele andere collega's in het Erasmus MC, met name bij de afdeling Epidemiologie en Biostatistiek (Dimitris Rizopoulos); en vele collega's daarbuiten, onder andere bij het Julius Centrum in Utrecht (Carl Moons), epidemiologen in Amsterdam (Patrick Bossuyt) en in Boston bij het Dana-Farber/Harvard Cancer Center (Sapna Syngal) en Tufts University (David Kent). Ik heb veel geleerd van de klinici met wie ik nu al vele jaren samenwerk, die ik hier niet allemaal kan noemen, maar waar zeker bij horen Henriëtte Moll, Peter Siersema en Andrew Maas.

Door de jaren heen heb ik altijd een zwak gehouden voor Leiden. In deze stad ging ik in 1985 studeren, eerst een jaar geneeskunde, daarna biomedische wetenschappen, waarbij ik het laatste jaar bij de Medische Statistiek doorbracht. Onnavolgbaar waren de lunchgesprekken met Hans van Houwelingen.

In mijn oratie in 2007 heb ik hem al expliciet bedankt, maar ik ben heel blij dat ik hem bij deze nogmaals kan bedanken voor alle wijze lessen en inspiratie. Uiteraard bedank ik ook mijn twee directe voorgangers hierbij: Theo Stijnen, voormalig afdelingshoofd van de afdeling Medische Statistiek en Bioinformatica; en Job Kievit, voormalig hoofd van de Medische Besliskunde groep. Dankzij jullie staan er nu grote en sterke onderzoeksgroepen, die een mooie toekomst tegemoet gaan in de afdeling Biomedical Data Sciences. Dank voor jullie grote inzet. Ik vind het een grote eer jullie op te mogen volgen. Ik hoefde niet lang na te denken om die kans te grijpen. Ik dank ook de Raad van Bestuur van het LUMC en het College van Bestuur van deze Universiteit voor het in mij gestelde vertrouwen, en de prettige ontvangst in de Leidse academische omgeving.

Tenslotte kom ik bij mijn familie. Ik ben dankbaar hier vandaag veel familieleden te zien, mijn broers Maarten en Rutger, zus Iris, schoonfamilie, en een flink aantal van de neven en nichten. Ik hoop nog veel mooie momenten met hen te mogen delen. Ik bedank mijn vader voor alle inzet voor zijn gezin, en het steunen van mijn wetenschappelijke vorming in de ruimste zin. Ik sta in gedachten ook stil bij mijn moeder, die zoveel van Leiden en de Leidse universiteit hield.

Op een dag als vandaag kijk ik uiteraard vooral naar de zonnige zijde; ik bedank, uit het diepst van mijn hart, Aleida voor al haar steun, liefdevolle zorgen voor ons gezin, flexibiliteit, en relativering en Matthijs, Laurens en Suzanne voor de gezelligheid en warmte die zij in ons huis brengen.

Ik dank u allen voor uw aandacht.

Ik heb gezegd.

