

La diabetes mellitus como modelo de sistema complejo

Chris Stephens, C3-Centro de Ciencias de la Complejidad y Instituto de Ciencias Nucleares, UNAM

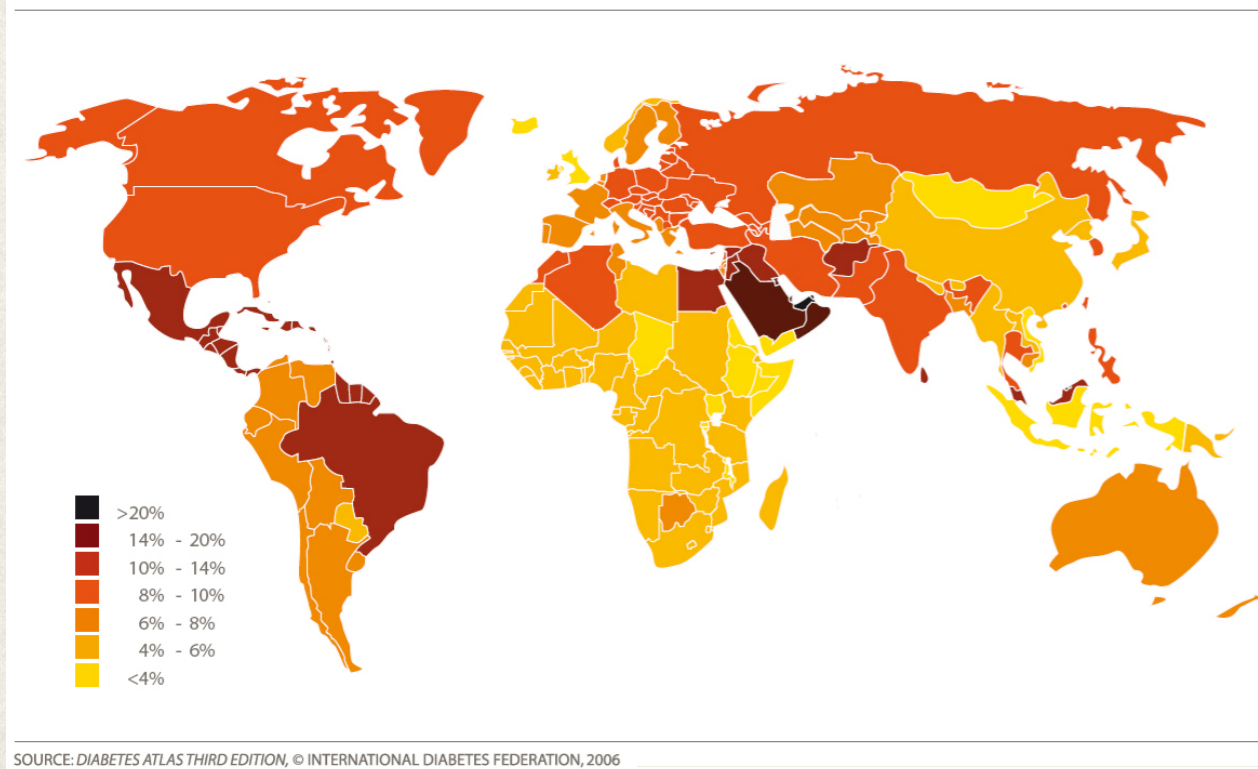


¿Quiénes somos?

