The Challenges of Modeling Human Health

Chris Stephens, Director of Data Science, C3 y ICN, UNAM

The principal purpose of living systems and the principal purpose of science - medicine, public health - is to…

Predict

for

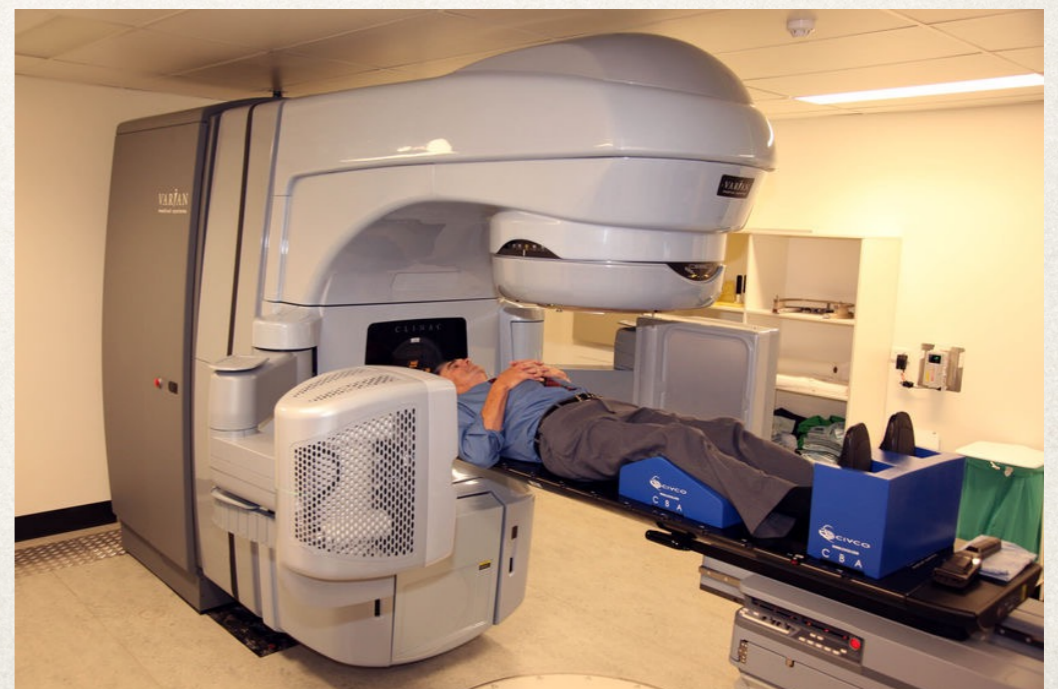**Decision making**

Predict for individuals…

Diagnostics

QBC Fluorescence Microscopy

- Upgrade your microscope to fluorescence
- Malaria diagnostics
- Tuberculosis diagnostics
- Other blood borne parasite diagnostics
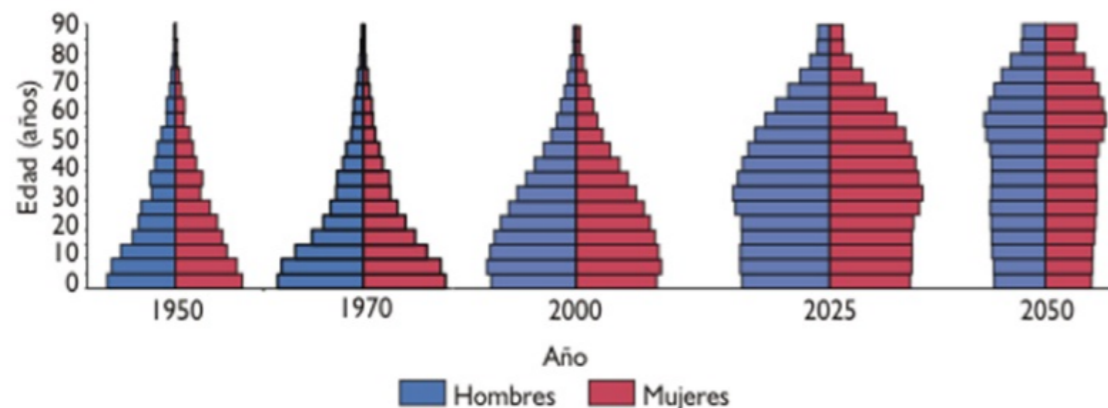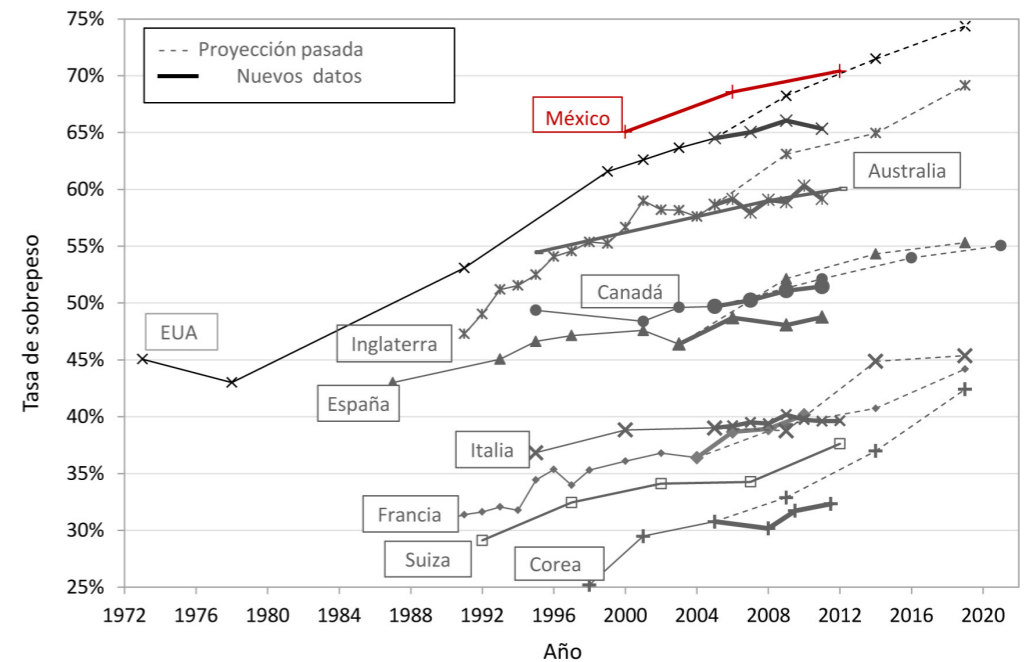
QBC *Europe*

Treatments

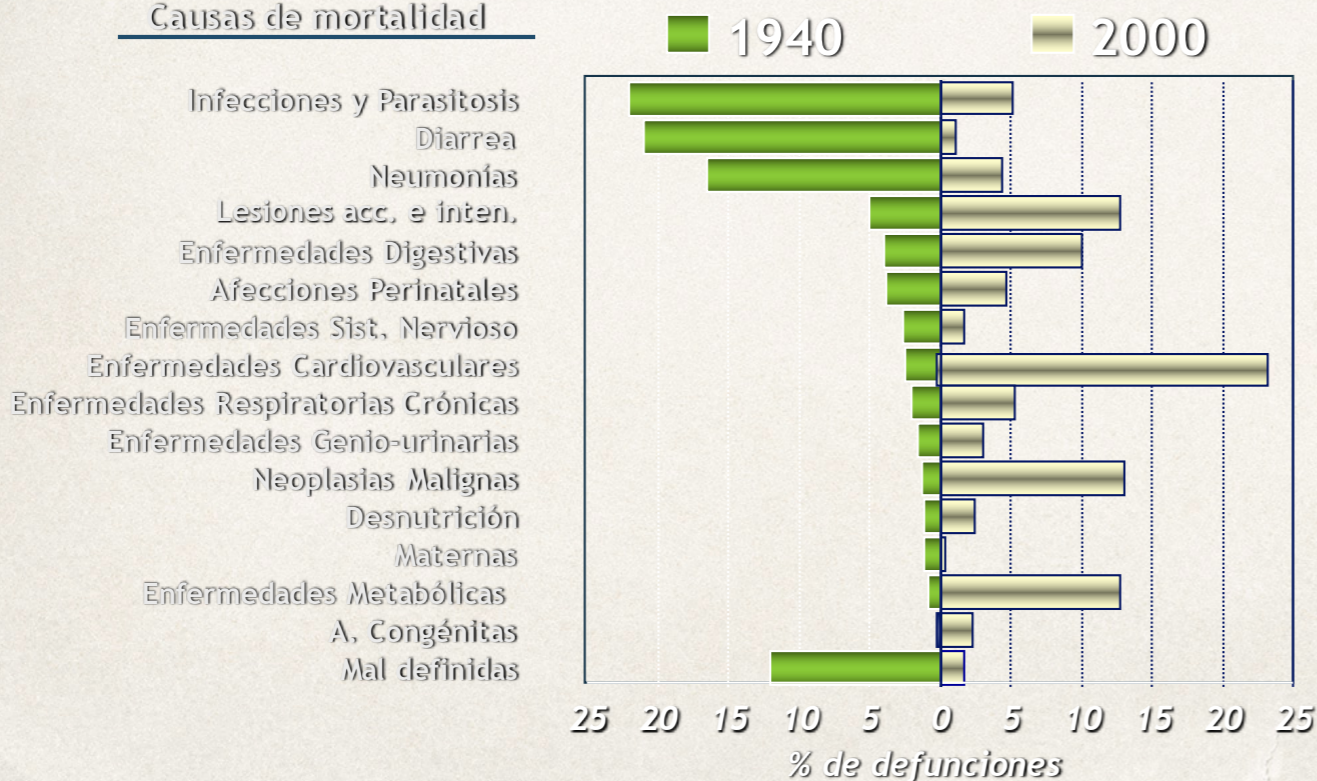# Predict for populations…

Demography, Epidemiology, Economics,…

Figura 1.5 Pirámides de edad en México, 1950-2050
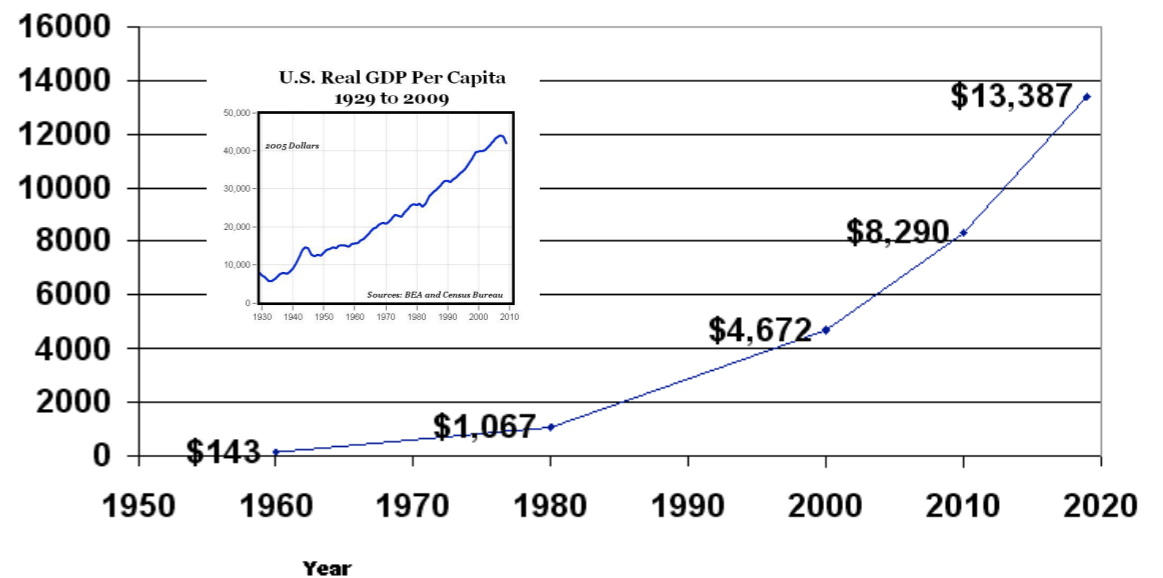
Edad (años) / Año

Hombres / Mujeres

Fuentes:
INEGI. *Estadísticas Históricas de México*. México. S/A.
INEGI. *XII Censo General de Población y Vivienda 2000*. México. 2001.
Conapo. *Proyecciones de la Población de México, 2000-2050*. México. 2002.

