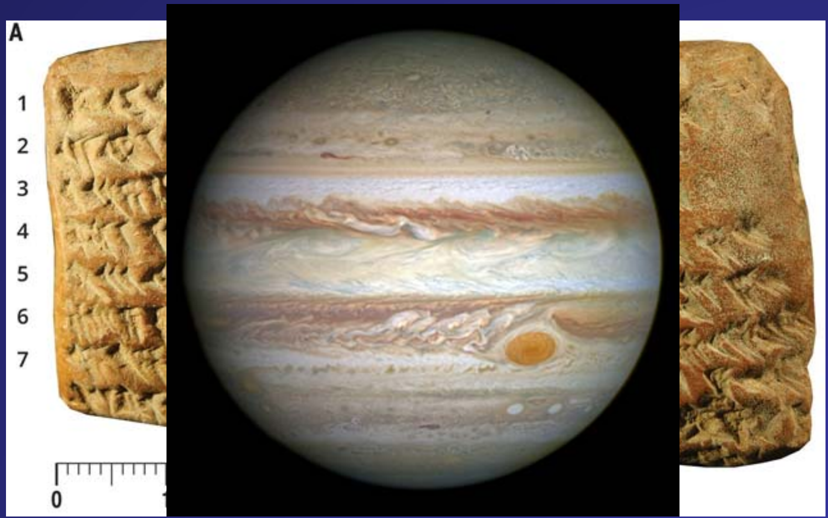
The end of Scientific Reasoning?

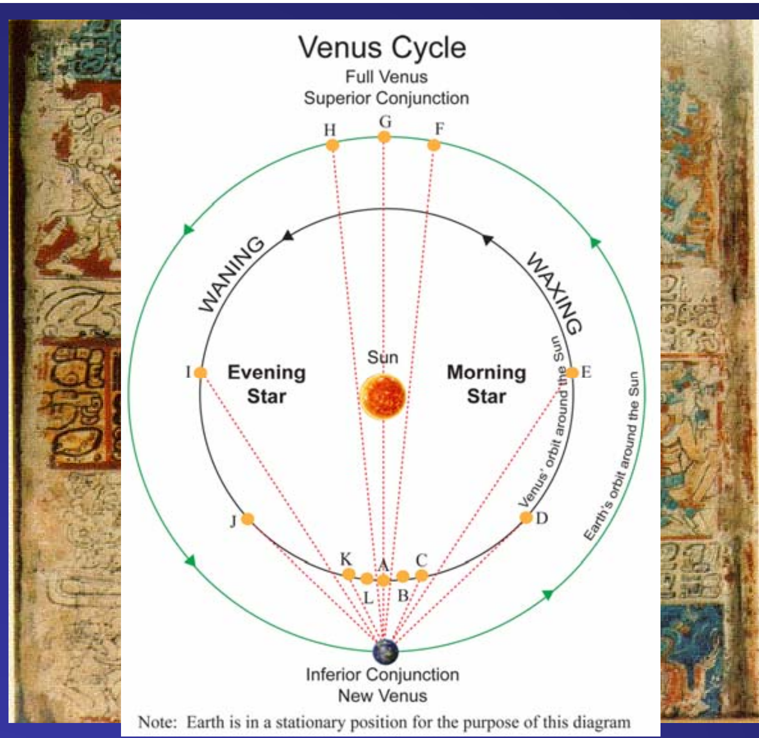
Peter M.A. Sloot

NANYANG TECHNOLOGICAL UNIVERSITY
Complexity Institute

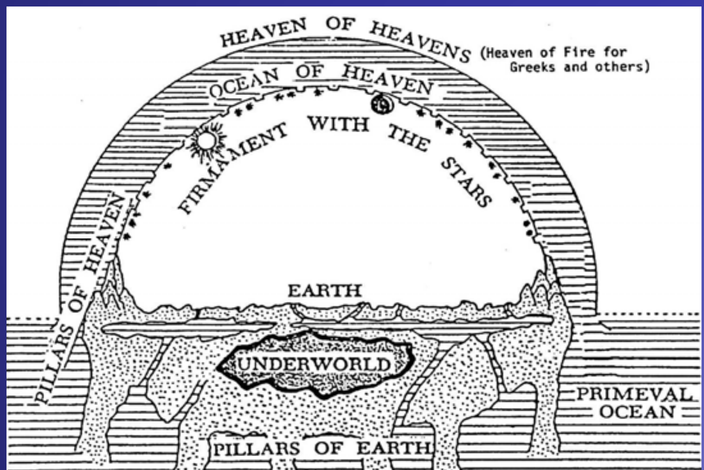
UNIVERSITY OF AMSTERDAM
Institute for Advanced Study

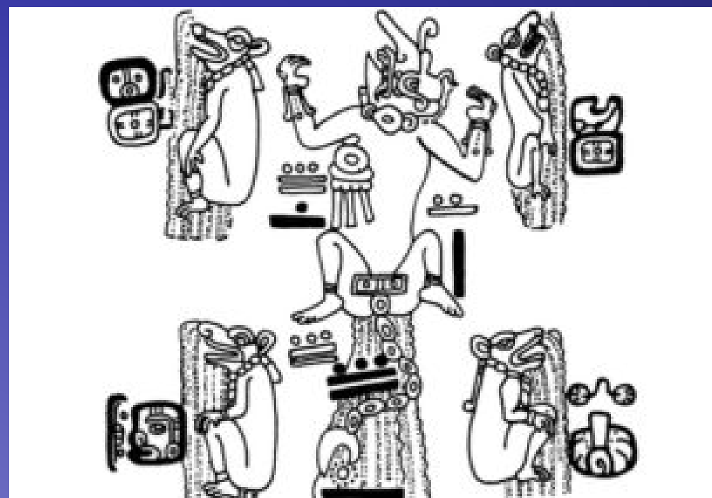


Babylonian Cuneiform tablet 350 BCE
Science: Vol. 351, Issue 6272 (2016) (Jupiter)



Mayan astronomy 500 BCE, Dresden Codex



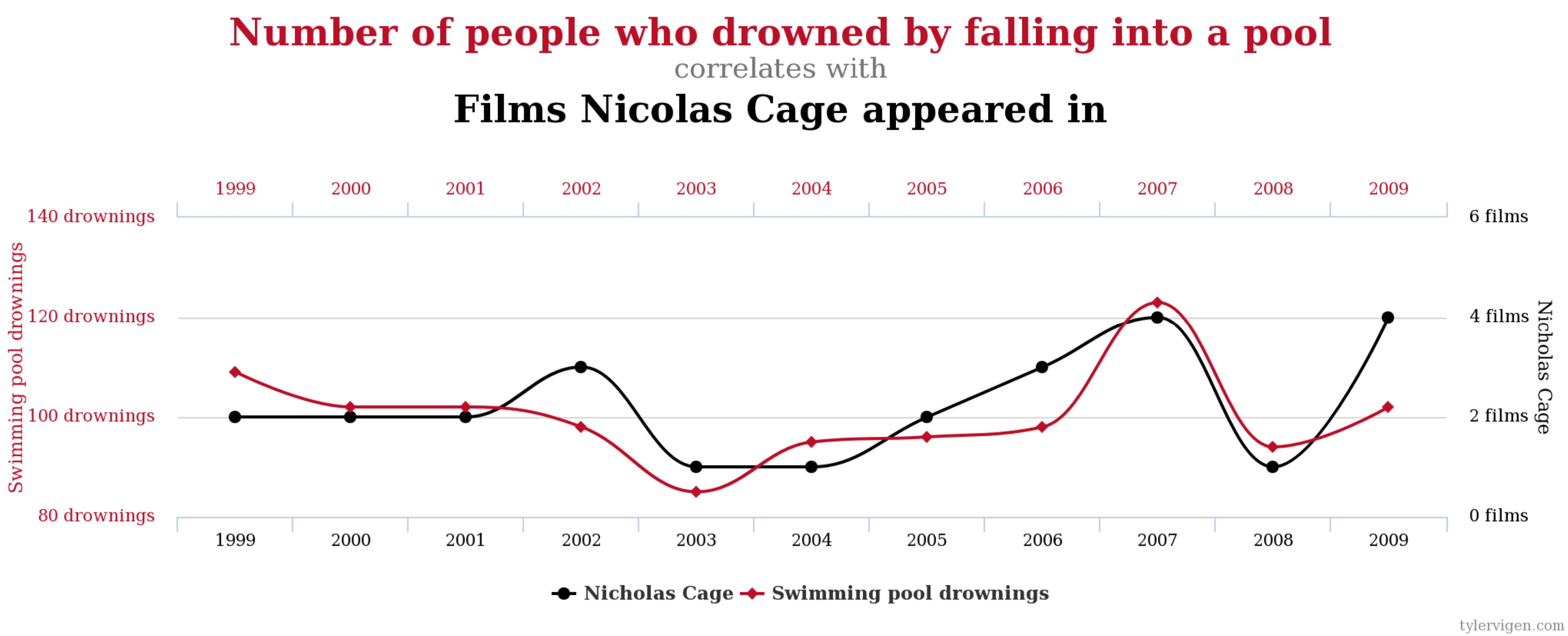


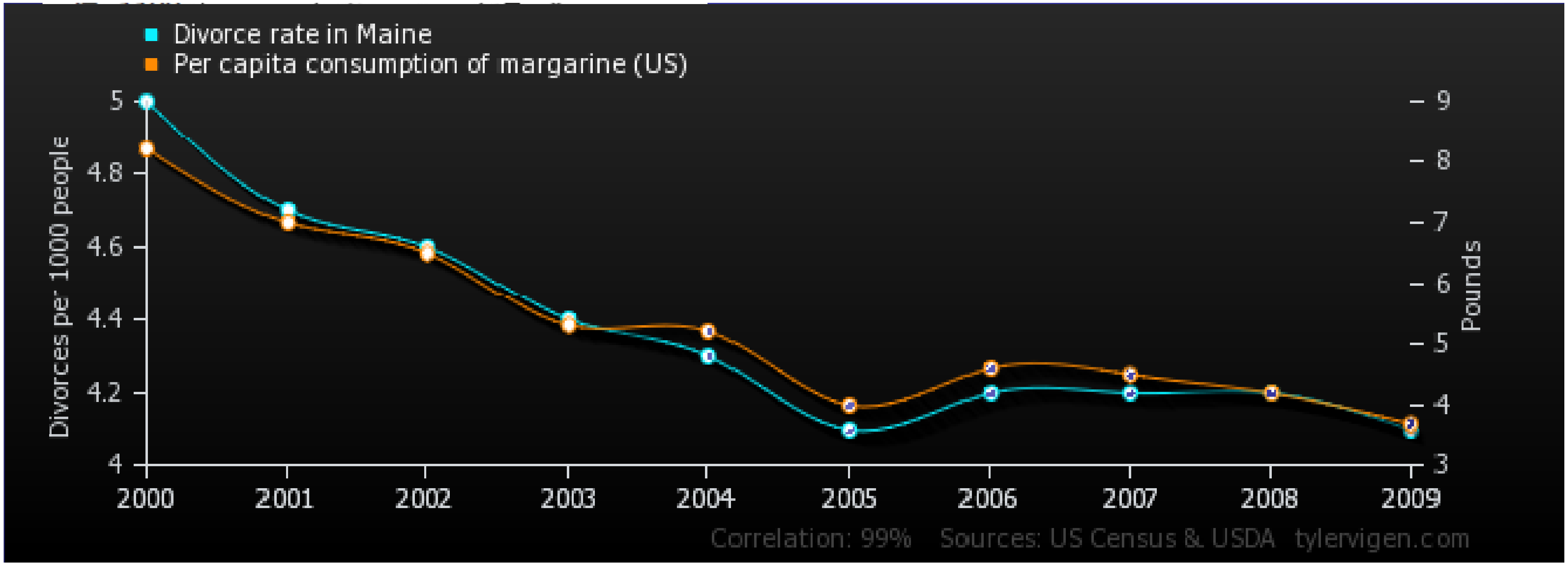
P. M. A. Sloot. '*More likely we would be ritually slaughtered*'. In: 43 Visions for Complexity. ISBN 9789813206847. pp. 65-66.

