Leiden University
Medical Center

Hacking Women's Stroke

Gebruik van big data voor vroegherkenning van verhoogd risico op beroerte bij vrouwen onder de 50 jaar

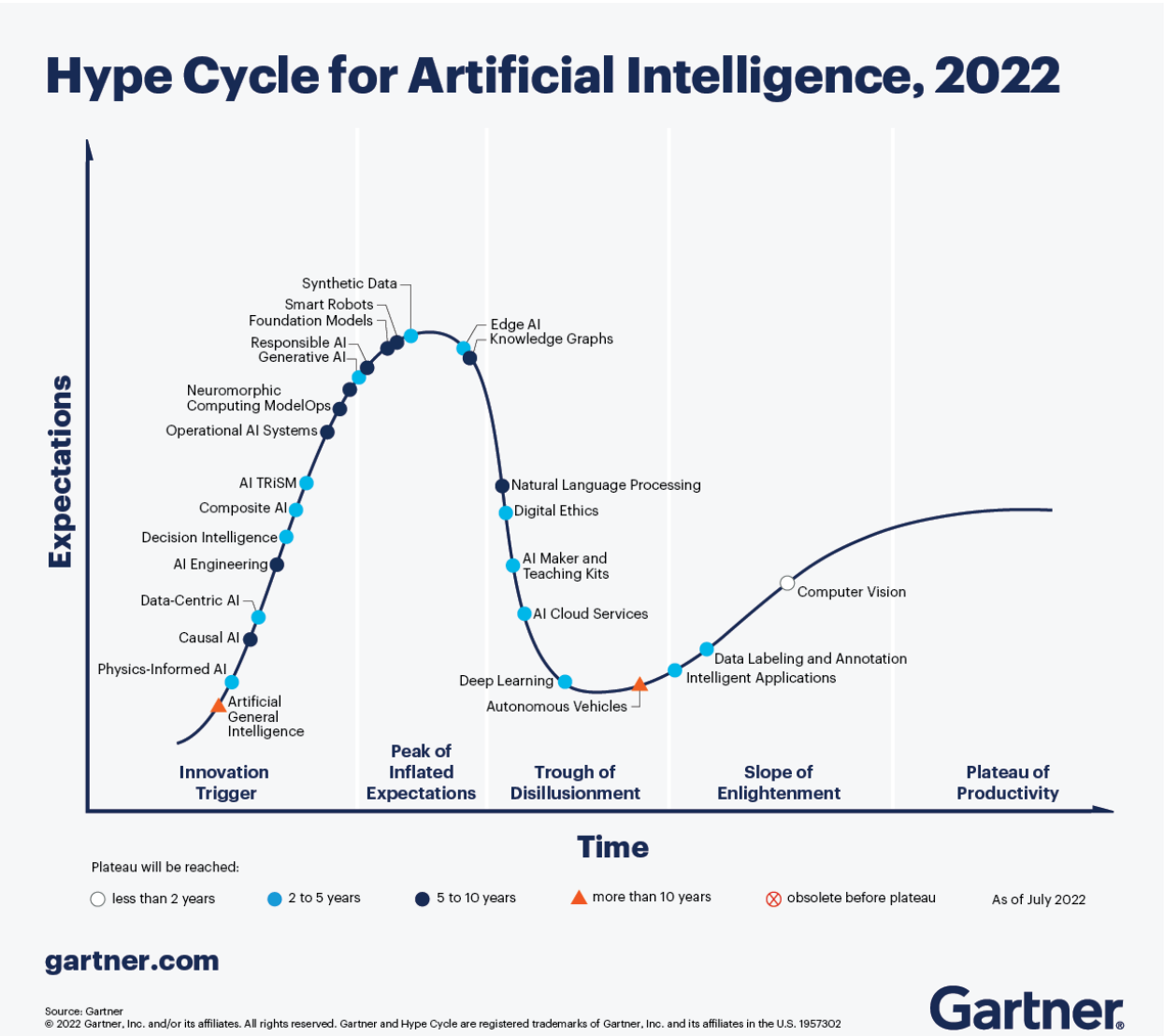
Hine van Os

General manager National eHealth Living Lab

Senior onderzoeker Public Health & Eerstelijns Geneeskunde en Neurologie

- Vroegherkenning beroerte risico bij vrouwen = voorbeeld
- (On)bruikbaarheid van ‘big data’ voor verbetering zorg
- (On)zin van geavanceerde analysemethoden (kunstmatige intelligentie) in de zorgpraktijk?

Kunstmatige intelligentie: hype or hope?



Klinisch probleem

Risico op beroerte bij vrouwen

- Vooral veel bekend over beroerte (en HVZ) bij mannen
- Beroerte bij vrouwen: prioriteit onderzoeksagenda Hartstichting
- Vrouw-specifieke risicofactoren te weinig onderkend





Wat zijn vrouw-specifieke risicofactoren?