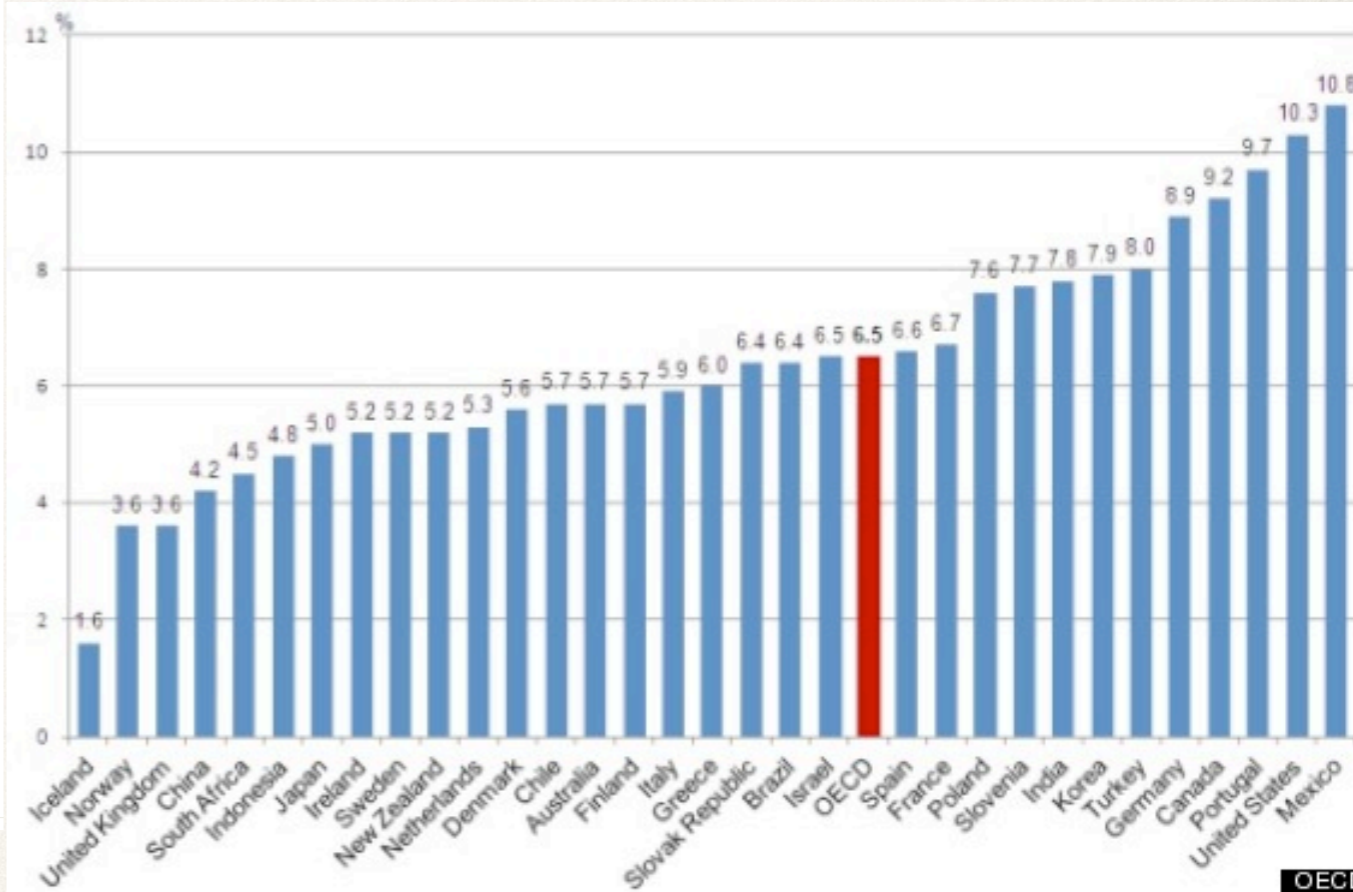
- * Grupo de trabajo en diabetes mellitus tipo 2 - Programa de Complejidad y Salud Pública del C3 - Centro de Ciencias de la Complejidad de la UNAM
 - * Chris Stephens, C3 y ICN-UNAM
 - * Marcia Hiriart, IFC-UNAM
 - * Samuel Canizales, FQ-UNAM
 - * Hugo Flores, C3 y IIMAS-UNAM
 - * Juan Arturo Herrera, Microsoft
 - * Jonathan Easton, University of Northumbria
 - * Germinal Cocho, C3 y IF-UNAM
 - * Carlos Villareal, C3 y IF-UNAM
 - * Carlos Gershenson, C3 y IIMAS-UNAM
 - * Gaby Arauz, IIMAS-UNAM