Proyección pasada / Nuevos datos

México / Australia / EUA / Inglaterra / Canadá / España / Italia / Francia / Suiza / Corea

Tasa de sobrepeso / Año

Causas de mortalidad

1940 / 2000

Infecciones y Parasitosis
Diarrea
Neumonías
Lesiones acc. e inten.
Enfermedades Digestivas
Afecciones Perinatales
Enfermedades Sist. Nervioso
Enfermedades Cardiovasculares
Enfermedades Respiratorias Crónicas
Enfermedades Genio-urinarias
Neoplasias Malignas
Desnutrición
Maternas
Enfermedades Metabólicas
A. Congénitas
Mal definidas

% de defunciones

Annual U.S. Healthcare Expenses per Person by Year

U.S. Real GDP Per Capita 1929 to 2009
2005 Dollars
Sources: BEA and Census Bureau

$143 $1,067 $4,672 $8,290 $13,387

Year

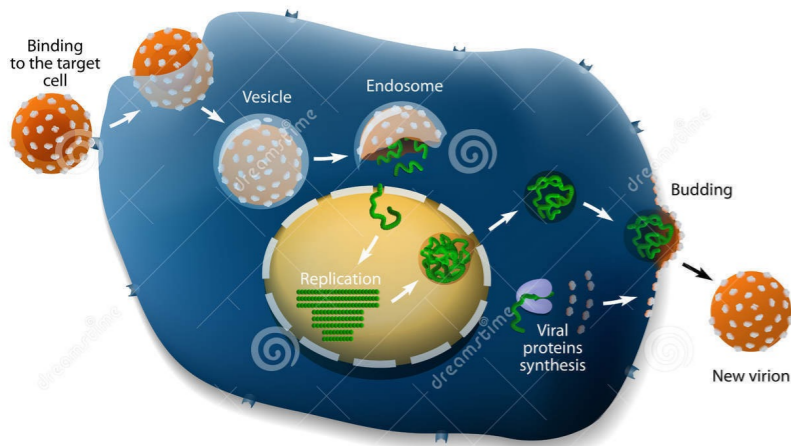Source: http://www1.cms.gov/NationalHealthExpendData/downloads/proj2009.pdf

Modelling
and "Cause and effect"

INFLUENZA VIRAL LIFE CYCLE
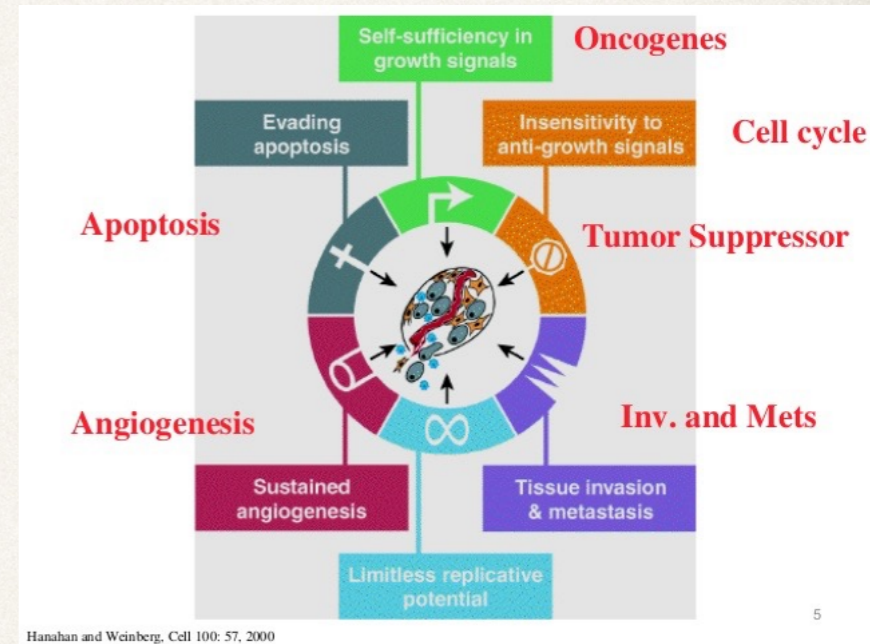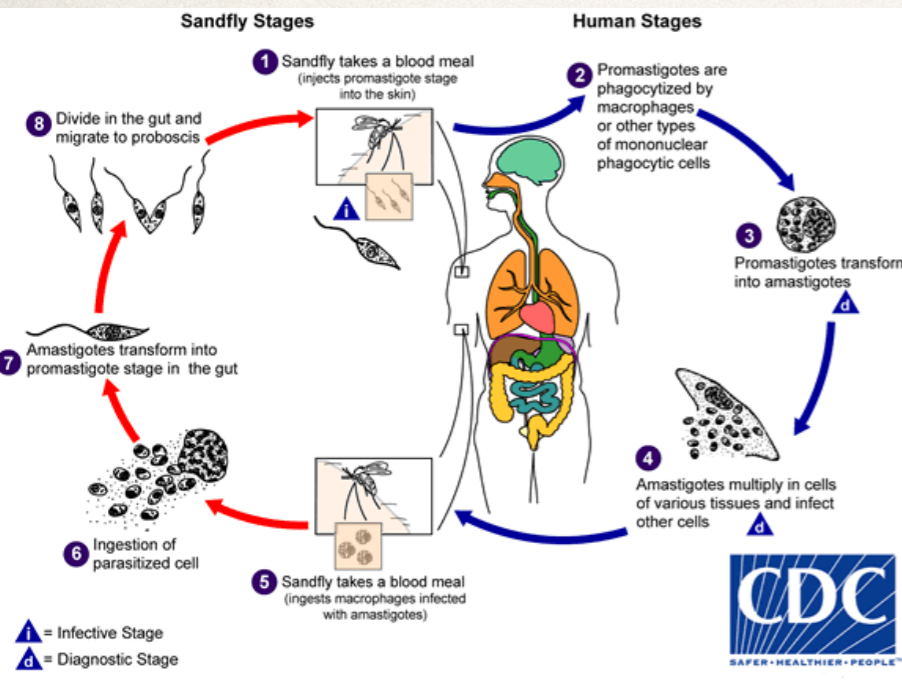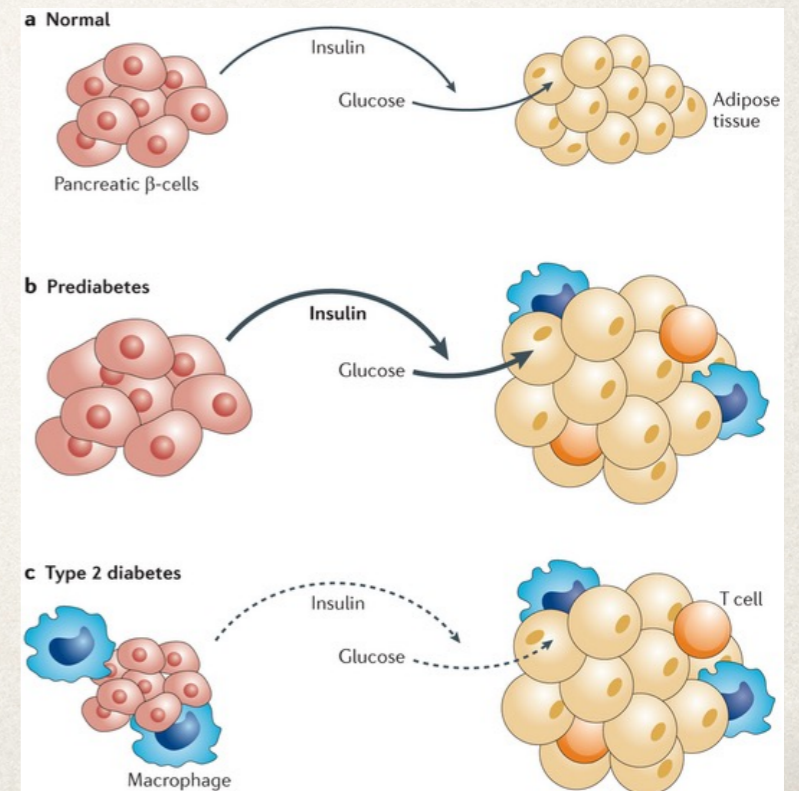
At the "micro" level

Modelling
and "Cause and effect"

Risk factors in males worldwide

Smoking | High blood pressure | High body-mass index | Alcohol | Childhood malnutrition | Diabetes | HIV

At the "macro" level

Flu-Like Epidemic

Heart Disease Rates as Reported on Death Certificates | Heart Disease Rates as Predicted By Twitter