….in context van andere risicofactoren

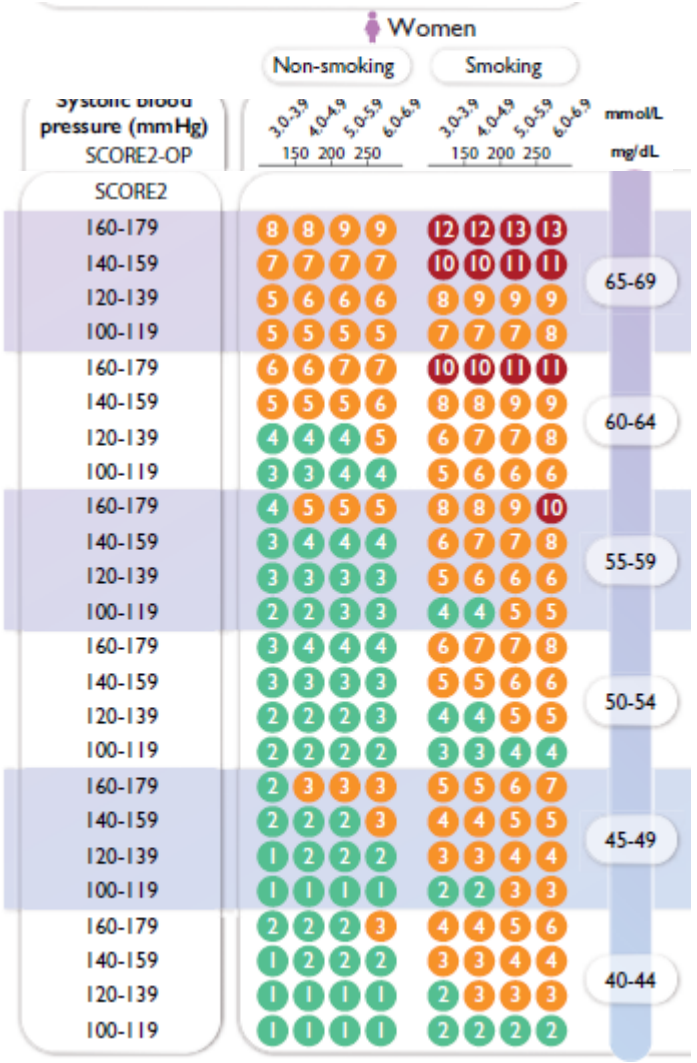


Klinische casus

- Vrouw
- Leeftijd: 48
- Migraine met aura
- Voorgeschiedenis van pre-eclampsie
- RR systolisch ~145 mmHg

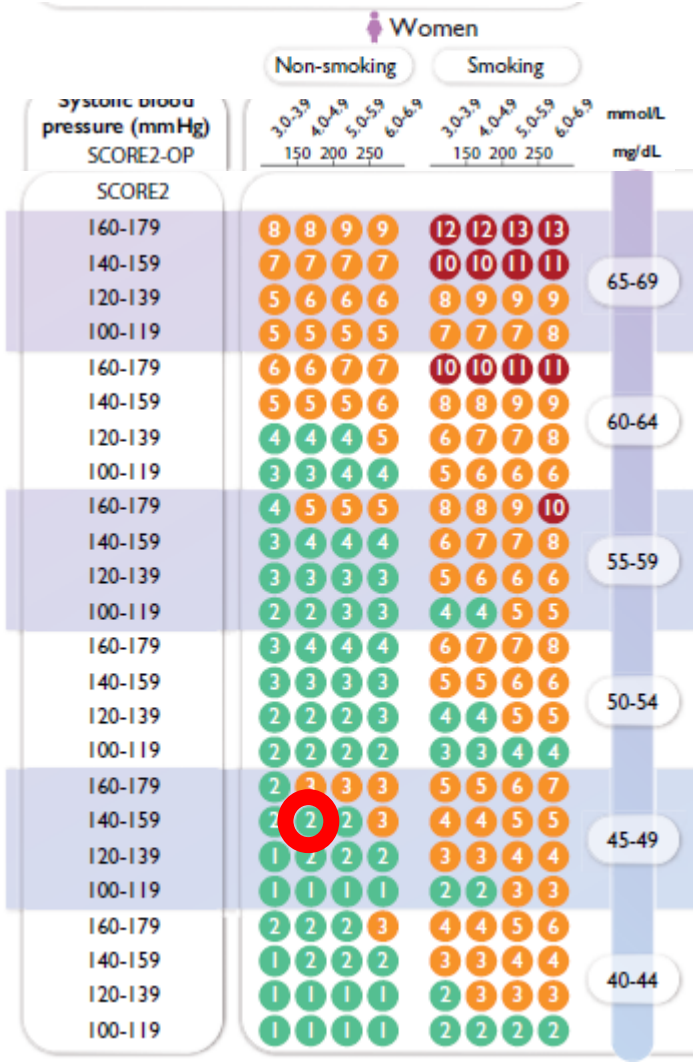
Huidige HVZ risicomodel: SCORE2

- 10-jaars risico op HVZ
- Traditionele risicofactoren
  - Leeftijd
  - Geslacht
  - Roken
  - Bloeddruk
  - Cholesterol
- Huisartspraktijk



Huidige risico model: SCORE2

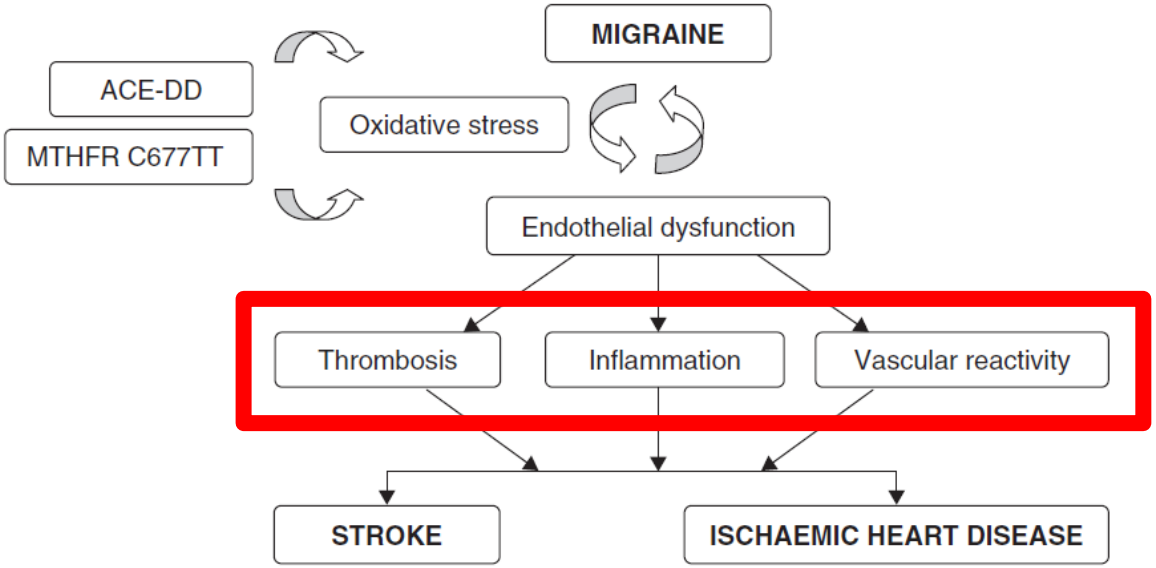
- Vrouw
- Leeftijd: 48
- Migraine met aura
- Voorgeschiedenis van pre-eclampsie
- RR systolisch ~145 mmHg



