



## **KANSRIJK BESLISSEN**

**EWOUT W. STEYERBERG**

---



## KANSRIJK BESLISSEN

Oplage 1000  
Omslagfoto Levien Willemse, Rotterdam  
Ontwerp Ontwerpwerk, Den Haag  
Drukwerk Demmenie Grafimedia, Alphen aan den Rijn

ISBN97-8907790-64-53

© Ewout W. Steyerberg, oratiereeks Erasmus MC  
7 december 2007

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).

# KANSRIJK BESLISSEN

REDE

In verkorte vorm uitgesproken  
ter gelegenheid van het aanvaarden  
van het ambt van hoogleraar  
in de Medische Besliskunde,  
in het bijzonder prognostische modellering,  
aan het Erasmus MC, faculteit van de  
Erasmus Universiteit Rotterdam  
op 7 december 2007

door

EWOUT WILLEM STEYERBERG



*Mijnheer de rector magnificus,  
geachte aanwezigen,*

## **Inleiding**

**S**tel: u, of één van uw naasten, wordt getroffen door een hartinfarct. In dat geval sterft een deel van het hart doordat er te weinig bloed naartoe stroomt. De oorzaak is vaak een stolsel in één van de kransslagaders. U wordt met spoed naar het ziekenhuis gebracht. Laten we aannemen dat dit het Erasmus MC is. De Eerste Hulp bevindt zich - heel toepasselijk- in het hart van dit complex.

U wilt uiteraard zo goed mogelijk behandeld worden. Wat betekent 'zo goed mogelijk'? U wilt die testen en behandelingen ondergaan die uiteindelijk leiden tot de beste prognose. Prognose kunnen we uitdrukken in de kans om het hartinfarct in relatief goede gezondheid te overleven. Omgekeerd geformuleerd willen we het risico op overlijden door het infarct zo klein mogelijk maken.

## Hoe kunnen we goede medische beslissingen nemen?

Het is tegenwoordig vrij algemeen geaccepteerd dat medische beslissingen ondersteund moeten worden door bewijs ('evidence'). Niet het oordeel van het afdelingshoofd met zijn ruime ervaring, of de eigen intuïtie staan centraal, maar empirisch bewijs is nodig. Dit noemen we 'evidence-based medicine'<sup>1</sup>. Beslissingen zijn te verdelen in twee hoofdgroepen: beslissingen over het doen van een diagnostische test om vast te stellen of een bepaalde ziekte aanwezig is, en beslissingen over een therapeutische interventie, zoals het geven van een medicijn of een andere behandeling. In het voorbeeld van een hartinfarct willen we dat de diagnose zo snel mogelijk wordt gesteld zodat een effectieve behandeling snel kan beginnen.

Voor een diagnostische test hopen we bewijs te hebben dat de test diegenen met de ziekte kan onderscheiden van diegenen zonder ziekte. In de praktijk zijn testen niet perfect. Soms suggereert een test bijvoorbeeld dat een patiënt een hartinfarct heeft, terwijl dat niet zo is. En omgekeerd, dat een patiënt geen hartinfarct heeft, terwijl dat wel zo is. Het vaststellen van een hartinfarct op de Eerste Hulp is niet eenvoudig, omdat de presentatie kan passen bij allerlei andere onderliggende aandoeningen. Maar laten we aannemen dat er in het Erasmus MC snel een adequate diagnose wordt gesteld.

	Sterfte < 30 dagen	Totaal
Nieuwe behandeling (tPA)	653 (6,3%)	10.348
Traditionele behandeling (SK)	1475 (7,3%)	20.162

Tabel 1. Belangrijkste resultaten van de GUSTO-I studie. In de groep die met de nieuwe medicatie werd behandeld ("tissue plasminogen activator", tPA) overleden 653 van de 10.348 patiënten binnen 30 dagen na een acuut hartinfarct (6,3%), tegenover 1475 van de 20.162 (7,3%) in de met het meer traditionele middel behandelde groep (streptokinase, SK)<sup>2</sup>.

Voor een therapeutische interventie willen we bewijs dat patiënten met de behandeling beter af zijn dan zonder. Dit soort bewijs komt het beste uit onderzoek waarbij één groep patiënten de behandeling heeft ondergaan, en een controle groep niet. Idealiter worden patiënten door het lot ingedeeld in de behandelings- of controlegroep, een zogenaamde gerandomiseerde trial.

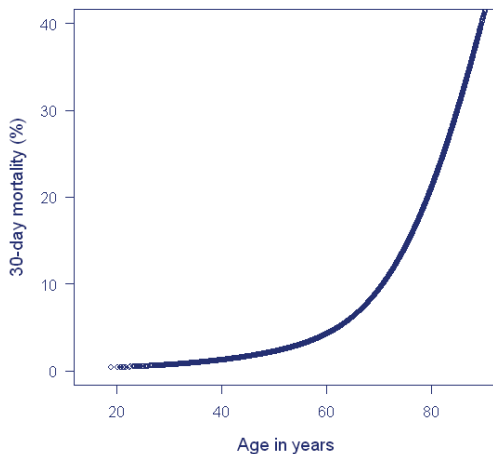
### Behandeling van een hartinfarct

In 1993 werden de resultaten gerapporteerd van een grote gerandomiseerde studie voor behandelingen van een acuut hartinfarct. Deze 'GUSTO-I' studie toonde aan dat de behandeling met een nieuw stolseloplossend medicijn ("tissue plasminogen activator", tPA), leidde tot minder sterfte binnen 30 dagen dan een ouder middel



(streptokinase, SK, Tabel 1)]<sup>2</sup>. Van de groep met het nieuwe medicijn overleed 6,3% binnen 30 dagen, tegenover 7,3% in de groep die met het oudere middel behandeld werd. Dit is een verschil van 1% in sterfte in absolute zin. Het is in de epidemiologie gebruikelijk om het verschil tussen de 2 behandelingsgroepen relatief uit te drukken. Vergelijken we 6,3% met 7,3%, dan is het zogenaamde 'relatieve risico' op sterfte  $6,3/7,3=0,86$ . We kunnen ook zeggen dat de sterfte relatief gezien 14% lager is in de tPA groep dan in de SK groep.

Deze resultaten zijn heel valide. Ten eerste door de degelijke studie opzet, namelijk een gerandomiseerde toewijzing van de behandeling, en ten tweede doordat de studie met grote aantallen werd uitgevoerd (meer dan 10.000 patiënten per groep).



Figuur 1. Relatie tussen leeftijd van de patiënt en sterfte binnen 30 dagen na een hartinfarct. De relatie werd geschat met een logistisch regressiemodel op basis van 40.830 patiënten uit de GUSTO-I studie, waarin een 'restricted cubic spline' functie met 4 knopen (3 vrijheidsgraden) werd gebruikt als transformatie van de leeftijd.<sup>3,4</sup>

### Van gemiddelden naar individuele patiënten

De GUSTO-I studie geeft bewijs voor wat gemiddeld het beste is om te doen. Er zijn echter grote verschillen tussen patiënten wat betreft hun risico op sterfte na een hartinfarct. Er bestaan eigenlijk geen gemiddelde patiënten, sommige hebben een laag risico, en andere een hoog risico.

Een belangrijk kenmerk voor het risico op sterfte is de leeftijd van de patiënt. Een relatief jonge patiënt heeft een veel lagere kans om te overlijden als gevolg

van het hartinfarct dan een oudere patiënt. Het risico is minder dan 1% voor een 40-jarige patiënt, 10% voor een 70-jarige, en zelfs 20% voor een 80-jarige patiënt (Figuur 1).

Maar er zijn natuurlijk veel meer voorspellers voor sterfte. In Tabel 2 zien we een prognostisch model, waarin de belangrijkste voorspellers gecombineerd zijn<sup>5</sup>. Een prognostisch model combineert meerdere kenmerken van de patiënt of de ziekte om tot een voorspelling van het beloop van de ziekte te komen, in dit geval sterfte binnen 30 dagen<sup>6</sup>.

We zien dat naast leeftijd, de conditie van het hart belangrijk is (de zogenaamde “Killip class”); de systolische bloeddruk; de hartslag; de plaats van het infarct; en of de patiënt al eerder een infarct gehad heeft. Al deze kenmerken kunnen we een bepaald gewicht toekennen voor het voorspellen van sterfte; de punten staan in tabel 2.

Als we alle punten optellen krijgen we een totaalscore, die we kunnen koppelen aan een risico op sterfte. Een man van 40 jaar, zonder enige negatieve risicofactoren, krijgt een score van 28 punten. Maar een man van 70 jaar, met een iets slechtere hartfunctie (Killip class II), en een wat verlaagde bloeddruk (80 mm Hg), krijgt 63 + 17 = 80 punten.

