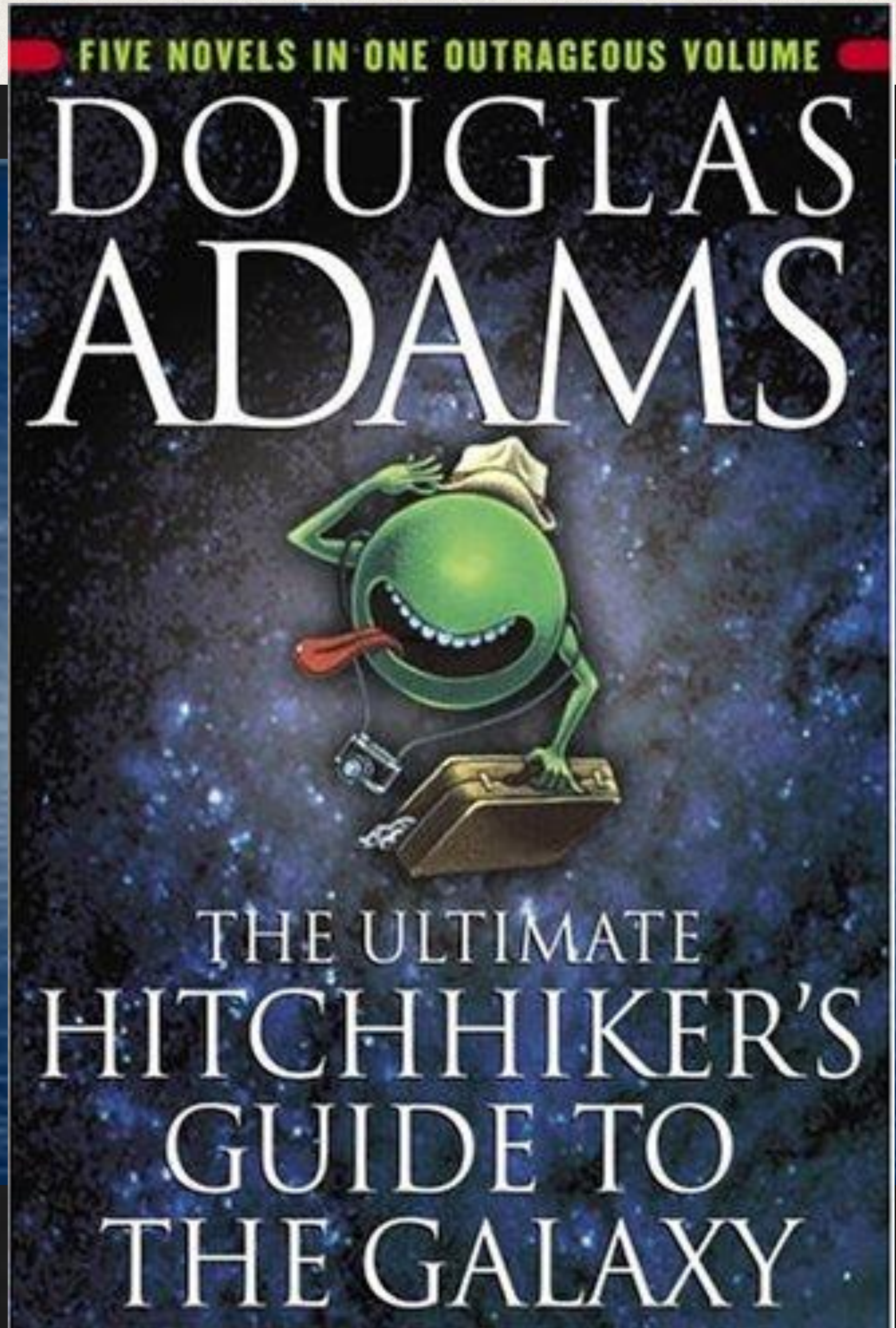
Nature versus Nurture:

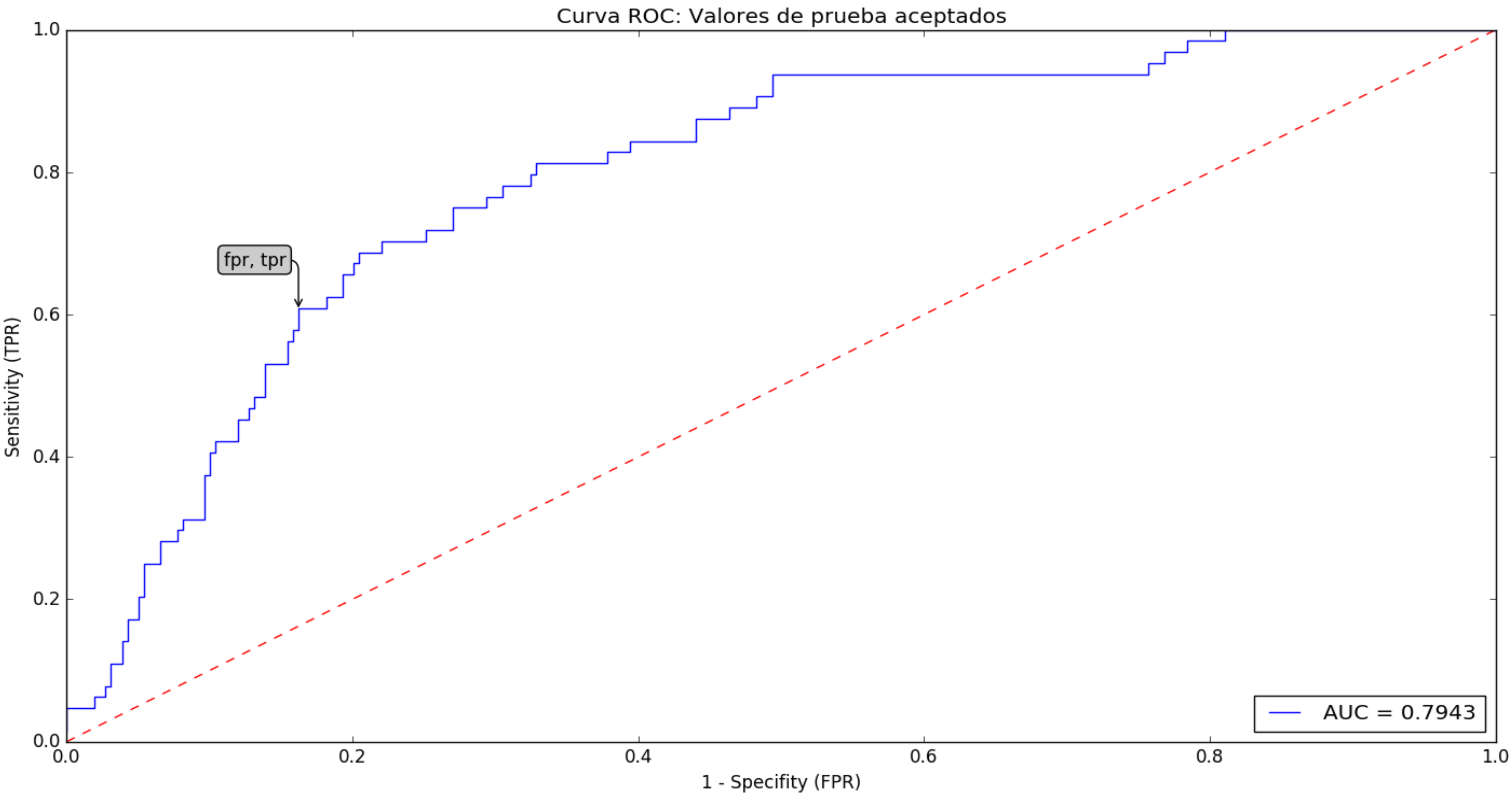
Confronting the Complexity of Obesity with Big Data

Chris Stephens, C3 y ICN, UNAM

INMEGEN Reunion de trabajo

19th June 2017


FIVE NOVELS IN ONE OUTRAGEOUS VOLUME
DOUGLAS ADAMS
THE ULTIMATE HITCHHIKER'S GUIDE TO THE GALAXY



= 42

Predictive model for obesity…

Results from predictive models * based on data from a study of 1,076 non-academics and academics from the UNAM:

2,524 variables - Genetic, epidemiological, physiological,…

- **Epidemiological**: Personal (81), Personal history (130), Family History (548), Self-health evaluation (226), Nutrition (220), Lifestyle (390), Health knowledge (293)
- **Genetic** (772)
- **Anthropometric and physiological** (49)

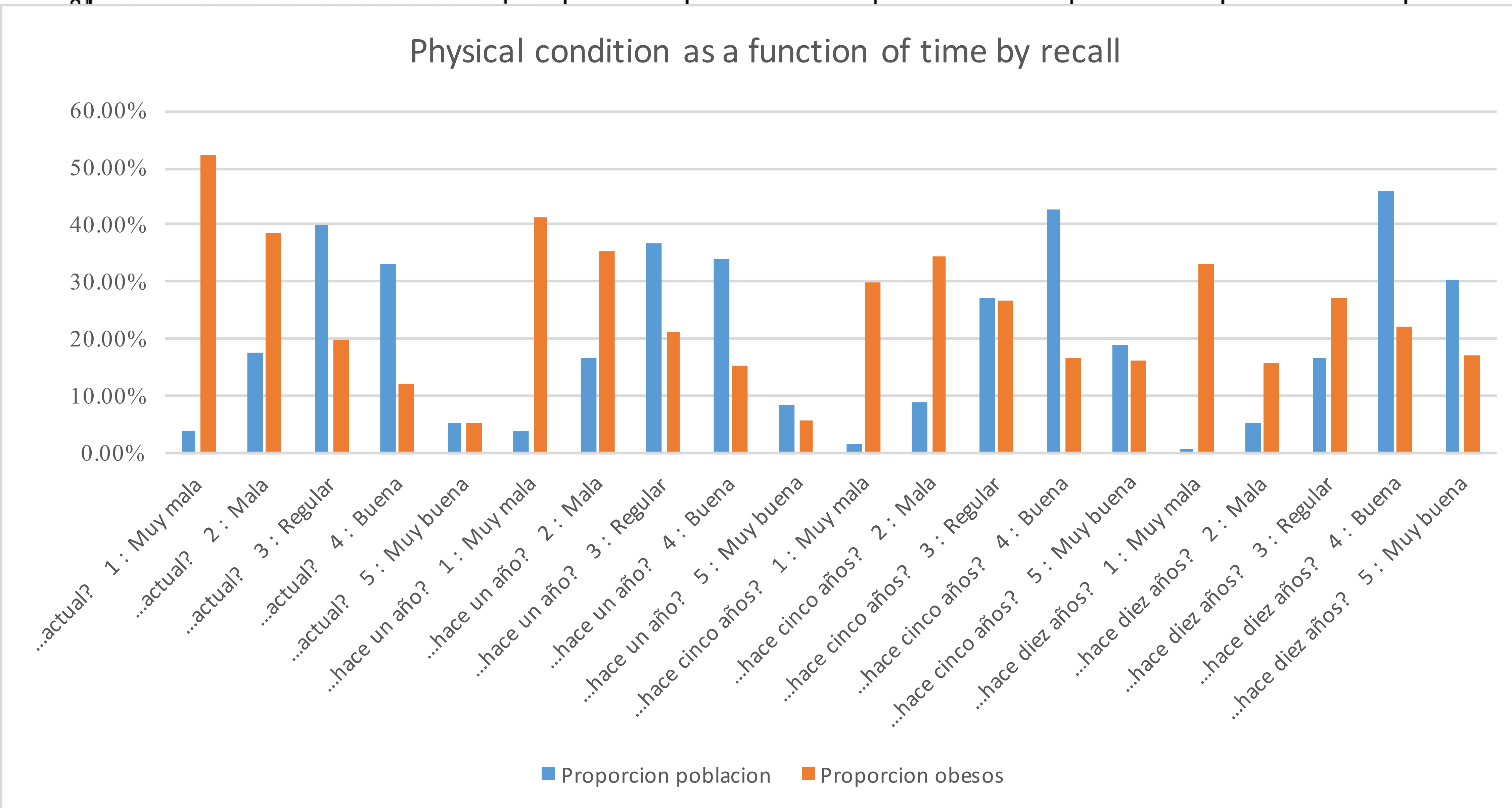
* Models are classification models of Naive Bayes type. Model performance is based on a 70/30 training/test split