Referenties

- 1 Definitie van het kansbegrip. [https://nl.wikipedia.org/wiki/Kans_\(statistiek\)](https://nl.wikipedia.org/wiki/Kans_(statistiek)).
- 2 Framingham risk calculator. <https://www.mdcalc.com/framingham-coronary-heart-disease-risk-score>.
- 3 Steyerberg EW, Keizer HJ, Zwartendijk J, Van Rijk GL, Van Groenigen CJ, Habbema JD and Stoter G. Prognosis after resection of residual masses following chemotherapy for metastatic nonseminomatous testicular cancer: a multivariate analysis. *British journal of cancer*. 1993; 68: 195-200.
- 4 Steyerberg EW, Keizer HJ, Fossa SD, Sleijfer DT, Toner GC, Schraffordt Koops H, Mulders PF, Messemer JE, Ney K, Donohue JP et al. Prediction of residual retroperitoneal mass histology after chemotherapy for metastatic non-seminomatous germ cell tumor: multivariate analysis of individual patient data from six study groups. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 1995; 13: 1177-87.
- 5 Steyerberg EW, Kievit J, De Mol Van Otterloo JC, Van Bockel JH, Eijkemans MJ and Habbema JD. Perioperative mortality of elective abdominal aortic aneurysm surgery. A clinical prediction rule based on literature and individual patient data. *Archives of internal medicine*. 1995; 155: 1998-2004.
- 6 Steyerberg EW, Eijkemans MJ, Van Houwelingen JC, Lee KL and Habbema JD. Prognostic models based on literature and individual patient data in logistic regression analysis. *Statistics in medicine*. 2000; 19: 141-60.
- 7 Maas AI, Marmarou A, Murray GD, Teasdale SG and Steyerberg EW. Prognosis and clinical trial design in traumatic brain injury: the IMPACT study. *Journal of neurotrauma*. 2007; 24: 232-8.
- 8 Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, Murray GD, Marmarou A, Roberts I, Habbema JD and Maas AI. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS medicine*. 2008; 5: e165; discussion e165.
- 9 Perel P, Arango M, Clayton T, Edwards P, Komolafe E, Pocock S, Roberts I, Shakur H, Steyerberg E and Yuthakasemsunt S. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ (Clinical research ed)*. 2008; 336: 425-9.
- 10 Balmana J, Stockwell DH, Steyerberg EW, Stoffel EM, Defenbaugh AM, Reid JE, Ward B, Scholl T, Hendrickson B, Tazelaar J, Burbidge LA and Syngal S. Prediction of MLH1 and MSH2 mutations in Lynch syndrome. *Jama*. 2006; 296: 1469-78.
- 11 Kastrinos F, Steyerberg EW, Mercado R, Balmana J, Holter S, Gallinger S, Siegmund KD, Church JM, Jenkins MA, Lindor NM, Thibodeau SN, Burbidge LA, Wenstrup RJ and Syngal S. The PREMM(1,2,6) model predicts risk of MLH1, MSH2, and MSH6 germline mutations based on cancer history. *Gastroenterology*. 2011; 140: 73-81.
- 12 Kastrinos F, Ojha RP, Leenen C, Alvero C, Mercado RC, Balmana J, Valenzuela I, Balaguer F, Green R, Lindor NM, Thibodeau SN, Newcomb P, Win AK, Jenkins M, Buchanan DD, Bertario L, Sala P, Hampel H, Syngal S and Steyerberg EW. Comparison of Prediction Models for Lynch Syndrome Among Individuals With Colorectal Cancer. *Journal of the National Cancer Institute*. 2016; 108.
- 13 Kastrinos F, Uno H, Ukaegbu C, Alvero C, McFarland A, Yurgelun MB, Kulke MH, Schrag D, Meyerhardt JA, Fuchs CS, Mayer RJ, Ng K, Steyerberg EW and Syngal S. Development and Validation of the PREMM5 Model for Comprehensive Risk Assessment of Lynch Syndrome. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2017; 35: 2165-2172.
- 14 Lynch syndrom prediction calculator. [http://premmdfci-harvard.edu/](http://premmdfci.harvard.edu/).
- 15 Het PRIAS project: Active Surveillance voor prostaatkanker. <https://www.prias-projectorg/>
- 16 De prostaatwijzer. <http://www.prostaatwijzernl/medical-risk-calculators>.
- 17 Sackett DL. Evidence-based medicine. *Seminars in Perina-*

tology. 1997; 21: 3-5.