Less Deaths — More Deaths

Influenza incidence forecasts for selected cities
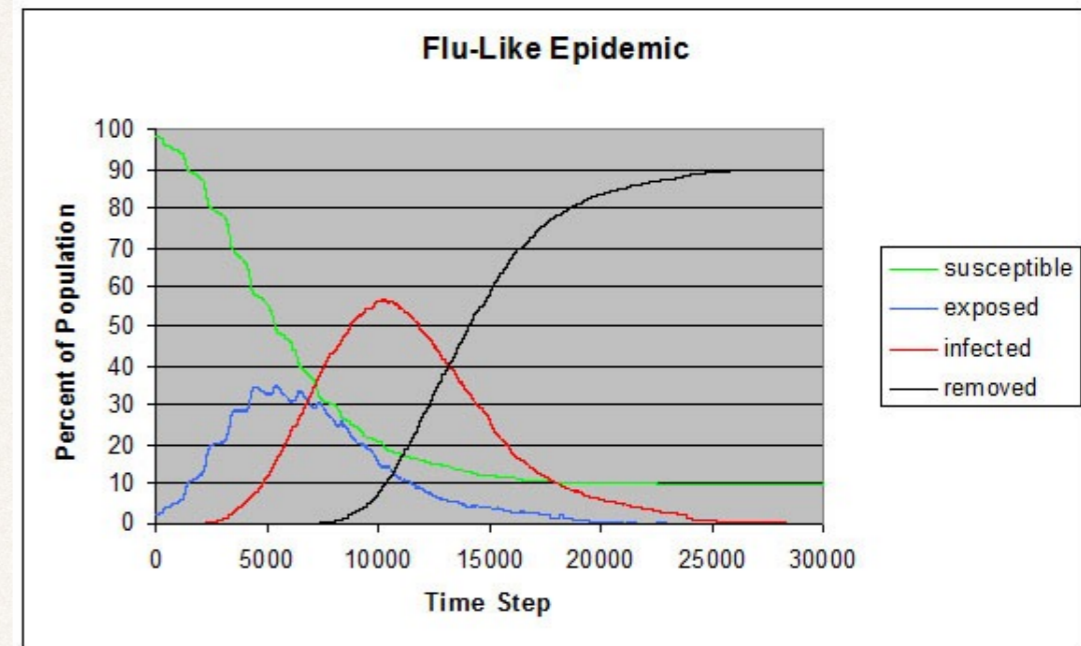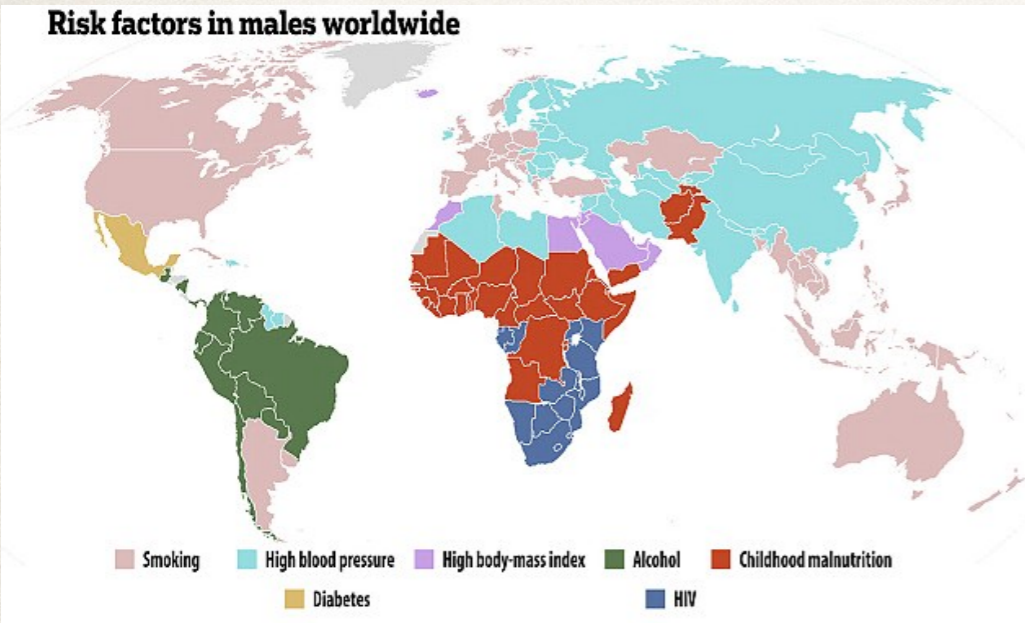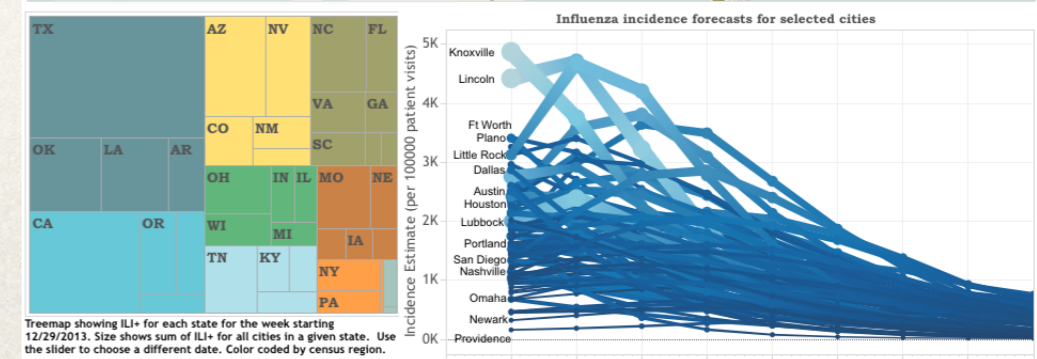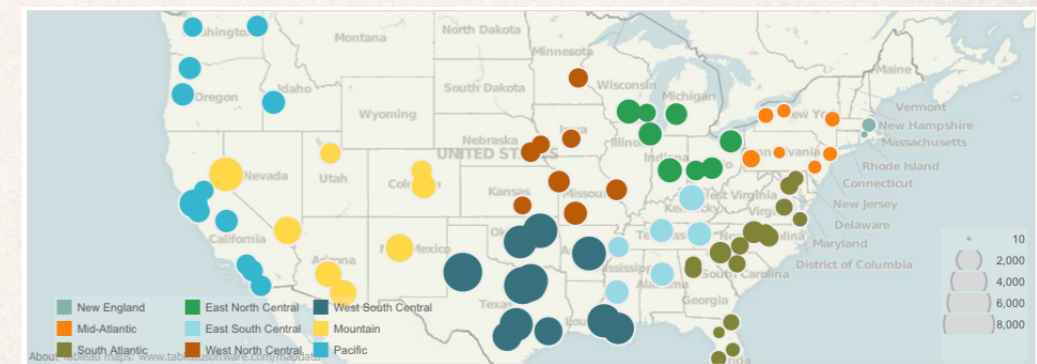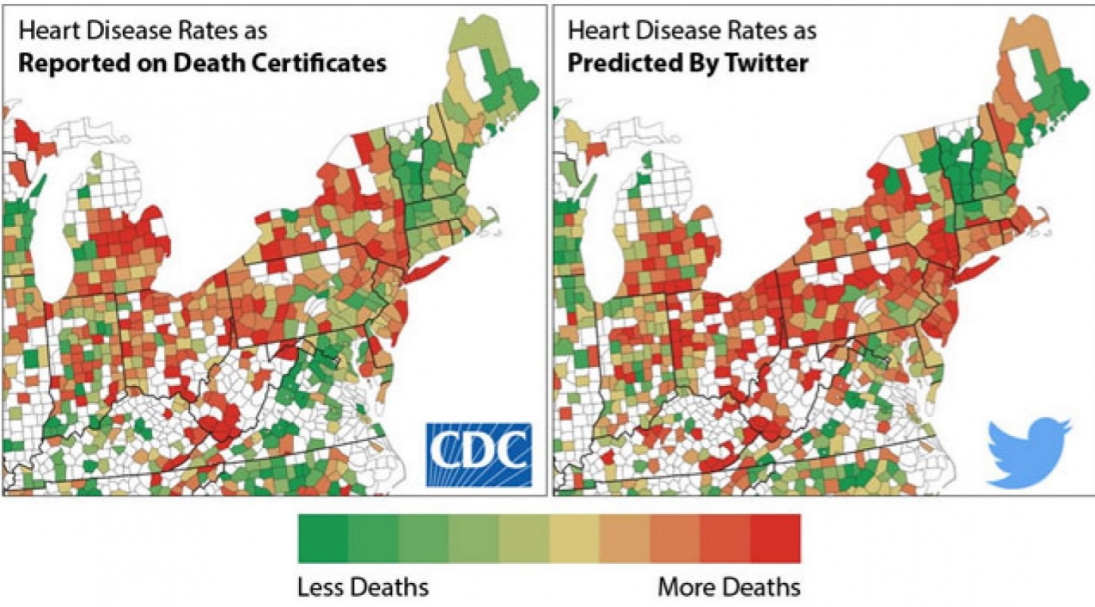
Modelling: "Cause and effect"