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

Variable	Valor	Epsilon	Nx	Nxc	N	Nc	Pc	Pxc	Descripción
Aestatura	1	4.801461	91	38	1076	228	0.2119	0.4176	Estatura que estima tener el encuestado < 1.5 : 1
Aestatura	2	-0.92449	399	77	1076	228	0.2119	0.193	Estatura que estima tener el encuestado [1.5, 1.6) : 2
Aestatura	3	-1.09413	366	69	1076	228	0.2119	0.1885	Estatura que estima tener el encuestado [1.6, 1.7) : 3
Aestatura	4	0.143796	185	40	1076	228	0.2119	0.2162	Estatura que estima tener el encuestado [1.7, 1.8) : 4
Aestatura	5	-1.63546	32	3	1076	228	0.2119	0.0938	Estatura que estima tener el encuestado [1.8, 1.9) : 5
Aestatura	6	-0.7333	2	0	1076	228	0.2119	0	Estatura que estima tener el encuestado [1.9, 2.0) : 6
Aestatura	7	1.928548	1	1	1076	228	0.2119	1	Estatura que estima tener el encuestado > 2.0) : 7
Apeso	1	-3.77209	62	1	1076	228	0.2119	0.0161	Peso que estima tener el encuestado <= 50 : 1
Apeso	2	-4.05811	79	2	1076	228	0.2119	0.0253	Peso que estima tener el encuestado (50, 55) : 2
Apeso	3	-5.74441	132	1	1076	228	0.2119	0.0076	Peso que estima tener el encuestado [55, 60) : 3
Apeso	4	-5.1211	172	9	1076	228	0.2119	0.0523	Peso que estima tener el encuestado [60, 65) : 4
Apeso	5	-1.86651	142	21	1076	228	0.2119	0.1479	Peso que estima tener el encuestado [65, 70) : 5
Apeso	6	-2.34173	138	18	1076	228	0.2119	0.1304	Peso que estima tener el encuestado [70, 75) : 6
Apeso	7	0.84116	106	26	1076	228	0.2119	0.2453	Peso que estima tener el encuestado [75, 80) : 7
Apeso	8	8.123762	143	70	1076	228	0.2119	0.4895	Peso que estima tener el encuestado [80, 90) : 8
Apeso	9	14.14686	102	80	1076	228	0.2119	0.7843	Peso que estima tener el encuestado >= 90 : 9
condi_act	1	5.045429	44	23	1076	228	0.2119	0.5227	¿Cómo consideras tu condición física actual? 1 : Muy mala
condi_act	2	5.865344	189	73	1076	228	0.2119	0.3862	¿Cómo consideras tu condición física actual? 2 : Mala
condi_act	3	-0.57931	429	86	1076	228	0.2119	0.2005	¿Cómo consideras tu condición física actual? 3 : Regular
condi_act	4	-4.18504	355	43	1076	228	0.2119	0.1211	¿Cómo consideras tu condición física actual? 4 : Buena
condi_act	5	-2.94241	57	3	1076	228	0.2119	0.0526	¿Cómo consideras tu condición física actual? 5 : Muy buena
condi_act	8	-0.7333	2	0	1076	228	0.2119	0	¿Cómo consideras tu condición física actual? 8 : No quiero re
condi1	1	3.176688	41	17	1076	228	0.2119	0.4146	¿Cómo consideras tu condición física hace un año? 1 : Muy m
condi1	2	4.71648	180	64	1076	228	0.2119	0.3556	¿Cómo consideras tu condición física hace un año? 2 : Mala
condi1	3	0.133941	396	85	1076	228	0.2119	0.2146	¿Cómo consideras tu condición física hace un año? 3 : Regula
condi1	4	-2.65254	367	57	1076	228	0.2119	0.1553	¿Cómo consideras tu condición física hace un año? 4 : Buena

Epsilon	# participantes	# obesos	Proporcion obesos	Puesto
-2.81	23			
1.23	74			
0.19	54			
-1.64	10			
-2.38	52			
-3.58	81			
-2.05				
2.03				
-0.53				
0.51				
2.41				
2.06				
3.53				
0.30				
2.85				

¿Cuántas calorías hay en un litro de agua?	Epsilon	# participantes	Proporcion poblacion	# obesos	Probabilidad obesidad	Proporcion obesos
No sabe el número de calorías que						