La importancia de la diabetes mellitus tipo 2

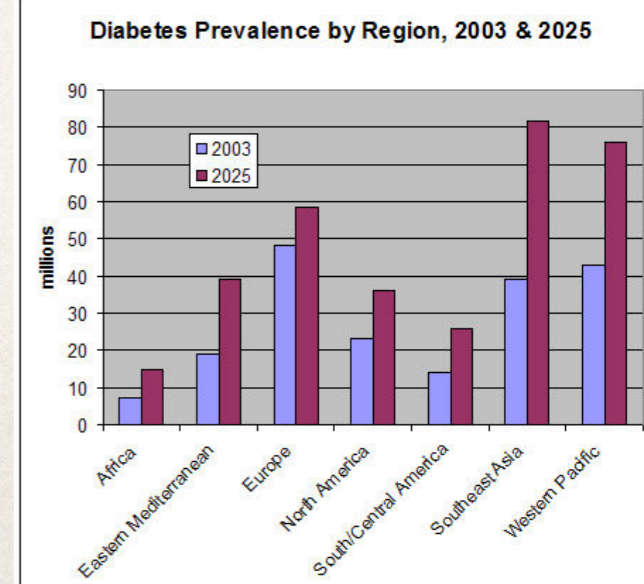
Prevalence estimates of diabetes, 2025



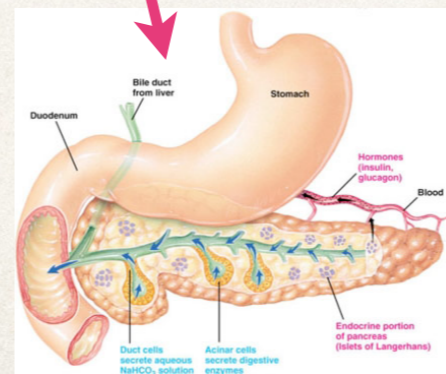
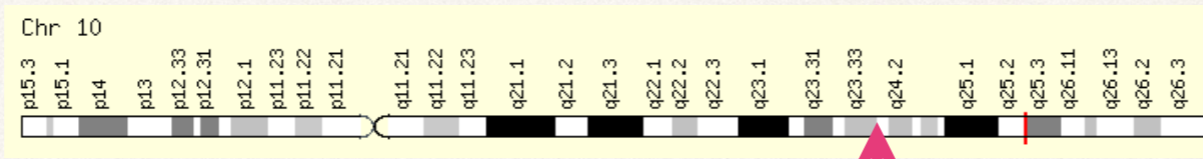
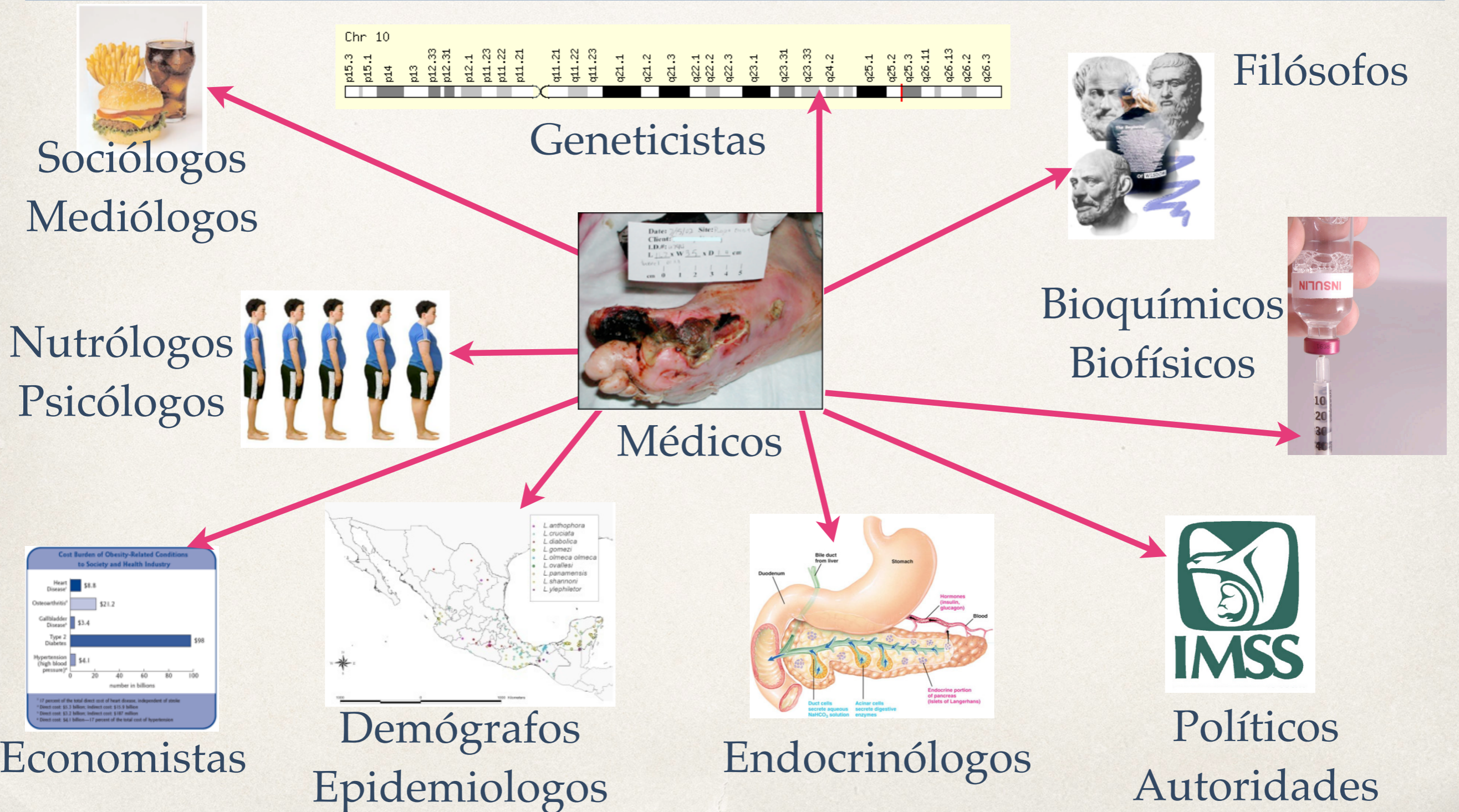
SOURCE: DIABETES ATLAS THIRD EDITION, © INTERNATIONAL DIABETES FEDERATION, 2006