- 18 Van Esch A, Van Steensel-Moll HA, Steyerberg EW, Of-
fringa M, Habbema JD and Derksen-Lubsen G. Antipyre-
tic efficacy of ibuprofen and acetaminophen in children
with febrile seizures. *Archives of pediatrics & adolescent
medicine*. 1995; 149: 632-7.
- 19 Van Stuijvenberg M, Derksen-Lubsen G, Steyerberg EW,
Habbema JD and Moll HA. Randomized, controlled trial
of ibuprofen syrup administered during febrile illnesses to
prevent febrile seizure recurrences. *Pediatrics*. 1998; 102:
E51.
- 20 Klomp HM, Spincemaille GH, Steyerberg EW, Habbema
JD and Van Urk H. Spinal-cord stimulation in critical
limb ischaemia: a randomised trial. ESES Study Group.
Lancet (London, England). 1999; 353: 1040-4.
- 21 Homs MY, Wahab PJ, Kuipers EJ, Steyerberg EW, Grool
TA, Haringsma J and Siersema PD. Esophageal stents with
antireflux valve for tumors of the distal esophagus and
gastric cardia: a randomized trial. *Gastrointestinal endo-
scopy*. 2004; 60: 695-702.
- 22 Albers MJ, Steyerberg EW, Hazebroek FW, Mourik M,
Borsboom GJ, Rietveld T, Huijmans JG and Tibboel D.
Glutamine supplementation of parenteral nutrition does
not improve intestinal permeability, nitrogen balance, or
outcome in newborns and infants undergoing digestive-
tract surgery: results from a double-blind, randomized,
controlled trial. *Annals of surgery*. 2005; 241: 599-606.
- 23 Roukema J, Steyerberg EW, Van der Lei J and Moll HA.
Randomized trial of a clinical decision support system:
impact on the management of children with fever without
apparent source. *Journal of the American Medical Informa-
tics Association : JAMIA*. 2008; 15: 107-13.
- 24 De Vos-Kerkhof E, Nijman RG, Vergouwe Y, Polinder S,
Steyerberg EW, Van der Lei J, Moll HA and Oostenbrink
R. Impact of a clinical decision model for febrile children
at risk for serious bacterial infections at the emergency
department: a randomized controlled trial. *PloS one*. 2015;
10: e0127620.
- 25 Geurts D, De Vos-Kerkhof E, Polinder S, Steyerberg E, Van
der Lei J, Moll H and Oostenbrink R. Implementation
of clinical decision support in young children with acute
gastroenteritis: a randomized controlled trial at the emer-
gency department. *European journal of pediatrics*. 2017;
176: 173-181.
- 26 Schasfoort F, Dallmeijer A, Pangalila R, Catsman C, Stam
H, Becher J, Steyerberg E, Polinder S and Bussmann J. Va-
lue of botulinum toxin injections preceding a comprehen-
sive rehabilitation period for children with spastic cerebral
palsy: A cost-effectiveness study. *Journal of rehabilitation
medicine*. 2018; 50: 22-29.
- 27 Schipper IB, Steyerberg EW, Castelein RM, Van der He-
ijden FH, Den Hoed PT, Kerver AJ and Van Vugt AB.
Treatment of unstable trochanteric fractures. Randomised
comparison of the gamma nail and the proximal femoral
nail. *The Journal of bone and joint surgery British volume*.
2004; 86: 86-94.
- 28 Homs MY, Steyerberg EW, Eijkenboom WM, Tilanus HW,
Stalpers LJ, Bartelsman JF, Van Lanschot JJ, Wijrdeman
HK, Mulder CJ, Reinders JG, Boot H, Aleman BM, Kuipers
EJ and Siersema PD. Single-dose brachytherapy versus
metal stent placement for the palliation of dysphagia from
oesophageal cancer: multicentre randomised trial. *Lancet
(London, England)*. 2004; 364: 1497-504.
- 29 Steyerberg EW, Homs MY, Stokvis A, Essink-Bot ML and
Siersema PD. Stent placement or brachytherapy for pal-
liation of dysphagia from esophageal cancer: a prognostic
model to guide treatment selection. *Gastrointestinal endo-
scopy*. 2005;62:333-40.
- 30 Jeurnink SM, Steyerberg EW, Van Hooft JE, Van Eijck CH,
Schwartz MP, Vleggaar FP, Kuipers EJ and Siersema PD.
Surgical gastrojejunostomy or endoscopic stent placement
for the palliation of malignant gastric outlet obstruction
(SUSTENT study): a multicenter randomized trial. *Gas-
trointestinal endoscopy*. 2010;71:490-9.
- 31 Walter D, Van Boeckel PG, Groenen MJ, Weusten BL,
Wittebman BJ, Tan G, Brink MA, Nicolai J, Tan AC, Alder-