47

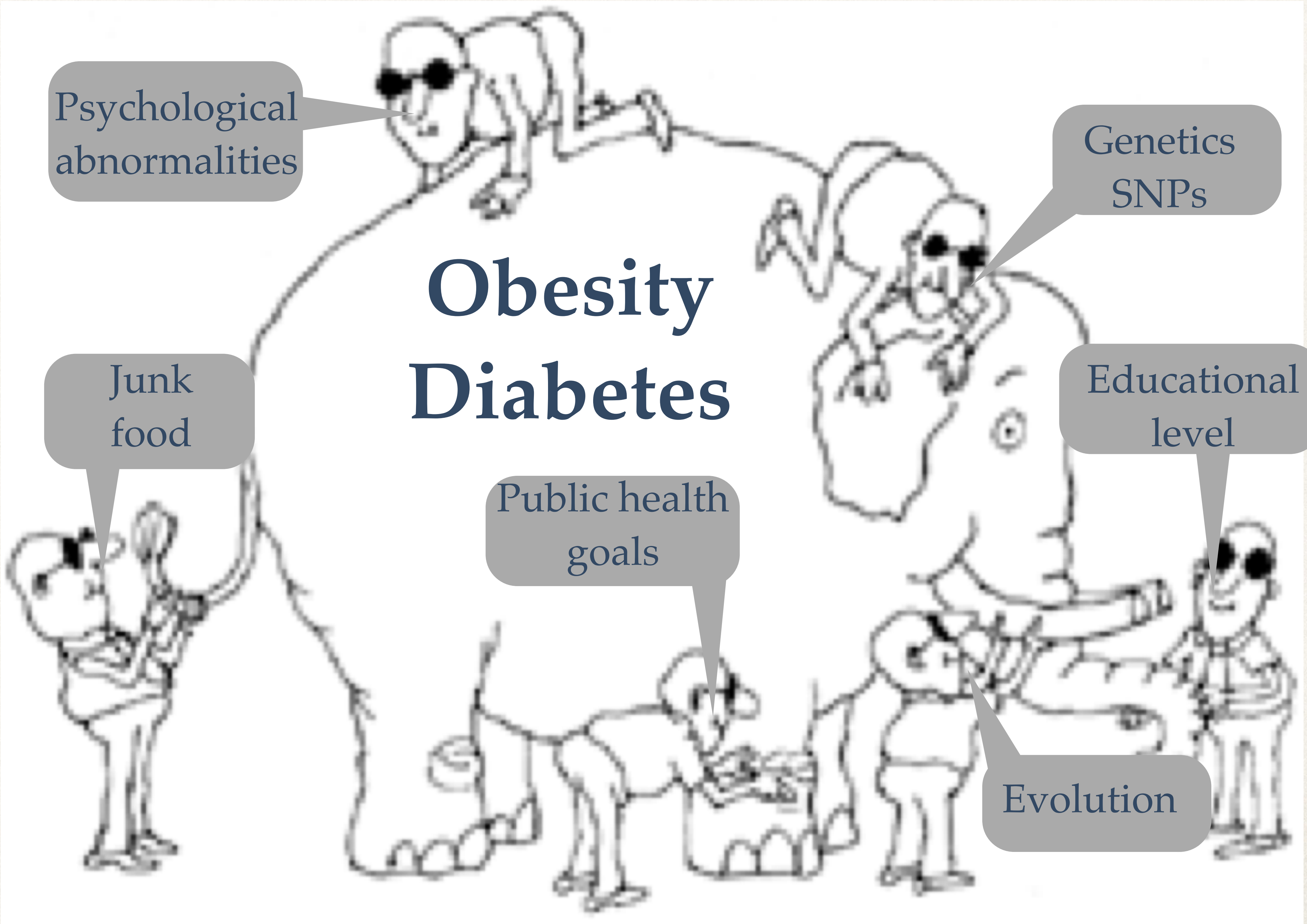
...cion ...os
31.18%
31.08%
19.38%

		Proporcion obesos
		17.81%
		22.36%
190	45	23.68%

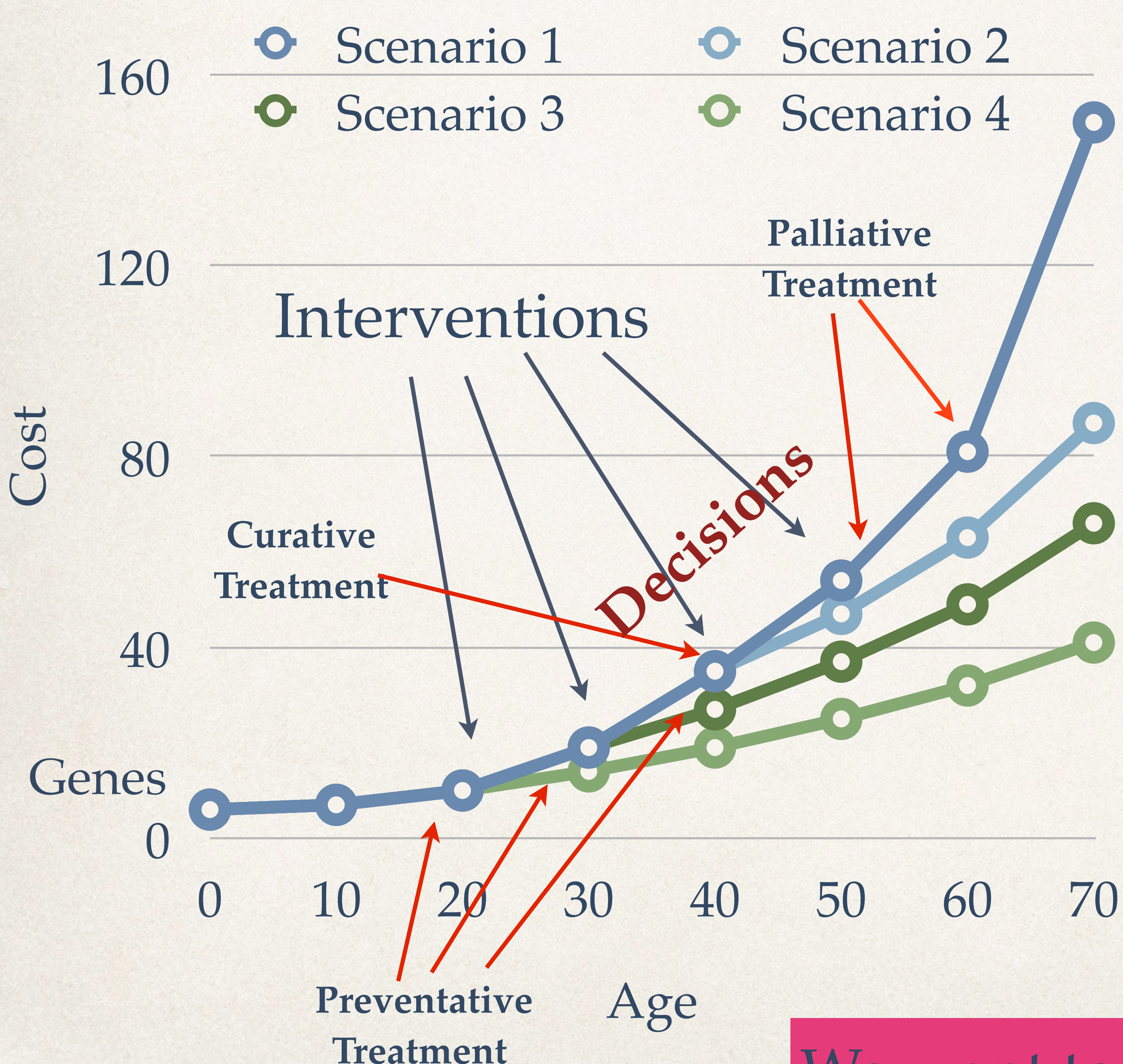
¿Cómo es tu s... actualmente?	Epsilon	# participantes	# obesos	obesos
1 : Muy mala	1.15	15	5	33.33%
2 : Mala				
3 : Regular				
4 : Buena				
5 : Muy buena				

¿Qué acciones le gustaría tomar resp	Epsilon	# participantes	Proporcion poblacion	# obesos	Probabilidad obesidad	Proporcion obesos
Bajar de peso	5.25	771	71.65%	223	28.92%	97.81%
Esta contento con su peso	-7.54	239	22.21%	3	1.26%	1.32%
Subir de peso	-3.50	63	5.86%	2	3.17%	0.88%
No sabe	-0.90	3	0.28%	0	0.00%	0.00%

Disease and the need to work in interdisciplinary groups



Disease is 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

ystem

s”

at b idual

d collective r

How plastic are those decisiones?

Your Prediction/Decision Heuristic/Algorithm depends on…

Who “you” are, e.g. your genes

What and how you “think”, e.g. your education

“What” and how you “feel”, e.g. your personality

Your prediction/decision heuristic/algorithm then determines your behaviour - what you do

What is a decision?

A "decision"

$$P(C \mid X(t))$$

Probability of C given X

Prediction

In the exact sciences, predictions tend to be **algoríthmic**

Curative Medicine
Less complex, less adaptative

Preventative Medicine
More complex, more adaptative

In medicine and public health, predictions tend to be **heurístic**

$X(t)$ = **the information used to make the decisión (predict)**

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

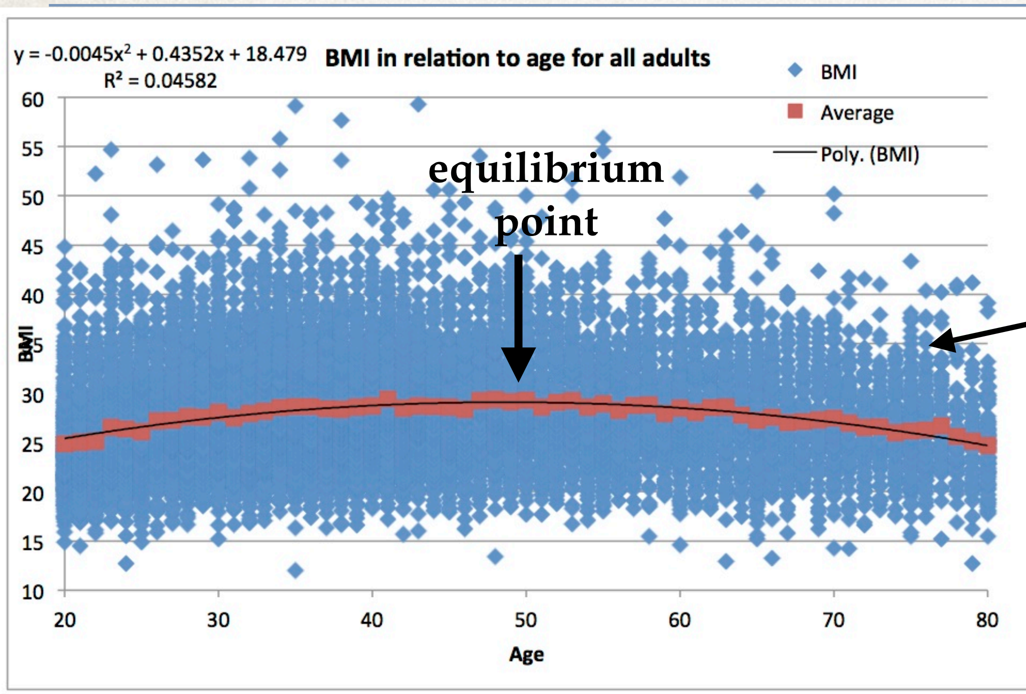
What degree of multi-factoriality is there?

Preventative medicine requires a lot more data.

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

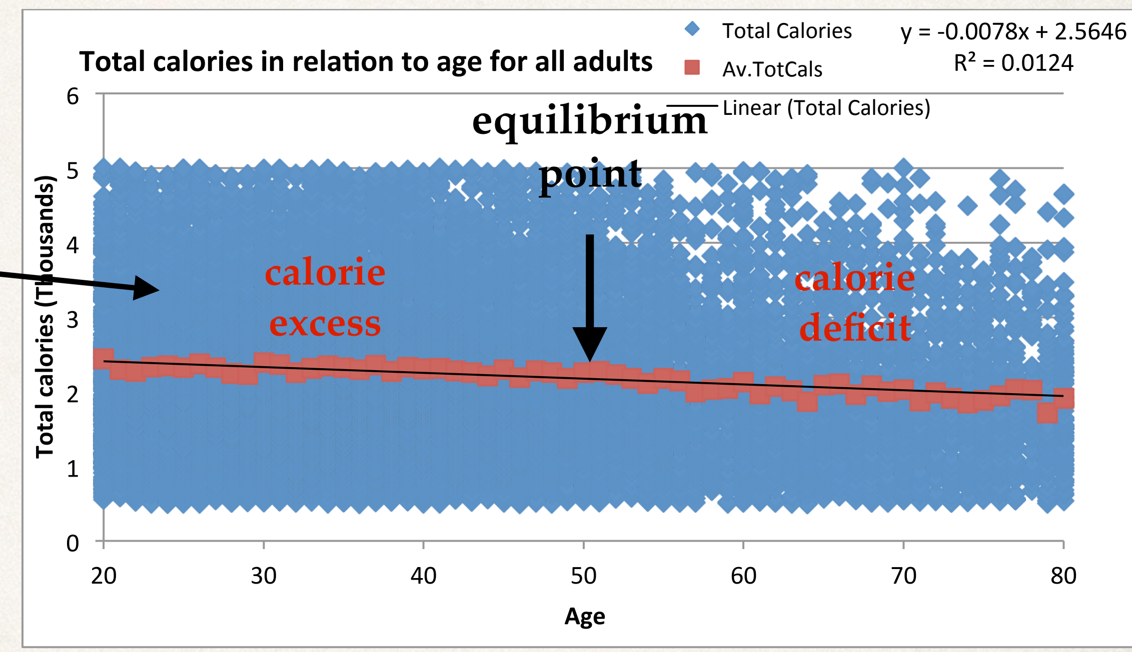
Obesity - risk factors
What you do
You aren't what you eat you become what you eat