The standard paradigm

The world is a machine
Its been very successful.
It is the paradigm of
But
curative medicine
including people!
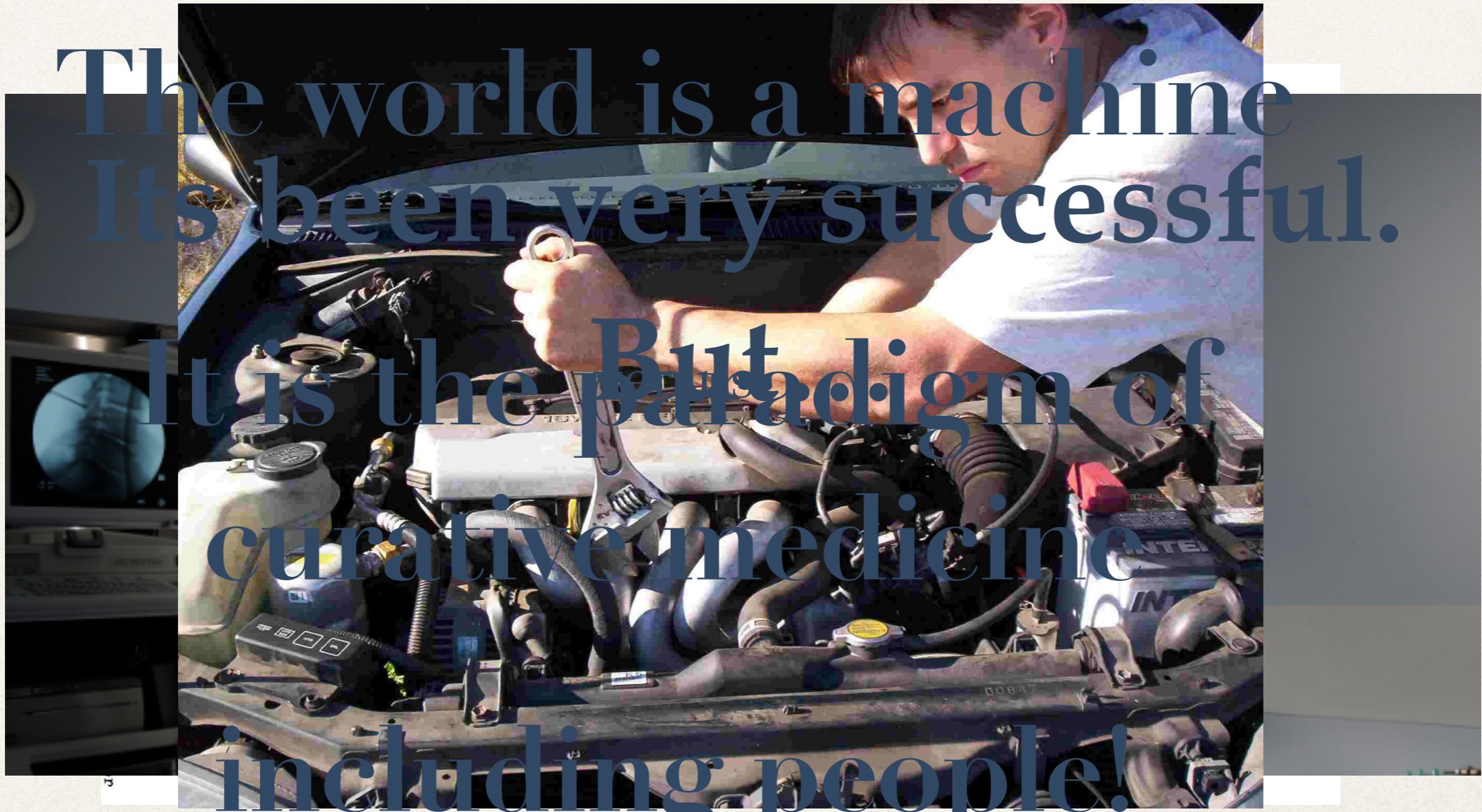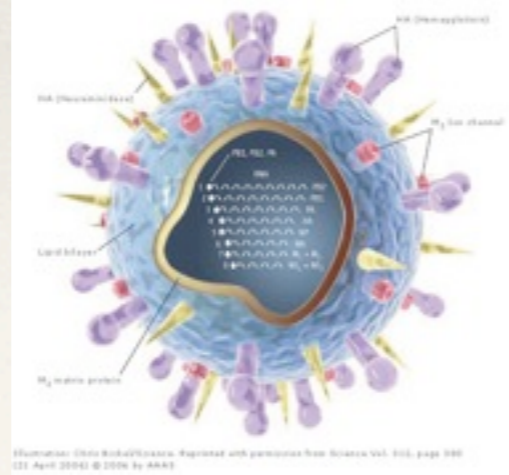
Diseases are
Complex Adaptive Systems
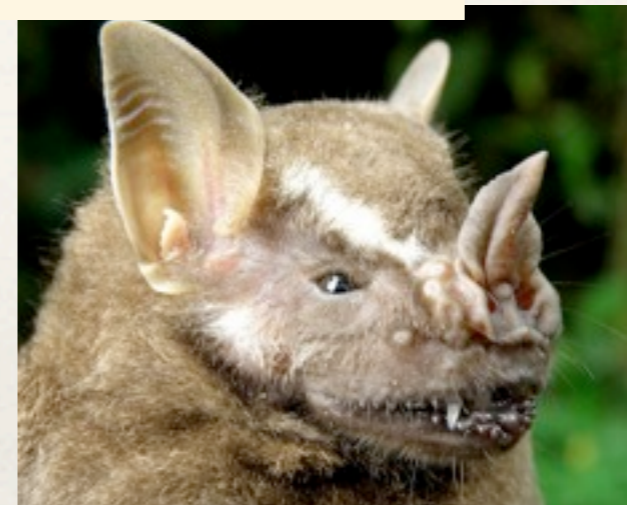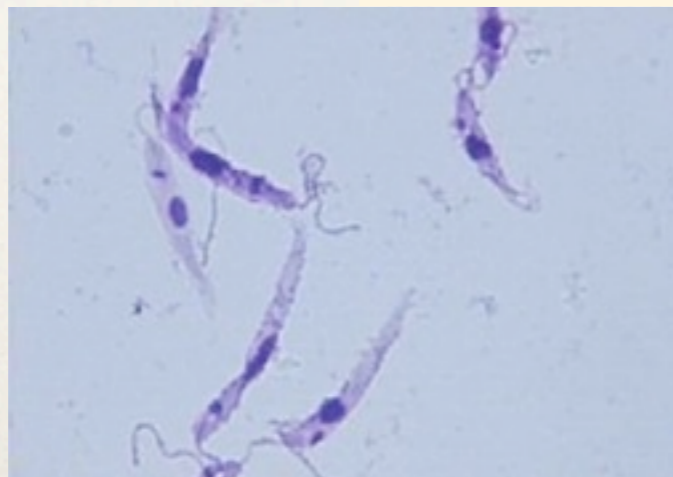
They are complex
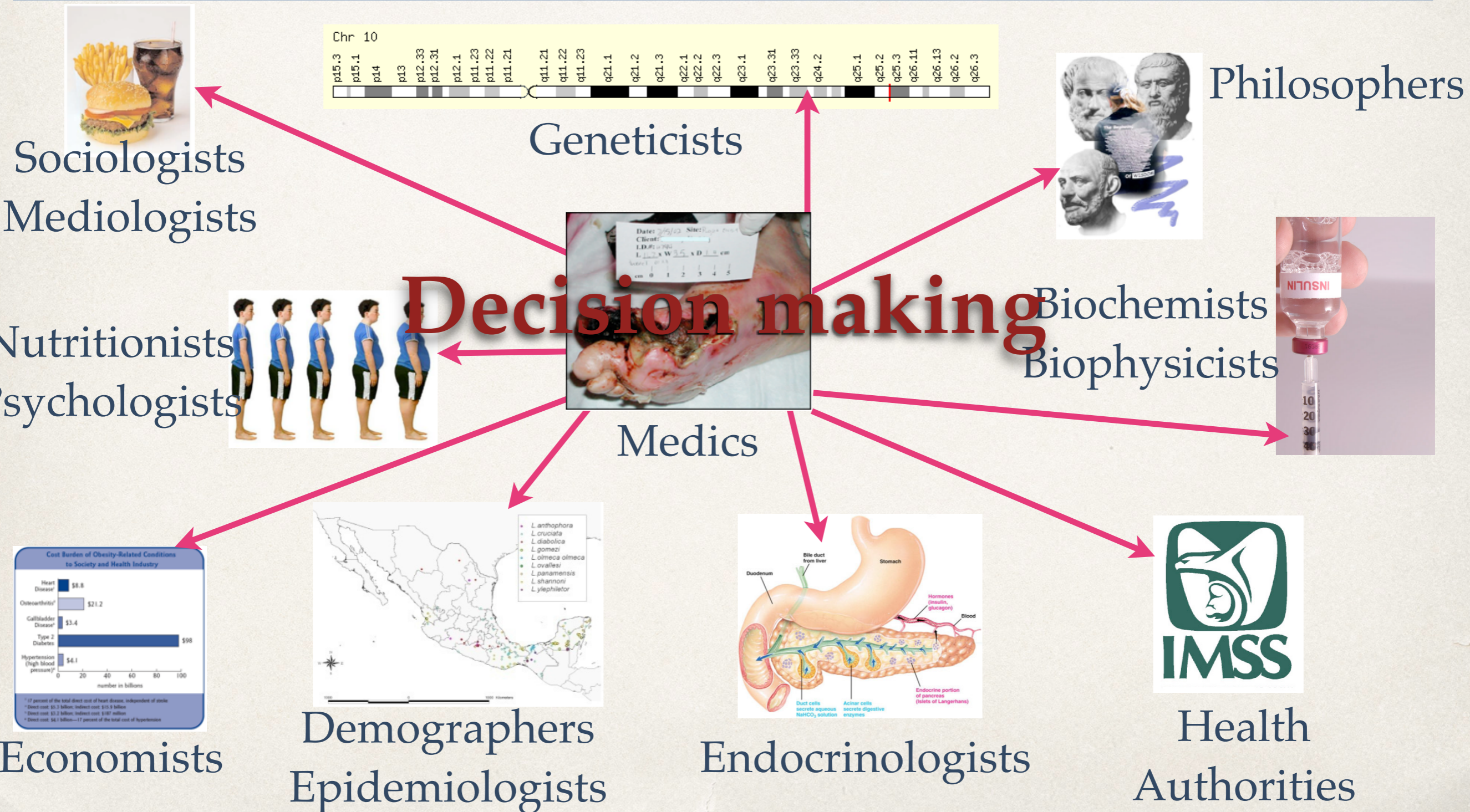
Sociologists
Mediologists
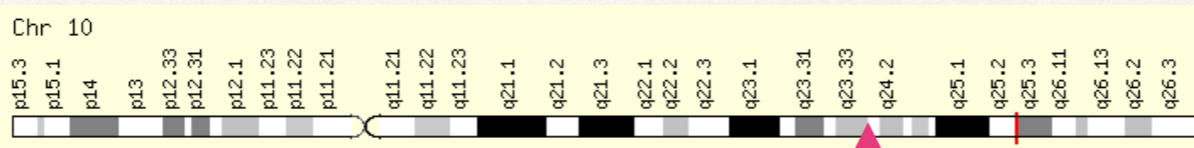
Geneticists

Philosophers

Nutritionists
Psychologists

Decision making

Biochemists
Biophysicists

Medics

Economists

Demographers
Epidemiologists
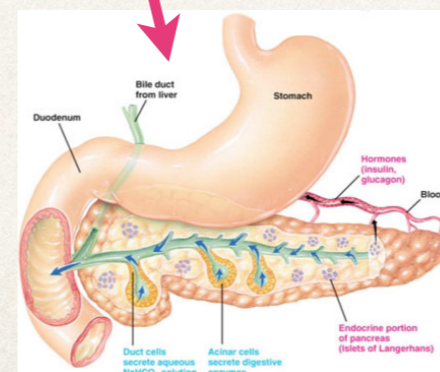
Endocrinologists

Health
Authorities

Disease and the need to work in interdisciplinary groups
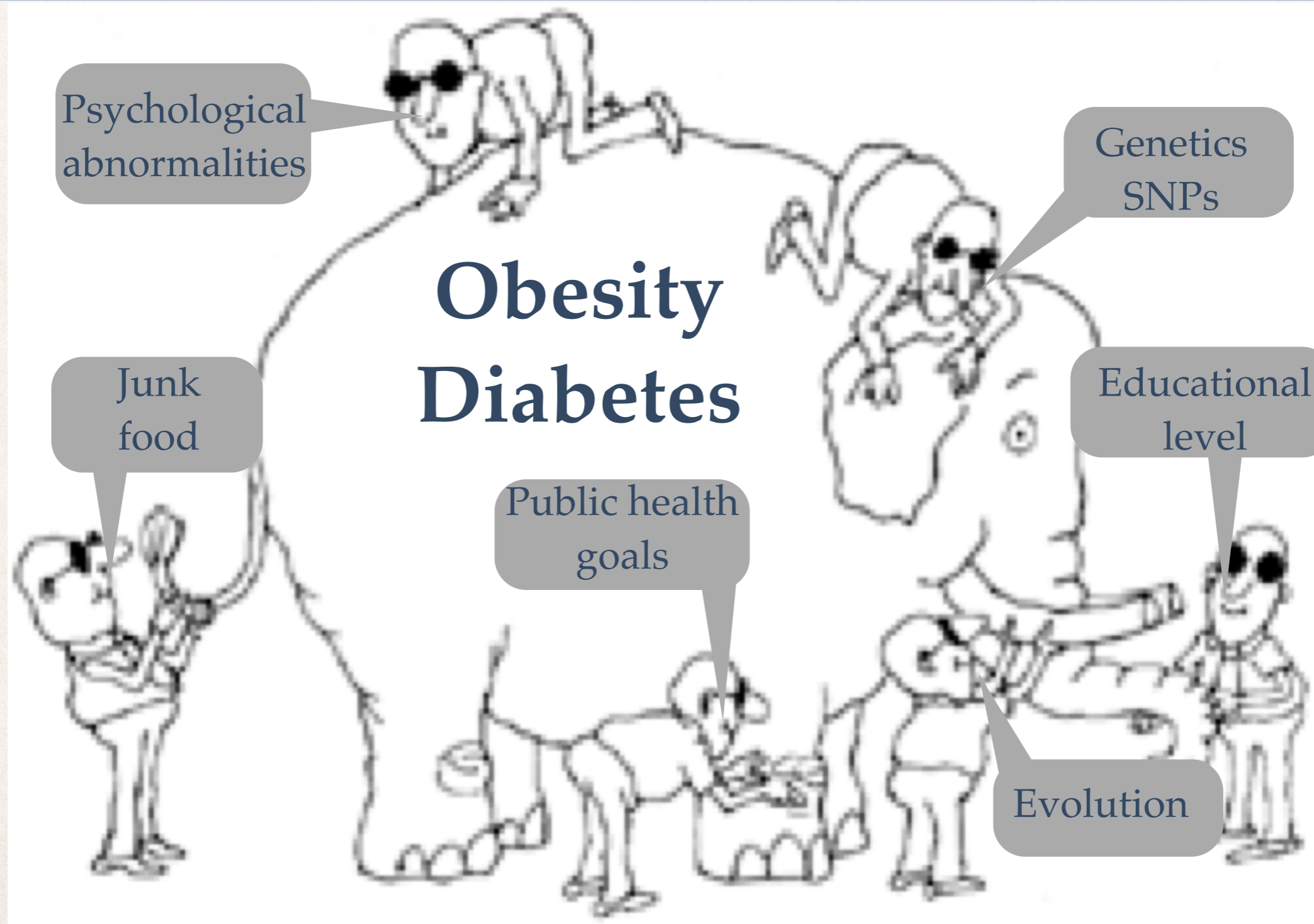
They are dynamical and adaptive

Scenario 1: Onset of obesity at 20
Continued obesity and onset of metabolic syndrome at 40
Onset of diabetes at 50
Onset of renal failure at 60
Death at 70

Scenario 2: Onset of obesity at 20
Continued obesity and onset of metabolic syndrome at 40
Onset of diabetes at 50
Adequate control and treatment of comorbilities at 60
Bad health and high cost at 70

Scenario 3: Overweight at 20
Obesity at 40
Onset metabolic syndrome at 50
Onset of diabetes at 60
Continued diabetes but no serious comorbilities at 60
Ill health and moderate cost at 70

Scenario 4: Overweight at 30
Obesity at 50
Onset of metabolic syndrome at 60
Onset of diabetes at 70 but relative health

We want to predict and understand "histories"

Adaptation, health and decision making

What is a decision?