‘if we uncover the pattern in the data then we understand’

Hans Rosling


BREXIT
55%
Get Out
vs
45%
Stay In
Low
Moderate
High
Intense
Google
Flu Trends
Jan
Feb
Mar
Apr
May
Jun
Jul
VS


Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in
140 drownings
120 drownings
100 drownings
80 drownings
Swimming pool drownings
6 films
4 films
2 films
0 films
Nicholas Cage
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
Nicholas Cage
Swimming pool drownings
tylervigen.com


Divorce rate in Maine
Per capita consumption of margarine (US)
Divorces per 1000 people
5
4.8
4.6
4.4
4.2
4
Pounds
9
8
7
6
5
4
3
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
Correlation: 99%
Sources: US Census & USDA
tylervigen.com

The recent case of Asthma

Rich Caruana et al:

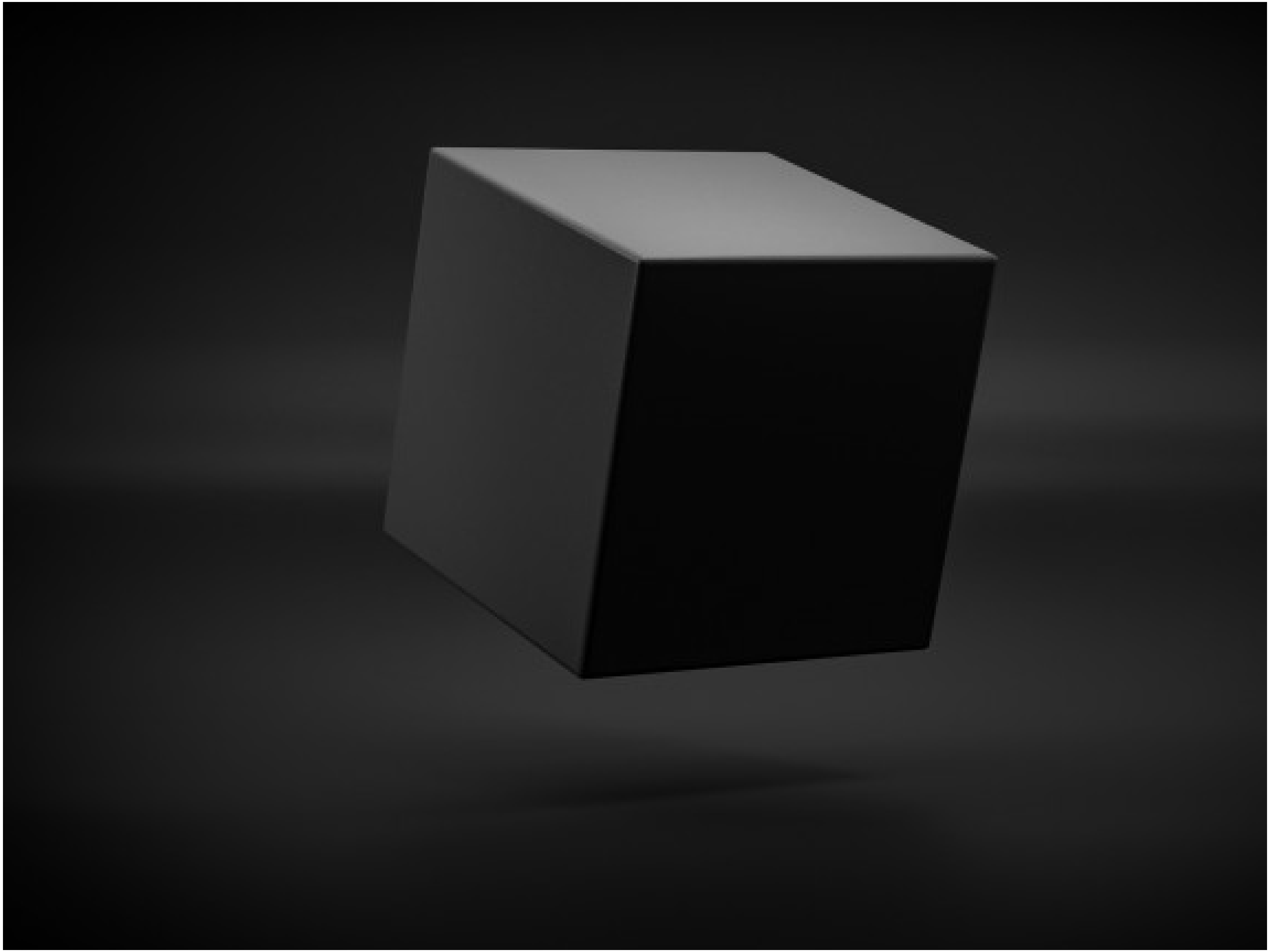
Applying Neural Nets and Rule Based Learning and Generalized Additive Models on 1000.000 Patients discovered the rule: ➔

'A History of Asthma Lowers a Patient's Chance of Dying From Pneumonia'

Hidden process: If asthmatic then much more vigilant and rapid early treatment for any pulmonary affection!

Imagine what would happen if the rule would have become a Hospital Regulation or GP's rule …

Rich Caruana et al., *Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission*. AAAS 2017



Even Worse: Most Published Research Findings Are False!

- Not only do we **interpret** the patterns in the data in the wrong way, and do we apply **black boxes** to distill meaning, we often even start out with ***wrong data****:*

- Our <u>most important dataset </u>– the conclusions of peer-reviewed studies – consists of predominantly bad (medical) data and cannot be relied upon!!

- As they miss evidence of good experimental design, first principle models and rigorous statistical analysis.

- Ioannidis JPA (2005) '*Why Most Published Research Findings Are False*'. PLoS Med 2(8)
- C. Glenn Begley and Lee M. Ellis, '*Drug development: Raise standards for preclinical cancer research*' Nature 483, 531–533 (2012)
- Peter C. Gøtzsche: '*Deadly Medicines and Organised Crime: How Big Pharma Has Corrupted Healthcare*'

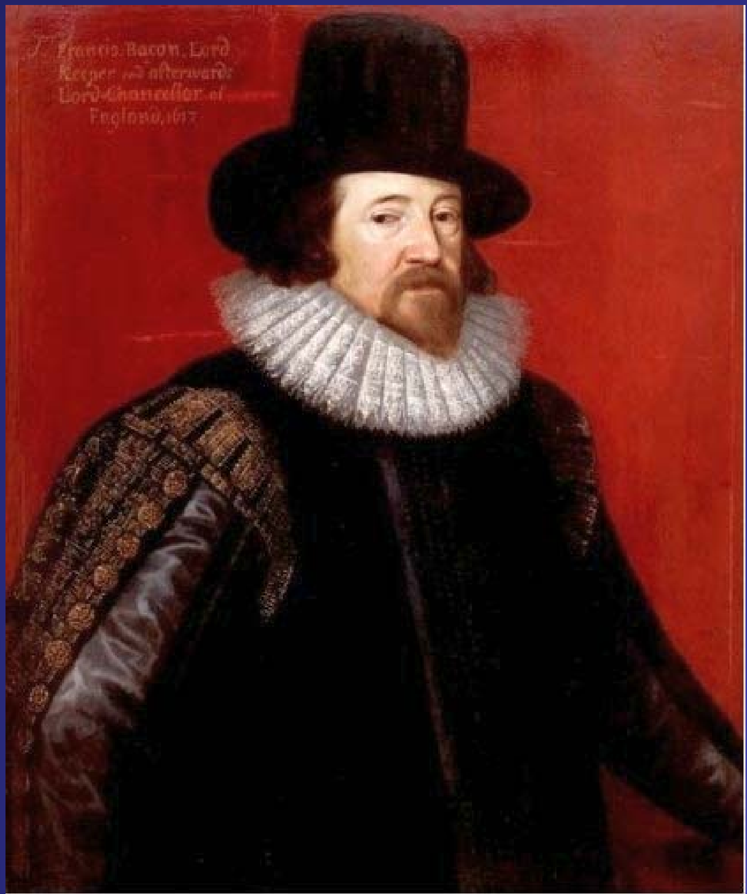
Data

‘The real purpose of the scientific method is to discover that ~~Nature~~ hasn’t misled you into thinking you know something you don’t actually know.’

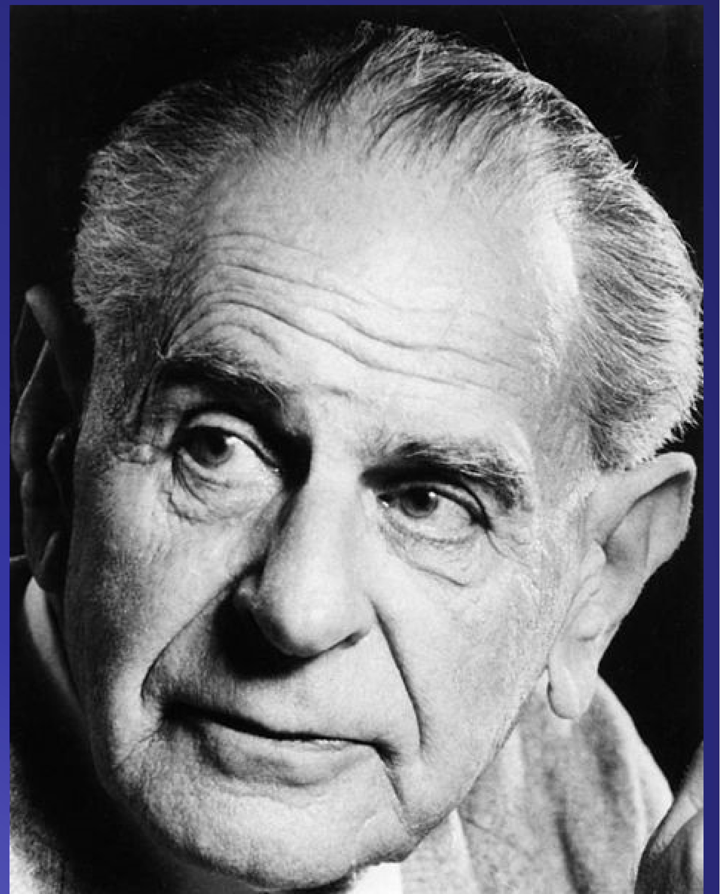
Robert M. Pirsig 1974

Zen and the Art of Motorcycle Maintenance

Clash of the Titans: Inductivism vs Deductivism



Sir Francis Bacon (1561 – 1626)



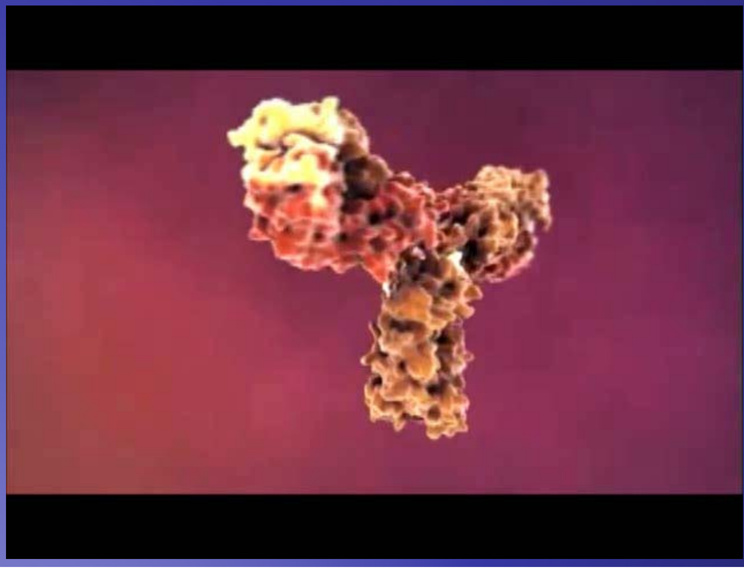
Sir Carl Popper (1902 - 1994)

P.M.A. Sloot, P. Coveney and J. Dongarra: Journal of Computational Science Vol. 1 (2010) 3–4;

Real World: Complex Systems

– Not simple ➔ not reducible to elements ➔ looses the aspects that make it complex

– Interacting elements that adapt to the environment they themselves create



Historical Perspective

– 17th, 18th, 19th Centuries: Newton, Faraday, Maxwell solved problems of 'simplicity' : Small number of variables

– 1800 onwards: Boltzmann, Gibbs: Problems of *Disorganized* Complexity with 10^{23} variables ➔ Microscopic from Microscopic

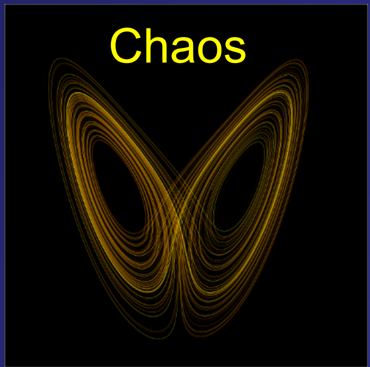
– 1948 Warren Weaver: Study *Organized* Complexity

– 1972 Philip Anderson: *The whole is more and also different from the sum of parts* (Science **177**, 4047, 1972)

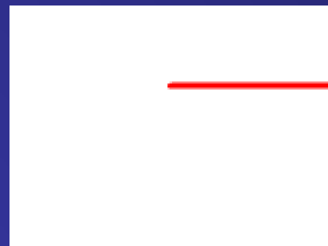
– 1984 Santa Fe Complexity Institute (Nobel Price Winners)

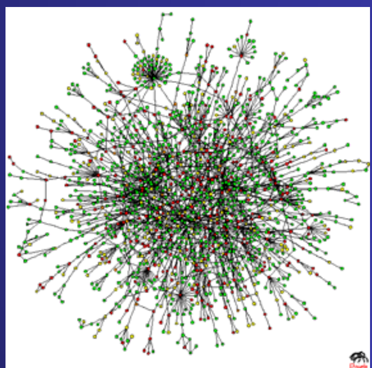
– 2000 Stephen Hawking, "*the 21st century will be the century of complexity*".