Predictor	Units	Points			
		Killip class			
		I	II	III	IV
Age (years)	40	28	42	53	59
	50	38	49	59	65
	60	47	56	64	70
	70	57	63	70	76
	80	66	70	75	82
	90	75	77	81	88
	100	94	91	92	100
Systolic BP (mm Hg)	40	34			
	80	17			
	120+	0			
Heart rate (beats/min)	10	10			
	30	5			
	50	0			
	90	8			
	130	16			
Infarct location	Anterior	6			
	Inferior	0			
	Other	3			
Previous MI	Yes	5			
<b>Total</b>	<b>Add points</b>	<b>...</b>			

Tabel 2. Scorekaart voor patiënten met een hartinfarct op basis van 40.830 patiënten uit de GUSTO-I studie. Logistische regressie analyse resulteerde in gewichten voor de verschillende predictoren<sup>5</sup>.

## Individuele winst van behandeling

De resultaten van de GUSTO-I studie gaven aan dat tPA gemiddeld de beste behandeling was. Maar moeten deze 2 hypothetische patiënten, met heel verschillende prognostische scores (28 en 80), allebei met het nieuwe middel (tPA) behandeld worden?

We kunnen eerst uitgaan van een relatief risico van 0,86: het centrale resultaat van de GUSTO-I trial. De tPA behandeling verlaagt het risico voor alle patiënten met een factor 0,86. Voor een patiënt met een lage score is het risico op sterfte laag met de traditionele SK behandeling, en de winst met de nieuwe tPA behandeling in absolute zin klein. Daarentegen heeft een patiënt met een hoge score een hoog risico op sterfte met de traditionele SK behandeling. De absolute winst met de nieuwe behandeling is voor deze patiënt veel groter. De winst in absolute zin verschilt sterk tussen patiënten, terwijl het relatief risico constant was voor alle patiënten.

Score	SK mortality	tPA mortality	tPA reduction
30	0.4%	0.4%	-
40	0.8%	0.8%	0.01%
50	1.7%	1.4%	0.3%
60	3,5%	2.8%	0.8%
70	10%	8.3%	1.7%
80	20%	17%	3%
90	40%	35%	5%

Tabel 3. Berekening van de individuele winst van behandeling voor patiënten met een hartinfarct op basis van de GUSTO-I studie. De mortaliteit met tPA en met SK behandeling werd geschat met een non-parametrische analyse van de sterfte binnen 30 dagen als functie van de score uit Tabel 2. Het verschil tussen deze 2 schattingen van de mortaliteit is de individuele winst door behandeling met tPA ('tPA reduction')<sup>5</sup>

Iets ingewikkelder wordt het als we rekening houden met bijwerkingen van de nieuwe behandeling. In het geval van dit nieuwe middel (tPA) is dat met name een verhoogd risico op bloedingen; het oplossen van een stolsel schiet soms wat door, en er ontstaat een ernstige bloeding. De patiënt kan dan aan een hersenbloeding overlijden, en niet aan zijn hartinfarct. We kunnen aannemen dat het nieuwe middel een sterke reductie geeft van de sterfte door een hartinfarct: bijvoorbeeld een relatief risico van 0,75. Maar daarbij komt een risico op fatale bijwerkingen. Dan blijkt dat we voor laag risico patiënten geen enkele winst meer behalen met behandelen. Integendeel, patiënten met een heel laag risico hebben meer risico op overlijden met de nieuwe behandeling dan met de traditionele behandeling, door het risico op fatale bijwerkingen<sup>7,8</sup>.

Voor de GUSTO-I studie is deze analyse van individuele winst gemaakt (Tabel 3)<sup>5</sup>. De hiervoor gepresenteerde 40-jarige patiënt had 28 punten (bijna 30); voor hem is

het risico op sterfte laag (namelijk rond de 0.4%), ongeacht de specifieke behandeling. Voor de 2<sup>e</sup> patiënt is het risico een stuk hoger. Een score van 80 punten betekent een risico van ongeveer 20% met de traditionele SK behandeling dat door tPA verlaagd kan worden naar 17%. Voor de laag risico patiënt is er geen winst, voor de hoog risico patiënt wel.

### Individueel risico op ziekte

**I**k heb nu een individuele benadering voor behandeling van patiënten met een hartinfarct laten zien. Voorkómen is echter beter dan genezen. We kunnen ook een op het individu gerichte benadering bij preventie van ziekten volgen.

Op internet kwam ik een 'risk calculator' tegen om mijn individuele risico te berekenen op hart- en vaatziekten<sup>|\*</sup>. Als ik dan in mijn naïviteit deze 'risk calculator' voor mijzelf ga invullen blijkt dat ik toch niet zo gezond ben als ik dacht. Bij een bezoek aan de bloedbank bleek mijn bloeddruk 130 over 70. Dat is volgens deze wijzer al een gevaarlijke situatie ('caution', Figuur 2). Mijn risico op een hartinfarct of een beroerte is ongeveer 2 keer verhoogd ten opzichte van een 'normal risk'; mij wordt geadviseerd om mijn leefstijl te veranderen. Daar kan mijn vrouw zich wel iets bij voorstellen: geen oratie meer schrijven, niet 's avonds laat aan artikelen of een boek werken, in het weekend niet werken: allemaal mogelijkheden om de bloeddruk wat te laten dalen wellicht. Wel doorgaan met hardlopen en fietsen denk ik.

Er zijn uiteraard wel enkele kanttekeningen te plaatsen bij deze berekening. Ten eerste wordt slechts naar 1 risicofactor gekeken voor het krijgen van hart- en vaatziekten, namelijk de bloeddruk. De moderne aanpak is om naar een combinatie van risicofactoren te kijken in een prognostisch model. In dat kader zijn bijvoorbeeld prognostische modellen te gebruiken die zijn ontwikkeld op basis van gegevens van de bevolking van het dorpje Framingham in Massachusetts, vlakbij Boston<sup>9</sup>.

Ten tweede wordt er een relatief risico gepresenteerd. Het absolute risico blijft laag als we met een heel laag risico beginnen: 2 x iets kleins blijft iets kleins. Dit fenomeen zagen we ook al bij de berekening van de individuele winst door tPA behandeling van een hartinfarct; als de score laag was, was de winst ook laag.

Tenslotte is de site gesponsord door industrie die bloeddrukverlagende medicijnen verkoopt. Die heeft er belang bij zoveel mogelijk mensen als behorend tot de hoog risicogroep te bestempelen.

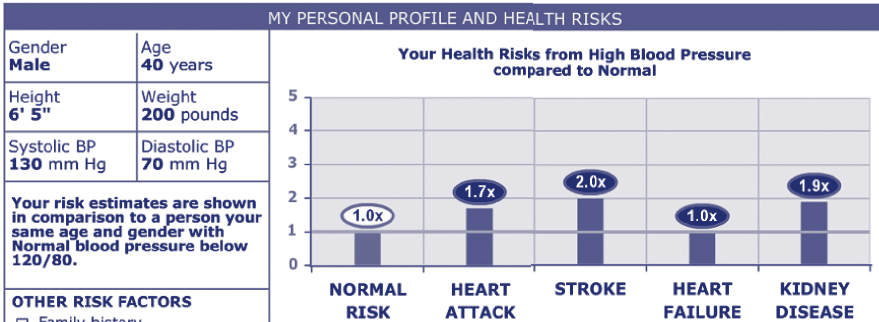
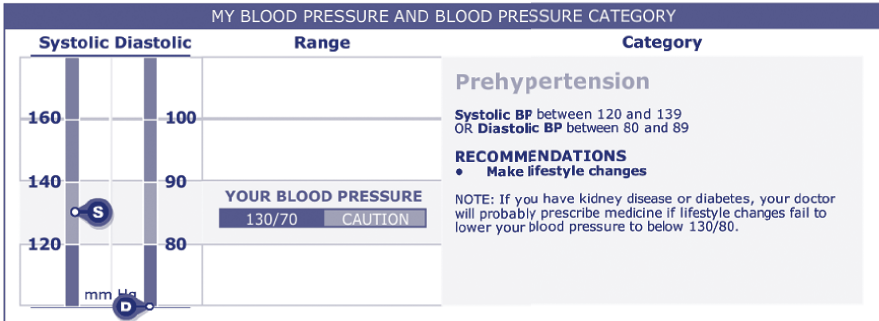
<sup>|\*</sup> De gebruikte calculator is te vinden op: [http://www.americanheart.org/beatyourrisk/en\\_US/main.html](http://www.americanheart.org/beatyourrisk/en_US/main.html). Een andere, meer uitgebreide 'risk assessment tool' is beschikbaar op: [http://post.americanheart.org/risk/chd\\_updates/flash\\_content/preloader.htm](http://post.americanheart.org/risk/chd_updates/flash_content/preloader.htm)

**HIGH BLOOD PRESSURE HEALTH RISK CALCULATOR**  
**ASSESSMENT PART 1: MY HEALTH RISKS**



NAME:

11/16/2007



Figuur 2. Berekening van risico op hart- en vaatziekten voor een 40-jarige man met een bloeddruk van 130 over 70, en geen andere risicofactoren.