A "decision"

Prediction

$$P(\textcolor{red}{C}\,|\,\textcolor{blue}{X(t)})$$

Probability of C given X

In the exact sciences, predictions

tend to be **algoríthmic**

In medicine and public health, predictions

tend to be **heurístic**

Curative Medicine
Less complex, less adaptative

Preventative Medicine
More complex, more adaptative

$\textcolor{blue}{X(t)} =$ **the information used to make the decisión (predict)**

How much information do you need or use to make a "good decision"?

What degree of multi-factoriality is there?

Preventative medicine requires a lot more data.

Where do we get that data…?   from the data revolution

Diagnoses

¿CUÁLES SON LOS SÍNTOMAS DE LA INFLUENZA? http://promocion.salud.gob.mx/dgps/interior1/influenza_informacion_sintomas.html

Fiebre de 38°C o más, tos y dolor de cabeza, acompañados de **uno o más** de los siguientes signos o síntomas:

- Escurrimiento nasal
- Enrojecimiento nasal
- Congestión nasal
- Dolor de articulaciones
- Dolor muscular
- Decaimiento (postración)
- Dolor al tragar
- Dolor de pecho
- Dolor de Estomago
- Diarrea

En menores de cinco años de edad, la irritabilidad es un signo que sustituye al dolor de cabeza. En mayores de 65 años no necesariamente fiebre.

Zika Symptoms https://www.cdc.gov/zika/symptoms/symptoms.html

Many people infected with Zika virus won't have symptoms or will only have mild symptoms. The most common symptoms of Zika are:

- Fever
- Rash
- Joint pain
- Conjunctivitis (red eyes)

Other symptoms include:

- Muscle pain
- Headache

Dengue Symptoms http://www.cdc.gov/dengue/symptoms/

The principal symptoms of dengue are:

- High fever <u>and</u> at least two of the following:
  - Severe headache
  - Severe eye pain (behind eyes)
  - Joint pain
  - Muscle and/or bone pain
  - Rash
  - Mild bleeding manifestation (e.g., nose or gum bleed, petechiae, or easy bruising)
  - Low white cell count

Generally, younger children and those with their first dengue infection have a milder illness than older children and adults.

Influenza Symptoms http://www.cdc.gov/flu/about/disease/complications.htm

Influenza (also known as the flu) is a contagious respiratory illness caused by flu viruses. It can cause mild to severe illness, and at times can lead to death. The flu is different from a cold. The flu usually comes on suddenly. People who have the flu often feel some or all of these symptoms:

- Fever* or feeling feverish/chills
- Cough
- Sore throat
- Runny or stuffy nose
- Muscle or body aches
- Headaches
- Fatigue (tiredness)

... y have vomiting and diarrhea, though this is ... children than adults. ... at not everyone with flu will have a fever.

Common Cold Symptoms

Symptoms mostly affect the nose. The most common cold symptoms are:

- Nasal congestion
- Runny nose
- Scratchy throat
- Sneezing

Adults and older children with colds generally have a low fever or no fever. Young children often run a fever around 100 to 102°F (37.7 to 38.8°C).

Depending on which virus caused your cold, you may also have:

- Cough
- Decreased appetite
- Headache
- Muscle aches
- Postnasal drip
- Sore throat

Most Americans will experience a diagnostic error at least once in th... these errors ar... year. Diagnost... the U.S. econo...

What is the age of the person?
What is ...
Whe...
Do they hav...
Where have they traveled to?

...eaths due to ...0 to 80,000 per ...fficiencies cost ...year...

https://w... Human-Cost-financial-impact-whitepaper.pdf
/download/? the ga... hitepaper.pdf

A combinatorially large set of possible symptoms!

And after a correct diagnosis...

Treatments and drugs

Treatment for gout usually involves medications. What medications you and your doctor choose will be based on your current health and your own preferences. Gout medications can be used to treat acute attacks and prevent future attacks as well as reduce your risk of complications from gout, such as the development of tophi from urate crystal deposits.

Medications to treat gout attacks

Drugs used to treat acute attacks and prevent future attacks include:

- **Nonsteroidal anti-inflammatory drugs (NSAIDs).** NSAIDs include over-the-counter options such as ibuprofen (Advil, Motrin IB, others) and naproxen sodium (Aleve, others), as well as more-powerful prescription NSAIDs such as indomethacin (Indocin) or celecoxib (Celebrex).
  Your doctor may prescribe a higher dose to stop an acute attack, followed by a lower daily dose to prevent future attacks.
  NSAIDs carry risks of stomach pain, bleeding and ulcers.
- **Colchicine.** Your doctor may recommend colchicine (Colcrys, Mitigare), a type of pain reliever that effectively reduces gout pain. The drug's effectiveness is offset in most cases, however, by intolerable side effects, such as nausea, vomiting and diarrhea.
  After an acute gout attack resolves, your doctor may prescribe a low daily dose of colchicine to prevent future attacks.
- **Corticosteroids.** Corticosteroid medications, such as the drug prednisone, may control gout inflammation and pain. Corticosteroids may be administered in pill form, or they can be injected into your joint.
  Corticosteroids are generally reserved for people who can't take either NSAIDs or colchicine. Side effects of corticosteroids may include mood changes, increased blood sugar levels and elevated blood pressure.

Medications to prevent gout complications

If you experience several gout attacks each year or if your gout attacks are less frequent but particularly painful, your doctor may recommend medication to reduce your risk of gout-related complications. Options include:

- **Medications that block uric acid production.** Drugs called xanthine oxidase inhibitors, including allopurinol (Aloprim, Lopurin, Zyloprim) and febuxostat (Uloric), limit the amount of uric acid your body makes. This may lower your blood's uric acid level and reduce your risk of gout.
  Side effects of allopurinol include a rash and low blood counts. Febuxostat side effects include rash, nausea and reduced liver function.
- **Medication that improves uric acid removal.** Probenecid (Probalan) improves your kidneys' ability to remove uric acid from your body. This may lower your uric acid levels and reduce your risk of gout, but the level of uric acid in your urine is increased. Side effects include a rash, stomach pain and kidney

A combinatorially large set of possible interventions!

And to cause a disease state...

C = patient suffers from diabetes mellitus type 2; X_1 = glucose level

X_2 = I have been obese 20 years, I don't exercise, I drink lots of coke, I urinate often, I am constantly thirsty, my parents suffer form diabetes,...

C = patient will be diabetic in 20 years;

X = X(sd)+X(se)+X(n)+X(ev)+X(g)+X(af)+X(hm)+X(i)+X(sp)+...

Socio-demographic factors
Age, Gender,...

Nutrition
How much you eat, what you eat,...

Genetic Factors
Rs7903146,...

Medical History
Obesity, metabolic syndrome,,...

Public health Factors
IMSS, ISSTE,...

Socio-economic factors
Educational achievement, income,...

Lifestyle
Smoke, drink, exercise...

Family History
Diabetic father, ...

Immunological Factors
HIV, stressed,...

A combinatorially large set of possible risk factors!

Deep data

Deep Data and the Data Revolution

A revolution in the generation of data

Human brain 10-100 Terrabytes

All the books in the world 30-50 Terrabytes
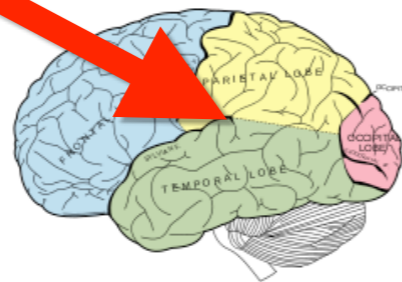
A revolution in data analysis
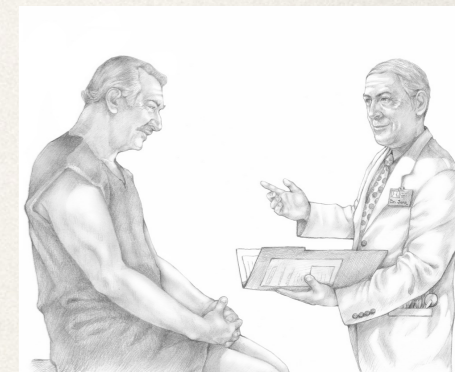
1 human genome
= 1GB (200)
CT image
= 10MB
MRI image
= 40MB
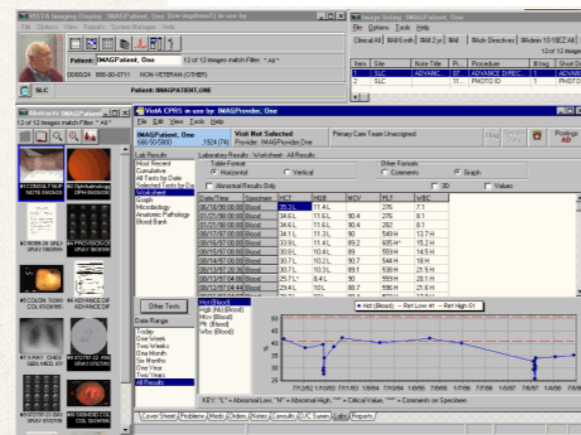
In electronic form
1 zettabyte

A revolution in data storage

Automated Data Analysis Using Excel

Brian D. Bissett

Chronic diseases

THE WORLD IS GETTING FATTER

THE SHAPE OF THINGS TO COME
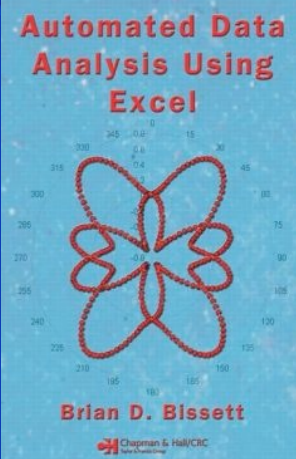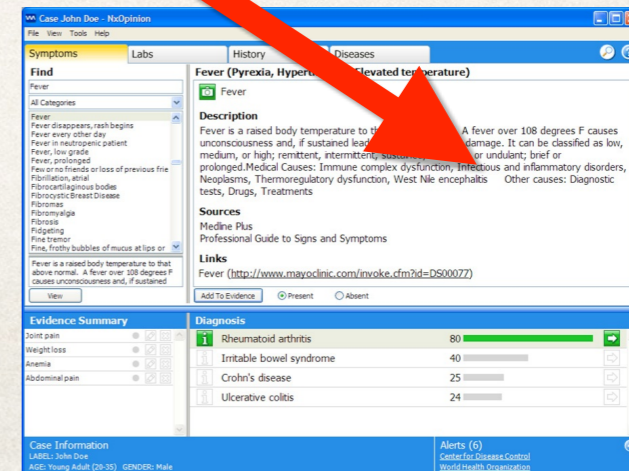
Obesity, type 2 diabetes, heart disease, strokes, cancer etc. are diseases associated with "lifestyle" and therefore are "preventible" (?)