liesten J, Venneman NG, Laleman W, Jansen JM, Bodelier A, Wolters FL, Van der Waaij LA, Breumelhof R, Peters FT, Scheffer RC, Leenders M, Hirdes MM, Steyerberg EW, Vleggaar FP and Siersema PD. Cost Efficacy of Metal Stents for Palliation of Extrahepatic Bile Duct Obstruction in a Randomized Controlled Trial. *Gastroenterology*. 2015; 149: 130-8.

- 32 Verschuur EM, Repici A, Kuipers EJ, Steyerberg EW and Siersema PD. New design esophageal stents for the palliation of dysphagia from esophageal or gastric cardia cancer: a randomized trial. *The American journal of gastroenterology*. 2008; 103: 304-12.
- 33 Rutten JH, Steyerberg EW, Boomsma F, Van Saase JL, Deckers JW, Hoogsteden HC, Lindemans J and Van den Meiracker AH. N-terminal pro-brain natriuretic peptide testing in the emergency department: beneficial effects on hospitalization, costs, and outcome. *American heart journal*. 2008; 156: 71-7.
- 34 Verschuur EM, Steyerberg EW, Tilanus HW, Polinder S, Essink-Bot ML, Tran KT, Van der Gaast A, Stassen LP, Kuipers EJ and Siersema PD. Nurse-led follow-up of patients after oesophageal or gastric cardia cancer surgery: a randomised trial. *British journal of cancer*. 2009; 100: 70-6.
- 35 Uitdehaag MJ, Van Putten PG, Van Eijck CH, Verschuur EM, Van der Gaast A, Pek CJ, Van der Rijt CC, De Man RA, Steyerberg EW, Laheij RJ, Siersema PD, Spaander MC and Kuipers EJ. Nurse-led follow-up at home vs. conventional medical outpatient clinic follow-up in patients with incurable upper gastrointestinal cancer: a randomized study. *Journal of pain and symptom management*. 2014; 47: 518-30.
- 36 Langeveld HR, Van 't Riet M, Weidema WF, Stassen LP, Steyerberg EW, Lange J, Bonjer HJ and Jeekel J. Total extraperitoneal inguinal hernia repair compared with Lichtenstein (the LEVEL-Trial): a randomized controlled trial. *Annals of surgery*. 2010; 251: 819-24.
- 37 Deerenberg EB, Harlaar JJ, Steyerberg EW, Lont HE, Van Doorn HC, Heisterkamp J, Wijnhoven BP, Schouten WR,

Cense HA, Stockmann HB, Berends FJ, Dijkhuizen FPH, Dwarkasing RS, Jairam AP, Van Ramshorst GH, Kleinrensink GJ, Jeekel J and Lange JF. Small bites versus large bites for closure of abdominal midline incisions (STITCH): a double-blind, multicentre, randomised controlled trial. *Lancet (London, England)*. 2015; 386: 1254-1260.