Difference Chaos and Complexity



In (deterministic) chaotic systems <u>uncertainty</u> arises from *practical inability* to know the initial conditions of a system





In complex systems <u>uncertainty</u> is *inherent* in the system because of emergence

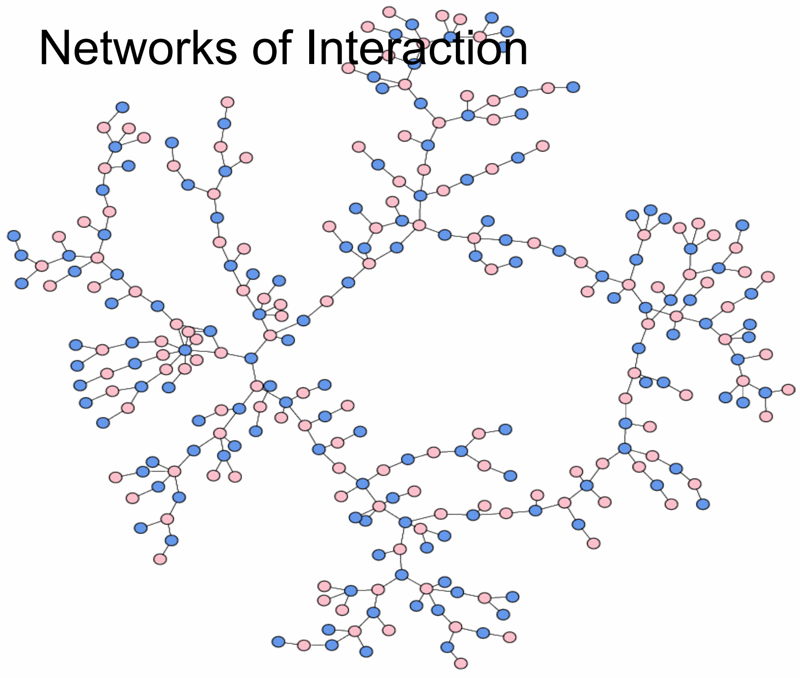
Example: complex financial world

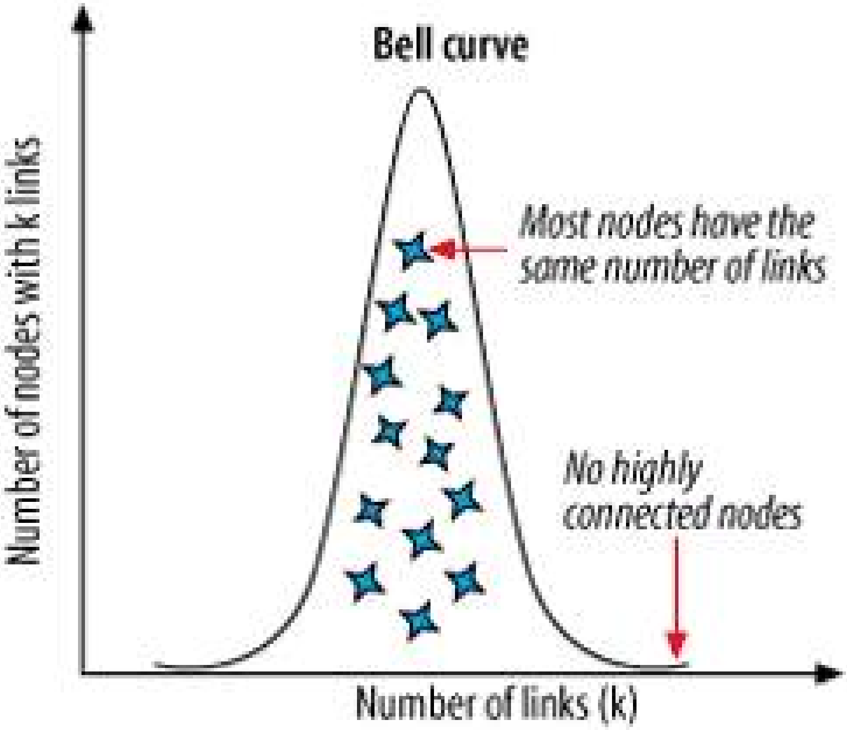
Dave Lauer: Quant.
Algorithms trade at speed of light
2010 Flash Crash: in minutes 862 Billion Dollars evaporated in Stock Markets

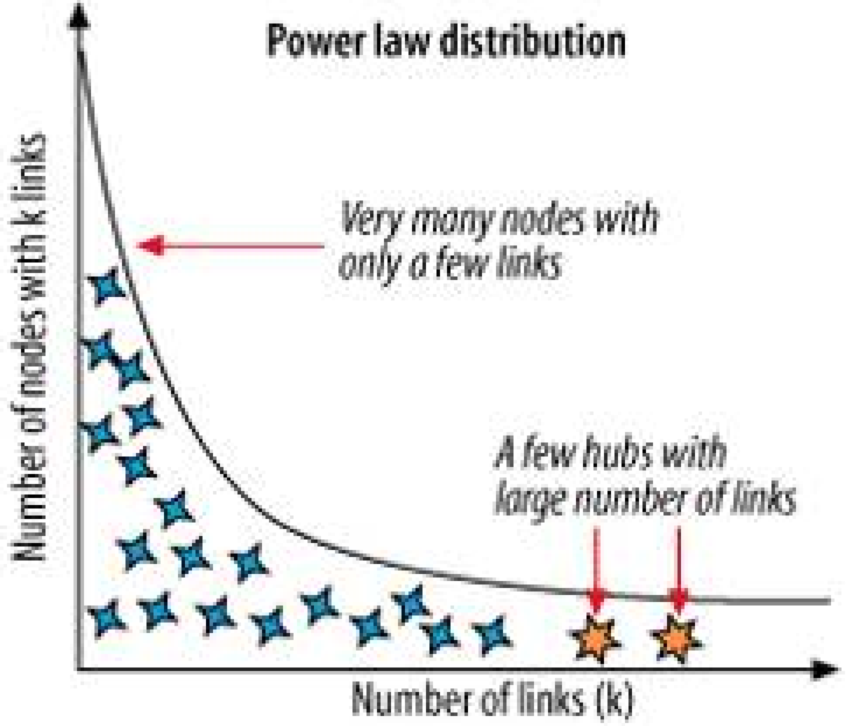
Flying Algorithms

R. Quax, D. Kandhai, and P. M. A. Sloot. *Information dissipation as an early-warning signal for the Lehman Brothers collapse in financial time series*. Scientific Reports 3, 1898. Nature Publishers

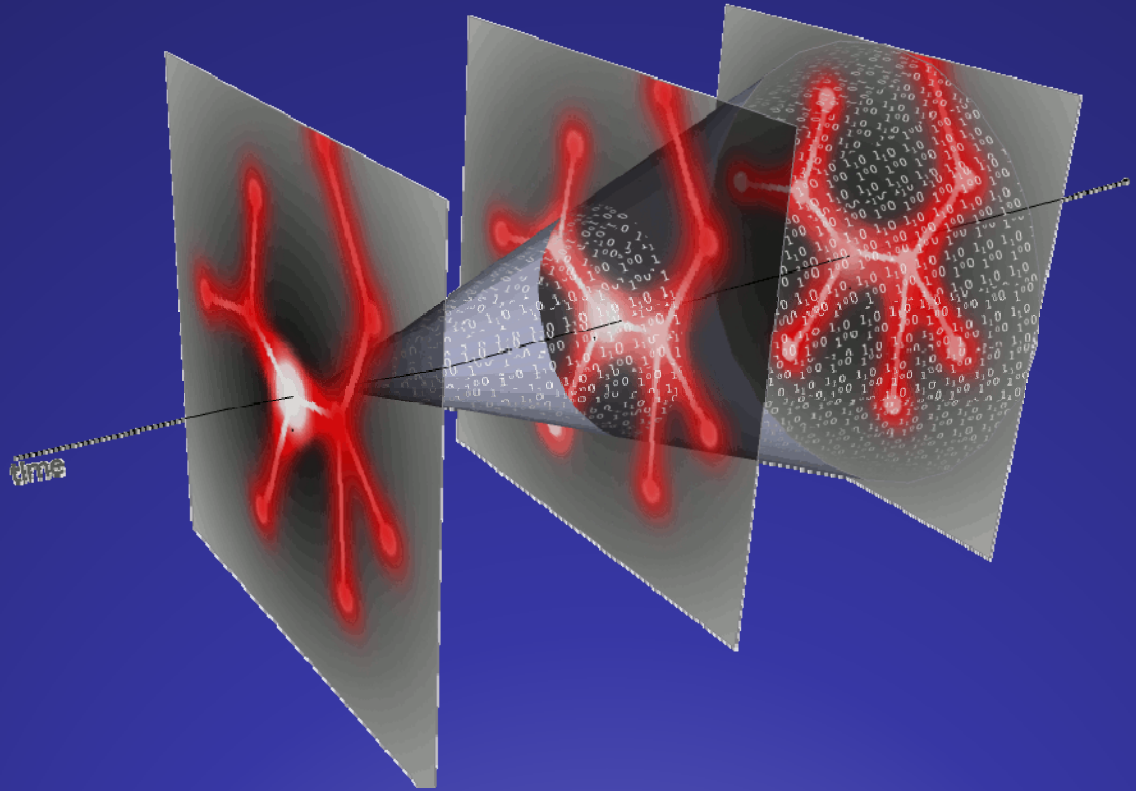
Networks of Interaction







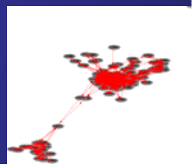
New Paradigm: Information processing in complex systems



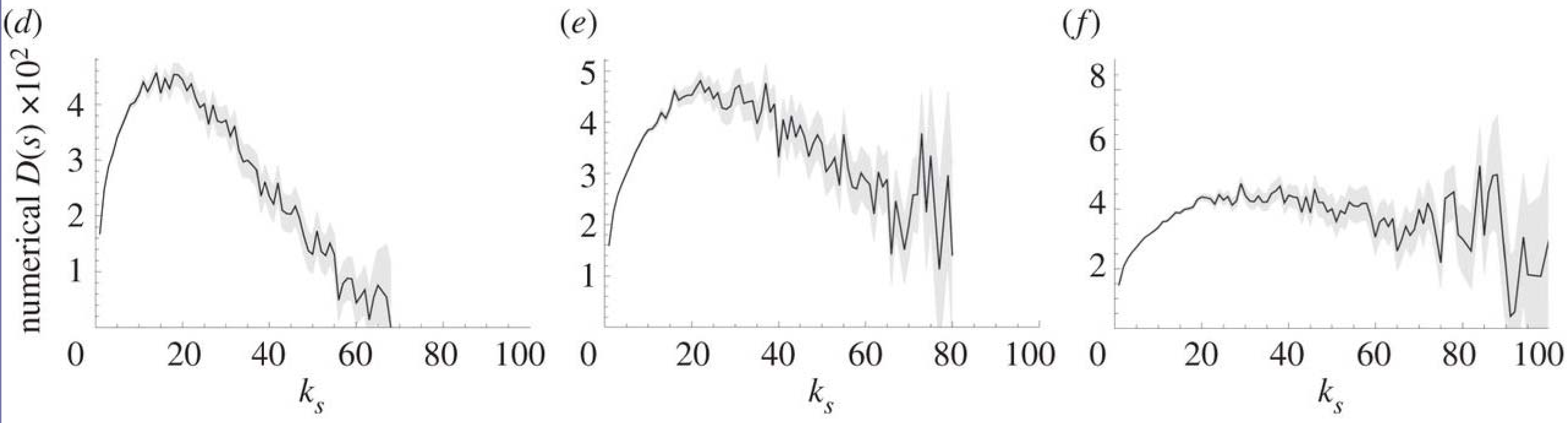
$$\mathcal{I}(\theta) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\Bigg|\,\theta\right] = \int \left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 f(x;\theta)\,\mathrm{d}x\,,$$

R. Quax and P.M.A. Sloot
‘understanding the behaviour of complex systems using information theory’

The European Physical Journal - Special Topics, vol. 222, nr 6 pp. 1389-1401. 2013.

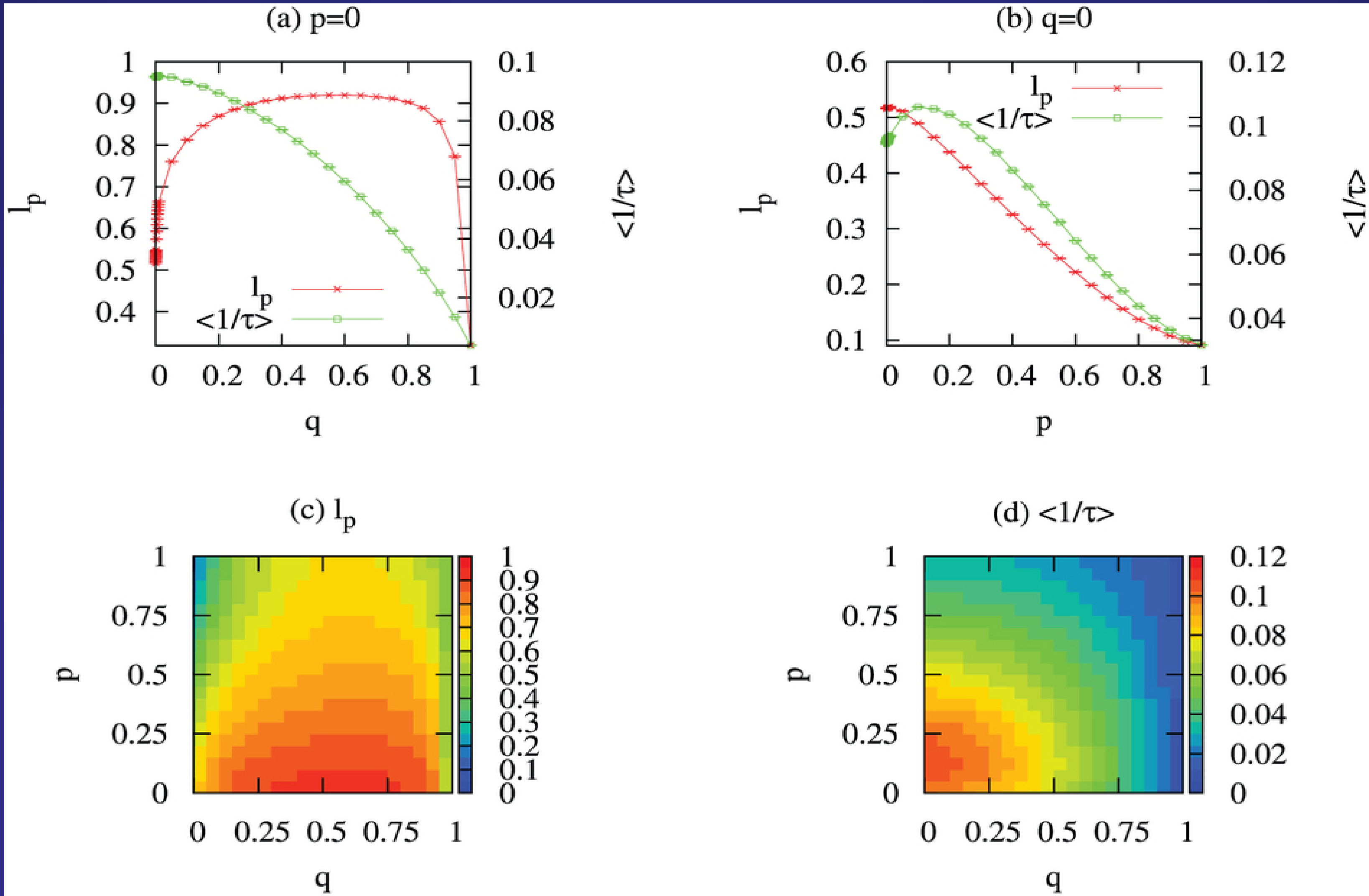


Not the ‘influentials’ but the ‘man in the street’ drives the change



Journal of the Royal Society Interface, vol. **10**, nr 88, 2013
Quax, Apolloni and Sloot