Epidemiological data from ENSANUT 2006

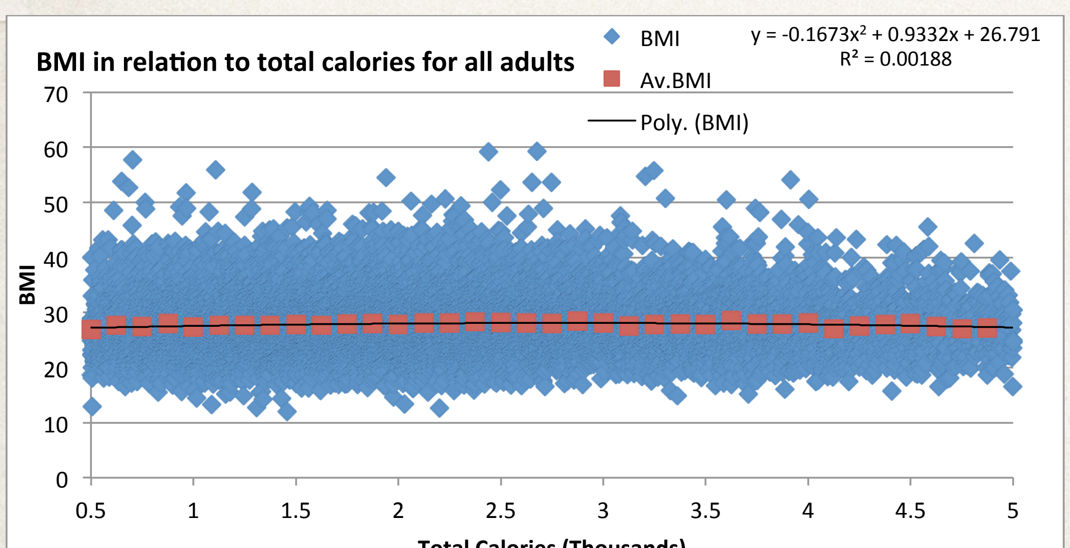


We get fatter then we get thinner

Its not "noise" its multifactoriality



We eat less the older we get



The obese eat as much as the thin

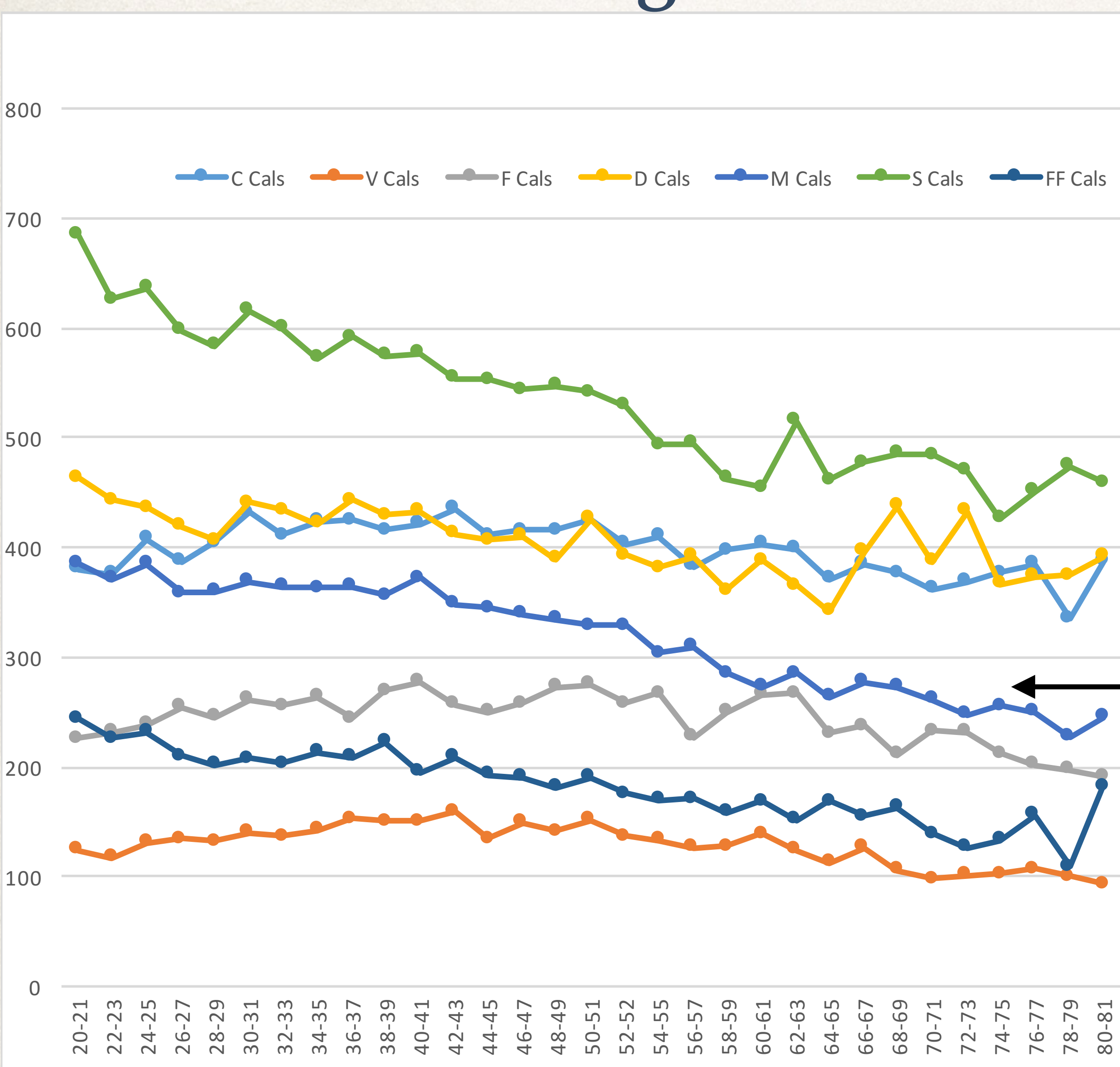
	Variable(s)	Unstd. B	Std. Error	t	f	R^2	Sig	Lower	Upper
Moving Av.					29.236	0.343	0		
BMI Change	Constant	-1.954	0.362	-5.392			0	-2.68	-1.228
ALL	Total_Cals	0.904	0.167	5.407			0	0.569	1.239
	Variable(s)	Unstd. B	Std. Error	t	f	R^2	Sig	Lower	Upper
Moving Av.					13.397	0.193	0.001		
BMI Change	Constant	-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
	Variable(s)	Unstd. B	Std. Error	t	f	R^2	Sig	Lower	Upper
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.

Obesity - risk factors
What you do

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 reduction in meat consumption in the aged

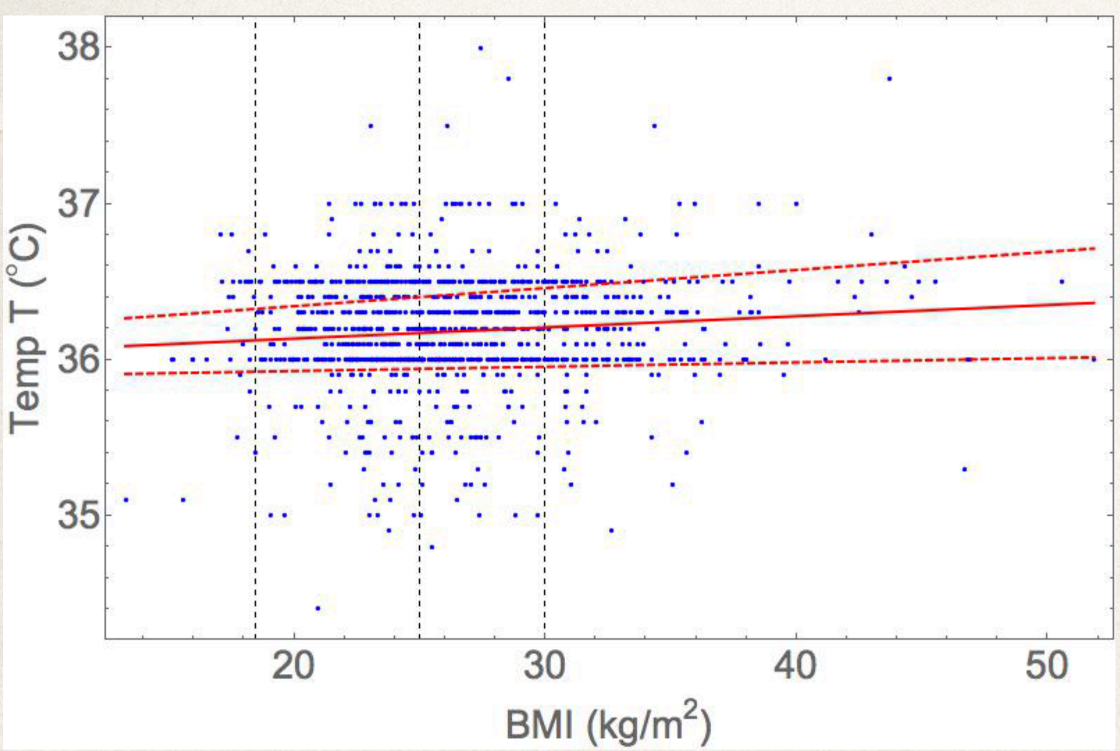
Do you become what you eat?

The data shows an overconsumption of 200-300 Cals/day at age 20-30. 8 Cal/day is enough (naively through the famous/infamous 3500 cal rule) to generate the observed increase in BMI. **Where do the other calories go?**

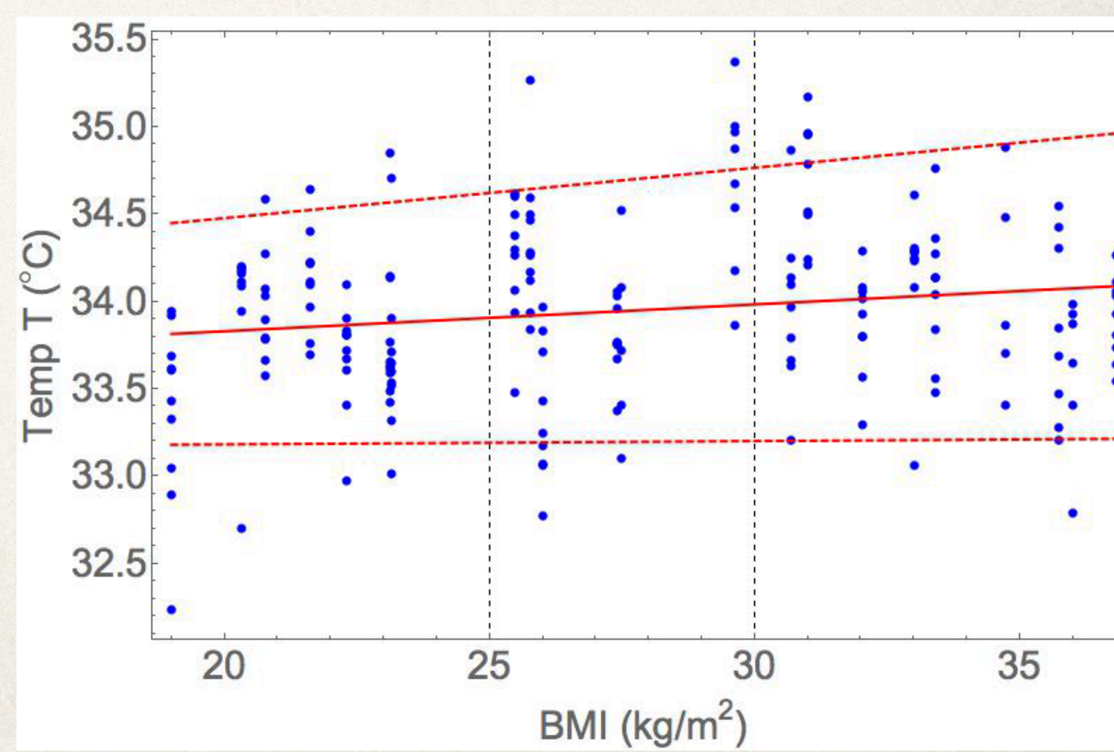
Why aren't we even fatter?

	Study 1		Study 2	
	points	**deciles**	**7-day mean**	**1-day mean**
slope	0.0072	0.0067	0.0093	0.015
intercept	35.99	36.00	33.69	33.524
CIslope	0.0028	0.0024	-0.019	0.0019
	0.012	0.011	0.038	0.029
CIintercept	35.88	35.89	32.88	33.15
	36.11	36.12	34.51	33.90
tslope	3.18	3.56	0.68	2.25
tintercept	590.34	708.93	86.9	174.92
F	10.15	12.64	0.46	5.06
p	0.0015 (*)	0.0074 (*)	0.50	0.026 (*)
R2	0.0094	0.61	0.022	0.027