Vrouw-specifieke risicofactoren

Relatief risico voor beroerte (ischemisch)

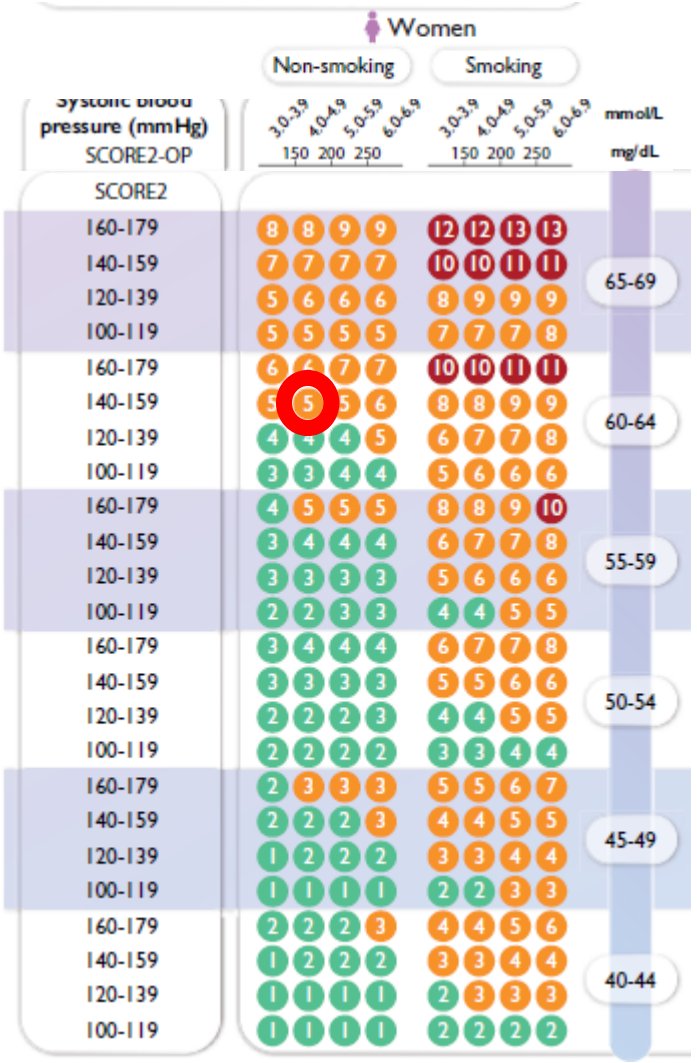
- Migraine met aura: ~ 2.0
- Pre-eclampsie: ~ 2.5
- Contraceptiva: ~ 1.5
- Contraceptiva + migraine: ~ 4 – 16
- Contraceptiva + migraine + roken: ~ 34



Huidige risico model: SCORE2

- Vrouw
- Leeftijd: 48
- Migraine met aura
- Voorgeschiedenis van pre-eclampsie
- RR systolisch ~145 mmHg

- Antihypertensiva?



Waarom is dit een belangrijk probleem?

- Pre-eclampsie: ~1 in 10 zwangerschappen
- Migraine met aura: ~6% van alle vrouwen

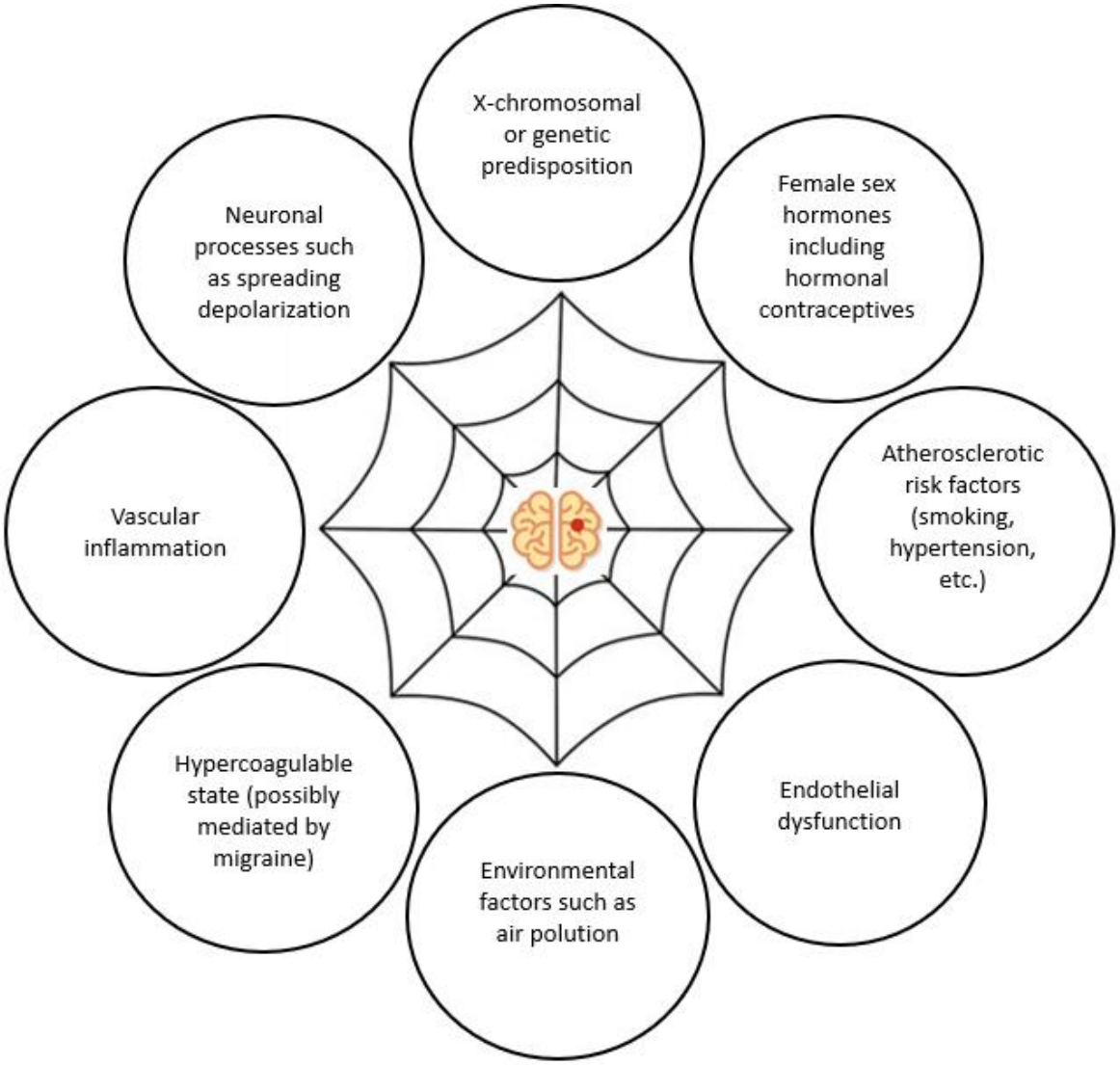
Oplossingsrichtingen?