Stochastic Resonance



A. Czaplicka.; J.A. Hołyst and P.M.A. Sloot:
Noise enhances information transfer in hierarchical networks
Scientific Reports, vol. 3, 2013. Nature publishers

Complex Systems are Scale Free

$$\boldsymbol{p(k) = Ck^{-\gamma}} \qquad \int_{k_{min}}^{\infty} p(k)dk = 1 \qquad \boldsymbol{C} = (\gamma - 1).\, k_{min}^{\gamma-1}$$

$$\langle k^n \rangle = \int_{k_{min}}^{k_{max}} k^n p(k)dk = C\,\frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}$$

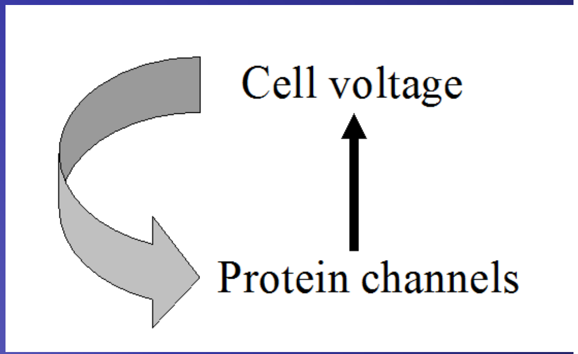
$$for\ k_{max} \to \infty \qquad \begin{cases} \langle k^n \rangle = \text{finite}\ \ if\ n \le \gamma - 1 \\ \langle k^n \rangle \to \infty\ if\ n > \gamma - 1 \end{cases}$$

$Most\ ScaleFree\ Networks\ have\ 2 < \gamma \le 3$

$\Rightarrow 2^{nd} and\ higher\ order\ moments\ diverge!$

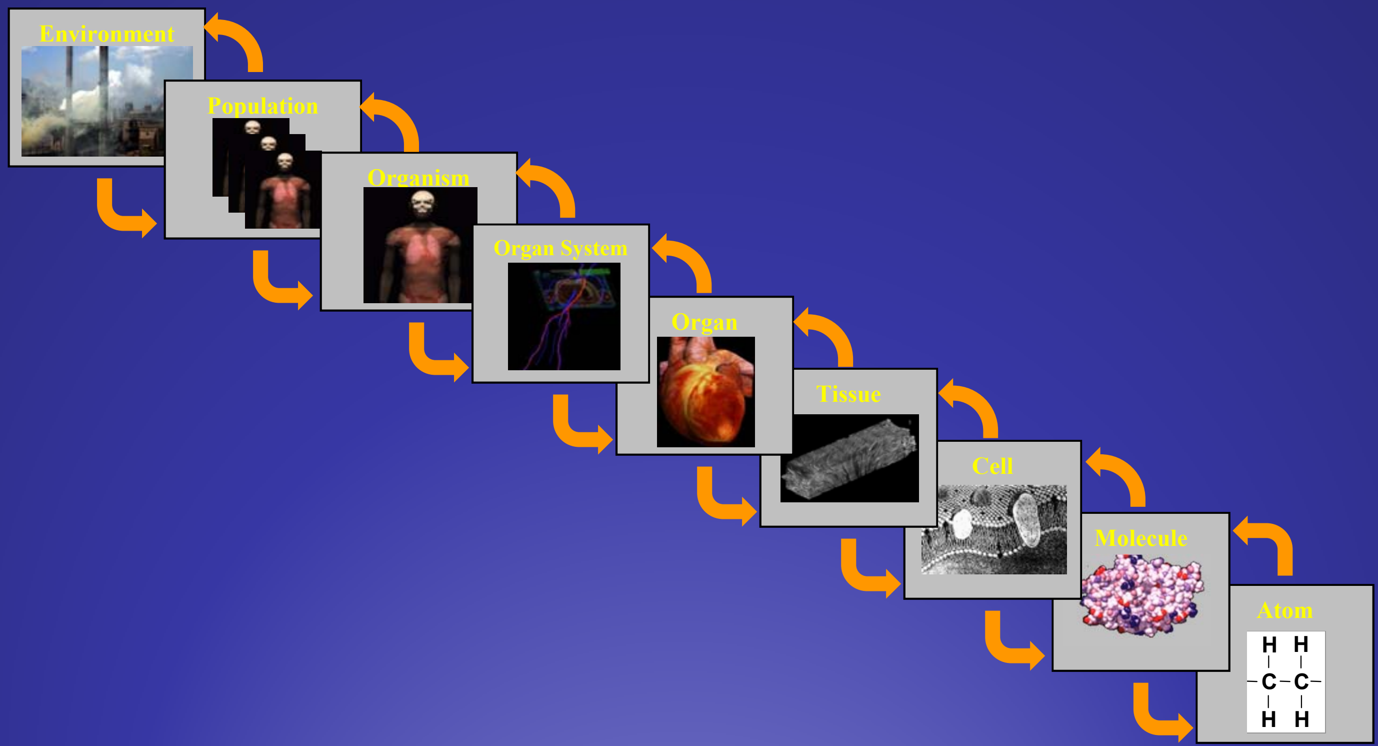
Consequence

- No average and/or infinite variance ➔
- More data ➔ More exceptions ➔ Learning algorithms won't converse
- In complex systems it often makes no sense to talk about cause and effect, they are intertwined

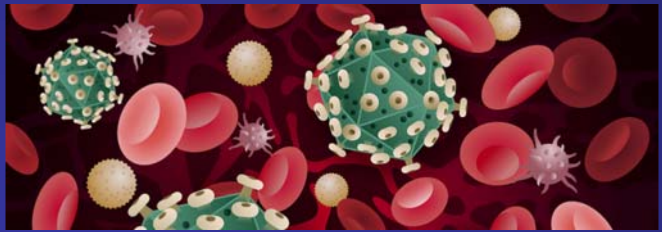


Hodgkin Cycle

Needs upwards and downwards causation: Integration across…

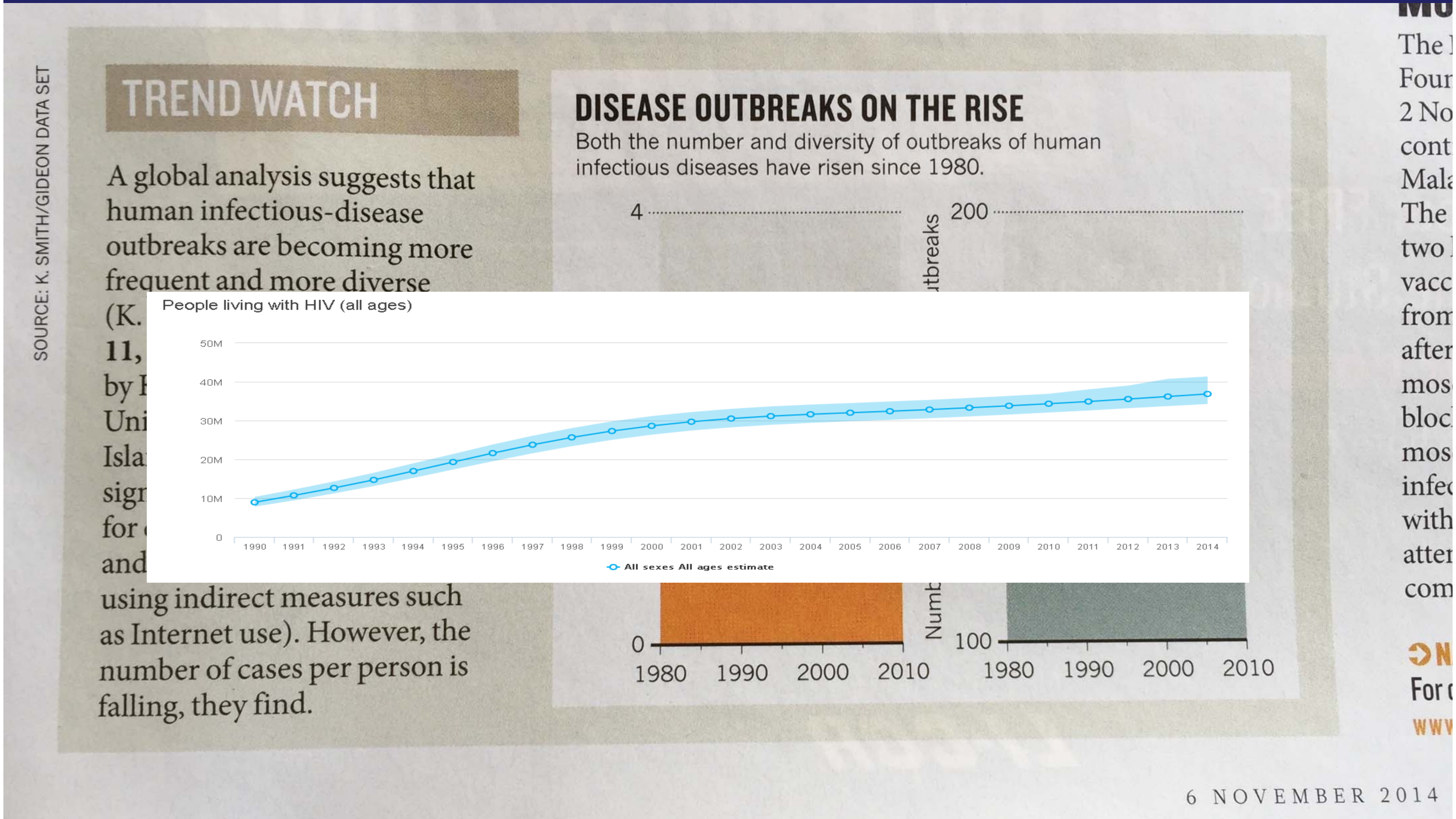


Two Examples:

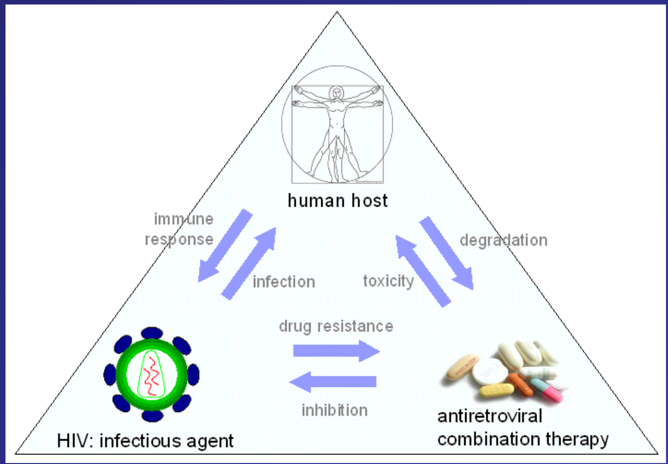


– The big question of Pandemics

– Diabetes Mellites (type 2)

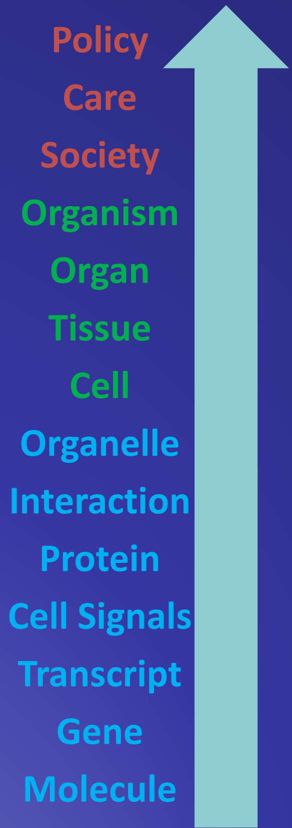

SOURCE: K. SMITH/GIDEON DATA SET
TREND WATCH
A global analysis suggests that human infectious-disease outbreaks are becoming more frequent and more diverse
(K.
11,
by
Uni
Isla
sign
for
and
using indirect measures such as Internet use). However, the number of cases per person is falling, they find.
DISEASE OUTBREAKS ON THE RISE
Both the number and diversity of outbreaks of human infectious diseases have risen since 1980.
4
200
Numb
0
100
1980 1990 2000 2010
1980 1990 2000 2010
People living with HIV (all ages)
50M
40M
30M
20M
10M
0
1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
All sexes All ages estimate
The
Four
2 No
cont
Mala
The
two
vacc
from
after
mos
bloc
mos
infec
with
atter
com
For
6 NOVEMBER 2014

HIV Truly complex



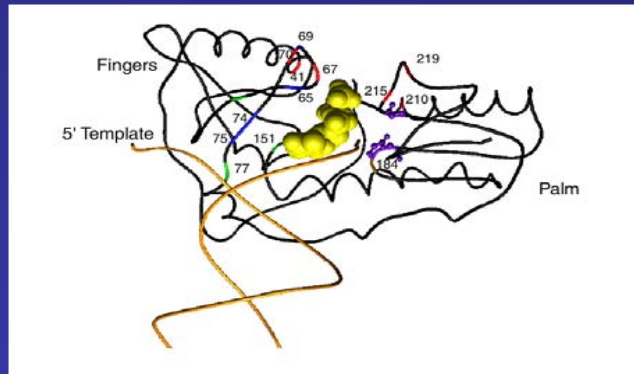
'Understanding, preventing and handling diseases requires a holistic approach'

Neil Ferguson
Nature 446, 2007



P.M.A. Sloot et al., *HIV Decision Support: From Molecule to Man*, Phil . Trans. R. Soc. A, vol. 367, nr 1898 pp. 2691-2703.