- 38 Jairam AP, Timmermans L, Eker HH, Pierik R, Van Klaveren D, Steyerberg EW, Timman R, Van der Ham AC, Dawson I, Charbon JA, Schuhmacher C, Mihaljevic A, Izbicki JR, Fikatas P, Knebel P, Fortelny RH, Kleinrensink GJ, Lange JF and Jeekel HJ. Prevention of incisional hernia with prophylactic onlay and sublay mesh reinforcement versus primary suture only in midline laparotomies (PRIMA): 2-year follow-up of a multicentre, double-blind, randomised controlled trial. *Lancet (London, England)*. 2017; 390: 567-576.
- 39 Van Hagen P, Hulshof MC, Van Lanschot JJ, Steyerberg EW, Van Berge Henegouwen MI, Wijnhoven BP, Richel DJ, Nieuwenhuijzen GA, Hospers GA, Bonenkamp JJ, Cuesta MA, Blaisse RJ, Busch OR, Ten Kate FJ, Creemers GJ, Punt CJ, Plukker JT, Verheul HM, Spillenaar Bilgen EJ, Van Dekken H, Van der Sangen MJ, Rozema T, Biermann K, Beukema JC, Piet AH, Van Rij CM, Reinders JG, Tilanus HW and Van der Gaast A. Preoperative chemoradiotherapy for esophageal or junctional cancer. *The New England journal of medicine*. 2012; 366: 2074-84.
- 40 Shapiro J, Van Lanschot JJB, Hulshof M, Van Hagen P, Van Berge Henegouwen MI, Wijnhoven BPL, Van Laarhoven HWM, Nieuwenhuijzen GAP, Hospers GAP, Bonenkamp JJ, Cuesta MA, Blaisse RJB, Busch ORC, Ten Kate FJW, Creemers GM, Punt CJA, Plukker JTM, Verheul HMW, Bilgen EJS, Van Dekken H, Van der Sangen MJC, Rozema T, Biermann K, Beukema JC, Piet AHM, Van Rij CM, Reinders JG, Tilanus HW, Steyerberg EW and Van der Gaast A. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *The Lancet Oncology*. 2015; 16: 1090-1098.