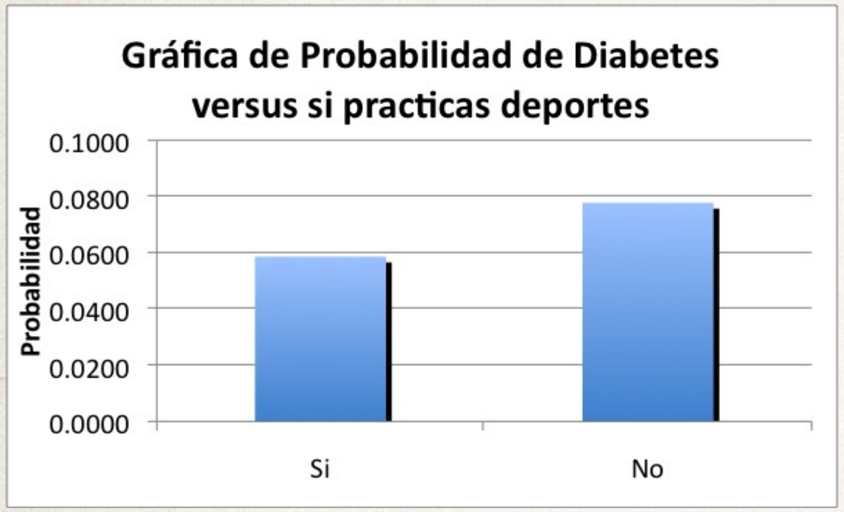
Relation between temperature and BMI

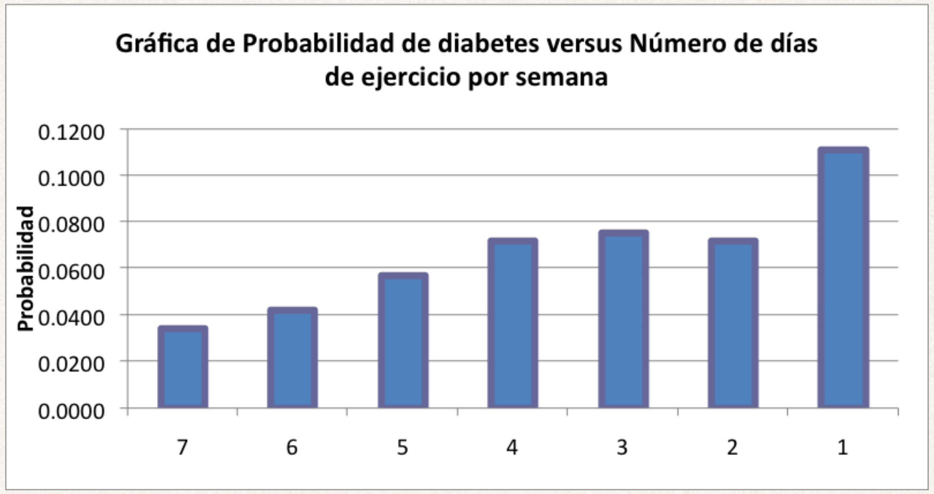


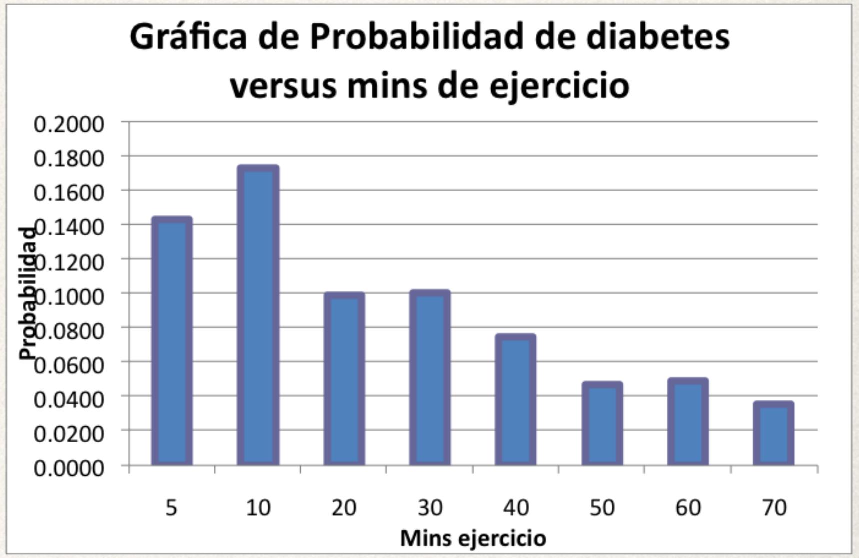
R. Fossion
DH17



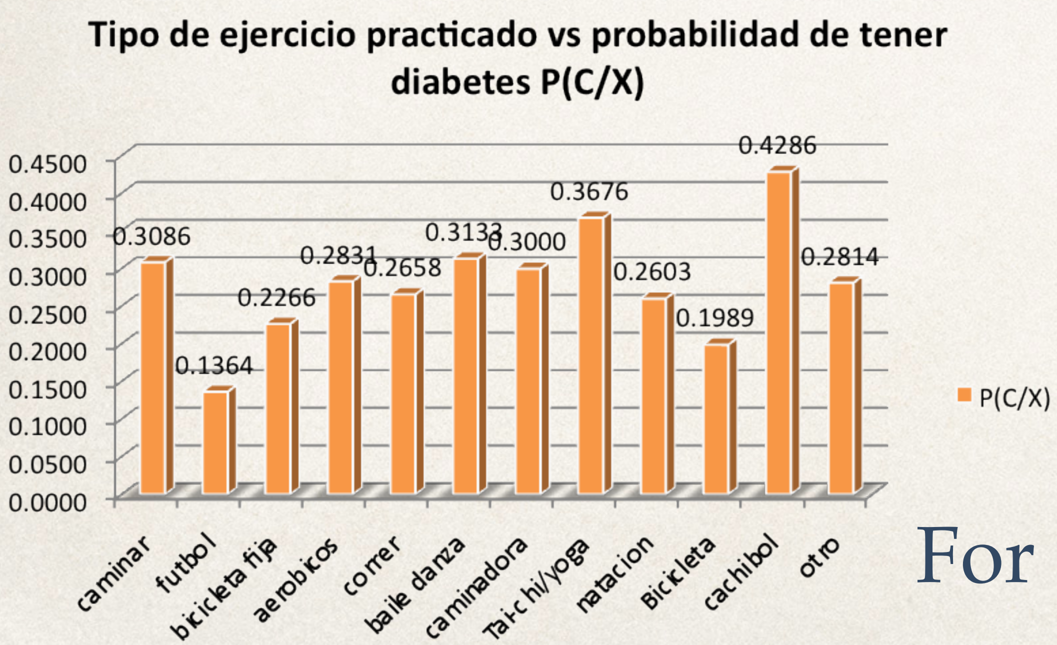
Chronic disease - Risk factors
What you do Exercise