Pharmaceutical Research, Vol. 25, No. 9, September 2008 (© 2008)
DOI: 10.1007/s11095-008-9661-9

Expert Review

Cancer is a Preventable Disease that Requires Major Lifestyle Changes
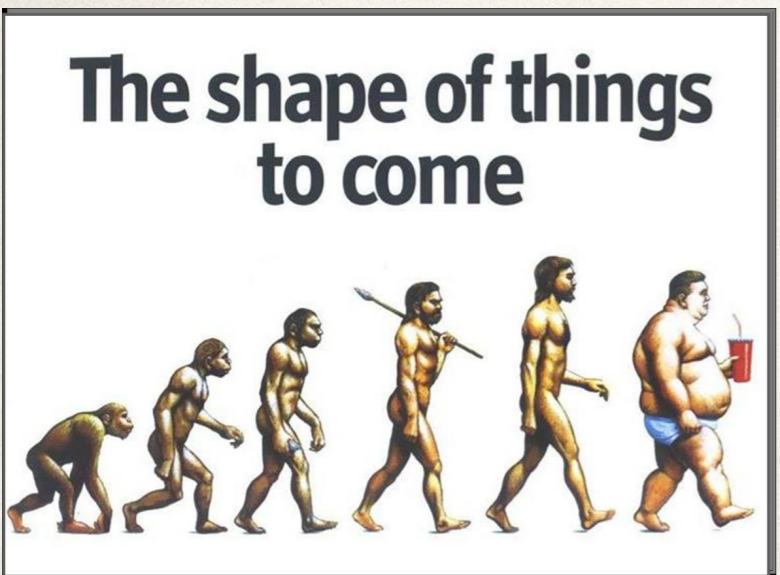
They are **behavioral** diseases,
i.e. diseases arising from **decision making**.

Human behavior is **complex**
and requires "**deep data**".

Environment
90-95%

Genes
5-10%

Chronic diseases - risk factors
You aren't what you eat you become what you eat

Gender	Male			Female			All Adults	
	#	% Males	% Age	#	% Females	% Age	#	% Adults
Age								
20-29	1170	20.66%	33.15%	2359	23.41%	66.85%	3529	22.42%
30-39	1511	26.69%	31.70%	3256	32.31%	68.30%	4767	30.29%
40-49	1250	22.08%	37.63%	2072	20.56%	62.37%	3322	21.11%
50-59	755	13.33%	41.26%	1075	10.67%	58.74%	1830	11.63%
60-69	545	9.63%	43.15%	718	7.13%	56.85%	1263	8.03%
70-80	431	7.61%	41.97%	596	5.92%	58.03%	1027	6.53%
Total	5662			10076			15738	

	Variable(s)	Unstd. B	Std. Error	t	f	R^2	Sig	Lower	Upper
BMI					103.539	0.035	0		
Men	Constant	20.06	0.493	40.666			0	19.093	21.027
	Age	0.321	0.022	14.347			0	0.277	0.364
	Age^2	-0.003		-14.326			0	-0.004	-0.003
BMI					290.452	0.055	0		
Women	Constant	17.399	0.46	37.821			0	16.497	18.301
	Age	0.504	0.021	23.794			0	0.463	0.546
	Age^2	-0.005		-22.767			0	-0.006	-0.005

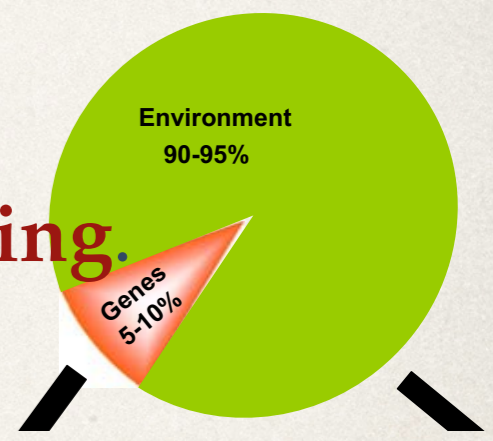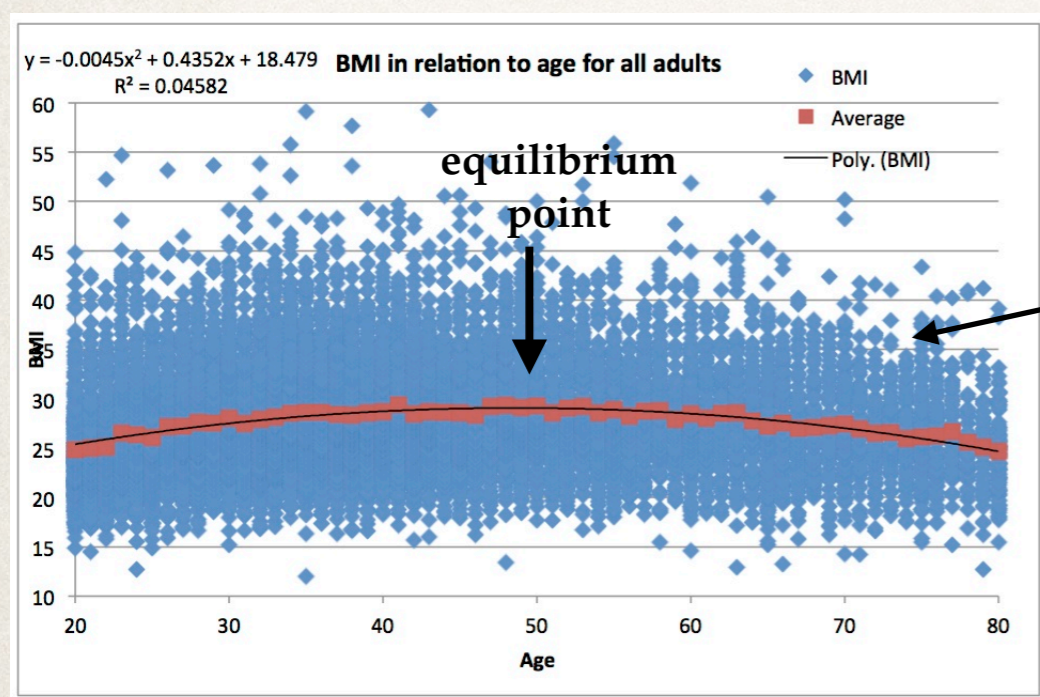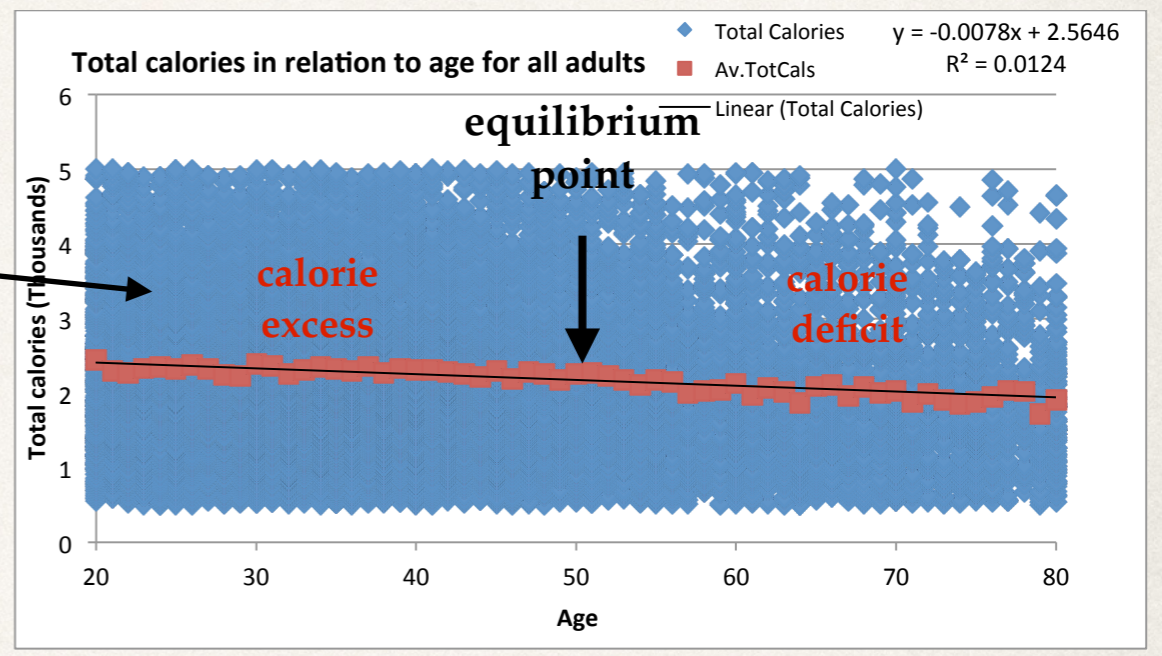|	Variable(s)	Unstd. B	Std. Error	t	f	R^2	Sig	Lower	Upper
Total Cals					197.52	0.012	0		
ALL	Constant	2.565	0.024	105.479			0	2.517	2.612
	Age	-0.008	0.001	-14.054			0	-0.009	-0.007
Total Cals					69.552	0.012	0		
Men	Constant	2.638	0.042	62.809			0	2.556	2.721
	Age	-0.008					0	-0.009	-0.006
Total Cals					144.087	0.014	0		
Women	Constant	2.544	0.03	85.301			0	2.485	2.602
	Age	-0.008	0.001	-12.004			0	-0.01	-0.007

	Variable(s)	Unstd. B	Std. Error	t	f	R^2	Sig	Lower	Upper
BMI Change					45.89	0.45	0		
ALL	Constant	0.513	0.08	6.428			0	0.353	0.674
	Age	-0.01	0.002	-6.774			0	-0.013	-0.007
BMI Change					14.05	0.401	0		
Men	Constant	0.387	0.11	3.581			0	0.168	0.607
	Age	-0.008	0.005	-3.748			0	-0.012	-0.004
BMI Change					41.91	0.428	0		
Women	Constant	0.595	0.097	6.163			0	0.402	0.789
	Age	-0.012	0.002	-6.473			0	-0.015	-0.008

BMI in relation to age for all adults
R² = 0.04582

Total calories in relation to age for all adults
R² = 0.0124

equilibrium point — Its not "noise" its multifactoriality — equilibrium point

calorie excess — calorie deficit

We get fatter then we get thinner

We eat less the older we get

BMI in relation to total calories for all adults
y = -0.1673x² + 0.9332x + 26.791
R² = 0.00188
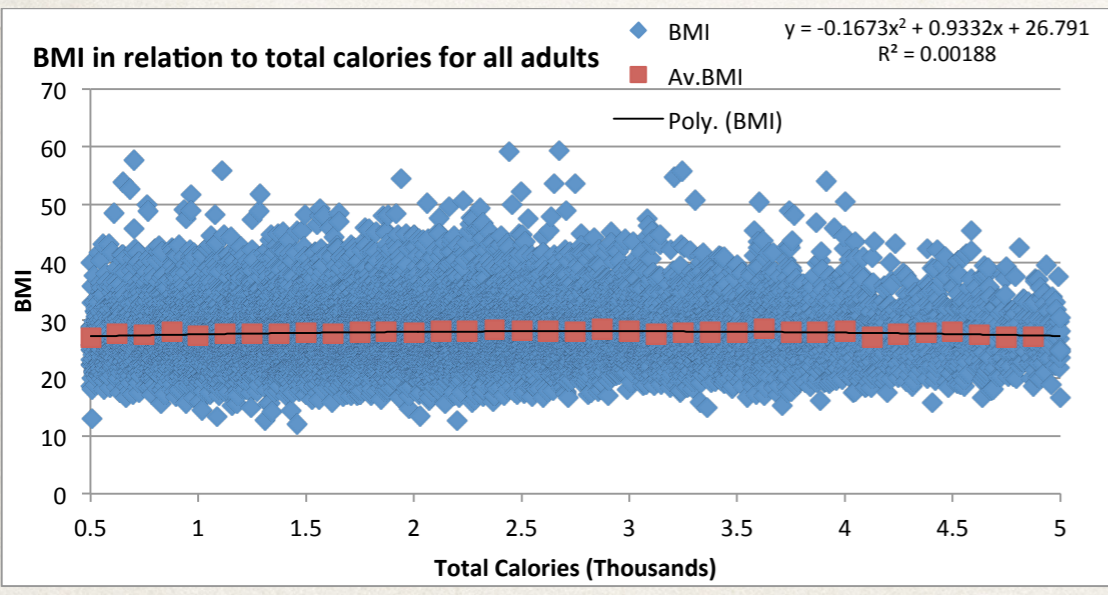
The obese eat as much as the thin

|	Variable(s)	Unstd. B	Std. Error	t	f	R^2	Sig	Lower	Upper
Moving Av.					29.236	0.343	0		
BMI Change	Constant	-1.954	0.362	-5.392			0	-2.68	-1.228
ALL	Total_Cals	0.904	0.167	5.407			0	0.569	1.239
Moving Av.					13.397	0.193	0.001		
BMI Change	Constant	-1.625	0.444	-3.656			0.001	-2.515	-0.734
Men	Total_Cals	0.724	0.198	3.66			0.001	0.328	1.121
Moving Av.					22.429	0.286	0		
BMI Change	Constant	-1.754	0.372	-4.711			0	-2.5	-1.008
Women	Total_Cals	0.833	0.176	4.736			0	0.481	1.185

Its the excess of calories that is the motor for obesity. The motor is more active at 20 and stops at 50 and then goes in reverse.