**Individuele preventie**

Op het individu gerichte preventie speelt sterk bij de discussie wie er cholesterolverlagende medicijnen (statines) moet gebruiken. Het ligt voor de hand dat deze middelen moeten worden voorgeschreven aan mensen met een verhoogd cholesterol. Cholesterol verlagende medicijnen leiden echter bij iedereen tot een relatieve verlaging van het risico op hart- en vaatziekten, ook als het cholesterol niet duidelijk verhoogd was. De richtlijnen voor cholesterolverlagers stellen dat mensen met een verhoogd risico deze medicijnen dienen te krijgen, en niet per se mensen met een verhoogd cholesterol. Verhoogd risico is belangrijker dan een verhoogd cholesterol.

Bij dit soort afwegingen spelen ook kosten een rol: als het risico op hart- en vaatziekten

beperkt is, is ook de winst van deze medicijnen beperkt, en zijn de kosten hoog voor wat we er mee bereiken. In deze tijd van toenemende preventieve mogelijkheden is een selectief beleid belangrijk. Prognostische modellen zijn daarbij essentieel.

---

## Samenvattend

- Gerandomiseerde studies geven sterk bewijs voor het gemiddelde effect van een behandeling; dit effect is goed uit te drukken in een relatief risico
- Het absolute risico kan goed geschat worden met een prognostisch model, waarmee individuele beslissingen ondersteund kunnen worden
- Goed beslissen vraagt om goed voorspellen

## Voorspellen

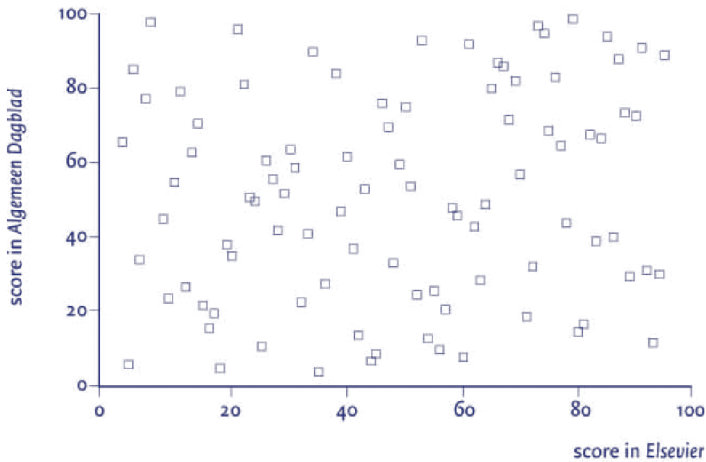
In tal van gebieden blijkt dat voorspellen heel moeilijk is. Denk aan het voorspellen van het weer; een Engelse weerkundige voorspelde een zogenaamde “superstorm”. Er zou tussen 24 en 28 november 2007 een superstorm met 80% waarschijnlijk over Nederland razen; de voorspelling was gebaseerd op zonne-activiteit. Deze voorspelling kwam niet uit. Misschien een geval van pech; of het nogal exotische model klopte gewoon echt niet, zoals het KNMI heeft gesteld.

Of denk aan aandelenkoersen: velen proberen rijk te worden door op het juiste moment te kopen of verkopen. Allerlei modellen worden door beursexperts gebruikt voor technische en fundamentele analyse. In dit verband refereer ik graag aan 2 apen, Jacko en Raven. Jacko is een gorilla in een Berlijnse dierentuin; hij eet elke veertien dagen één banaan uit een mandje, waarbij elke banaan een beursfonds vertegenwoordigt dat verkocht of juist aangekocht moet worden. De keus van Jacko is uiteraard volstrekt willekeurig, maar zijn beleggingsresultaten zijn verbazingwekkend. De chimpansee Raven uit New York is vermoedelijk nog iets bekender. Hij gooide voor het eerst in januari 1999 tien pijltjes naar een dartbord waar alle 133 vakken een aandeel vertegenwoordigden. Aan het eind van het jaar was zijn selectie 213 procent meer waard geworden. Op de ranglijst van circa 6000 Amerikaanse beursexperts bereikte hij daarmee de 22e plaats.

## *Kwaliteit van ziekenhuizen*

Laat ik in verband met toevalligheden een kort uitstapje maken naar het ordenen van de kwaliteit van Nederlandse ziekenhuizen. Hierbij spelen vele problemen: wat is kwaliteit, hoe meten we dat, en wat is de rol van toeval. In het Nederlands Tijdschrift voor Geneeskunde van 2006 werd de score van de kwaliteit van een ziekenhuis volgens Elsevier vergeleken met de score volgens het Algemeen Dagblad (Figuur 3)<sup>10</sup>. Er werden verschillende methoden gebruikt voor het ordenen van kwaliteit, maar het is opmerkelijk dat de ranglijsten geen enkele overeenkomst vertoonden; we zien een

puntenwolk. Deze ranglijsten moeten we voorlopig dus niet al te serieus nemen, ze kunnen in ieder geval niet alle 2 gelijk hebben.



Figuur 3. Correlatie tussen de ziekenhuisranglijsten van het Algemeen Dagblad en Elsevier. Er is geen overeenstemming van de beoordelingen (correlatiecoëfficiënt  $r=0,14$ )<sup>10</sup>.

### Samenvattend

- Voorspellen is moeilijk
- De rol van toeval (geluk of pech) is niet te onderschatten
- Ranglijsten van ziekenhuizen zijn niet te vertrouwen; ze hadden net zo goed door een pijltjes gooierende aap samengesteld kunnen worden

## Individualized evidence-based medicine

Terug naar kansen en beslissen. In het ideale geval is de toekomstige geneeskunde er één van 'Individualized evidence-based medicine'. Hiervoor moet bekend zijn hoe effectief elke te overwegen medische interventie is. Ik merk hierbij op dat dit voor vele testen en behandelingen helemaal niet het geval is. Er is voor slechts weinig behandelingen in de moderne geneeskunde onderbouwing door een gerandomiseerde studie. We moeten dan terugvallen op minder sterk bewijs. Dit zien we bijvoorbeeld terug in richtlijnen van de medische beroepsgroepen, waar vaak op de opinie van de expert moet worden vertrouwd. Hier is nog veel werk te doen. Ik wil echter met name ingaan op 2 andere punten: het voorspellen van individuele risico's, en het nemen van beslissingen op basis van de verwachte individuele winst.

Ik bespreek hierbij 3 grote uitdagingen. Ten eerste moeten we sterke voorspellers hebben. Ten tweede moeten we ze combineren met de juiste gewichten in een model, en ten derde moeten we de modellen zien te implementeren in de praktijk.

### 1. Sterke voorspellers

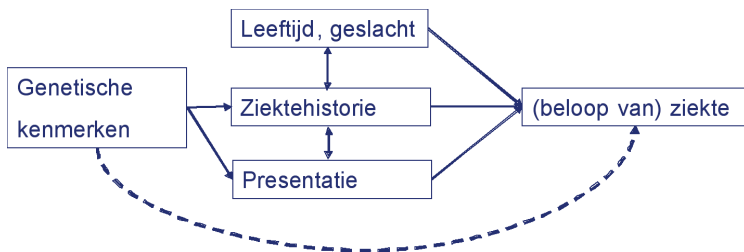
Om goed te voorspellen hebben we voorspellende kenmerken nodig ('predictoren'). Onderzoek naar nieuwe voorspellers is heel populair; iedere medisch wetenschapper wil graag een nieuwe 'onafhankelijke risicofactor' ontdekken. Het betreft dan vaak onderzoek naar relatieve effecten: hoe is de uitkomst van de groep mensen met aanwezigheid van de predictor vergeleken met degenen zonder de predictor. Denk bijvoorbeeld aan een verhoogde bloeddruk die bepaalde risico's 2 x verhoogt, zoals eerder geïllustreerd.