- Alle vrouwen met pre-eclampsie of migraine uitnodigen voor screening?
- Tijdsinvestering voor zorgverleners?
- Bereidheid van zorgverzekeraar om te betalen?

Rationale van onze aanpak

- Screening met behulp van een predictiemodel voor cardiovasculair risico
  - Traditionele + non-traditionele risicofactoren
- Gebruik van informatie in het elektronisch medisch dossier van de eerstelijnszorg
  - Zeer grote (n ~ 3M) gekoppelde routinezorgdatabase
- Gebruik van flexibele analytische methode om voorspellers voor eindmodel te selecteren
  - Cox elastic net survival analyse

Risico op beroerte bij vrouwen: multifactorieel

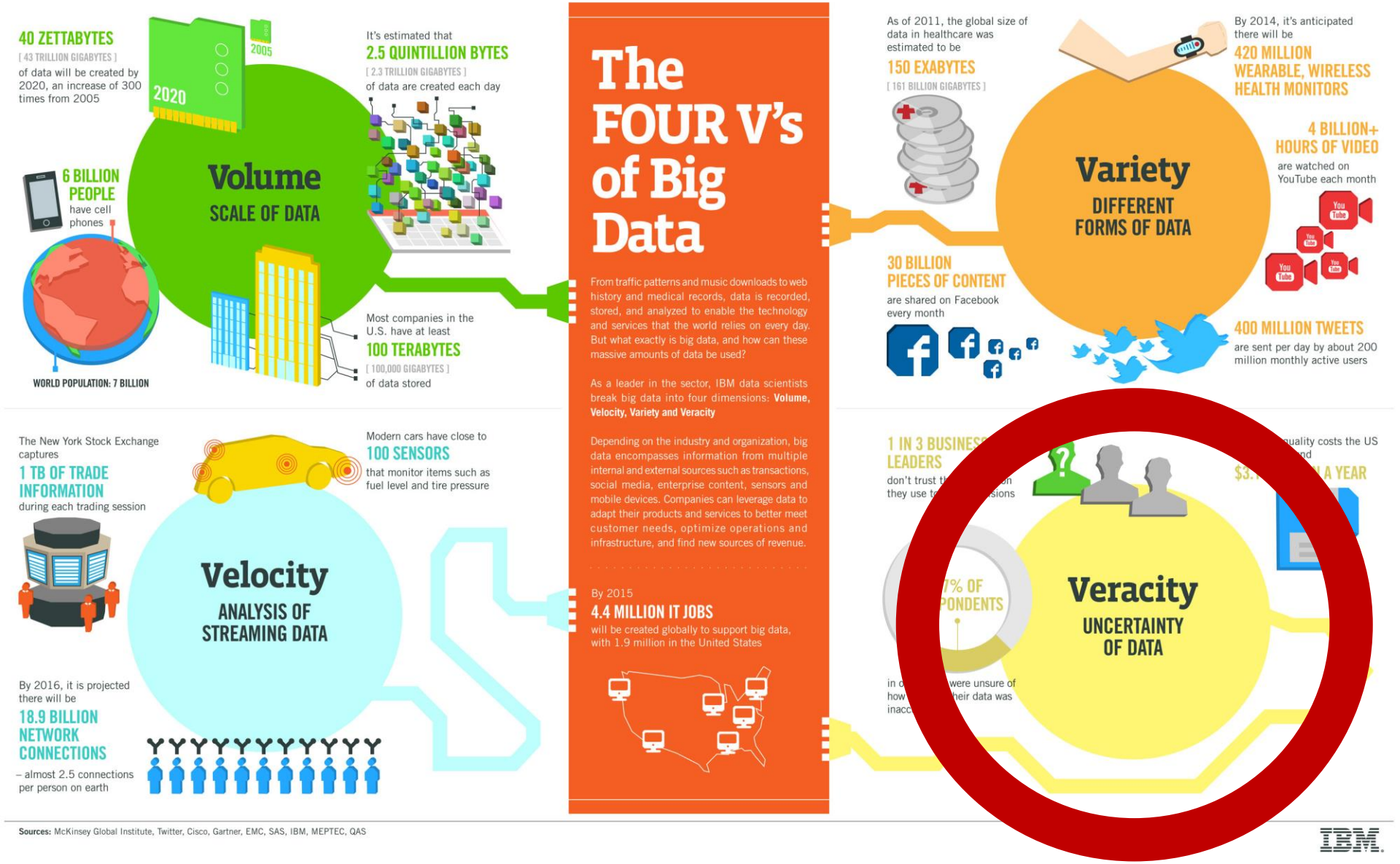


Wat hebben we nodig? Veel, en veel soorten data!