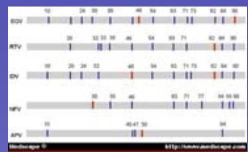
I Stoica, et al., *Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases.* J. Am. Chem. Soc., 130 (8), 2639-2648, 2008



Protein
Structure
& Binding
Affinity

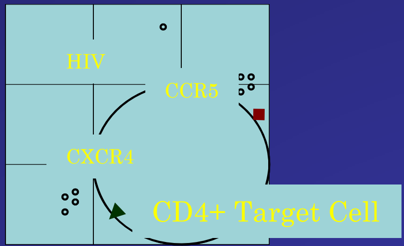
Molecular Dynamics
Binding Affinity

Protease and RT
mutations

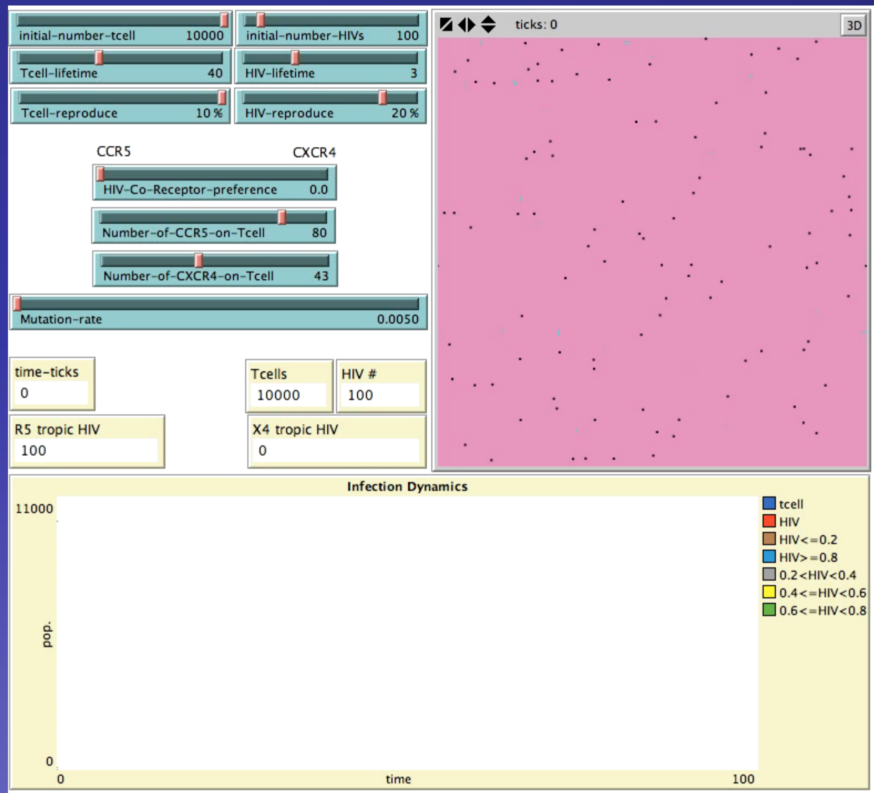


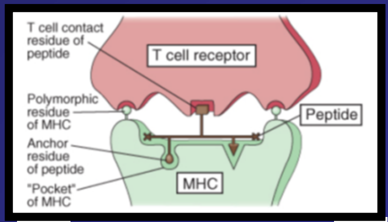
S. Jaeger, G. Ertaylan, D. van Dijk, U. Leser, and P.M.A. Sloot. *Inference of surface membrane factors of HIV-1 infection through functional interaction networks*. PLoS One 5(10), p. e13139. 2010

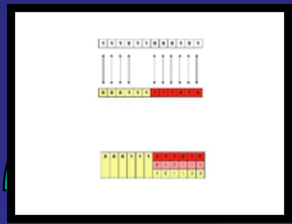
G. Ertaylan and P. M. A. Sloot. *A complex automata model of HIV-1 co-receptor tropism: Understanding mutation rate pressure*. In: Reviews in Antiretroviral Therapy. 2007

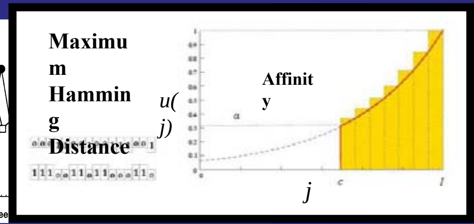


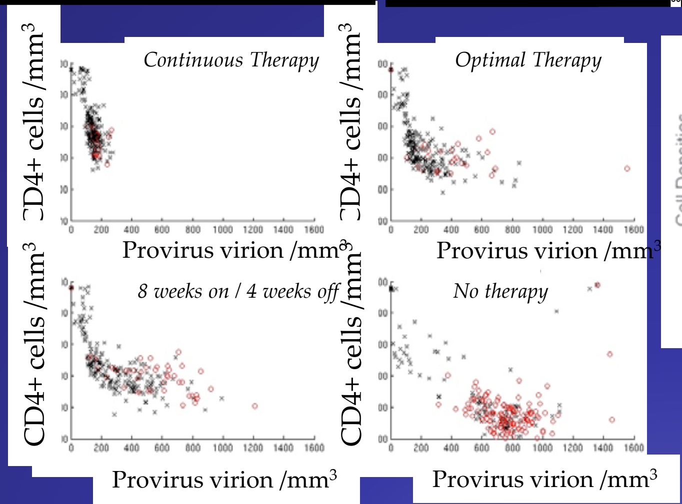
Agent-Based
Entry
Simulation

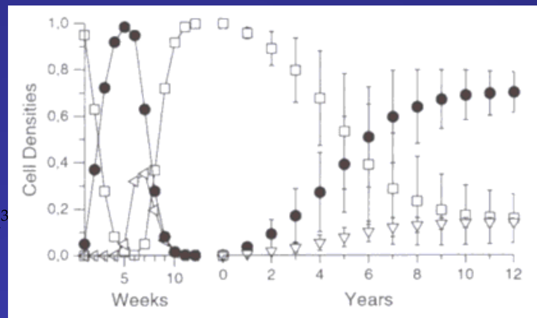










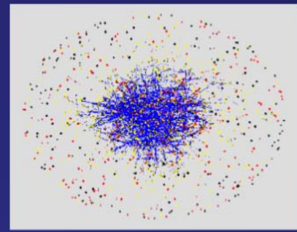


P. M. A. Sloot, F. Chen, and C. A. B. Boucher. *Cellular Automata Model of Drug Therapy for HIV Infection*. ACRI 2002. 2493, pp. 282-293.

E. Mancini; F. Castiglione; M. Bernaschi; A. de Luca and P.M.A. Sloot: *HIV Reservoirs and Immune Surveillance Evasion Cause the Failure of Structured Treatment Interruptions*, PLoS ONE, vol. 7, nr 4 pp. e36108. 2012.

R. Quax, D. A. M. C. van de Vijver, D. Frentz, and P. M. A. Sloot. *Inferring epidemiological parameters from phylogenetic information for the HIV-1 epidemic among MSM*. The European Physical Journal 222(6), pp. 1347-1358.

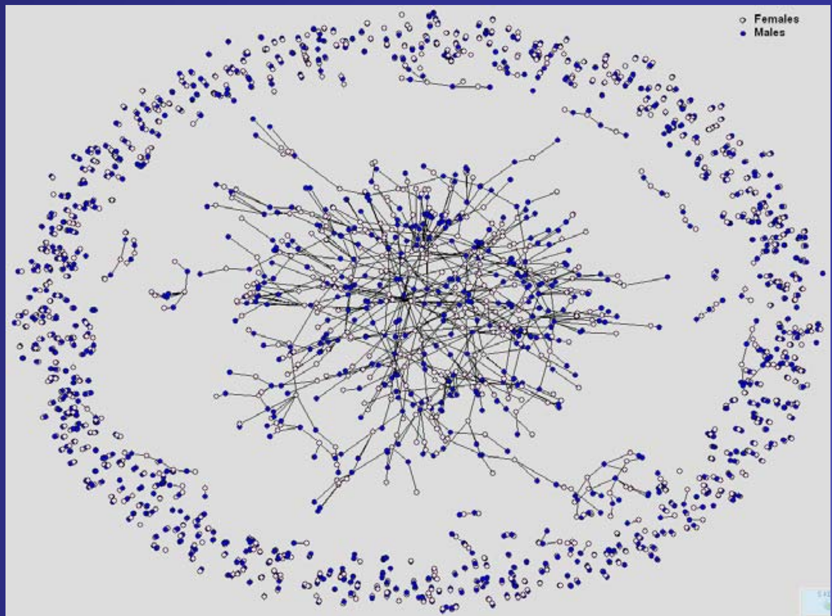
S. Mei, P. M. A. Sloot, R. Quax, Y. Zhu, and W. Wang. *Complex agent networks explaining the HIV epidemic among homosexual men in Amsterdam*. Mathematics and Computers in Simulation 80(5), pp. 1018-1030.

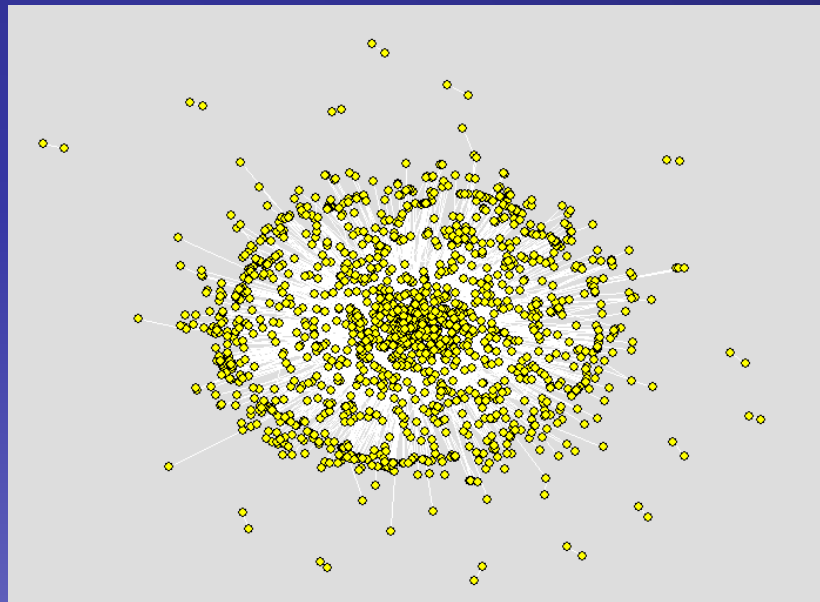


Complex Networks Epidemics

Heterosexual $\gamma = 2.7,\ k_{\max} = 70$

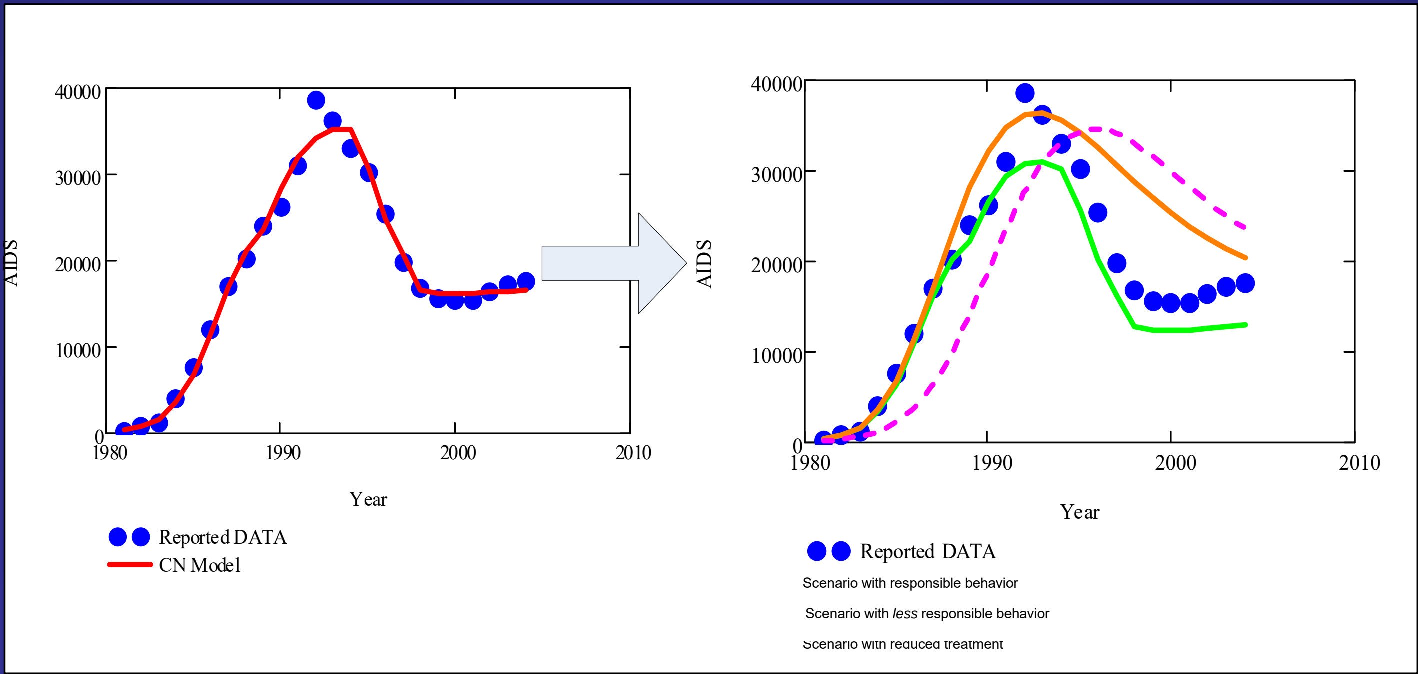
Homosexual $\gamma = 1.6,\ k_{\max} = 250$





A. Mei Shan, R. Quax, D. van de Vijver, Y. Zhu, and P. M. A. Sloot. *Increasing risk behaviour can outweigh the benefits of antiretroviral drug treatment on the HIV incidence among men-having-sex-with-men in Amsterdam*. BMC Infectious Diseases 11(118).