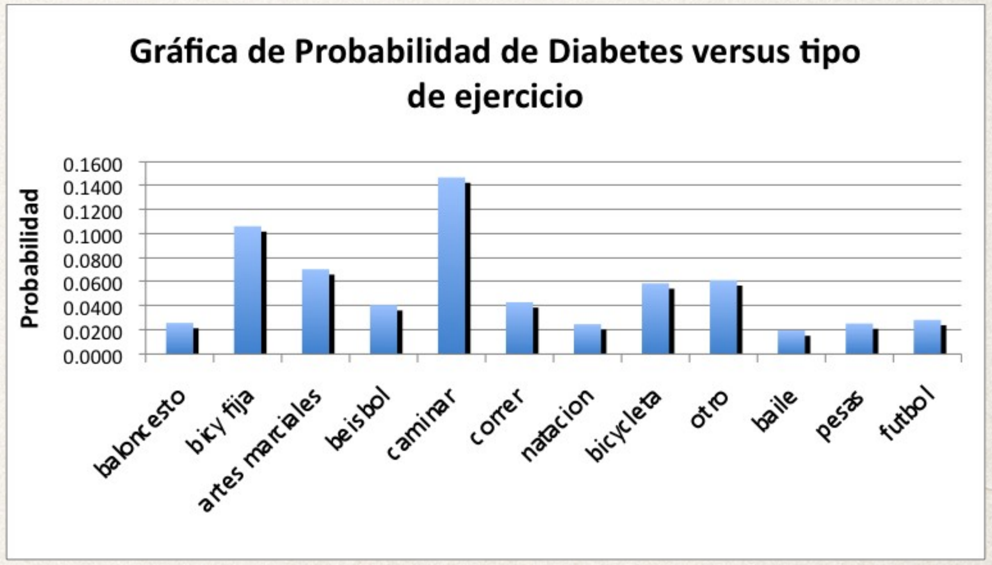




For men 20-59 de PREVENIMSS 2006



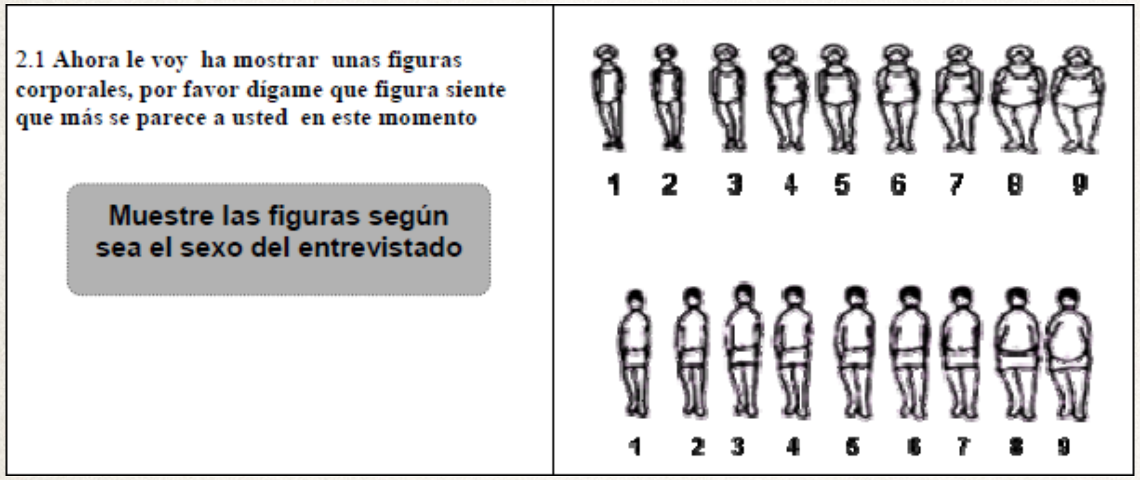
For seniors > 59

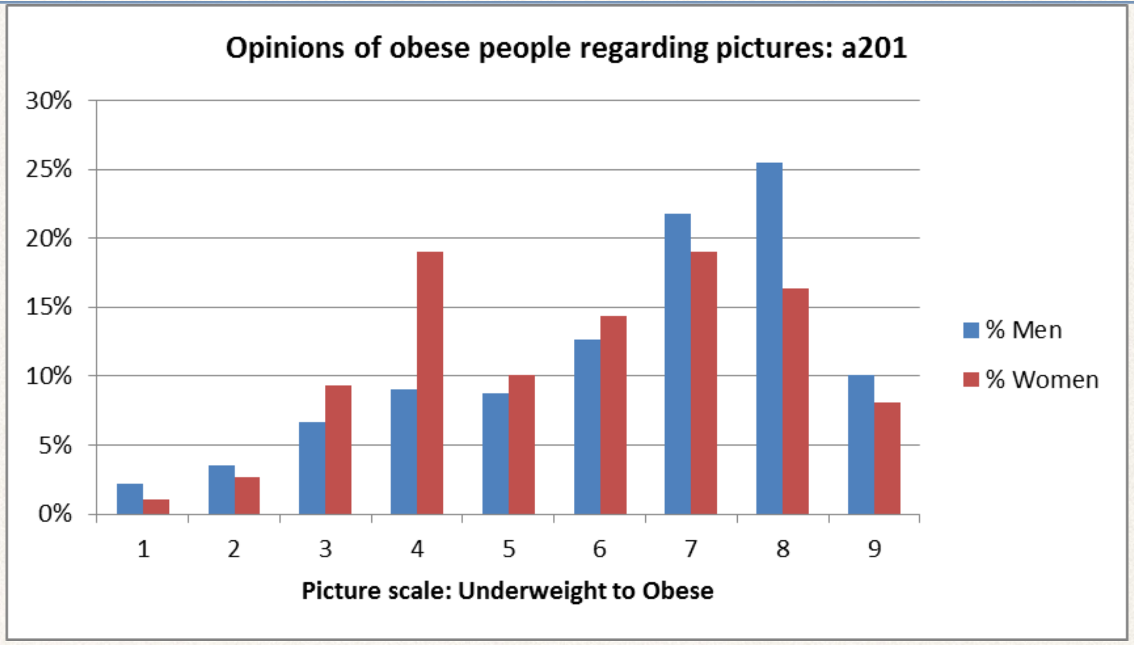


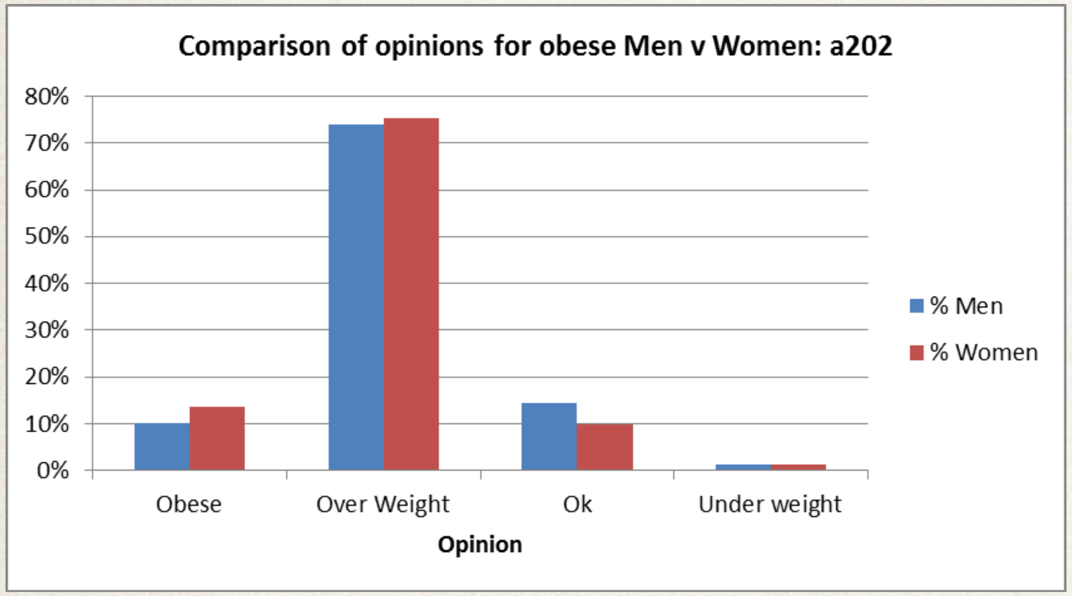
Is it riskier to walk than do nothing?

Obesity- risk factors
What you think/feel
Obesity is unrecognised by the sufferer in spite of the symptoms

Epidemiological data from ENSANUT 2006







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

Fundamental question: Why do we "lie" to ourselves?

J. Easton, H. Sicilia - BMC, Frontiers in Public Health

Perception of weight and Cognitive Biases - What you think/feel

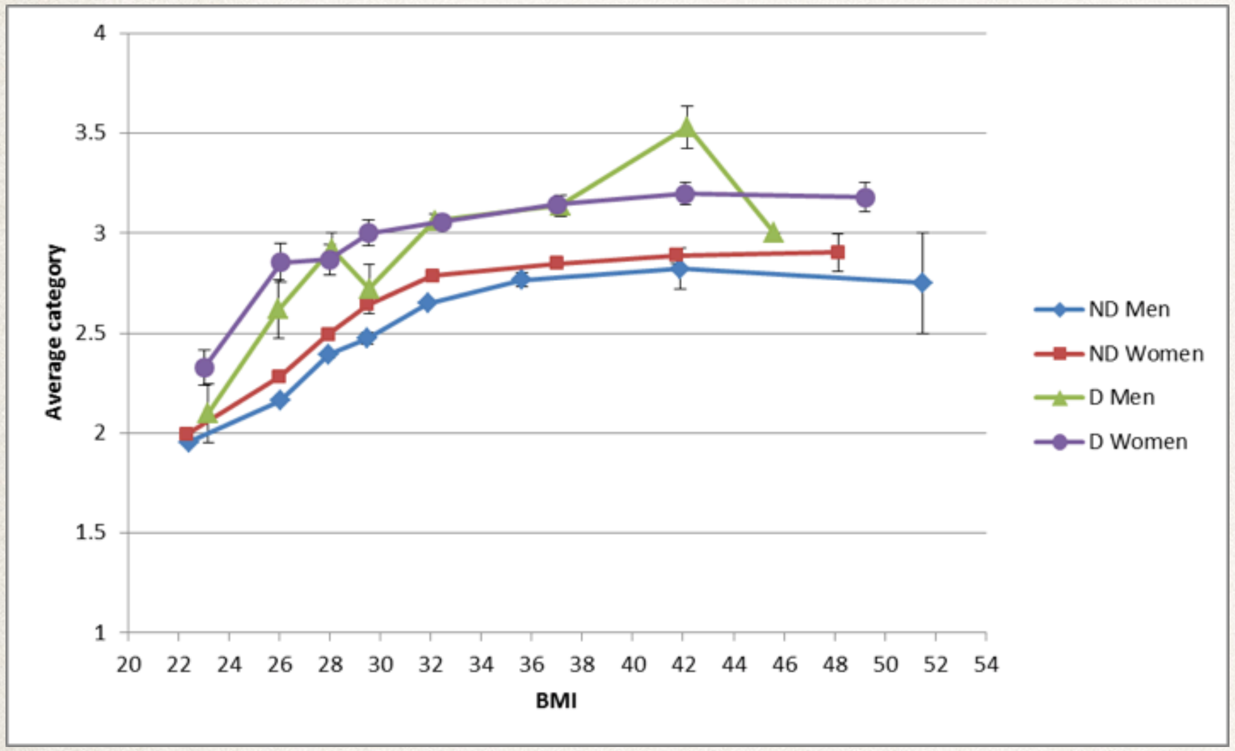


Figure 2. Comparison of non-diagnosed (ND) versus diagnosed (D) obese mean responses for the category self-perception question by gender.

Slopes in the linear range are 35-50% less than one would expect if people could gauge their weight accurately! The lobster in the pot syndrome

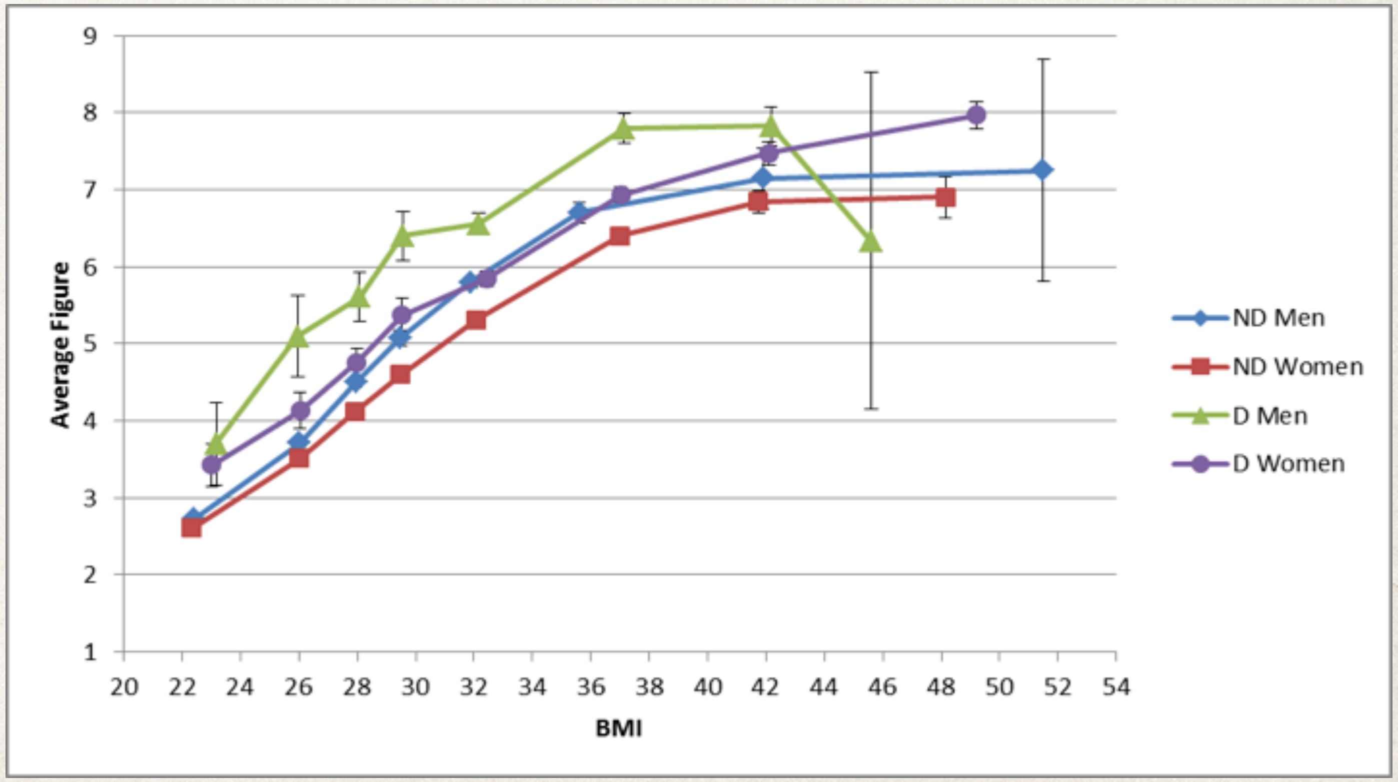


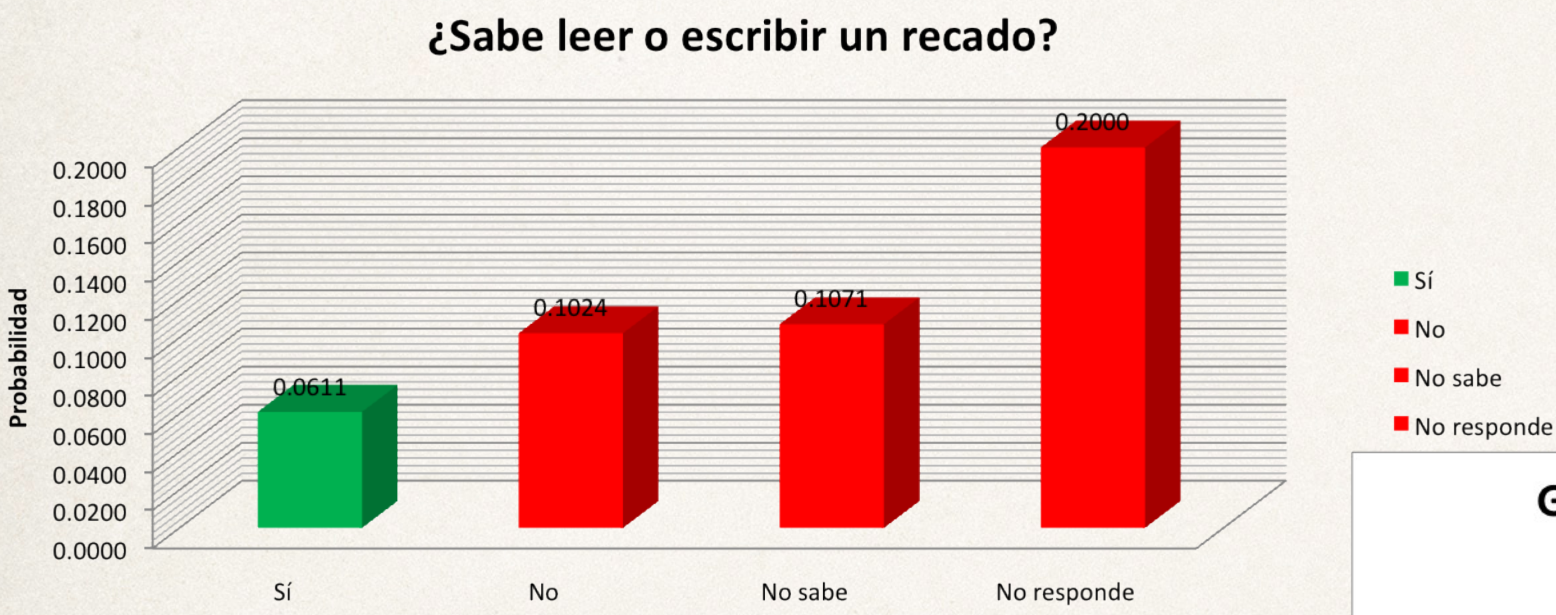
Figure 3. Comparison of non-diagnosed (ND) versus diagnosed (D) obese mean responses for the Stunkard figure rating scale question by gender.