- N huisartspatiënten ~3 miljoen
- Demografische gegevens
  - Leeftijd & geslacht
  - 4-cijferige postcode
- ICPC-codes (‘huisartsdiagnose’ medische episode) ~250 miljoen
- ATC-codes: afhaalrecept apotheek ~500 miljoen
- Laboratorium: cholesterol, nierfunctie, etc. ~250 miljoen
- Bloeddruk: thuismeting, 24-uurs, praktijk, etc. ~250 miljoen
- Verrichtingen/verwijzingen
- BMI, informatie diëtiste, roken

Big data

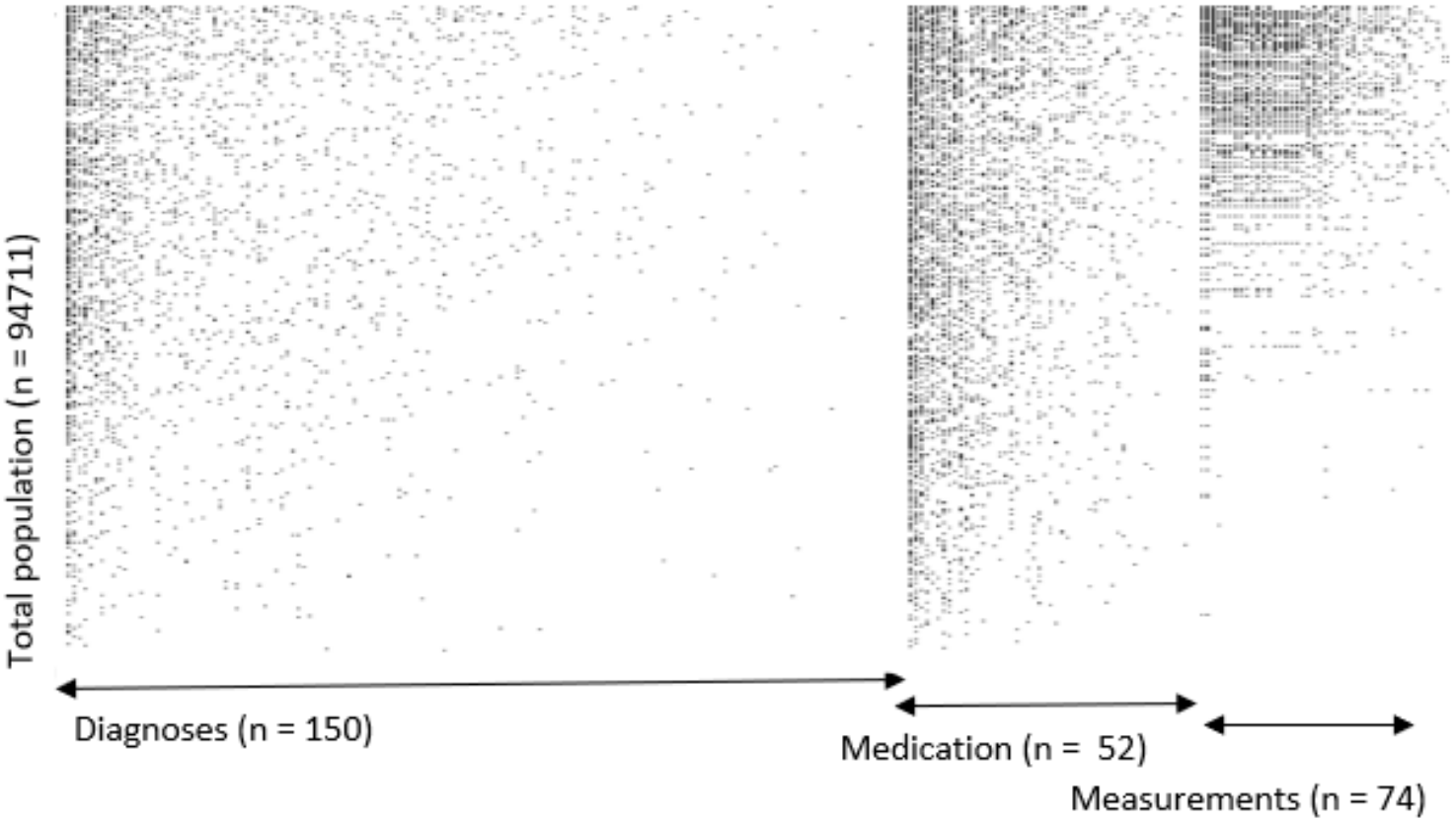
Is dit Big Data?



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

IBM

Hoe bruikbaar zijn onze EPDs?