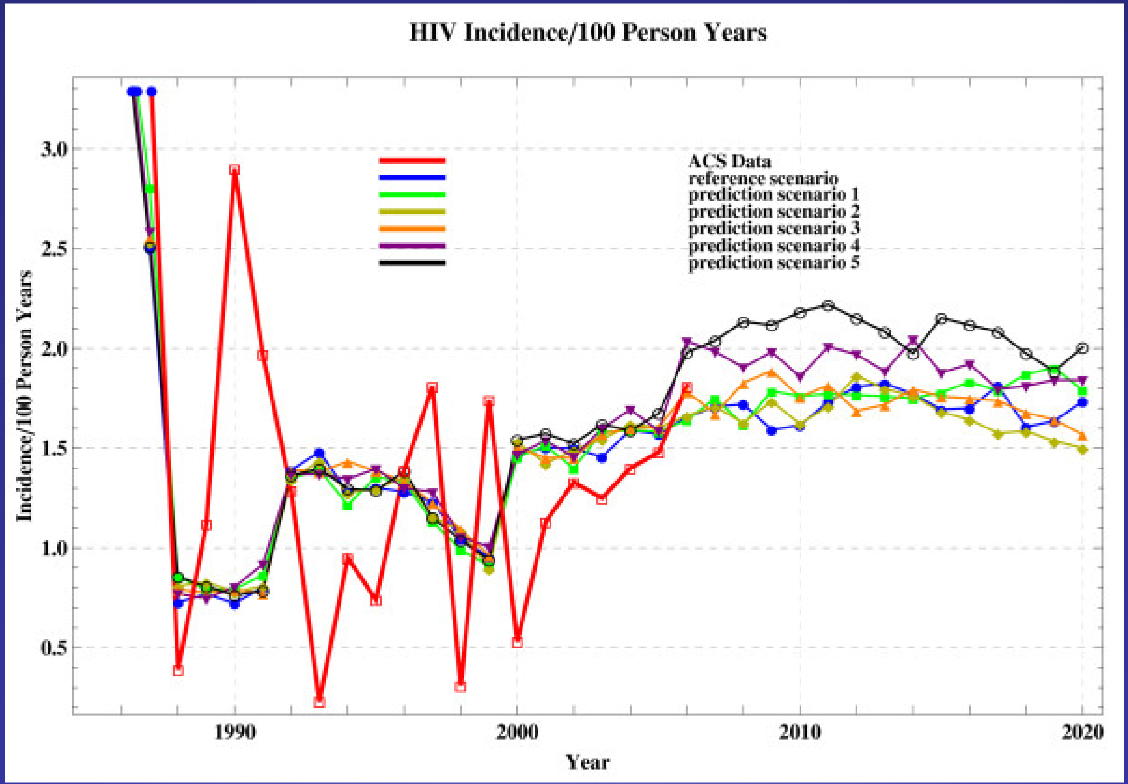
Calibration for Total Pandemic



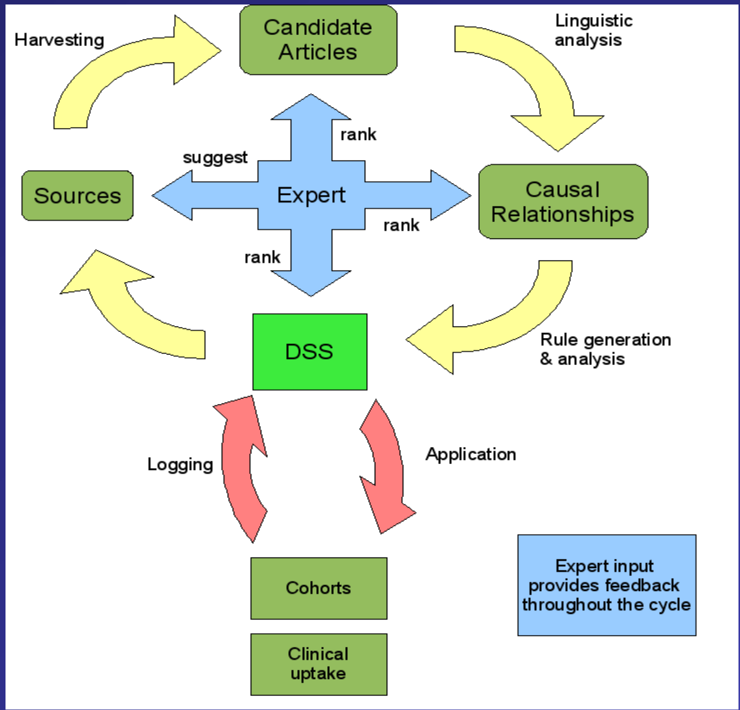
P.M.A. Sloot; S.V. Ivanov; A.V. Boukhanovsky et al., 2008

S. Mei; P.M.A. Sloot; R. Quax; Y. Zhu and W. Wang: 2010.

Prediction of MSM HIV incidence in Big Cities



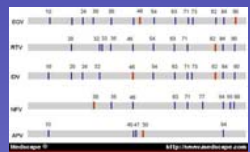
S. Mei; R. Quax; D.A.M.C. van de Vijver; Y. Zhu and P.M.A. Sloot: *Increasing risk behaviour can outweigh the benefits of antiretroviral drug treatment on the HIV incidence among men-having-sex-with-men in Amsterdam*, BMC Infectious Diseases, vol. 11, 2011.



C. Bui; Ó Nualláin; Boucher, P.M.A. Sloot: *Extracting causal relations on HIV drug resistance from literature*, BMC Bioinformatics, vol. 11, nr 1 pp. 101+11. 2010.

Q. C. Bui, P. M. A. Sloot, E. M. van Mulligen, J. A. Kors. *A novel feature-based approach to extract drug-drug interactions from biomedical text*. Bioinformatics 30(23), pp. 3365-3371.

Protease and RT mutations



Clinical Parameters:
-weight
- opportunistic infections and tumors
-survival

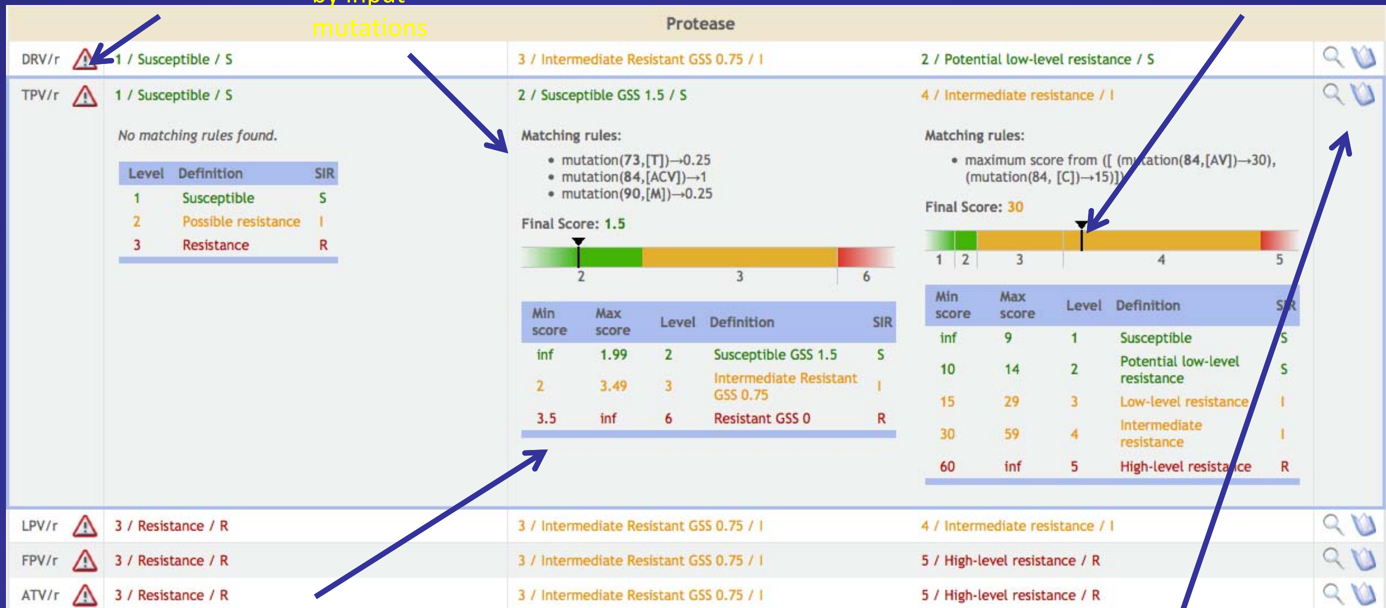
Text Mining →
Drugranking ← 1^{st} order logic

HiV Drug Ranking

Alert icon to highlight discordances

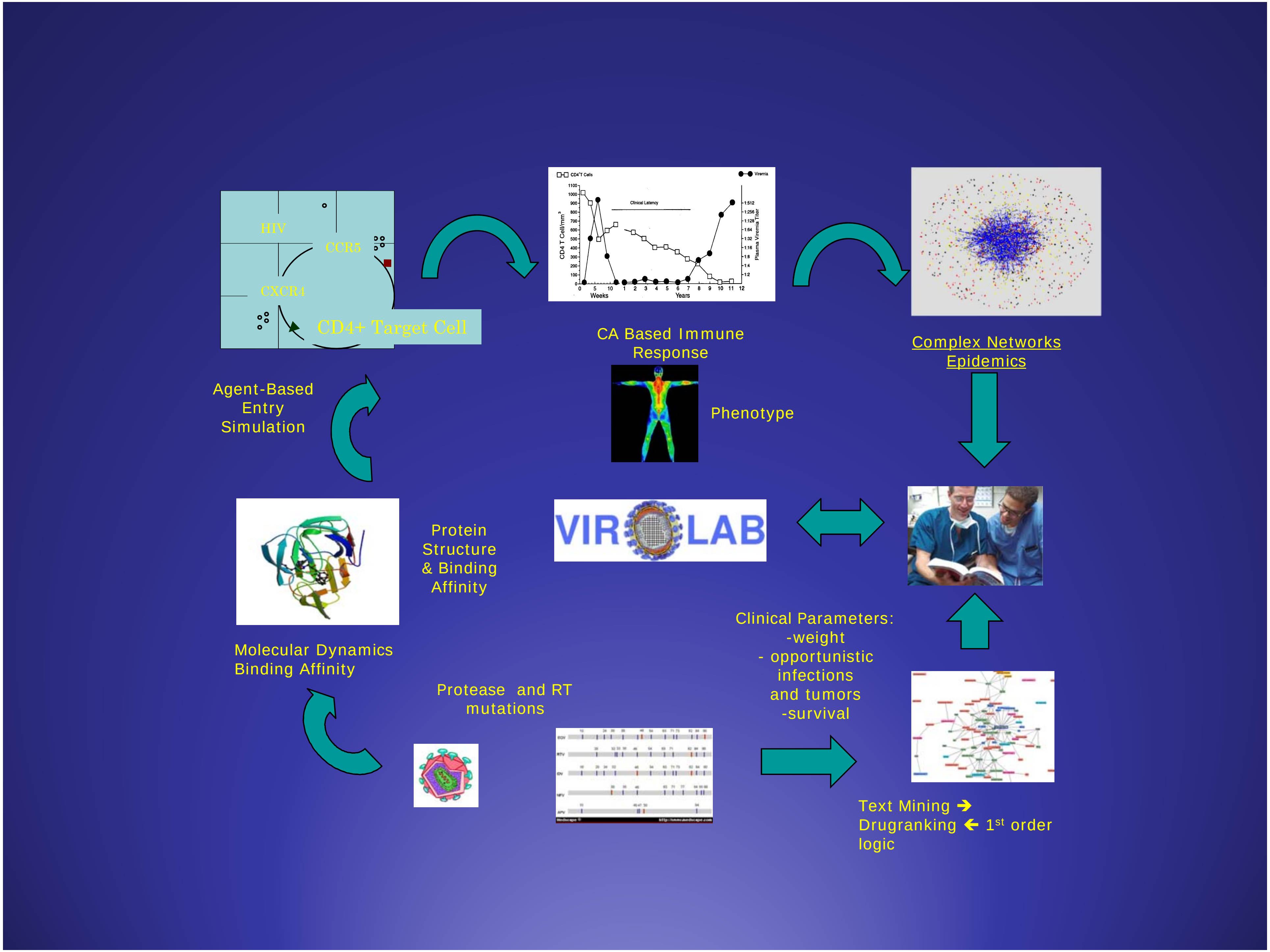
Display specific rules triggered by input mutations