- 41 Berkhemer OA, Fransen PS, Beumer D, Van den Berg LA, Lingsma HF, Yoo AJ, Schonewille WJ, Vos JA, Nederkoorn PJ, Wermer MJ, Van Walderveen MA, Staals J, Hofmeijer J, Van Oostayen JA, Lycklama a Nijeholt GJ, Boiten J, Brouwer PA, Emmer BJ, de Bruijn SF, Van Dijk LC, Kappelle LJ, Lo RH, Van Dijk EJ, De Vries J, De Kort PL, Van Rooij WJ, Van den Berg JS, Van Hasselt BA, Aerden LA, Dallinga RJ, Visser MC, Bot JC, Vroomen PC, Eshghi O, Schreuder TH, Heijboer RJ, Keizer K, Tielbeek AV, Den Hertog HM, Gerrits DG, Van den Berg-Vos RM, Karas GB, Steyerberg EW, Flach HZ, Marquering HA, Sprengers ME, Jenniskens SF, Beenen LF, Van den Berg R, Koudstaal PJ, Van Zwam WH, Roos YB, Van der Lugt A, Van Oostenbrugge RJ, Majoie CB and Dippel DW. A randomized trial of intraarterial treatment for acute ischemic stroke. *The New England journal of medicine*. 2015; 372: 11-20.
- 42 Kalincik T, Manouchehrinia A, Sobisek L, Jokubaitis V, Spelman T, Horakova D, Havrdova E, Trojano M, Izquierdo G, Lugaresi A, Girard M, Prat A, Duquette P, Grammond P, Sola P, Hupperts R, Grand'Maison F, Pucci E, Boz C, Alroughani R, Van Pesch V, Lechner-Scott J, Terzi M, Bergamaschi R, Iuliano G, Granella F, Spitaleri D, Shayannejad V, Oreja-Guevara C, Slee M, Ampapa R, Verheul F, McCombe P, Olascoaga J, Amato MP, Vucic S, Hodgkinson S, Ramo-Tello C, Flechter S, Cristiano E, Rozsa C, Moore F, Luis Sanchez-Menoyo J, Laura Saladino M, Barnett M, Hillert J and Butzkueven H. Towards personalized therapy for multiple sclerosis: prediction of individual treatment response. *Brain : a journal of neurology*. 2017; 140: 2426-2443.
- 43 NIH definition of Precision Medicine. <https://ghrnlmmnih-gov/primer/precisionmedicine/definition>.
- 44 Steyerberg EW and Claggett B. Towards personalized therapy for multiple sclerosis: limitations of observational data. *Brain : a journal of neurology*. 2018.
- 45 Van Houwelingen JC. Shrinkage and Penalized Likelihood as Methods to Improve Predictive Accuracy. *Statistica Neerlandica*. 2001; 55: 17-34.
- 46 Goeman JJ. CRAN - Package penalized. <https://cranr-projectorg/package=penalized>.
- 47 Steyerberg EW, Van der Ploeg T and Van Calster B. Risk prediction with machine learning and regression methods. *Biometrical journal Biometrische Zeitschrift*. 2014; 56: 601-6.
- 48 Van der Ploeg T, Nieboer D and Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. *Journal of clinical epidemiology*. 2016; 78: 83-89.
- 49 Cochrane collaboration. <http://www.cochranelibrary.com/>.
- 50 Wallach JD, Sullivan PG, Trepanowski JF, Steyerberg EW and Ioannidis JP. Sex based subgroup differences in randomized controlled trials: empirical evidence from Cochrane meta-analyses. *BMJ (Clinical research ed)*. 2016; 355: i5826.
- 51 Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW and Ioannidis JP. Evaluation of Evidence of Statistical Support and Corroboration of Subgroup Claims in Randomized Clinical Trials. *JAMA internal medicine*. 2017; 177: 554-560.
- 52 Burke JF, Sussman JB, Kent DM and Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ (Clinical research ed)*. 2015; 351: h5651.
- 53 Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG and Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *International journal of epidemiology*. 2016; 45: 2075-2088.
- 54 Farooq V, Van Klaveren D, Steyerberg EW, Meliga E, Vergouwe Y, Chieffo A, Kappetein AP, Colombo A, Holmes DR, Jr., Mack M, Feldman T, Morice MC, Stahle E, Onuma Y, Morel MA, Garcia-Garcia HM, Van Es GA, Dawkins KD, Mohr FW and Serruys PW. Anatomical and clinical characteristics to guide decision making between coronary artery bypass surgery and percutaneous coronary intervention for individual patients: development and validation of SYNTAX score II. *Lancet (London, England)*.

2013; 381: 639-50.

- 55 Venema E, Mulder M, Roozenbeek B, Broderick JP, Yeatts SD, Khatri P, Berkhemer OA, Emmer BJ, Roos Y, Majoie C, Van Oostenbrugge RJ, Van Zwam WH, Van der Lugt A, Steyerberg EW, Dippel DWJ and Lingsma HF. Selection of patients for intra-arterial treatment for acute ischaemic stroke: development and validation of a clinical decision tool in two randomised trials. *BMJ (Clinical research ed)*. 2017; 357: j1710.
- 56 Costa F, Van Klaveren D, James S, Heg D, Raber L, Feres F, Pilgrim T, Hong MK, Kim HS, Colombo A, Steg PG, Zanchin T, Palmerini T, Wallentin L, Bhatt DL, Stone GW, Windecker S, Steyerberg EW and Valgimigli M. Derivation and validation of the predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet therapy (PRECISE-DAPT) score: a pooled analysis of individual-patient datasets from clinical trials. *Lancet (London, England)*. 2017; 389: 1025-1034.
- 57 Van Klaveren D, Vergouwe Y, Farooq V, Serruys PW and Steyerberg EW. Estimates of absolute treatment benefit for individual patients required careful modeling of statistical interactions. *Journal of clinical epidemiology*. 2015; 68: 1366-74.
- 58 Jakola AS, Myrnes KS, Kloster R, Torp SH, Lindal S, Unsgard G and Solheim O. Comparison of a strategy favoring early surgical resection vs a strategy favoring watchful waiting in low-grade gliomas. *Jama*. 2012; 308: 1881-8.
- 59 Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF and Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine*. 2015; 162: W1-73.
- 60 Steyerberg EW, Uno H, Ioannidis JPA and Van Calster B. Poor performance of clinical prediction models: the harm of commonly applied methods. *Journal of clinical epidemiology*. 2017.
- 61 Steyerberg EW and Harrell FE, Jr. Prediction models need