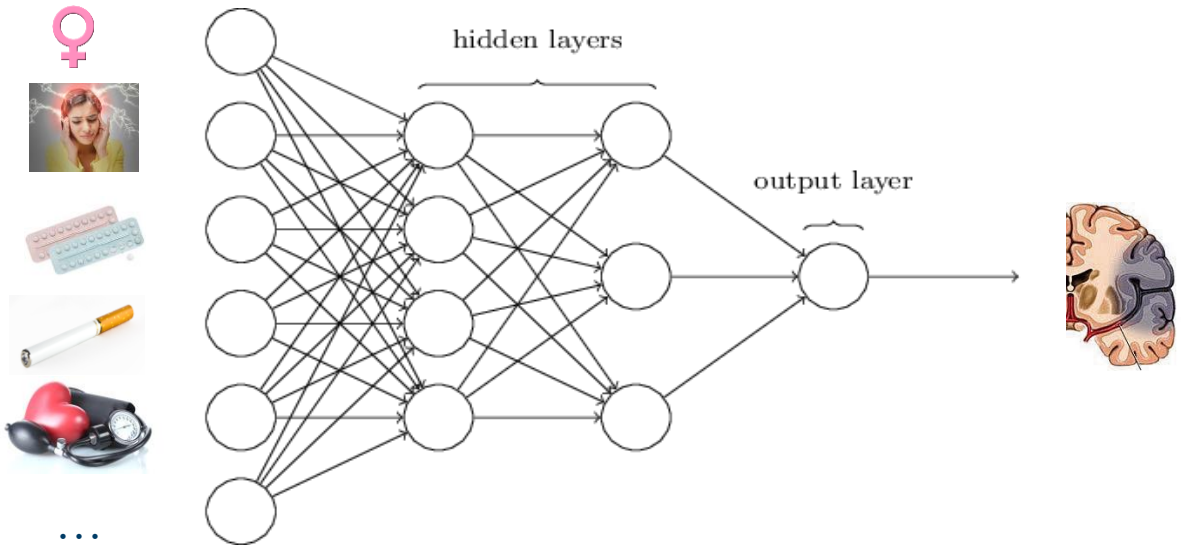
Methoden

- Laat de data spreken!
  - Niet vooraf bepalen
  - Nieuwe data omgeving: mogelijk ook nieuwe verbanden?
  - Verder kijken dan traditionele factoren (bloeddrukmetingen, cholesterol)?
- Optimaal gebruik maken van de rijkdom in patiëntendossiers
  - In totaal 500 potentiële factoren onderzocht, inclusief interacties


De oplossing: kunstmatige intelligentie
?

Waar hebben we het precies over?

- Kunstmatige intelligentie: ‘computersystemen die taken kunnen uitvoeren waar normaliter menselijke intelligentie voor nodig is’
- Machine learning: ‘het proces waarbij computers de manier waarop zij taken uitvoeren veranderen door te leren van nieuwe gegevens, zonder dat een mens instructies hoeft te geven in de vorm van een programma’



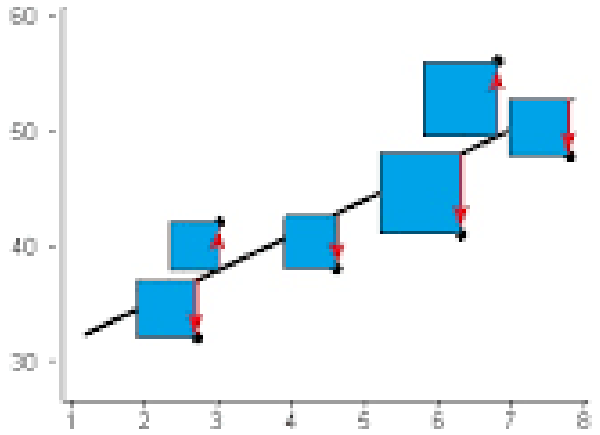
- Interacties automatisch meegenomen in veel machine learning modellen