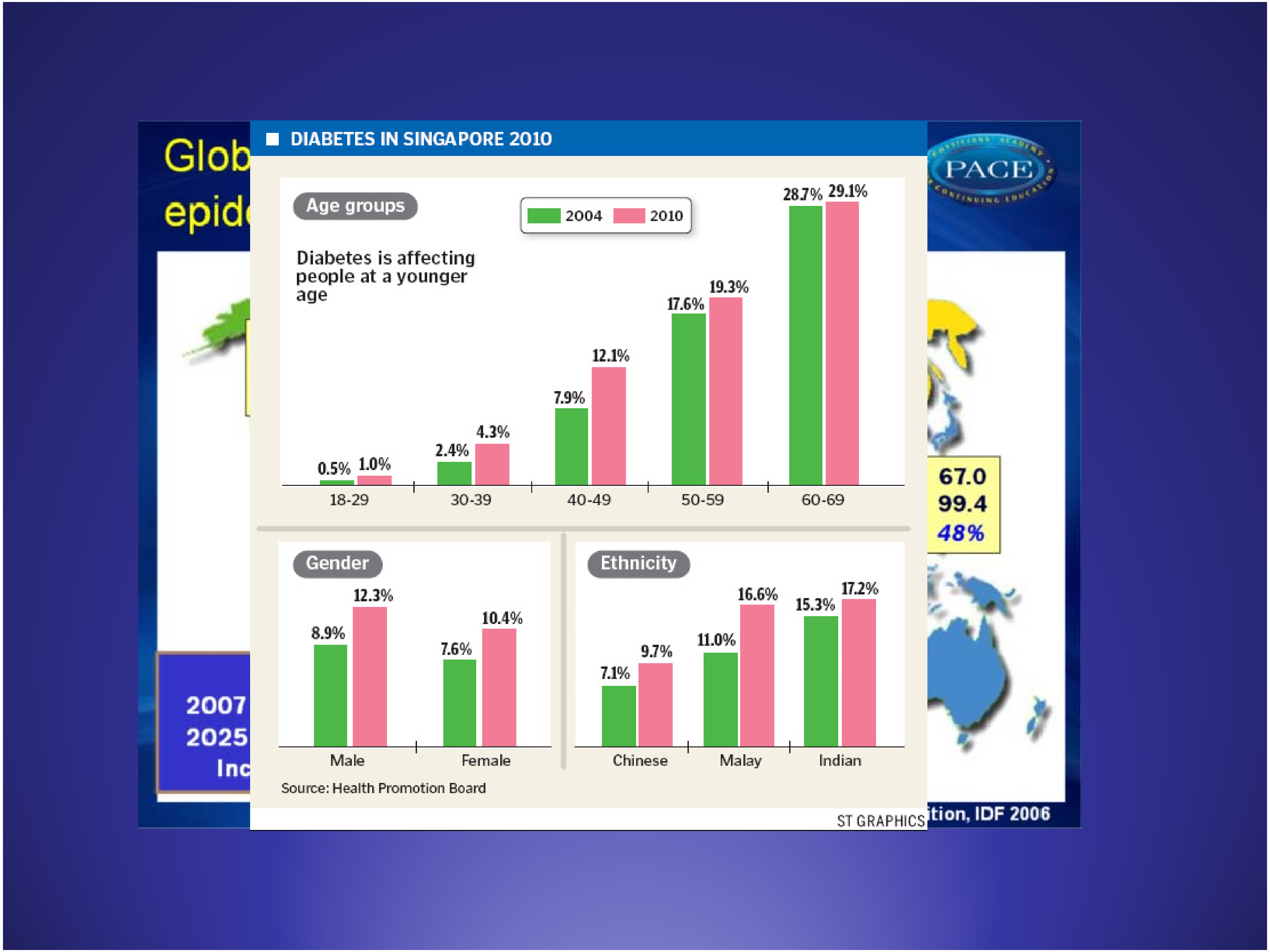
Intuitive visualization of the score



Legend of the drug resistance ranking

Literature Mining Tool

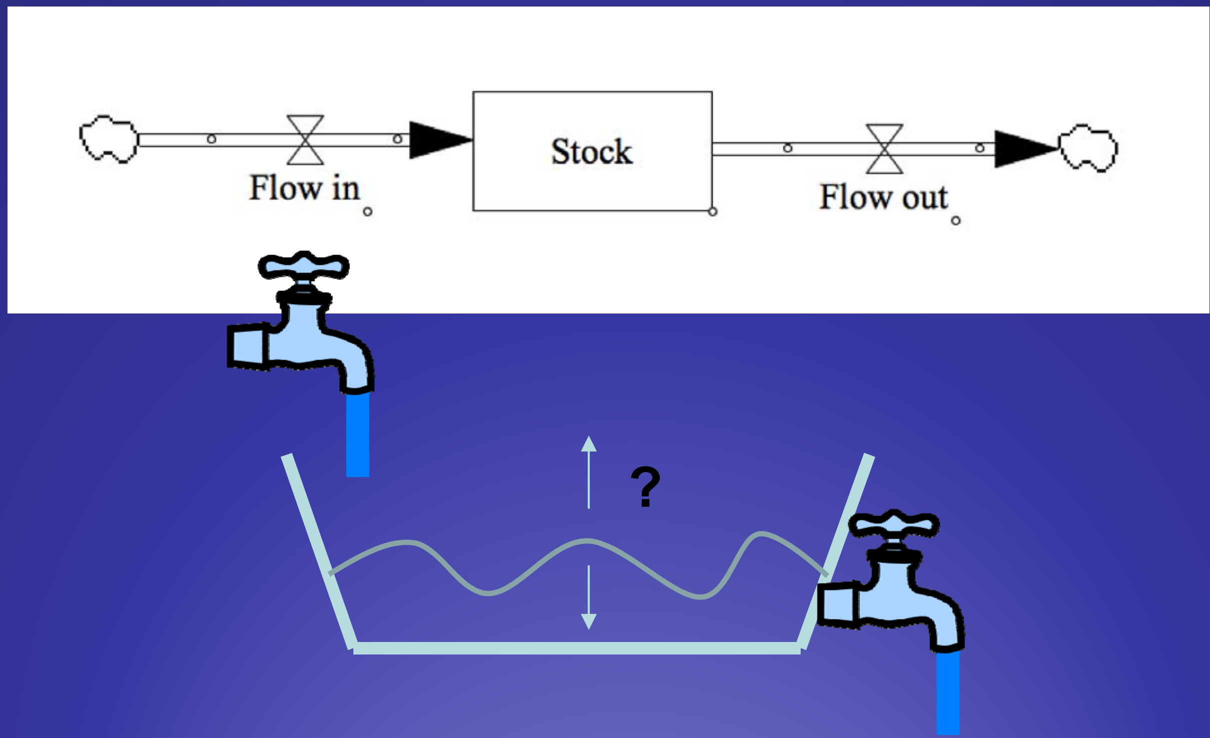

HIV
CCR5
CXCR4
CD4+ Target Cell
Agent-Based
Entry
Simulation
CA Based Immune
Response
Phenotype
Complex Networks
Epidemics
VIRLAB
Protein
Structure
& Binding
Affinity
Molecular Dynamics
Binding Affinity
Protease and RT
mutations
Clinical Parameters:
-weight
- opportunistic
infections
and tumors
-survival
Text Mining →
Drugranking ← 1st order
logic

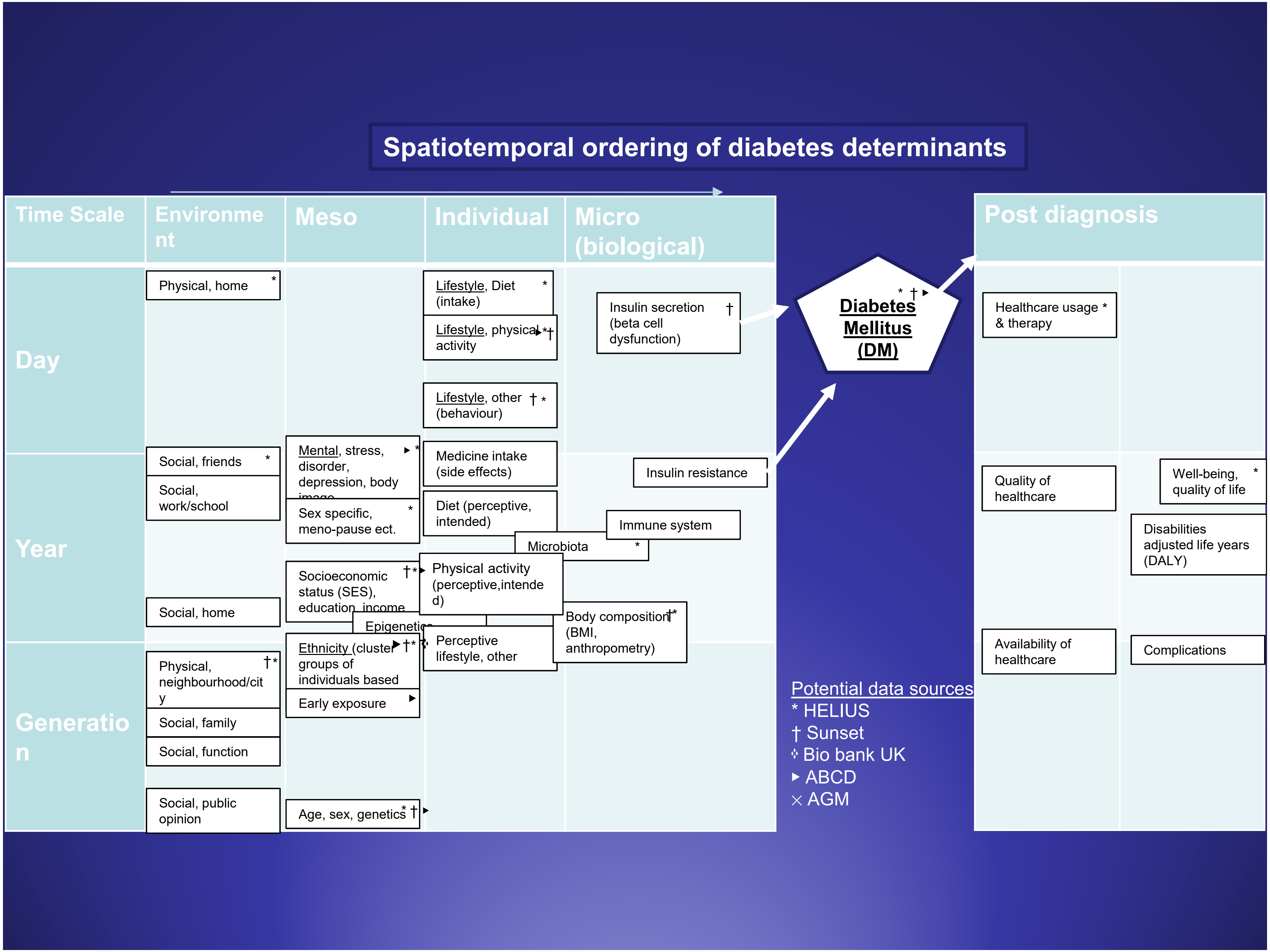


Glob
epid
2007
2025
Inc
DIABETES IN SINGAPORE 2010
Age groups
2004
2010
Diabetes is affecting people at a younger age
0.5%
1.0%
2.4%
4.3%
7.9%
12.1%
17.6%
19.3%
28.7%
29.1%
18-29
30-39
40-49
50-59
60-69
Gender
8.9%
12.3%
7.6%
10.4%
Male
Female
Ethnicity
7.1%
9.7%
11.0%
16.6%
15.3%
17.2%
Chinese
Malay
Indian
Source: Health Promotion Board
ST GRAPHICS
PACE
67.0
99.4
48%
ition, IDF 2006