De specifieke voorspellers van het hebben van een ziekte of van de uitkomst van een ziekte hangen uiteraard af van het bestudeerde ziektebeeld. Maar er zijn wel een paar algemene uitgangspunten te noemen. Figuur 4 geeft een schematisch voorbeeld. Eenvoudig te verkrijgen kenmerken zoals leeftijd en geslacht zijn vaak heel belangrijk. Gegevens uit de ziektehistorie komen daarbij, en de presentatie van de patiënt. Deze kenmerken hangen met elkaar samen. Deze samenhang moet in acht genomen worden in statistische modellen; hier kom ik straks op terug.

#### *Genetische kenmerken om te voorspellen*

Tegenwoordig is er veel interesse in 'genetische markers' voor betere individuele predictie. Het menselijk DNA verschilt tussen individuen, en zou daarom gebruikt kunnen worden om het risico op het optreden van een ziekte te kunnen voorspellen. Toch zou de bijdrage van kennis over genetische markers aan beter voorspellen nog wel eens tegen kunnen vallen. We kunnen ons namelijk voorstellen dat genetische markers een verhoogd risico geven op een ziekte, of een slechte uitkomst van een ziekte, maar dat dit al tot uiting komt via allerlei makkelijk te bepalen klinische klachten en symptomen. Het gaat om het aanvullende effect van een genetisch kenmerk, zoals aangegeven met de gestippelde pijl in Figuur 4.

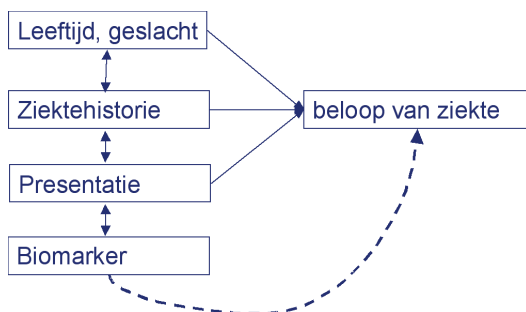




Figuur 4. Schematische weergave van de bijdrage van genetische kenmerken aan het voorspellen van het optreden of beloop van een ziekte. Het onafhankelijke effect van een genetisch kenmerk wordt weergegeven door de gestippelde pijl.

### *Biomarkers om te voorspellen*

**D**aarentegen zijn nieuwe zogenaamde biomarkers die de ernst van de ziekte weergeven wellicht wel bruikbaar om beter te voorspellen (Figuur 5). Bij deze biomarkers kunnen we bijvoorbeeld denken aan allerlei stoffen in het bloed, of aan kenmerken van kankercellen. Hier is verder, veelal fundamenteel, onderzoek nodig. Nieuwe markers moeten na de laboratoriumfase geanalyseerd worden op hun meerwaarde voor de medische praktijk. De vraag die beantwoord moet worden is of een nieuwe biomarker ons echt helpt om beter te screenen op een verhoogd risico op ziekte in de algemene bevolking, om sneller een diagnose te stellen, of om beter te voorspellen wat de uitkomst van ziekte zal zijn? Hiervoor moeten we de vergelijking maken tussen een voorspellend model met en zonder de marker; het verschil tussen deze modellen geeft de meerwaarde van een marker aan.



Figuur 5. Schematische weergave van de bijdrage van een biomarker aan het voorspellen van het van een ziekte. Het onafhankelijke effect van de biomarker wordt weergegeven door de gestippelde pijl.

---

## Samenvattend

- Sterke predictoren zijn nodig voor het doen van voorspellingen
- Nieuwe predictoren moeten aanvullende informatie bieden boven traditionele predictoren; dit kan goed onderzocht worden met een prognostisch model

## 2. Combineren van predictoren

Laten we ervan uit gaan dat er een aantal sterke voorspellers bekend is. Voor een individuele voorspelling moeten we predictoren op een verstandige manier combineren. In de medische wetenschap worden hiervoor multivariabele regressiemodellen het meest gebruikt. De kans op sterfte binnen 30 dagen na een hartinfarct kwam uit een logistisch regressiemodel, waarin verschillende combinaties van prognostische factoren werden geanalyseerd bij ruim 40.000 patiënten<sup>3</sup>. Dit grote aantal is echter uitzonderlijk, en vaak zijn slechts relatief kleine datasets beschikbaar om een model te ontwikkelen.

Over het verstandig en betrouwbaar combineren van predictoren zijn slechts enkele boeken beschikbaar; ik noem het boek van Michael Katz, dat makkelijk toegankelijk is voor klinici<sup>11</sup>. Een stuk moeilijker is Frank Harrell's *Regression Modeling Strategies*<sup>6</sup>. Dit is een verplicht werk voor onderzoekers die een voorspelmodel willen ontwikkelen. Heel goed, maar bijna onleesbaar voor klinici is 'the Elements of Statistical Learning'<sup>12</sup>.

### *Nog een boek over prognostische modellen?*

Uit overzichtsstudies blijkt dat er veel is aan te merken op prognostische modellen die in de medische literatuur verschijnen. Daarom is er volgens mij behoefte aan een praktisch boek over prognostisch modelleren, qua moeilijkheid gepositioneerd tussen de boeken van Harrell en Katz in. Dit boek, 'Clinical Prediction Models', waaraan ik nu al enkele jaren werk, verschijnt medio 2008 bij Springer<sup>7</sup>. U bent natuurlijk, net als ik, reuze nieuwsgierig naar de inhoud van dat boek; het is nog niet helemaal af.

Wat staat er zoal in het boek? Het begint met een beschrijving van de achtergrond van modellering. De kern bestaat uit 7 stappen voor het ontwikkelen van een valide predictiemodel, samengevat in een praktische checklist (Tabel 4). Daarnaast zult u veel aantreffen over modelvalidatie, dat wil zeggen het toetsen of een ontwikkeld model ook echt werkt. Een belangrijk probleem van prognostische modellen is dat ze vaak gebaseerd worden op relatief kleine data sets; er is dan slechts een beperkte statistische kracht voor het toetsen van hypothesen en voor het betrouwbaar schatten van de parameters in een model<sup>13</sup>. Het is vaak wel mogelijk om een model te maken dat goed lijkt te werken voor de patiënten in de data set waarop het model is ontwikkeld. Het gaat echter niet om retrodictie (achteruit voorspellen), maar predictie (vooruit voorspellen), en dat moet goed getoetst worden.

|\*\* Achtergrond informatie over het boek is te vinden op [www.clinicalpredictionmodels.org](http://www.clinicalpredictionmodels.org)

Deze problemen en oplossingsrichtingen worden geïllustreerd met vele voorbeelden. Het boek bevat veel figuren, en weinig formules. Hopelijk zal het bijdragen aan het maken van betere voorspelmodellen, waar mensen in de algemene bevolking en patiënten in het ziekenhuis echt wat aan hebben.

Step	Specific issues
<b>General considerations</b>	
Research question	Aim: predictors / prediction?
Intended application?	Clinical practice / research?
Outcome	Clinically relevant?
Predictors	Reliable measurement? Comprehensiveness
Study design	Retrospective/prospective? Cohort; case-control
Statistical model	Appropriate for research question and type of outcome?
Sample size	Sufficient for aim?
<b>7 modeling steps</b>	
1. Preliminary	Missing values
2. Coding of predictors	Continuous predictors Combining categorical predictors Combining predictors with similar effects
3. Model specification	Appropriate selection of main effects? Assessment of assumptions (distributional, linearity and additivity)?
4. Model estimation	Shrinkage included? External information used?
5. Model performance	Appropriate measures used?
6. Model validation	Internal validation including model specification and estimation? External validation?
7. Model presentation	Format appropriate for audience
<b>Validity</b>	
Internal: overfitting	Sufficient attempts to limit and correct for overfitting?
External: generalizability	Predictions valid for plausibly related populations?

Tabel 4. Checklist voor het ontwikkelen van een betrouwbaar prognostisch model. Details in het in 2008 te verschijnen boek 'Clinical Prediction Models'.

### *Methodologische uitdagingen*

**N**aast regressieanalyse zijn er allerlei andere statistische technieken om gewichten aan predictoren te geven. Moderne technieken zijn bijvoorbeeld ‘Support vector machines’, ‘Neural Networks’, of een zogenaamde ‘Random forest’. Het toenemend gebruik van deze modernere methoden hebben we grotendeels te danken aan de toegenomen rekenkracht van onze computers. Op basis van de beschikbare vergelijkende studies is echter geen groot voordeel te verwachten van alleen dit soort nieuwe statistische methoden. Beter gebruik van bestaande methoden is belangrijker dan gebruik van geheel nieuwe methoden.