Machine learning vs. ‘classical’ statistics

- Machine learning = statistisch leren

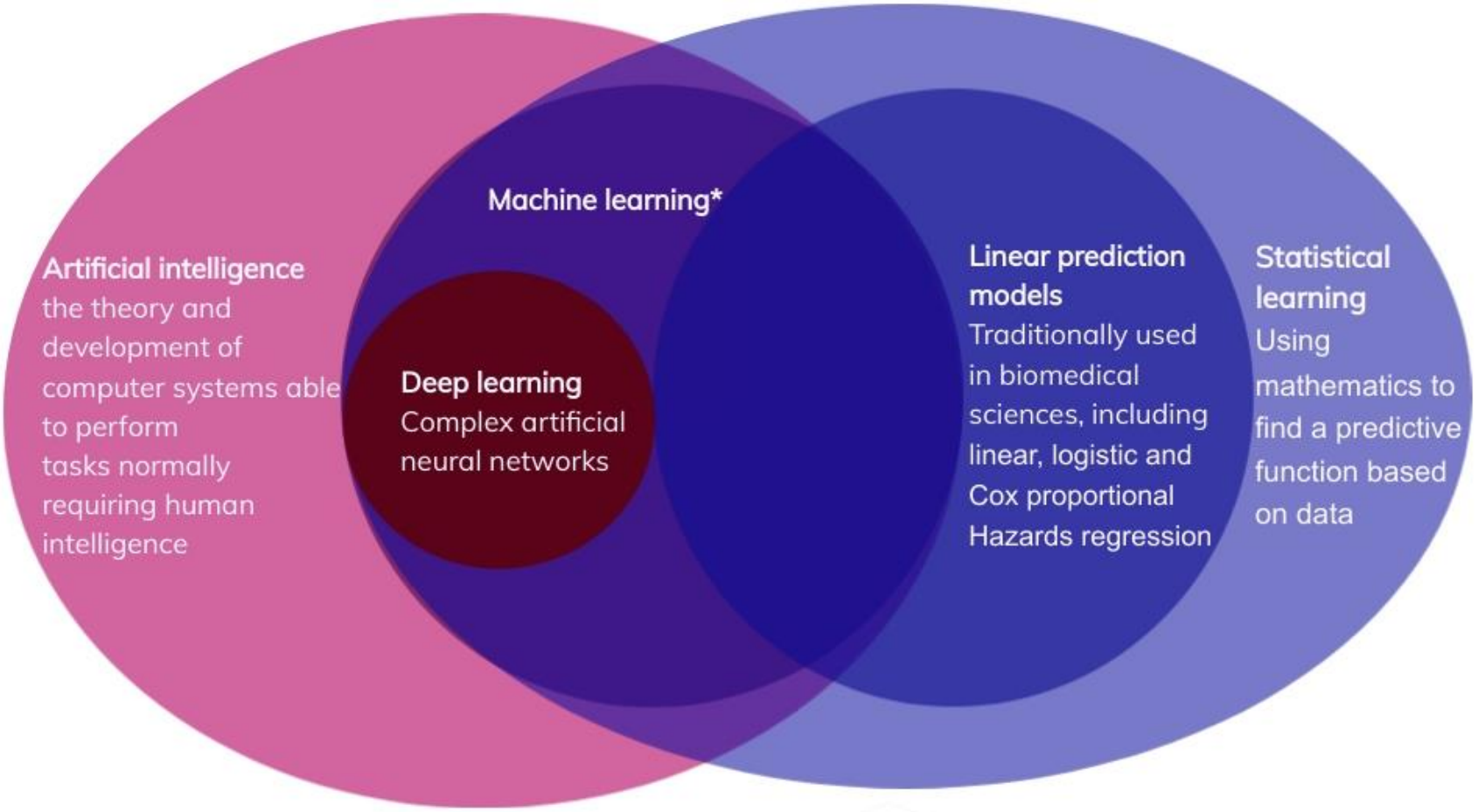
$$Y = f(X) + \epsilon.$$

- Least squares aanpak voor lineaire regressie (Legendre, 1805)



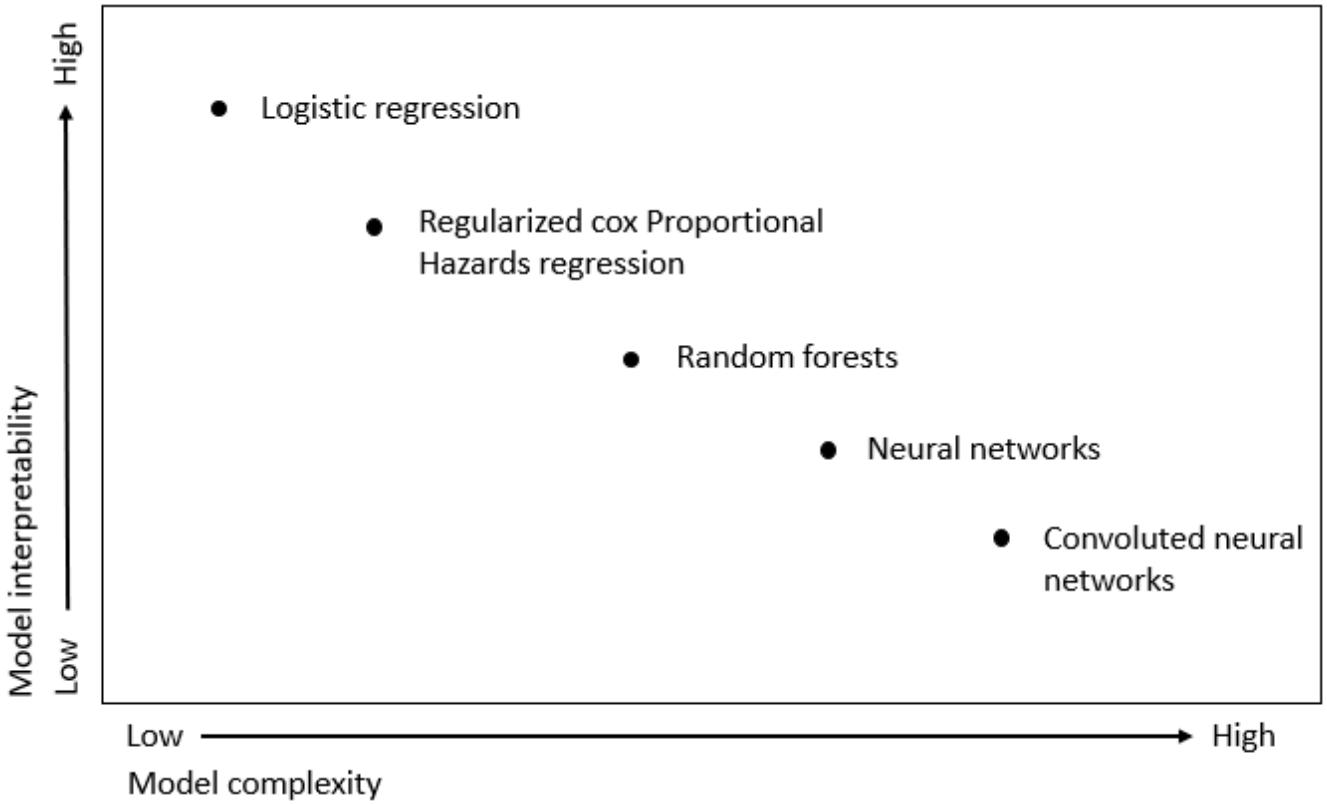


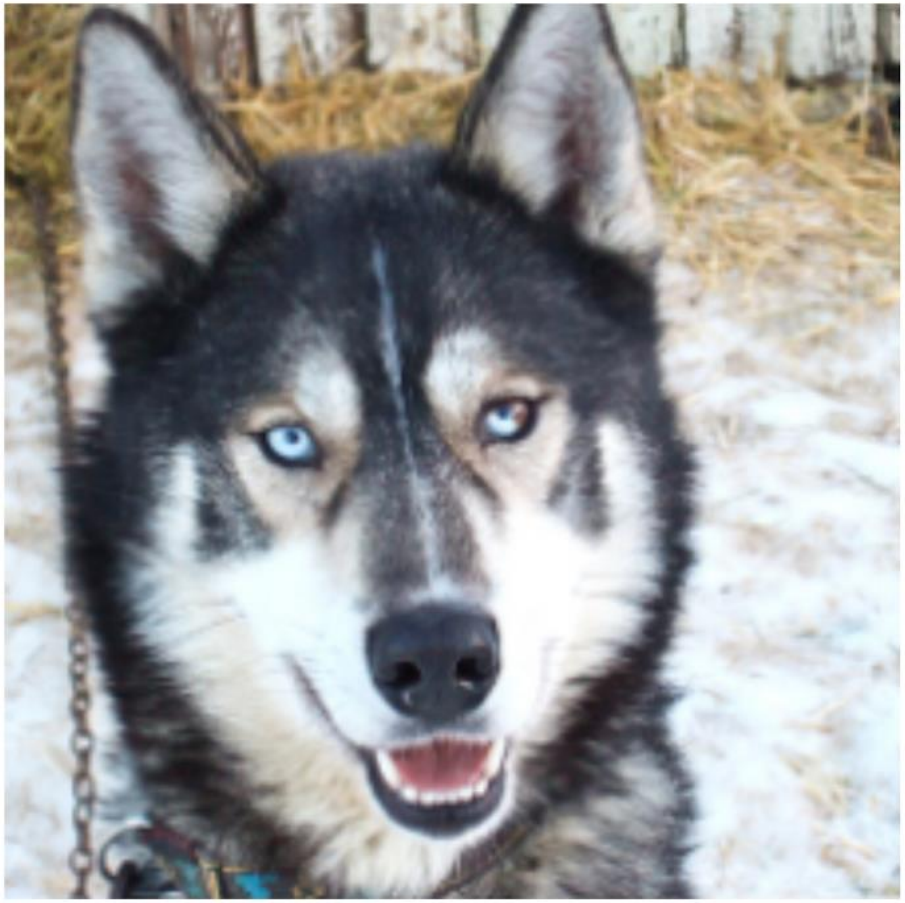
Overlap tussen machine learning & traditionele statistiek?