Self-serving bias
Anchoring bias

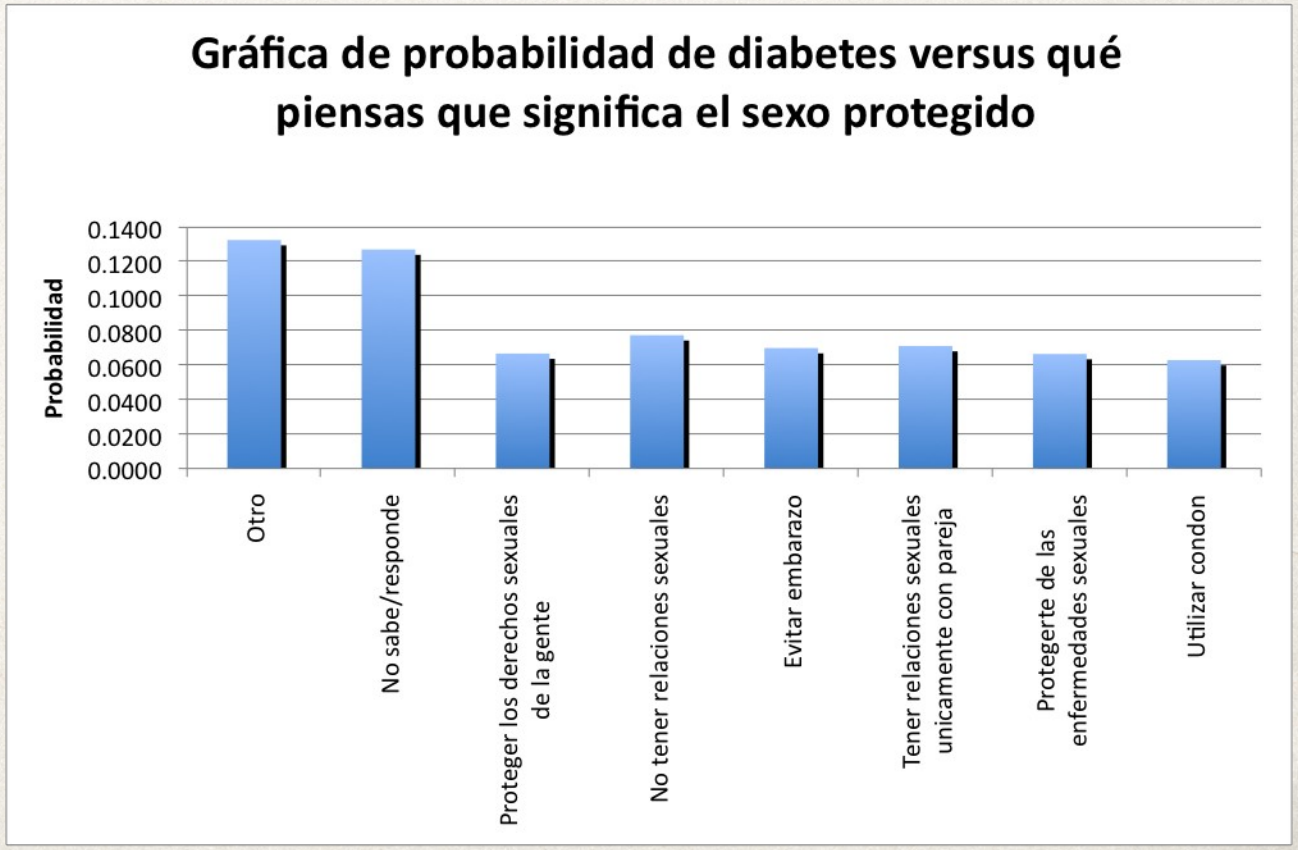
Chronic disease - risk factors

What you think (know): Ignorance can kill

Epidemiological data from ENCOPREVENIMSS 2006



For men 20-59 from PREVENIMSS 2006



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

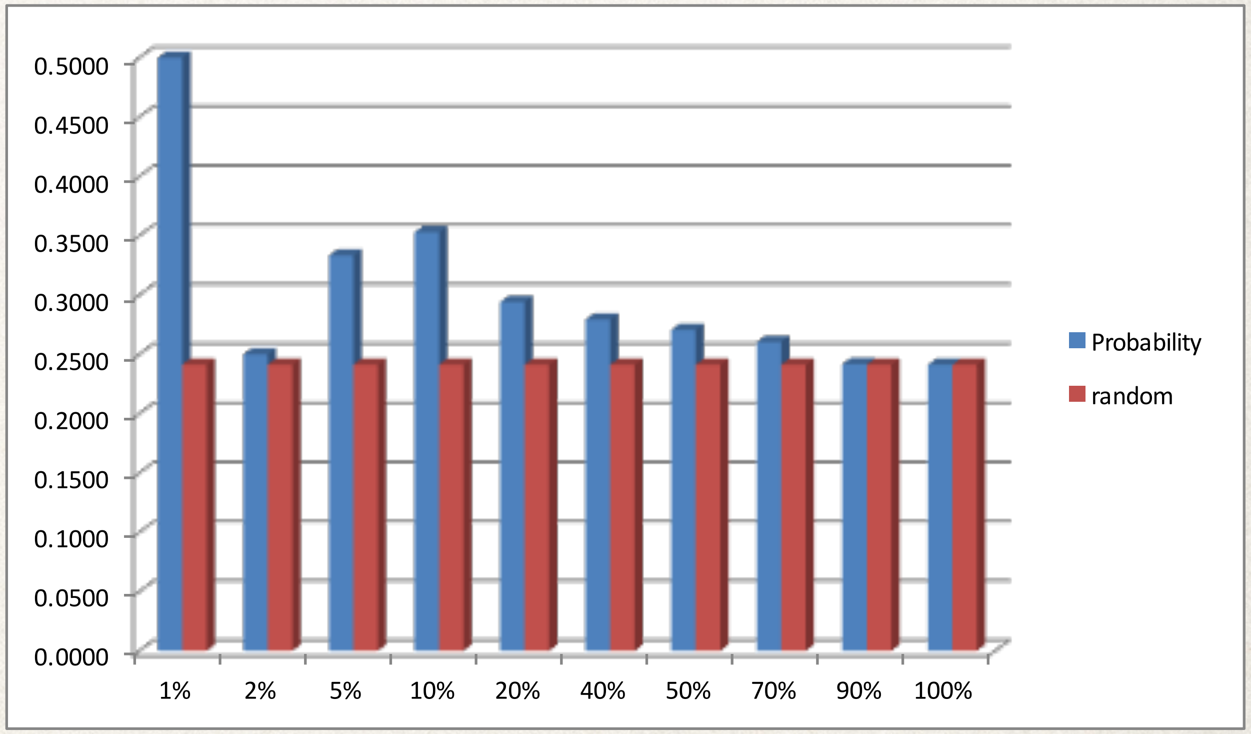
Obesity -risk factors
Who you are - genes

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)