Er is nog wel een aantal andere grote methodologische uitdagingen op het gebied van prognostisch modelleren. Een eerste uitdaging is hoe we gegevens uit verschillende bronnen verstandig combineren. Eerder heb ik gewerkt aan het ontwikkelen en toepassen van methoden voor het combineren van gegevens uit gepubliceerde literatuur met gegevens van individuele patiënten<sup>14</sup>, en aan methoden voor het up-to-date maken van een bestaand model<sup>15</sup>, maar hier is nog veel meer werk te doen.

Ten tweede noem ik het aanbieden van predictieregels in een geautomatiseerde omgeving. Ideaal zou zijn om een lerend voorspelmodel te creëren in een elektronisch patiënten dossier. Het systeem zou steeds gevoed kunnen worden met nieuwe modellen, die meer toegesneden worden naarmate er meer gegevens van de lokale setting beschikbaar komen.

Ten derde zijn er ontwikkelingen rond het meten van de kwaliteit van een model. Hoe goed zijn de voorspellingen voor het ondersteunen van beslissingen? Sommige maten zoals ‘reclassificatie’ ontberen een beslistkundige inbedding<sup>16,17</sup>, maar beslistkundige maten zijn vaak weer moeilijk te volgen door de clinicus practicus<sup>18</sup>.

Tenslotte noem ik de generaliseerbaarheid van voorspellingen van een model. Vaak zien we systematische verschillen tussen de voorspellingen van een model en de gevonden uitkomsten bij nieuwe patiënten<sup>19,20</sup>. Deze zogenaamde miscalibratie kunnen we achteraf soms wel verklaren; we hebben dan misschien een belangrijke voorspeller gemist in ons model. Maar we moeten meer zicht krijgen op de oorzaken van miscalibratie om generaliseerbare modellen te maken, of aan te kunnen geven wat de grenzen zijn van de toepasbaarheid van een model.

Ik hoop de komende jaren verder aan deze methodologische uitdagingen te werken.

---

### **Samenvattend**

- Regressieanalyse biedt goede mogelijkheden voor het ontwikkelen van voorspelmodellen; beter gebruik van bestaande methoden is belangrijker dan het ontwikkelen van geheel nieuwe methoden
- Er zijn diverse methodologische uitdagingen op het gebied van prognostisch modelleren, waaronder het combineren van informatiebronnen, continu updaten, meten van de kwaliteit van een model, en inzicht in generaliseerbaarheid

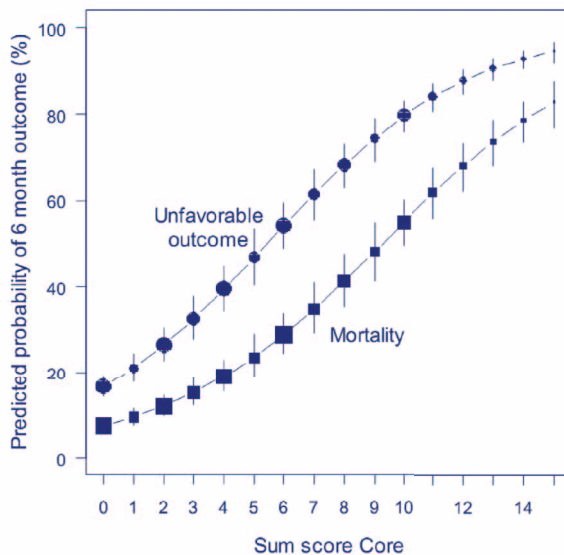
Characteristics	Value	Score	Sum
Age (years)	<= 30	0	
	30 – 39	1	
	40 – 49	2	
	50 – 59	3	
	60 – 69	4	
	70 +	5	
Motor score	None/Extension	6	
	Abnormal flexion	4	
	Normal flexion	2	
	Localizes/obeys	0	
	Unstable/missing	3	
Pupillary reactivity	Both pupils reacted	0	
	One pupil reacted	2	
	No pupil reacted	4	
<b>Sum score Core model</b>			

Figuur 6. Scorekaart voor prognose na ernstig schedelhersensletsel. Een 40-jarige patiënt in coma, maar zonder andere heel ernstige problemen, krijgt een score van 2 punten.

### 3. Implementatie: betere beslissingen?

Stel nu dat we een regressiemodel hebben ontwikkeld, dan willen we ook dat dit model in de medische praktijk gebruikt gaat worden. Ik ga ervan uit dat er redelijk gemodelleerd is, bijvoorbeeld volgens de aanpak zoals beschreven in Frank Harrell's of in mijn boek; het model is gevalideerd, dat wil zeggen getest in op zijn minst dezelfde patiëntengroep (interne validatie), en liefst ook met nieuwe patiënten (externe validatie). Maar als aan deze methodologische eisen is voldaan, is het ook essentieel dat er een eenvoudige presentatie plaatsvindt.

Op papier zijn er vele mogelijkheden. Figuur 6 laat een voorbeeld zien van een score kaart, in dit geval voor het voorspellen van een slechte uitkomst 6 maanden na opname in het ziekenhuis vanwege ernstig schedelhersensletsel, bijvoorbeeld na een auto-ongeval. In Figuur 6 wordt een score berekend, die past bij een kans op sterfte en een kans op een slechte uitkomst (Figuur 7). We kunnen dit prognostische model ook via een spreadsheet aanbieden, en op het internet. In de toekomst zal een geautomatiseerde aanbieding steeds meer opgang doen; niet alleen spreadsheets en webcalculators, maar ook programmaatjes voor de PDA, mobiele telefoon, of integratie in het al eerder genoemde elektronisch patiënten dossier.



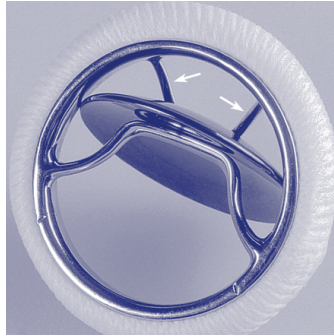
Figuur 7. Voorspelde kansen op sterfte en een slechte uitkomst na ernstig schedelhersenletsel. Een score van 2 in Figuur 6 past bij een kans op sterfte van 10%, en een kans op slechte uitkomst van 30%. De onzekerheid in de voorspellingen wordt weergegeven met een 95% betrouwbaarheidsinterval.

### Worden er betere beslissingen genomen?

Het eerder besproken model voor patiënten met een acuut hartinfarct is eenvoudig en overtuigend. In dezelfde geest is er gerekend door onderzoekers van de afdeling cardiologie van het Erasmus MC en zijn er protocollen in gebruik genomen waarbij patiënten behandeld werden op basis van hun individuele risico<sup>21</sup>. Dit is een mooi voorbeeld van succesvolle implementatie in een academisch centrum zoals het Erasmus MC. Inmiddels is de behandeling al weer anders, met dotteren van de kransslagader als behandeling van eerste keus, in plaats van de stolseloplossende medicijnen. In het algemeen is de cardiologie behoorlijk kwantitatief ingesteld, en nemen voorspelmodellen er een belangrijke plaats in.

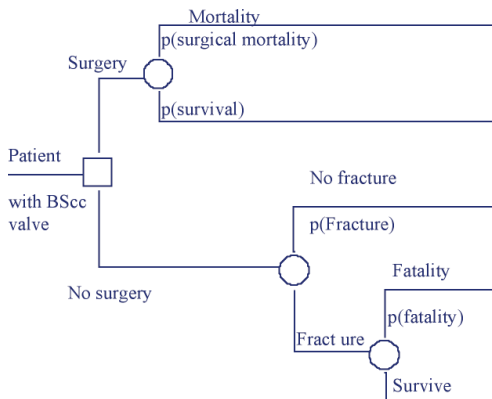
### Gevaarlijke hartkleppen

Een ander succesvol voorbeeld is het volgende. Eind jaren 80 werden er mechanische hartkleppen bij patiënten geïmplant, waarbij later bleek dat er een risico was op breuk (Figuur 8). In 1979 kwam de klep op de markt, en in de jaren 80 werd deze er weer vanaf gehaald.



Figuur 8. Een Björk-Shiley convexo-concave hartklep, waarbij de kleine beugel (aangegeven met de pijlen) kan afbreken.

In 1992 werden epidemiologische gegevens gepubliceerd; met name een Nederlandse studie van collega Van der Graaf uit Utrecht liet zien dat het gemiddelde risico op breuk substantieel was, en dat de risico's behoorlijk verschilden tussen verschillende typen kleppen<sup>22</sup>. Wij voerden hier in Rotterdam besliskundige analyses uit om specifieke richtlijnen te geven voor thoraxchirurgen over welke kleppen vervangen zouden moeten worden. De eerste besliskundige publicatie was in 1993<sup>23</sup>, en er werd een meer genuanceerde update gemaakt in 1998<sup>24</sup>.