‘Spectrum van statischisch leren’

- Maak het zo moeilijk als noodzakelijk, en niet moeilijker!





(a) Husky classified as wolf



(b) Explanation

Hacking Women's Stroke studie

Onze onderzoekspopulatie

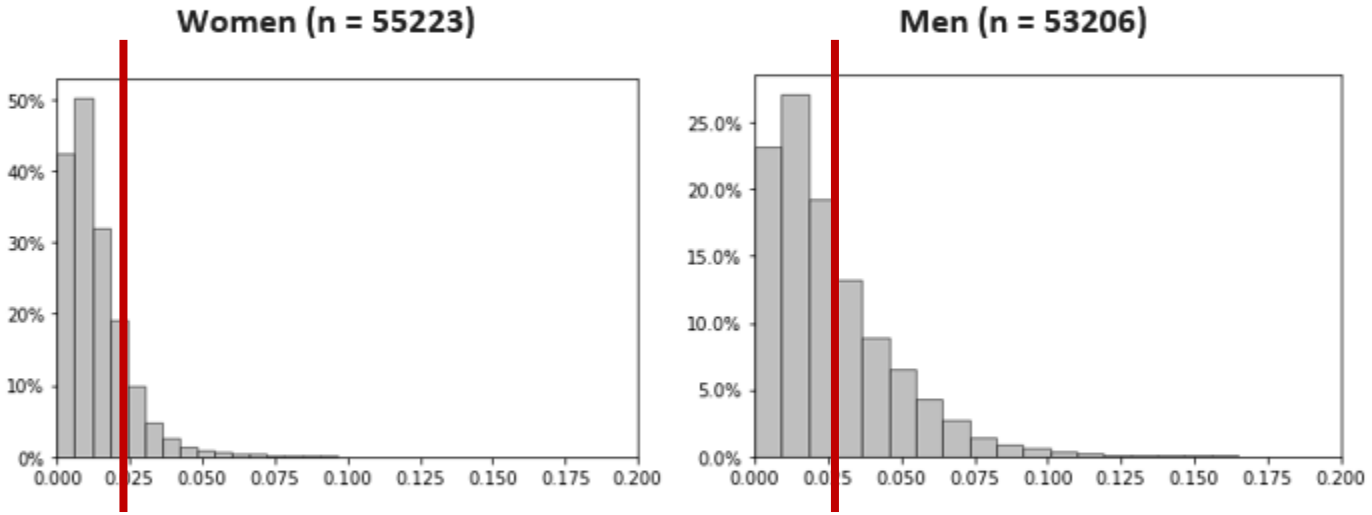
- 276116 vrouwen, 266013 mannen
- Gemiddelde leeftijd: 39.5 jaar
- 10715 myocardinfarcten en beroertes gedurende gemiddeld 7 jaar
- Bloeddrukmeting bij maar 8% van patiënten
- Bepaling cholesterol bij maar 5% patiënten

Model resultaten

- Acceptabele prestatie van modellen
- Complex versus minder complex model: 2% meer cases kunnen identificeren

Most important predictors (data-driven selection)

Figure. Absolute ten-year risk predictions of models including the 50 most important predictors, for women and men



On the X-axis predicted probabilities from prediction models including the 50 most important predictors. On the Y-axis the fraction of the population of each bin. All histograms have a bin size of 80.

Conclusie

Conclusie

- Data-gedreven selectie van predictors leidt tot acceptabele modellen voor predictie van cardiovasculaire ziekten in vrouwen onder 50 jaar
  - Complexere versus minder complexe modellen: kleine winst
  - Afweging modelprestatie versus -complexiteit: afhankelijk van setting

Hoe nu verder?

- Meerstappenplan
  - ‘Screenen’ met algoritme op populatieniveau
  - O.b.v. screening uitnodigen in de praktijk
    - Meten van factoren
    - Toepassen reeds gevalideerde ‘traditionele’ predictiemodellen
  - Gebruik informatie over verhoogd risico voor leefstijl counceling, medicatie zo nodig

- Vervolgstappen de komende twee jaar
  - Externe validatie
  - Voorbereiding grote implementatiestudie in ZH

Leiden University
Medical Center

Met dank aan:

LUMC
- Marieke J. H. Wermer
- Mattijs A. Numans
- Tobias N. Bonten
- Janet Kist
- Rolf H. H. Groenwold
- Niels H. Chavannes

VUMc Arificial intelligence
- Mark Hoogendoorn
- Ioannis Pantazis

Gezonde Zorg, Gezonde Regio
- Eric Hiddink
- Ron Herings

Leiden Institute for Advanced Computer Sciences
- Matthijs de Leeuw
- Anneloes Louwe
- Suzan Verberne
- Ted Etoeharnowo