Chronic diseases - risk factors

Epidemiological data from ENSANUT 2006

The motor changes its fuel…
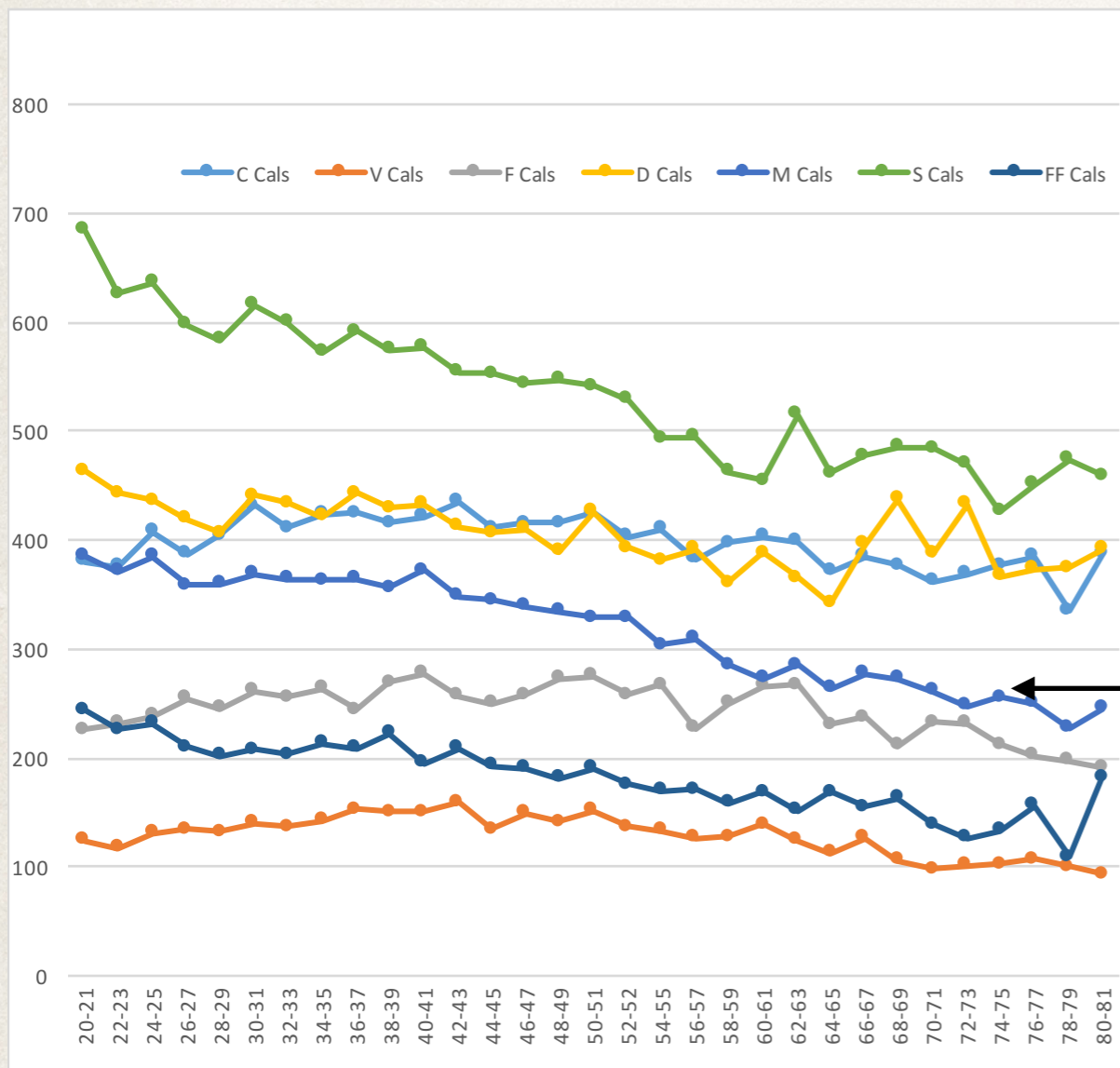
	Edad 20	Edad 50	Edad 80	Diff 50 20	Diff 80 20	Diff 80 50	Edad 20	Edad 50	Edad 80
S	650	540	460	16.92%	29.23%	14.81%	26.75%	23.38%	24.73%
FF	230	185	140	19.57%	39.13%	24.32%	9.47%	8.01%	7.53%
M	370	330	240	10.81%	35.14%	27.27%	15.23%	14.29%	12.90%
D	450	415	370	7.78%	17.78%	10.84%	18.52%	17.97%	19.89%
F	230	270	200	-17.39%	13.04%	25.93%	9.47%	11.69%	10.75%
V	120	150	90	-25.00%	25.00%	40.00%	4.94%	6.49%	4.84%
C	380	420	360	-10.53%	5.26%	14.29%	15.64%	18.18%	19.35%
	2430	2310	1860	4.94%	23.46%	19.48%			

The fuel mix at age 20 consists of 51.5% sugars, junk food and meat and 30% fruit, vegetables and cereals. At age 50 its 45.5% and 36.5%.
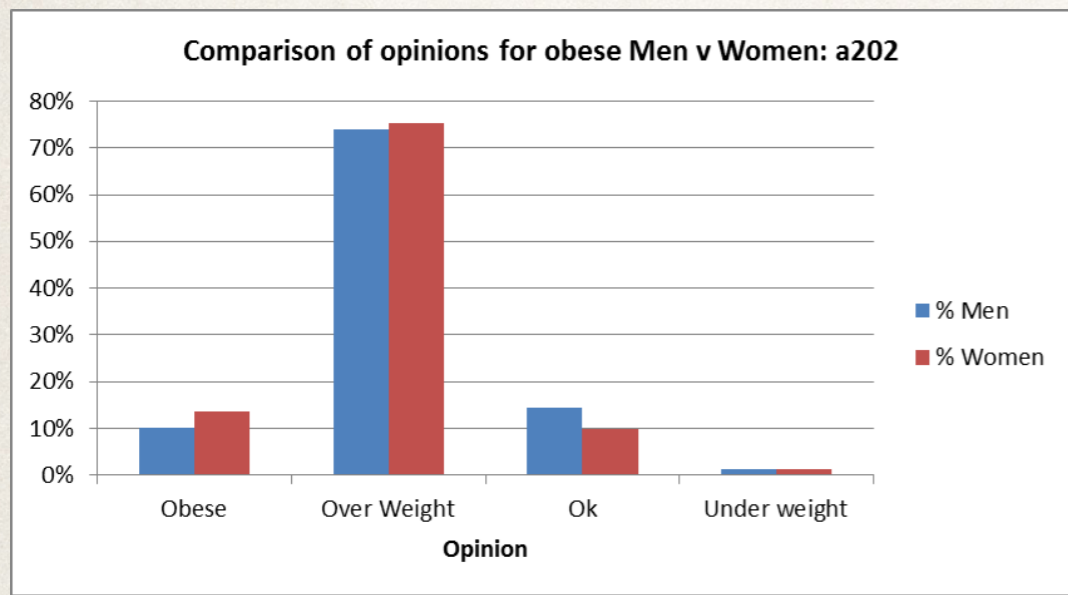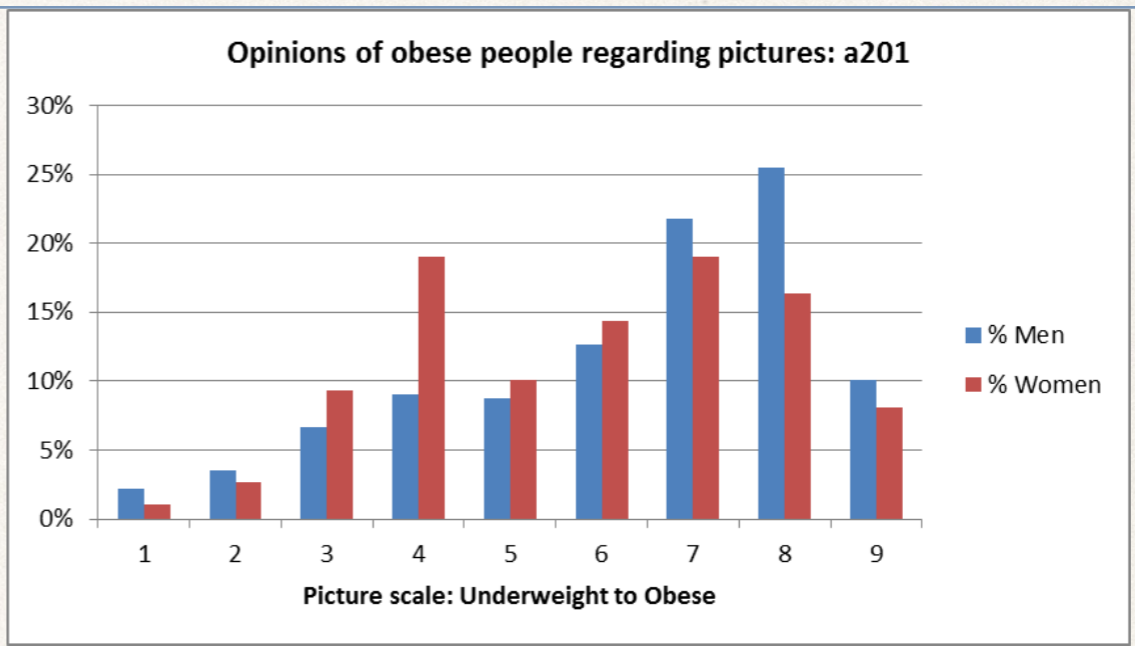
Accelerated red consumption in

Chronic diseases

People think they're less overweight/obese than they are. Symptom severity is under-estimated.

Fundamental question: Why do we lie to ourselves?