Figuur 9. Schematische weergave van de beslisboom voor de keuze voor opereren of afwachten bij een patiënt met een Björk-Shiley convexo-concave hartklep.

We gebruikten een aantal prognostische modellen in onze beslissingsanalyse: voor de operatiesterfte bij vervangen van de gevaarlijke klep door een ander type, voor klepbreuk en sterfte bij klepbreuk, en tenslotte voor de overlevingskansen als er geen klepbreuk op zou treden (Figuur 9)|<sup>25</sup>. Voor individuele patiënten kon zo het optimale beleid worden bepaald: opereren of afwachten.

Uit een evaluatie bleek dat de patiënten bij wie een klep vervangen was door de thoraxchirurgen ook werkelijk tot de groep behoorden die besliskundig gezien in aanmerking kwam, dat wil zeggen, de jongere patiënten met een nog relatief lange levensverwachting, en risicovolle hartkleppen|<sup>24</sup>. Dit is een prachtig voorbeeld van hoe epidemiologen, besliskundigen en artsen kunnen samenwerken om optimale beslissingen te nemen.



Figuur 10. Voorbeeld van een restmassa (aangegeven met de pijlen) na behandeling met chemotherapie vanwege uitgezaaide testiskanker. Operatie is alleen nuttig indien de restmassa kwaadaardig is.

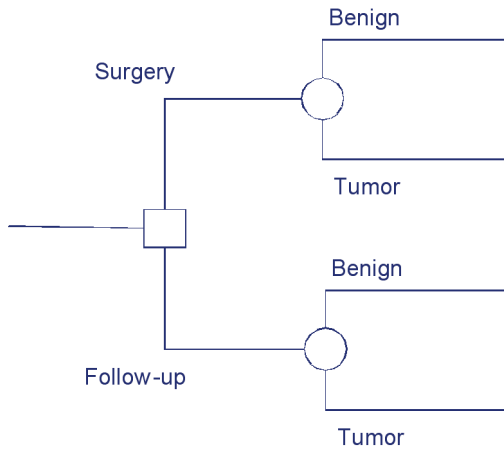
### *Restmassa's bij testistumoren*

Rond de implementatie van prognostische modellen heb ik helaas ook wel minder positieve ervaringen. Sinds het eind van mijn studie in Leiden heb ik mij beziggehouden met analyses bij mannen met testistumoren|<sup>26</sup>. Chemotherapie is een zeer effectieve behandeling van deze kanker. Na behandeling met chemotherapie zijn



vaak restmassa's aanwezig die operatief verwijderd kunnen worden. Figuur 10 geeft een voorbeeld van een grote restmassa.

De massa's kunnen goed- of kwaadaardig zijn. Voor kwaadaardige massa's is een operatie nuttig, maar voor goedaardige massa's is een operatie overbodig (Figuur 11). We ontwikkelden prachtige modellen om de kans op goedaardig weefsel te voorspellen, valideerden die, en publiceerden die in topklinische tijdschriften<sup>27, 28, 29, 30</sup>. Maar ik krijg de indruk dat zelfs mijn co-auteurs de modellen niet echt gebruiken om te beslissen over opereren.



Figuur 11. Schematische weergave van de beslisboom voor de keuze voor opereren of afwachten bij een patiënt met een restmassa na behandeling met chemotherapie vanwege uitgezaaide testiskanker.

### *Gebruik van een beslismodel*

**H**oe komt het dat dit model niet veel gebruikt wordt? Misschien is het model niet goed genoeg. Het blijkt echter dat het model redelijk goed voorspelt naar gebruikelijke statistische maatstaven, zoals discriminerend vermogen<sup>16, 30</sup>.

Een andere verklaring is de moeilijkheid om 'kansrijk', of probabilistisch, te denken. Wat kunnen we met een kans van 75% op goedaardig weefsel? Het is geen 100%, dan wisten we zeker wat we moesten doen. Het klinisch redeneren van de oncologen is meer klassiek mechanisch. Het gebruikelijke klinisch beleid is hedentendage nog steeds dat massa's die radiologisch gezien afwijkend zijn worden verwijderd, dat wil zeggen dat de diameter meer is dan 1 cm. Normale lymfklieren zijn namelijk maximaal 1 cm in grootte. Deze manier van redeneren in termen van pathofysiologie zal moeten

veranderen naar een evidence-based, kwantitatief probabilistische redenering. Tot die tijd worden onnodig veel mannen met goedaardig restweefsel geopereerd, en aan de andere kant mannen met kwaadaardig restweefsel niet geopereerd terwijl ze dat wel nodig hebben. Het is duidelijk dat ik dat een onbevredigende situatie vind.

### Implicaties

**E**r is een aantal implicaties van deze voorbeelden, ten eerste voor het medisch onderwijs. Er moet aan onze medische studenten een kwantitatieve denkwijze worden bijgebracht. Dit is voor de meeste mensen onnatuurlijk, en daarom moeilijk. Mensen denken nu eenmaal makkelijker in simpele heuristieken, in regeltjes, dan in expliciete kansen<sup>31</sup>. Toch vraagt beter beslissen om een meer expliciete afweging van de voor- en nadelen van een keuze.

Ten tweede moeten we besliskundige berekeningen zo makkelijk mogelijk aanbieden voor de medische praktijk, als we willen dat er iets mee gedaan wordt. De toenemende automatisering van de gezondheidszorg biedt hier vele kansen.

---

### Samenvattend

- Prognostische modellen moeten zo eenvoudig mogelijk gepresenteerd worden
- Prognostische modellen kunnen een directe invloed hebben op medische beslissingen, met name als de gebruiker open staat voor een kwantitatieve, probabilistische manier van redeneren.

## Toekomst

In het laatste deel van deze oratie kijk ik naar de toekomst. Hoe gaan wij dit alles voor elkaar krijgen? Moeten we streven naar prognostische modellen voor elke ziekte, en diagnostische voorspelmodellen voor alle differentiaal diagnoses?

Een voorspelmodel is met name relevant als er predictoren bekend zijn. Als er geen predictoren bekend zijn op het moment van beslissen, valt er niets te voorspellen, en ook niets te individualiseren.

Een tweede voorwaarde is dat de besluitvorming van voorspellingen af moet hangen. Als we een weinig belastende behandeling hebben, die heel effectief is, zullen we die altijd geven, en geen prognostisch model nodig hebben. Verder moet er een duidelijke therapeutische consequentie zijn van de diagnose als we een diagnostisch voorspelmodel maken.

## Toepassingsgebieden

Er zijn diverse min of meer klassieke toepassingsgebieden van prognostische modellen. Voor de heelkunde is bijvoorbeeld sterfte na een operatie relevant voor chirurgische besluitvorming. De winst van een operatie moet worden afgewogen tegen de individuele risico's, zoals in het voorbeeld van de hartklepvervangingsoperatie. Voor patiënten met kanker is de kans op progressie van ziekte en de verwachte overleving vaak cruciaal voor keuzes rond type en intensiteit van behandeling.

Het aantal toepassingsgebieden zal toenemen door de toename in kennis over predictoren, de toenemende behandelingsmogelijkheden (meer keuzes), en de toenemende automatisering die implementatie van prognostische modellen makkelijker mogelijk maakt.

Nieuwe toepassingen zullen er bijvoorbeeld komen in de preventieve gezondheidszorg.