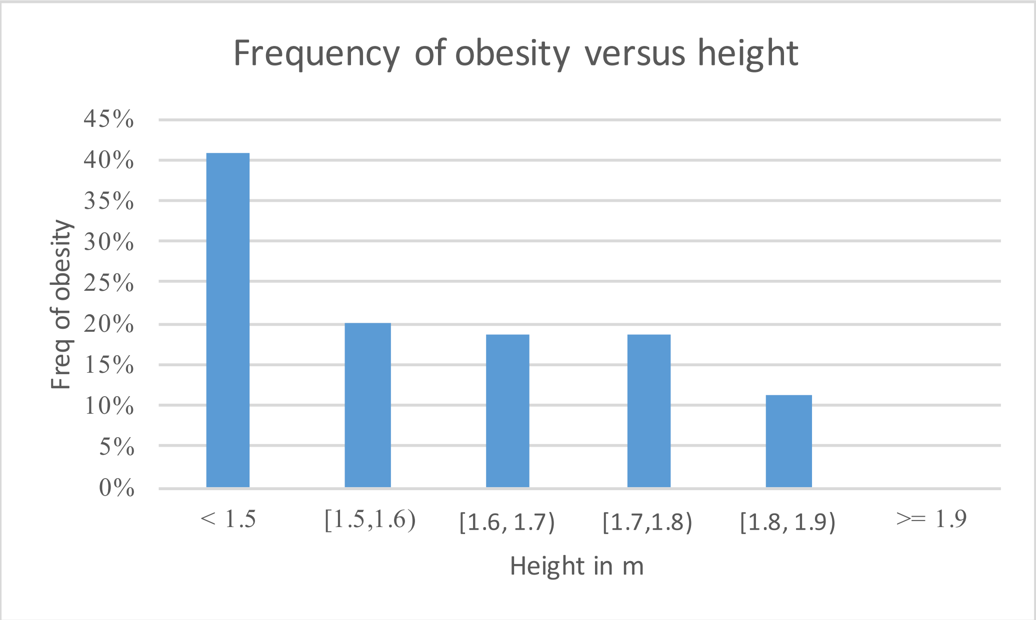


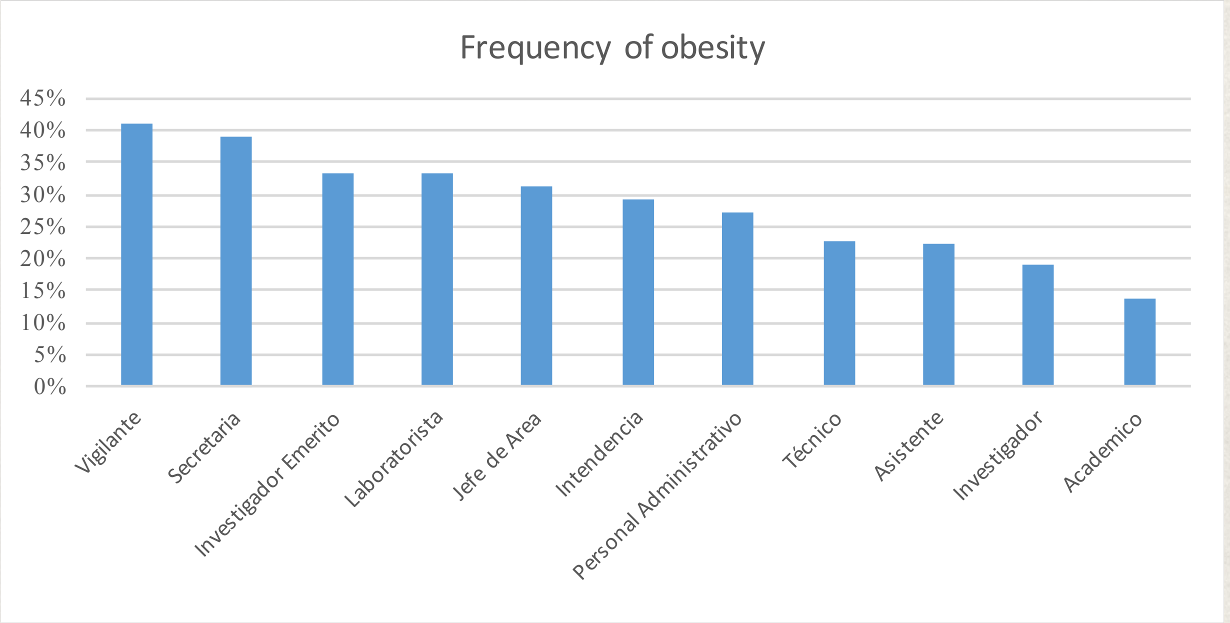
Doesn't give a good model on its own

UNAM Study 2014: Genetic analysis

Obesity -risk factors

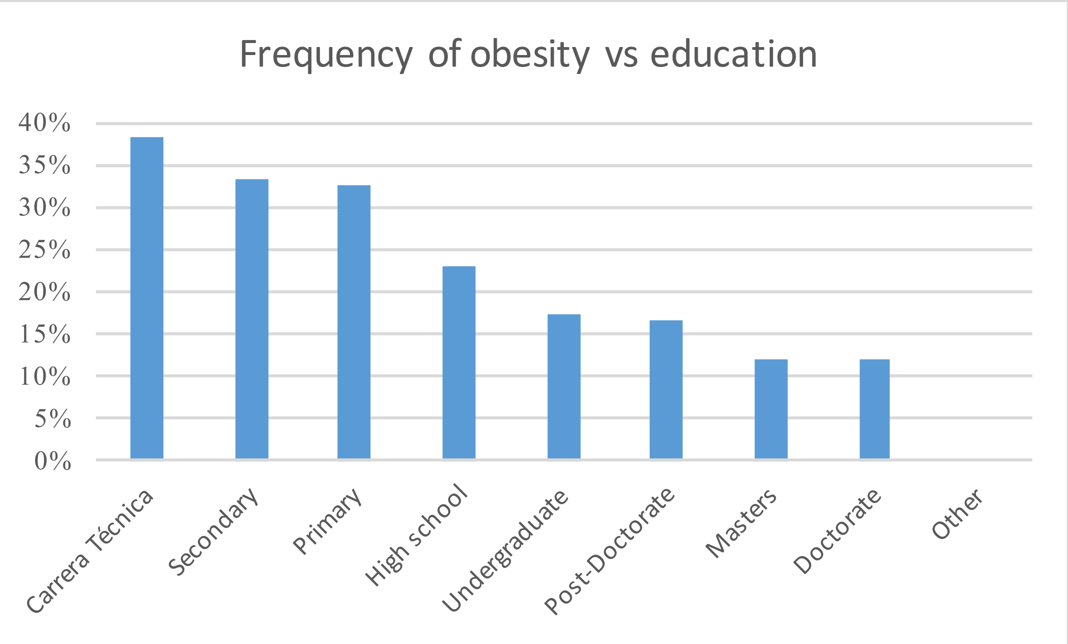
Who you are, what you think, what you do





Why are short people so prone to obesity?
Unit bias?*

The crucial role played by "education"
But what does it really mean?



* Katherine Stephens

Chronic diseases

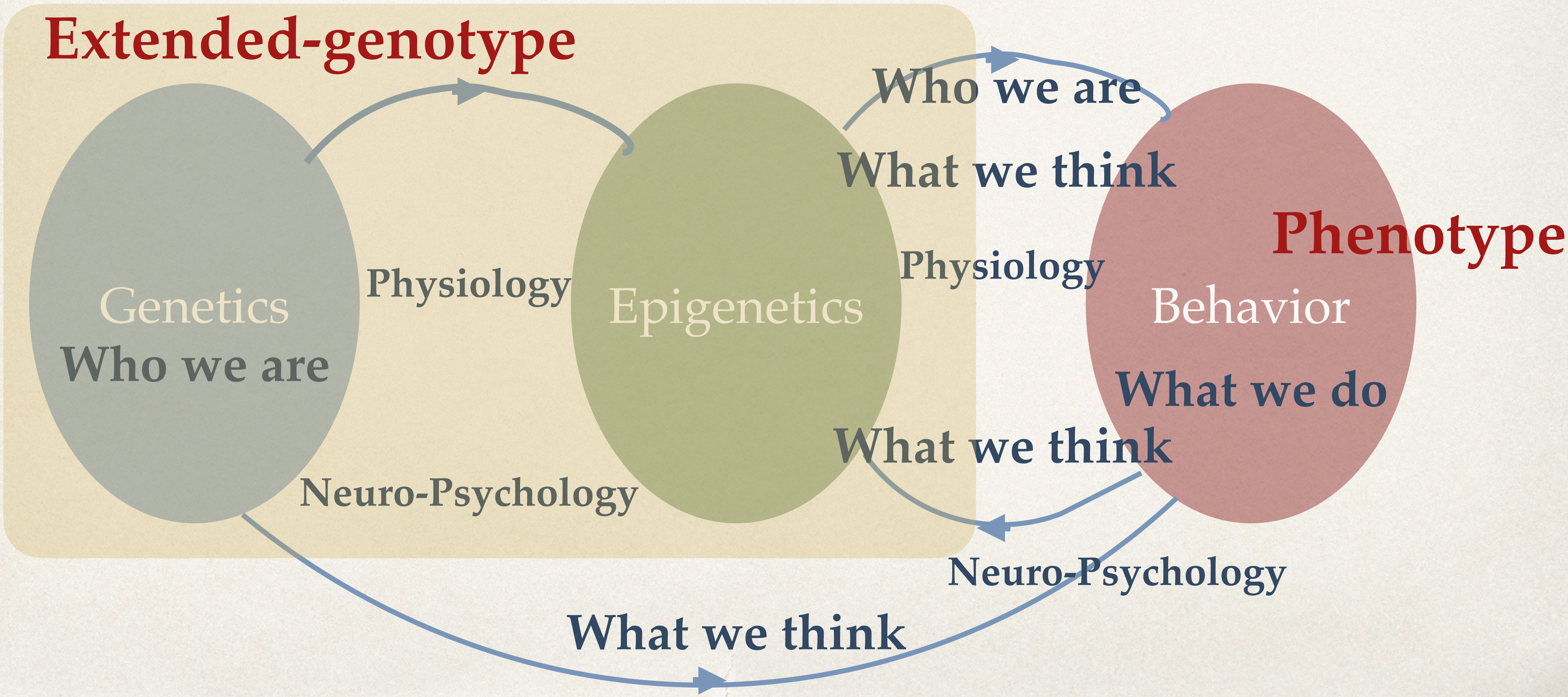
To understand the physiology and genetics of such diseases is important. However, these diseases are predominantly “behavioural” diseases, associated with “bad” decisions.
Why do we make “bad” decisions? What behaviour is plastic?

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

Observations

Obesity is COMMON and maybe not as extreme as it should be

The complex (behavioral) phenotype is a window into the extended genotype but the physiological genotype-phenotype map is much simpler than the behavior-genotype map



Questions

1. What is the appropriate taxonomy of those “universal” tendencies in human physiology/behaviour that are associated with the obesity pandemic?

2. What are the genetic/epigenetic underpinnings of these “universal” tendencies?

3. What are the phenotypic variables that will most help to identify these tendencies? (Stop looking for only high signal to noise relations)

4. How have the consequences of those tendencies changed due to environmental changes (and how has the environment objectively changed?)

5. How do we quantify the effect of a given variable/class of variables?

6. What is the impact of time horizon on a given variable (e.g., the difference between being obese for one year versus 20)

7. How do we disentangle the cause-effect relationships?

8. What is actionable? What factors are plastic and what is their degree of plasticity?

Oportunidad para el INMEGEN y el C3

- Crear la base de datos más profunda en el mundo para la investigación de obesidad y sus consecuencias
  - Multiples poblaciones con gran heterogeneidad entre si (UNAM - educacion/ edad/ …)
  - Multiples instrumentos y mediciones
    - Conducta - nutrición, estilo de vida, historia de vida
    - Socio-economico y socio-demográfico, historia de vida
    - Psicologia y neuropsicologia
    - Secuenciación y expresión
  - Estudio transversal y longitudinal - estudios fisiológicos anuales
- Analizar esa base en forma distinta a la actual - Sistemas Complejos Adaptativos
- Retos: Recursos financieros, recursos humanos, capital político,…
- Recursos actuales: C3 - proyecto de Fronteras (approx $4 millones); surtido de colaboraciones con otras instituciones - INGer, INNSZ, INMEGEN etc.

Partial list of members of the C3 research program in Obesity and Diabetes

Chris Stephens C3 y ICN, UNAM
Marcia Hiriart C3 y IFC, UNAM
Enrique Hernández-Lemus INMEGEN
Martha Käufer INNSZ
Eduardo Garcia INNSZ
Alejandro Frank C3 y ICN, UNAM
Bruno Estañol INNSZ
Guillermo Melendez Hospital General
Ruben Fossion C3 y ICN, UNAM
Ali Ruíz Coronel C3, UNAM
Samuel Canizales INMEGEN
Emmanuel Landa C3 y ICN, UNAM
Irving Morales C3 y ICN, UNAM
Joel Mendoza C3 y ICN, UNAM
Jose Figueroa UNAM
Ana Leonor Rivera ICN, UNAM
Natalia Mantilla C3 y FC, UNAM
Sergio Hernández C3 y FC, UNAM
Jonathan Easton C3, UNAM
Hugo Flores Huerta C3 y IIMAS, UNAM
Luis Miguel Gutierrez INGer
Ulises Perez INGer
Roberto Carlos Castrejon INGer
Diana de la Cruz FM, UNAM
Concepción García FC, UNAM
Francisco Fernández de Miguel IFC, UNAM
Dagmara Wrzecionkowska FCP, UNAM
José Antonio Rivera FC, UNAM
Heriberto Sicilia FC, UNAM

CONACyT Fronteras-2015-2-1093
UNAM PAPIIT
CONACyT Programa de REDES