appropriate internal, internal-external, and external validation. *Journal of clinical epidemiology*. 2016; 69: 245-7.

- 62 Beaver DD. Reflections on Scientific Collaboration (and its study): Past, Present, and Future. *Scientometrics*. 2001; 52: 365.
- 63 Gigerenzer G and Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ*. 2003; 327: 741-4.

PROF.DR. EWOUT STEYERBERG (DELFT, 1967)



- 2017- Hoogleraar Klinische Biostatistiek en Medische Besliskunde,
 Hoofd van de afdeling *Biomedical Data Sciences*,
 Leids Universitair Medisch Centrum, Leiden
- 2006- Hoogleraar Medische Besliskunde, afdeling
 Maatschappelijke Gezondheidszorg, Erasmus MC,
 Rotterdam
- 2002-2006 Universiteit hoofddocent, afdeling Maatschappelijke
 Gezondheidszorg, Erasmus MC, Rotterdam
- 1998-2002 *Fellow* van de Koninklijke Nederlandse Academie voor
 Wetenschappen
- 1991-2002 Wetenschappelijk onderzoeker bij Centrum voor
 Medische Besliskunde, afdeling Maatschappelijke
 Gezondheidszorg, Erasmus MC, Rotterdam
- 1990-1991 Student-assistent Medische Statistiek, Universiteit
 Leiden
- 1996 *PhD*, Medische faculteit van de Erasmus Universiteit
 Rotterdam
- 1991 *MSc*, Biomedische Wetenschappen, Universiteit Leiden
 (*cum laude*)
- 1986 Propedeuse geneeskunde, Universiteit Leiden

Het schatten van kansen op ziekte, het beloop van ziekte en het behandel-effect voor individuele patiënten krijgt steeds meer aandacht, ook in het kader van *Precision Medicine*: welke preventieve of therapeutische behandeling werkt voor welke mensen? De gerandomiseerde klinische trial is de hoeksteen van onderzoek naar het gemiddelde effect van een behandeling. Het is moeilijk om verschillen in het behandel-effect te schatten in subgroepen binnen een trial, bijvoorbeeld of het effect voor mannen of vrouwen in relatieve zin anders is. Dit komt doordat de aantallen meestal te klein zijn voor betrouwbare statistische analyse. Met grotere aantallen zouden we meer succes kunnen hebben, bijvoorbeeld in een meta-analyse waarin we de resultaten van verschillende trials combineren. Grotere aantallen staan ook centraal in *Big Data*, maar dan gaat het om niet-gerandomiseerde patiënten en is de kwaliteit van de data niet altijd duidelijk. Dit soort gegevens kan ons een vertekend beeld van het behandel-effect geven, ook als we gebruik maken van hippe *Machine Learning* technieken. *Big Data* zal ons wel goed kunnen helpen om betrouwbare prognostische modellen te maken die de uitkomst voor individuele patiënten schatten. Verder onderzoek is nodig naar geschikte methoden om gegevens uit gerandomiseerde en niet-gerandomiseerde studies optimaal te combineren voor het schatten van geïndividualiseerde behandel-effecten. Hiermee kunnen we nog betere beslissingen nemen en nog betere zorg verlenen.



Universiteit
Leiden