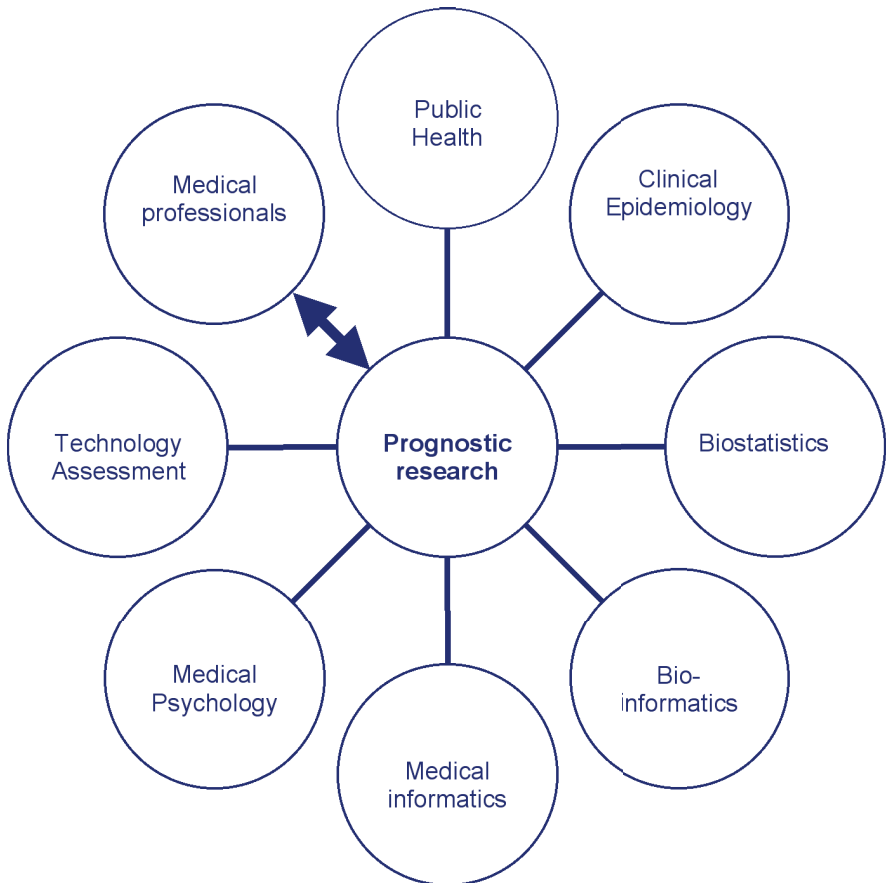
Met voorspelmodellen kan preventieve behandeling gericht worden op personen met een hoog risico, zoals we bijvoorbeeld proberen rond het optreden van hart- en vaatziekten. Ook kunnen we zogenaamde 'health checks' specifiek aanbieden aan die personen die er mogelijk baat bij hebben, en niet aan iedereen. Verder kunnen we de intervallen van kankerscreening op maat gaan maken als we individuele risicokenmerken kennen, net als het opvolgen van voorstadia van kanker (bijvoorbeeld het colonoscopie beleid na verwijderen van een darmpoliep, en het endoscopie beleid voor mensen met een zogenaamde Barrett slokdarm). Interessant wordt ook of we bepaalde kankers kunnen volgen in de tijd ('Active Surveillance') in plaats van ze direct agressief te behandelen (bijvoorbeeld vermoedelijk onschuldige prostaat-kankers).

Het slagen van ontwikkelingen in deze richting hangt af van vele factoren. Voor een goed prognostisch model zijn ten eerste goede data nodig. De mogelijkheden voor dataverzameling voor onderzoek nemen toe met de beschikbaarheid van steeds meer elektronische databases en registraties. De kwaliteit van de data in deze systemen is een belangrijk aandachtspunt, maar in potentie kan dit interessante

onderzoeksmogelijkheden opleveren.

Goede modellering vereist ook véél data. Hiervoor zal vaak samenwerking tussen onderzoeksgroepen nodig zijn. We zien dit mooi van de grond komen in de vorm van consortia voor onderzoek naar genetische markers<sup>32</sup>. Iets dergelijks moet ook voor andere, meer traditionele predictoren vaker nagestreefd worden.

---



---

Figuur 12. Schematische weergave van belangrijke disciplines die kunnen bijdragen aan succesvol prognostisch onderzoek.

## Organisatie

**E**r zijn ook organisatorische consequenties. Onderzoek aan prognostische modellen moet plaatsvinden in samenwerking (Figuur 12). Ten eerste kijk ik naar de sterke onderzoekslijnen binnen mijn eigen afdeling, Maatschappelijke Gezondheidszorg, of 'Public Health'. Prognostisch onderzoek kan hier nog uitgebreid worden in allerlei toepassingen.

Bij de methodologen, met name de gezondheidswetenschappers, zijn er raakvlakken met vele groepen. Ik noem de in mijn perceptie belangrijkste disciplines:

- Klinische epidemiologie, bijvoorbeeld voor de aansluiting bij klinische studies
- Biostatistiek, bijvoorbeeld voor het in de diepte verder ontwikkelen van methoden
- Bioinformatica, bijvoorbeeld voor het ontdekken van nieuwe markers
- Medische Informatica, bijvoorbeeld voor het ontsluiten van data en geautomatiseerd aanbieden van voorspelmodellen
- Medische Psychologie, bijvoorbeeld voor de psychologische aspecten van medische beslissingen
- Technology assessment, zoals in Rotterdam geconcentreerd in het 'institute for medical technology assessment' (iMTA), bijvoorbeeld voor meer ingewikkelde economische vraagstellingen rond kosten en kosten-effectiviteit.

Verder is het essentieel om aan relevante vraagstellingen te werken; voor medische vraagstellingen is input nodig van medici die in het ziekenhuis of in de publieke gezondheidszorg werken. Zij kunnen aangeven op welk gebied er sterke predictoren bekend aan het raken zijn, en waar de moeilijke beslissingen liggen die ondersteund zouden kunnen worden door een prognostisch model. Ik doe dan ook een beroep op mijn meer medisch geschoolde collega's om actief over deze vragen na te denken. Ik verheug mij op verdere samenwerking met diverse onderzoekers en onderzoeksgroepen in en buiten Nederland.

### Prognostisch onderzoek in het Erasmus MC

**D**it soort samenwerkingen passen goed in de strategienota van het Erasmus MC, Koers '013. Hierin wordt samenwerking benadrukt. De methodologen van het Erasmus MC werken al jaren samen in het Consultatiecentrum voor Patiëntgebonden Onderzoek. Dit is een belangrijk aanspreekpunt voor klinici en andere onderzoekers die een onderzoek hebben uitgevoerd, of willen gaan uitvoeren en daarvoor een subsidieaanvraag voorbereiden. De samenwerking zal kunnen worden versterkt in een 'Center for Quantitative Methods', waarin de verschillende methodologische groepen participeren voor hun eigen methodologisch georiënteerd onderzoek. Ik vertrouw erop dat dit de ontwikkeling van prognostisch modelleren ten goede komt, zowel in theoretisch als meer toegepast opzicht.

---

**Samenvattend:**

- De toepassingsgebieden voor prognostisch onderzoek zullen zich uitbreiden, zowel in de klinische als preventieve gezondheidszorg
- Goed prognostisch onderzoek vereist samenwerking tussen methodologische onderzoeksgroepen en met medische professionals

Dames en heren, ik ben deze oratie begonnen met u mee te nemen naar individuele behandelingen van hartinfarcten, en het gericht voorkómen van hartinfarcten, als voorbeelden van ‘individualized evidence-based medicine’. Ik heb laten zien dat prognostische modellen, of meer in het algemeen ‘voorspelmodellen’, hier en in vele andere situaties, een essentiële rol vervullen. Dit soort modellen biedt kansen om beter te beslissen.

Ik wil graag afsluiten met een kort dankwoord.

## Dankwoord

Ik wil allen bedanken die hebben bijgedragen aan mijn wetenschappelijke ontwikkeling tot nu toe, en die dat hopelijk ook nog geruime tijd blijven doen.

Daarbij denk ik aan collega's van de afdeling Maatschappelijke Gezondheidszorg, specifiek mijn directe collega's bij het Centrum voor Medische Besliskunde, waaronder Cecile Janssens, Marie-Louise Essink-Bot, en René Eijkemans; collega's in het Erasmus MC, met name bij de afdeling Epidemiologie en Biostatistiek; en vele collega's daarbuiten, onder andere bij het Julius Centrum in Utrecht, de afdeling Medische Statistiek in Leiden, epidemiologen in Amsterdam, en in Boston bij het Dana-Farber/Harvard Cancer Center. Ik heb veel geleerd van de klinici met wie ik nu al vele jaren samenwerk, waaronder Henriëtte Moll, Andrew Maas, Ernst Kuipers, en Peter Siersema. Ik kan slechts een enkeling hier verder specifiek noemen.

Dik Habbema, hoogleraar Medische Besliskunde aan deze universiteit. Dik: we lijken in veel opzichten op elkaar; we houden van rekenen; de titels van onze proefschriften verschillen maar in 1 woord, onze leeropdrachten zijn gelijk behalve dat er bij mij wat 'bijzonders' aan is en bij jou wat 'gewoons'. Ik heb veel van je geleerd sinds ik in 1991 solliciteerde bij jouw afdeling, toen nog het zelfstandige Centrum voor Klinische Besliskunde, tegenwoordig het Centrum voor Medische Besliskunde. Dank voor de grote vrijheid die je mij gaf om me te kunnen ontwikkelen op die gebieden die mij aantrokken, met name de prognostische modellen.

Verder wil ik Johan Mackenbach bedanken, die de afdeling MGZ nu alweer een aantal jaar met vaste hand als afdelingshoofd leidt. Johan, ook van jou heb ik de laatste jaren veel geleerd, met name over hoe onderzoek te organiseren, en ik bedank je voor de voordracht om hoogleraar te mogen worden. Ik dank ook de Raad van Bestuur van het Erasmus MC en het College van Bestuur van deze Universiteit voor het in mij gestelde vertrouwen.