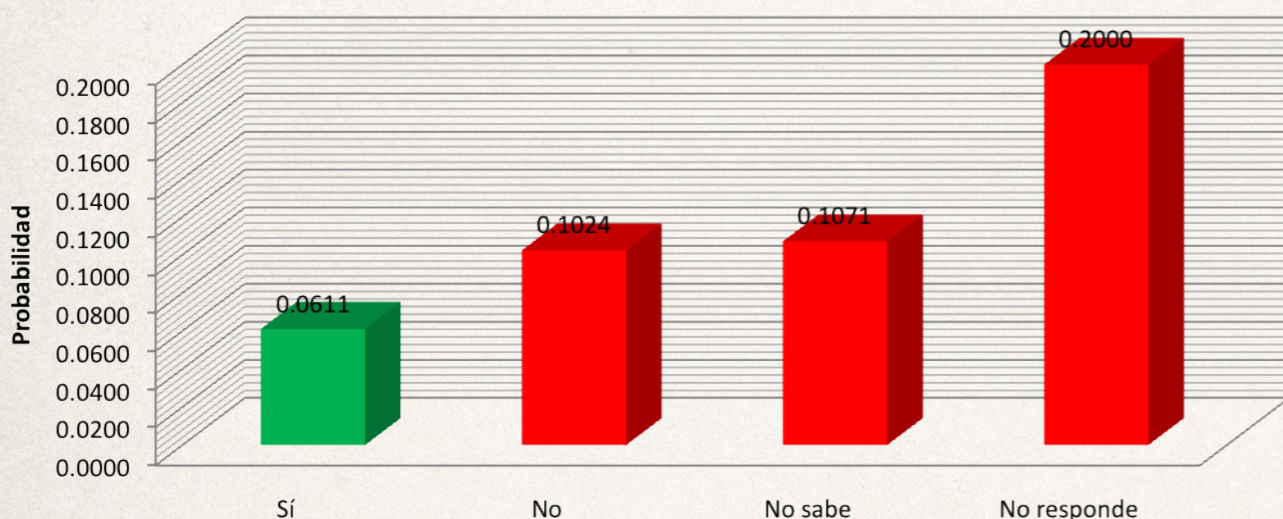# Chronic diseases - risk factors
Ignorance can kill
Epidemiological data from ENCOPREVENIMSS 2006

¿Sabe leer o escribir un recado?

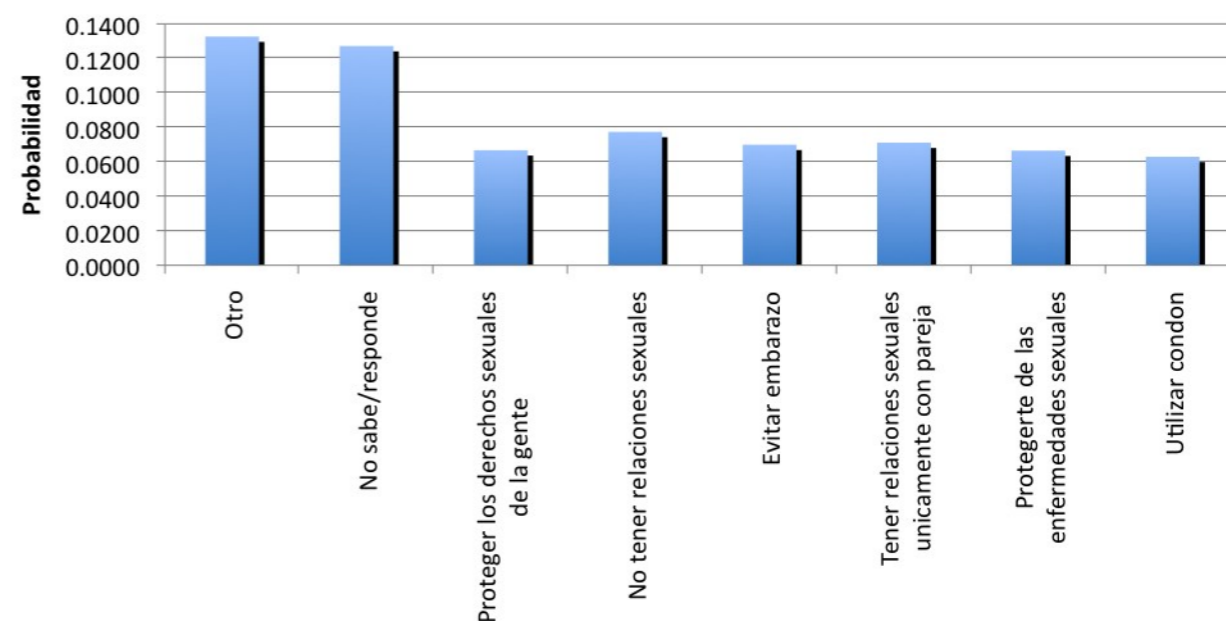For men 20-59 from PREVENIMSS 2006

Legend:
- Sí
- No
- No sabe
- No responde

Gráfica de probabilidad de diabetes versus qué piensas que significa el sexo protegido

Ignorance and especially about health issues is as important a risk factor as obesity

Beyond Epidemiological Data
UNAM Study 2014: Genetic analysis

772 SNPs considered
Subsets with obesity,
DM2, lipids, hepatic

Driver	Value	Epsilon	P(C/X)	P(C)	N(X/C)	N(X)	N(C)	NTotal
rs2943641_A	2	2.9391	0.6000	0.2169	6	10	123	567
rs2972146_C	2	2.9391	0.6000	0.2169	6	10	123	567
rs2943650_G	2	2.9391	0.6000	0.2169	6	10	123	567
rs12629908_A	2	2.6981	0.3116	0.2169	43	138	123	567
rs870347_C	2	2.2200	0.2914	0.2169	44	151	123	567
rs1407434_G	0	2.1617	0.2841	0.2169	50	176	123	567
rs972283_A	2	2.1543	0.3085	0.2169	29	94	123	567
rs10496971_C	2	1.9688	0.3011	0.2169	28	93	123	567
rs2241766_C	1	1.9472	0.2741	0.2169	54	197	123	567
rs10885122_A	2	1.9426	0.5000	0.2169	4	8	123	567
rs2986742_G	2	1.9121	0.4545	0.2169	5	11	123	567
rs1799884_A	2	-2.0385	0.0000	0.2169	0	15	123	567
rs3943253_A	2	-2.0502	0.1364	0.2169	15	110	123	567
rs4607517_A	2	-2.1053	0.0000	0.2169	0	16	123	567
rs4880436_A	2	-2.1388	0.0870	0.2169	4	46	123	567
rs174537_C	2	-2.1927	0.0851	0.2169	4	47	123	567
rs174546_G	2	-2.1927	0.0851	0.2169	4	47	123	567
rs174550_A	2	-2.1927	0.0851	0.2169	4	47	123	567
rs972283_A	0	-2.3181	0.1521	0.2169	33	217	123	567
rs2073821_A	2	-2.3502	0.1170	0.2169	11	94	123	567
rs1513181_G	2	-2.3605	0.1250	0.2169	14	112	123	567
rs2237895_A	2	-2.3836	0.1308	0.2169	17	130	123	567
rs7803075_G	2	-2.4635	0.0847	0.2169	5	59	123	567
rs896854_A	0	-2.5528	0.1398	0.2169	26	186	123	567
rs7809589_C	2	-2.5964	0.1231	0.2169	16	130	123	567
rs1111875_A	0	-3.2065	0.1211	0.2169	23	190	123	567

obesity (score = 0.904, predictive but scarce)

obesity (score = 0.105, not so predictive but common)
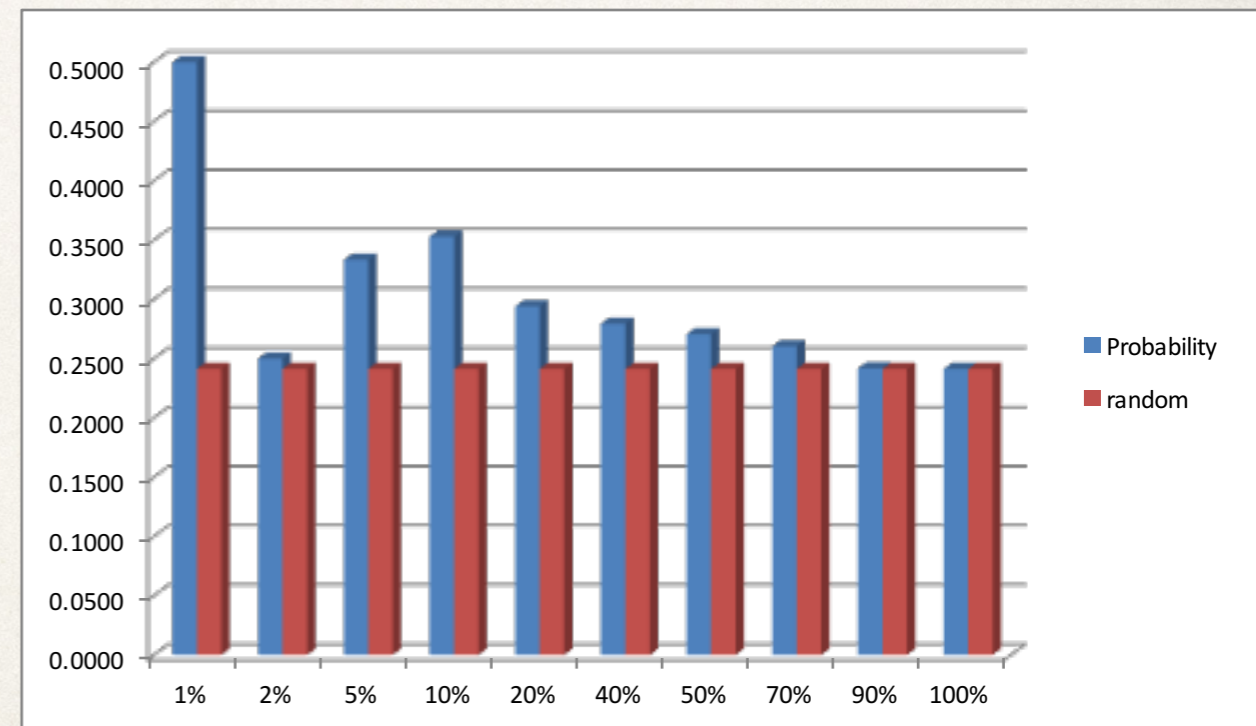
Doesn't give a good model on its own

Putting it all together…

Nutrition	
Specificity (TNR)	83.40%
1 – Specificity (SPC)	16.60%
Sensitivity (FPR)	29.69%
Accuracy (ACC)	72.76%
AUC ROC	0.63
Lifestyle	
Specificity (TNR)	84.17%
1 – Specificity (SPC)	15.83%
Sensitivity (FPR)	31.25%
Accuracy (ACC)	73.68%
AUC ROC	0.70
Lifestyle and Nutrition	
Specificity (TNR)	78.38%
1 – Specificity (SPC)	21.62%
Sensitivity (FPR)	46.88%
Accuracy (ACC)	72.14%
AUC ROC	0.71
Lifestyle and Nutrition and Personal and Family History	
Specificity (TNR)	81.08%
1 – Specificity (SPC)	18.92%
Sensitivity (FPR)	51.56%
Accuracy (ACC)	75.23%
AUC ROC	0.76

3,524 variables

Genetic, epidemiological, physiological,…

Epidemiological: Personal (81), Anthropometry (49),

Personal history (130), Family History (548),

Self-health evaluation (226), Nutrition (220),

Lifestyle (390), Health knowledge (293).

There are predictive variables in all categories.
The more variables you put together the more predictability you get.

PROOF that it is VERY multifactorial.

Chronic diseases

To understand the physiology and genetics of such diseases is important. However, these diseases are predominantly "behavioural" diseases, associated with "bad" decisions.
Why do we make "bad" decisions? What behaviour is plastic?

Establishing and untangling causal chains is very difficult. Causality must be respected…e.g.,
overeating —> overweight —> inflammation…
                    Not
inflammation —> overeating…

The Challenges of Modelling Human Health

Human health, and any disease, is a CAS. To model such systems is on the very forefront of science. We don't do it well.

* CAS are extraordinarily multifactorial, requiring big data across multiple scales: genetics, epigenetics, physiology, psychology, neuroscience, epidemiology, sociology,… We don't have it.
* CAS require appropriate frameworks for generating data and sharing data. We don't have them.
* CAS require interdisciplinary teams to analyse and model the data. We don't have them.
* We need a more data science centered medicine and health science, requiring a shift in emphasis from curative medicine to preventative medicine

We have the technology to do the data "plumbing" but not the data semantics.
We have a lot of interesting work to do over the coming months, years, decades,…

You're all invited!