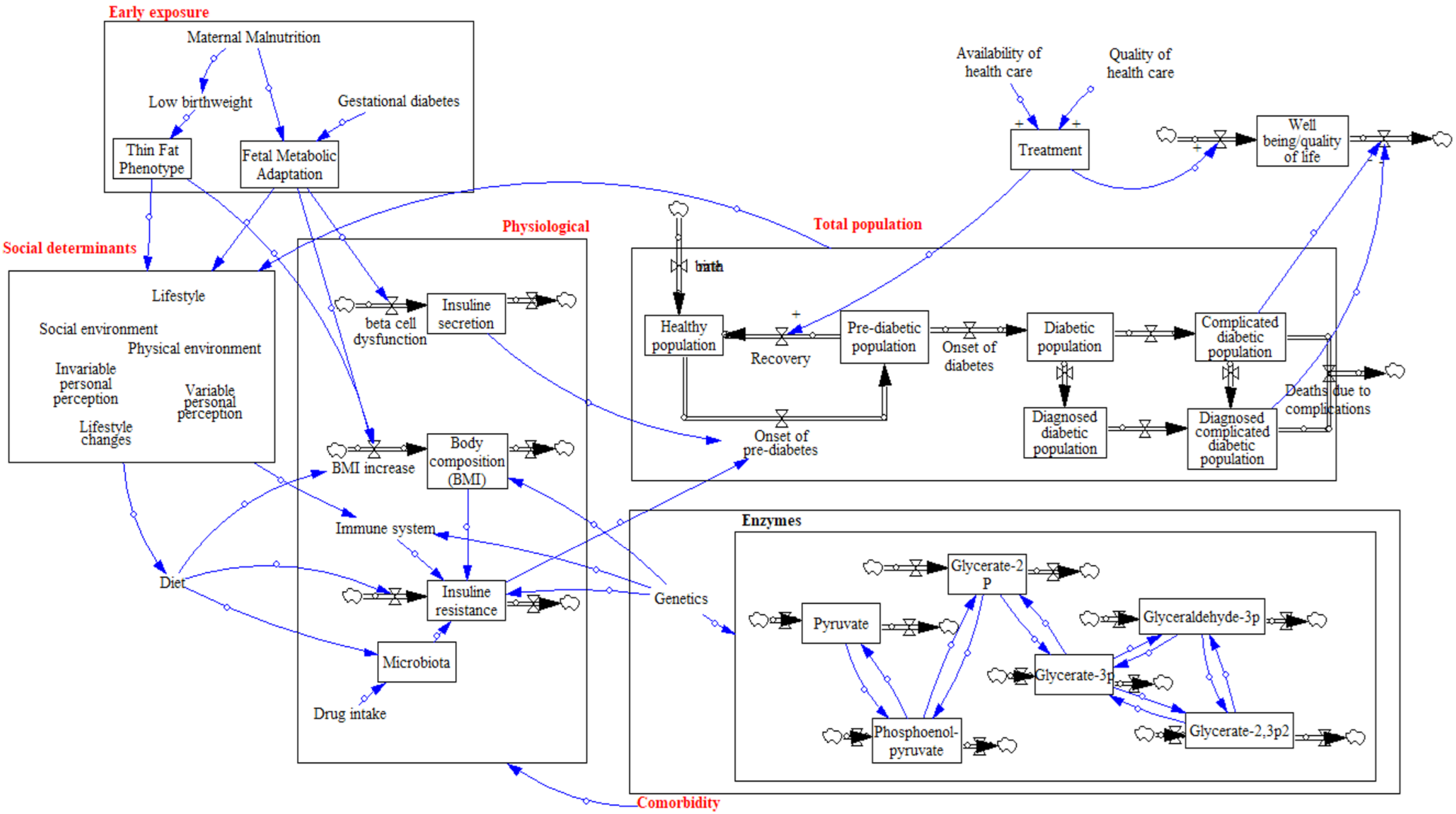
System Dynamics Approach

Stock-flow diagrams

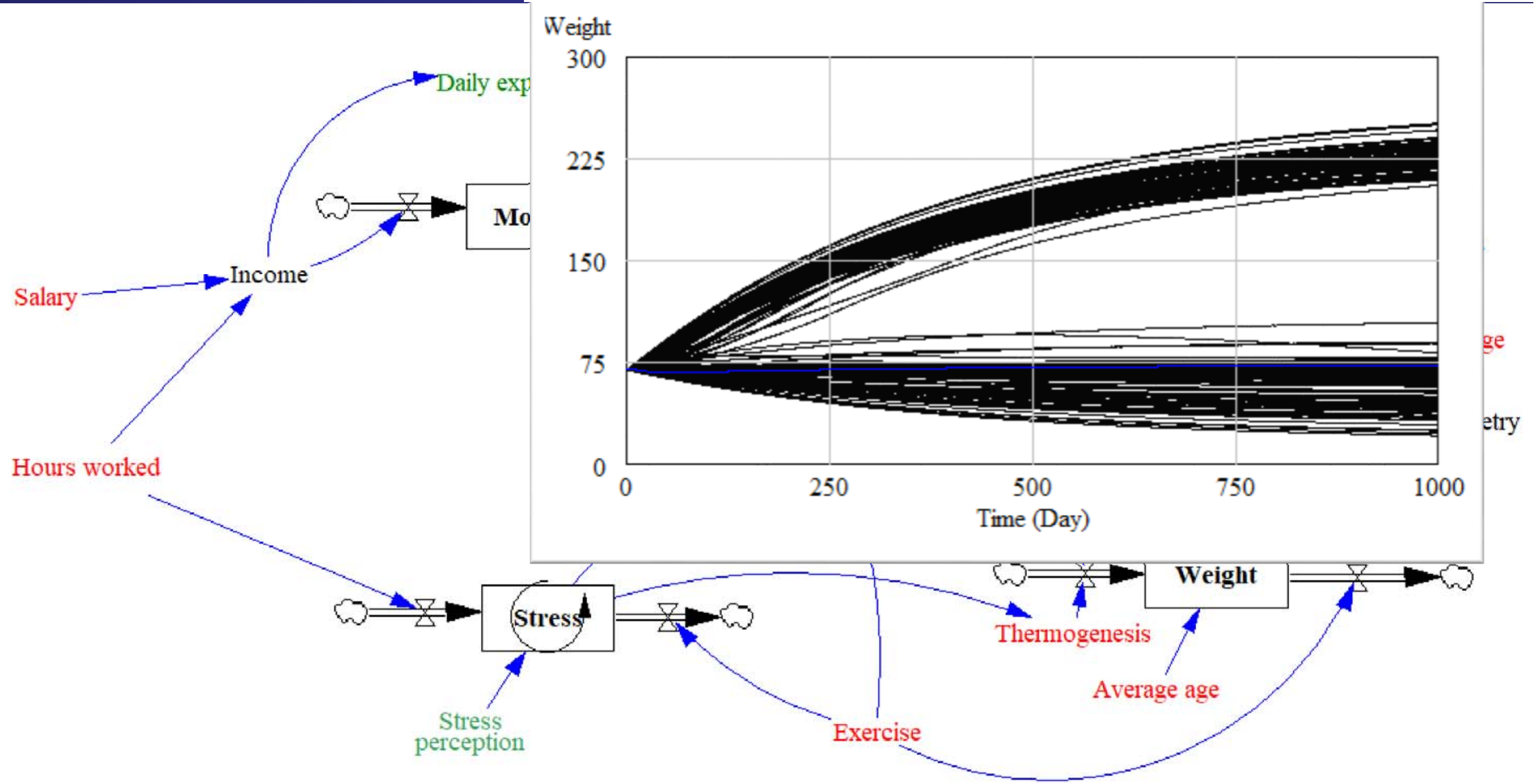




Spatiotemporal ordering of diabetes determinants
Time Scale
Environment
Meso
Individual
Micro (biological)
Post diagnosis
Day
Year
Generation
Physical, home *
Social, friends *
Social, work/school
Social, home
Physical, neighbourhood/city †*
Social, family
Social, function
Social, public opinion
Mental, stress, disorder, depression, body image
Sex specific, meno-pause ect. *
Socioeconomic status (SES), education, income †*
Epigenetics
Ethnicity (cluster groups of individuals based †*
Early exposure
Age, sex, genetics * †
Lifestyle, Diet (intake) *
Lifestyle, physical activity *†
Lifestyle, other (behaviour) † *
Medicine intake (side effects)
Diet (perceptive, intended)
Physical activity (perceptive,intended)
Perceptive lifestyle, other
Insulin secretion (beta cell dysfunction) †
Insulin resistance
Immune system
Microbiota *
Body composition (BMI, anthropometry) †*
Diabetes Mellitus (DM) * †
Healthcare usage & therapy *
Quality of healthcare
Well-being, quality of life *
Disabilities adjusted life years (DALY)
Availability of healthcare
Complications
Potential data sources
* HELIUS
† Sunset
Bio bank UK
ABCD
AGM


Early exposure
Maternal Malnutrition
Low birthweight
Gestational diabetes
Thin Fat Phenotype
Fetal Metabolic Adaptation
Social determinants
Lifestyle
Social environment
Physical environment
Invariable personal perception
Variable personal perception
Lifestyle changes
Physiological
beta cell dysfunction
Insuline secretion
BMI increase
Body composition (BMI)
Immune system
Diet
Insuline resistance
Microbiota
Drug intake
Genetics
Comorbidity
Total population
birth
Healthy population
Recovery
Pre-diabetic population
Onset of diabetes
Diabetic population
Complicated diabetic population
Onset of pre-diabetes
Diagnosed diabetic population
Diagnosed complicated diabetic population
Deaths due to complications
Availability of health care
Quality of health care
Treatment
Well being/quality of life
Enzymes
Glycerate-2 P
Pyruvate
Glyceraldehyde-3p
Glycerate-3p
Phosphoenol-pyruvate
Glycerate-2,3p2

Two stable weights emerge due to feedback and adaptation in the system:
This simple example already behaves as a complex system!

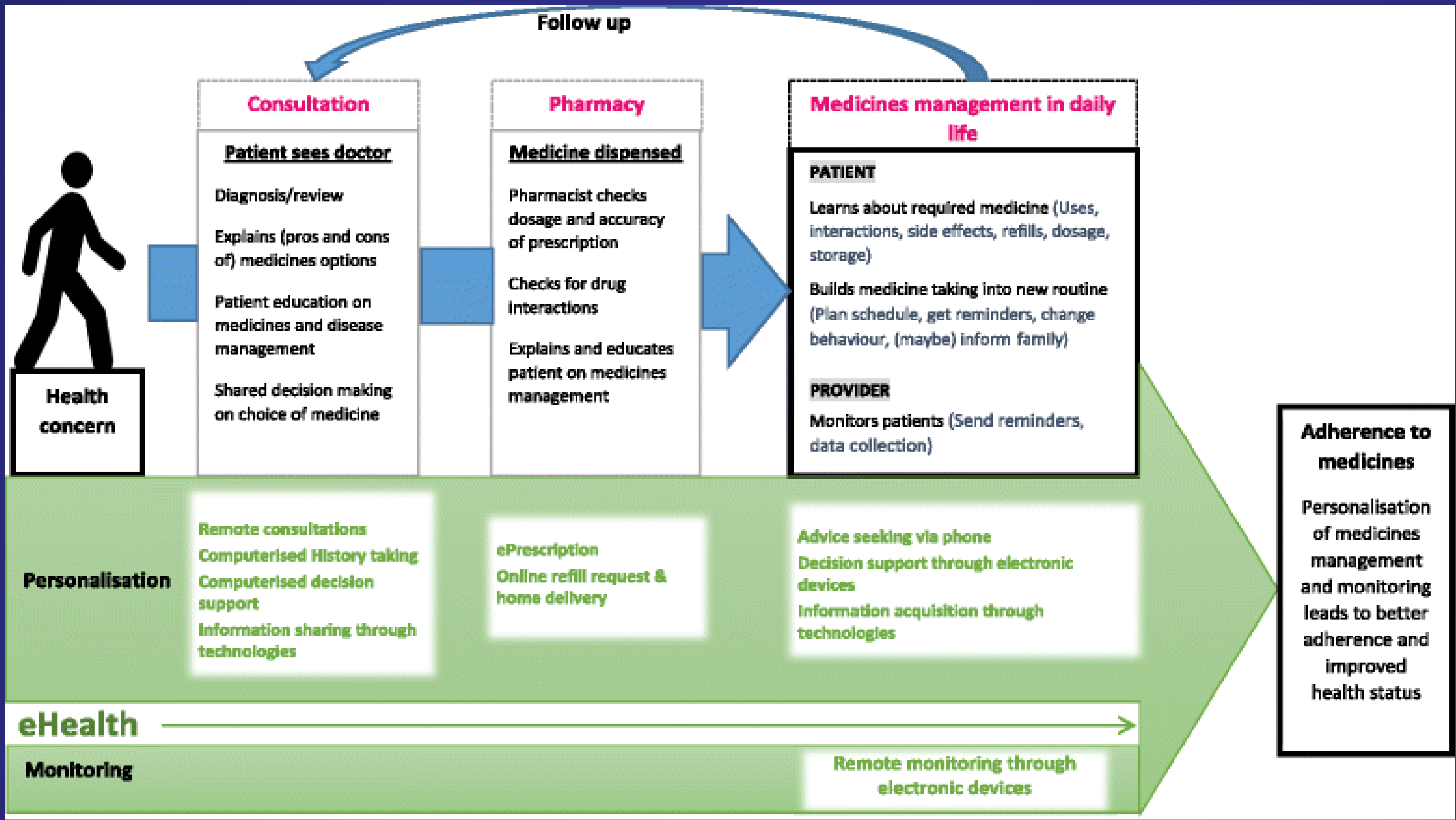


‘The future of medications management: personalisation, monitoring and adherence’

Josip Car, Woan Shin Tan, Zhilian Huang and Peter Sloot

BMC Medicine 2017 **15**:73

Regulation through eHealth provides even more feedback loops!



Conclusions…

- Data is dead (without what-if models[1]):
  - Data is a record, not a conclusion or an insight or a solution. Data Analytic predictions work well only if the future is fundamentally like the past.
- Big Data leads to Radical Empiricism (data without reason)
  - 'Denies the possibility of knowledge': Reichenbach, H. 1971 The Rise of Scientific Philosophy
- We need to: … 'stop and think about the complexity, the inconceivable nature of Nature' (R. Feynman)

[1]See e.g.: Peter J. Haas et al., VLDB, 2011