It was a privilege to have several foreign guests this week, among whom Jay Piccirillo and Ed Spitznagel. I hope they survived the test of the Dutch language that this lecture must have been for them, and look forward to further collaboration.

Special words of thanks go to 2 eminent methodologists who were influential in arousing my interest in prognostic research: Frank Harrell and Hans van Houwelingen. To me they are the fathers of modern prognostic research.

Frank: I still remember reading some of your papers in 1992, when we were working on prognostic models for the heart valve replacement problem. These papers were very different from other biostatistical papers that I had read before; this was on strategies, on 'how to make a model' instead of describing a statistical procedure with its assumptions<sup>33</sup>. Your work on the 'Design library' is of tremendous importance to my more applied work. Thank you so much for your presence here, and all the inspiration you have given modelers here in the Netherlands and abroad. We share a passion for Russian composers - such as Shostakovich -, and I hope you appreciate my choice of Prokofiev's 7<sup>th</sup> symphony as musical surrounding for this occasion.

Hans: ik weet nog goed dat ik in 1990 bij jou wel eens aan de lunchtafel zat, toen

ik mijn afstudeeronderzoek deed bij Ronald Brand in Leiden. Meestal begreep ik niet veel van wat je nou eigenlijk zei. Dat is wel een beetje zo gebleven; je beroemde artikel uit 1990 met Saskia le Cessie heb ik 3 keer moeten lezen om de essentie ervan half te begrijpen<sup>34</sup>. Voor mij ben je de theoretisch-statistische goeroe. Dank voor al je tijd die je aan mijn soms wat vage plannen hebt besteed om ze biostatistisch goed doortimmerd te maken. Ik hoop nog veel meer van je te leren.

Tenslotte kom ik bij mijn familie. Ik ben dankbaar hier vandaag zovele familieleden te zien, op dit moment zonder de zich steeds uitbreidende kinderschare, en ik hoop nog veel van hun warme gezelligheid te mogen meegenieten. Ik bedank mijn vader voor het steunen van mijn wetenschappelijke vorming in de ruimste zin. Ik sta in gedachten ook stil bij enkelen die hier niet bij zijn, met name bij mijn moeder, die helaas vele jaren geleden al is overleden.

Echter 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 en relativering, en Matthijs, Laurens, en Suzanne voor hun drukke maar vrolijke noten in ons huis.

Ik dank u allen voor uw aandacht.

Ik heb gezegd.



## Referenties

- <sup>1</sup> Sackett DL, Rosenberg WM. On the need for evidence-based medicine. *J Public Health Med.* 1995;17(3):330-334.
- <sup>2</sup> An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO investigators. *N Engl J Med.* 1993;329(10):673-682.
- <sup>3</sup> Lee KL, Woodlief LH, Topol EJ, Weaver WD, Betriu A, Col J, Simoons M, Aylward P, Van de Werf F, Califf RM. Predictors of 30-day mortality in the era of reperfusion for acute myocardial infarction. Results from an international trial of 41,021 patients. GUSTO-I Investigators. *Circulation.* 1995;91(6):1659-1668.
- <sup>4</sup> Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst.* 1988;80(15):1198-1202.
- <sup>5</sup> Califf RM, Woodlief LH, Harrell FE, Jr., Lee KL, White HD, Guerci A, Barbash GI, Simes RJ, Weaver WD, Simoons ML, Topol EJ. Selection of thrombolytic therapy for individual patients: development of a clinical model. GUSTO-I Investigators. *Am Heart J.* 1997;133(6):630-639.
- <sup>6</sup> Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis.* New York: Springer; 2001.
- <sup>7</sup> Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ.* 1995;311(7016):1356-1359.
- <sup>8</sup> Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA.* 2007;298(10):1209-1212.
- <sup>9</sup> Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation.* 1998;97(18):1837-1847.
- <sup>10</sup> Giard RW. [Top 100 hospitals: varying lists, varying reputations]. *Nederlands tijdschrift voor geneeskunde.* 2006;150(43):2355-2358.
- <sup>11</sup> Katz MH. *Multivariable analysis: a practical guide for clinicians.* 2nd ed. Cambridge: Cambridge University Press; 2006.
- <sup>12</sup> Hastie T, Tibshirani R, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction.* New York: Springer; 2001.
- <sup>13</sup> Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making.* 2001;21(1):45-56.
- <sup>14</sup> Steyerberg EW, Eijkemans MJ, Van Houwelingen JC, Lee KL, Habbema JD. Prognostic models based on literature and individual patient data in logistic regression analysis. *Stat Med.* 2000;19(2):141-160.
- <sup>15</sup> Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med.* 2004;23(16):2567-2586.
- <sup>16</sup> Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Ann Intern Med.* 2006;145(1):21-29.
- <sup>17</sup> Pencina MJ, D'Agostino RB S, D'Agostino RB J, Vasan RS. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med.* 2008;27(2):157-72.
- <sup>18</sup> Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26(6):565-574.
- <sup>19</sup> Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999;130(6):515-524.
- <sup>20</sup> Steyerberg EW, Roobol MJ, Kattan MW, van der Kwast TH, de Koning HJ, Schroder FH. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. *J Urol.* 2007;177(1):107-112.
- <sup>21</sup> Boersma E, Steyerberg EW, Van der Vlugt MJ, Simoons ML. Reperfusion therapy for acute myocardial infarction. Which strategy for which patient? *Drugs.* 1998;56(1):31-48.
- <sup>22</sup> van der Graaf Y, de Waard F, van Herwerden LA, Defauw J. Risk of strut fracture of Bjork-Shiley valves. *Lancet.* 1992;339(8788):257-261.
- <sup>23</sup> van der Meulen JH, Steyerberg EW, van der Graaf Y, van Herwerden LA, Verbaan CJ, Defauw JJ, Habbema JD. Age thresholds for prophylactic replacement of Bjork-Shiley convexo-concave heart valves. A clinical and economic evaluation. *Circulation.* 1993;88(1):156-164.
- <sup>24</sup> van Gorp MJ, Steyerberg EW, Van der Graaf Y. Decision guidelines for prophylactic replacement of Bjork-Shiley convexo-concave heart valves: impact on clinical practice. *Circulation.* 2004;109(17):2092-2096.
- <sup>25</sup> Steyerberg EW, Kallewaard M, van der Graaf Y, van Herwerden LA, Habbema JD. Decision analyses for prophylactic replacement of the Bjork-Shiley convexo-concave heart valve: an evaluation of assumptions and estimates. *Med Decis Making.* 2000;20(1):20-32.
- <sup>26</sup> Steyerberg EW, Keizer HJ, Zwartendijk J, Van Rijk GL, Van Groenigen CJ, Habbema JD, Stoter G. Prognosis after resection of residual masses following chemotherapy for metastatic nonseminomatous testicular cancer: a multivariate analysis. *Br J Cancer.* 1993;68(1):195-200.

- <sup>27</sup> 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 nonseminomatous germ cell tumor: multivariate analysis of individual patient data from six study groups. *J Clin Oncol.* 1995;13(5):1177-1187.
- <sup>28</sup> Steyerberg EW, Gerl A, Fossa SD, Sleijfer DT, de Wit R, Kirkels WJ, Schmeller N, Clemm C, Habbema JD, Keizer HJ. Validity of predictions of residual retroperitoneal mass histology in nonseminomatous testicular cancer. *J Clin Oncol.* 1998;16(1):269-274.
- <sup>29</sup> Steyerberg EW, Keizer HJ, Sleijfer DT, Fossa SD, Bajorin DF, Gerl A, de Wit R, Kirkels WJ, Koops HS, Habbema JD. Retroperitoneal metastases in testicular cancer: role of CT measurements of residual masses in decision making for resection after chemotherapy. *Radiology.* 2000;215(2):437-444.
- <sup>30</sup> Vergouwe Y, Steyerberg EW, Foster RS, Sleijfer DT, Fossa SD, Gerl A, de Wit R, Roberts JT, Habbema JD. Predicting retroperitoneal histology in postchemotherapy testicular germ cell cancer: a model update and multicentre validation with more than 1000 patients. *Eur Urol.* 2007;51(2):424-432.
- <sup>31</sup> Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ.* 2003;327(7417):741-744.
- <sup>32</sup> Ioannidis JP, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am J Epidemiol.* 2006;164(7):609-614.
- <sup>33</sup> Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med.* 1984;3(2):143-152.
- <sup>34</sup> Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. *Stat Med.* 1990;9(11):1303-1325.



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

ISBN 97-8907790-64-53