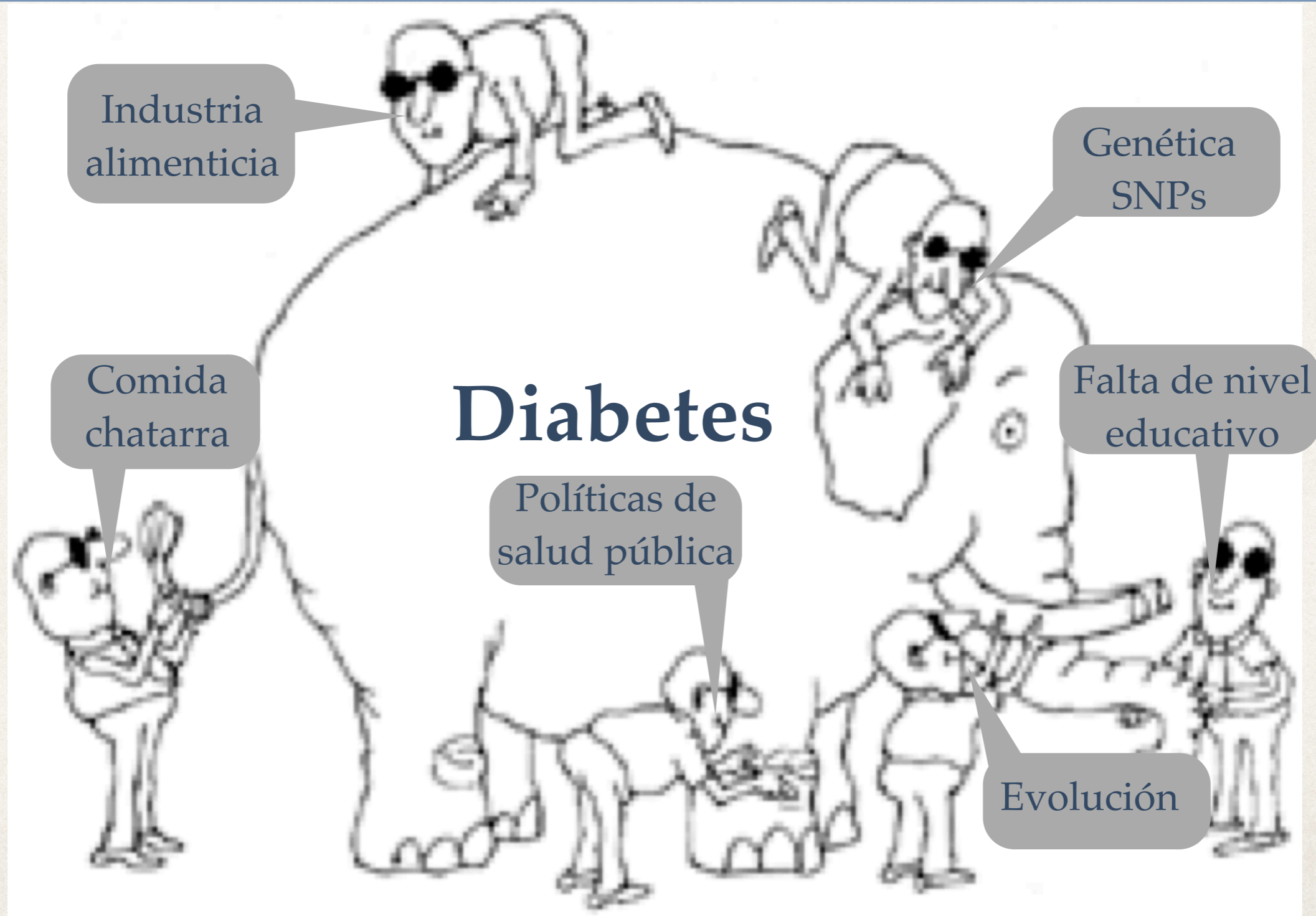
- ❖ Diabetes es una enfermedad donde el riesgo aumenta como función de tiempo dependiente del estilo de vida
- ❖ Un factor que es muy importante es la nutrición
- ❖ Costo en México \$200 mil millones por año, en EU, \$2 millones millones



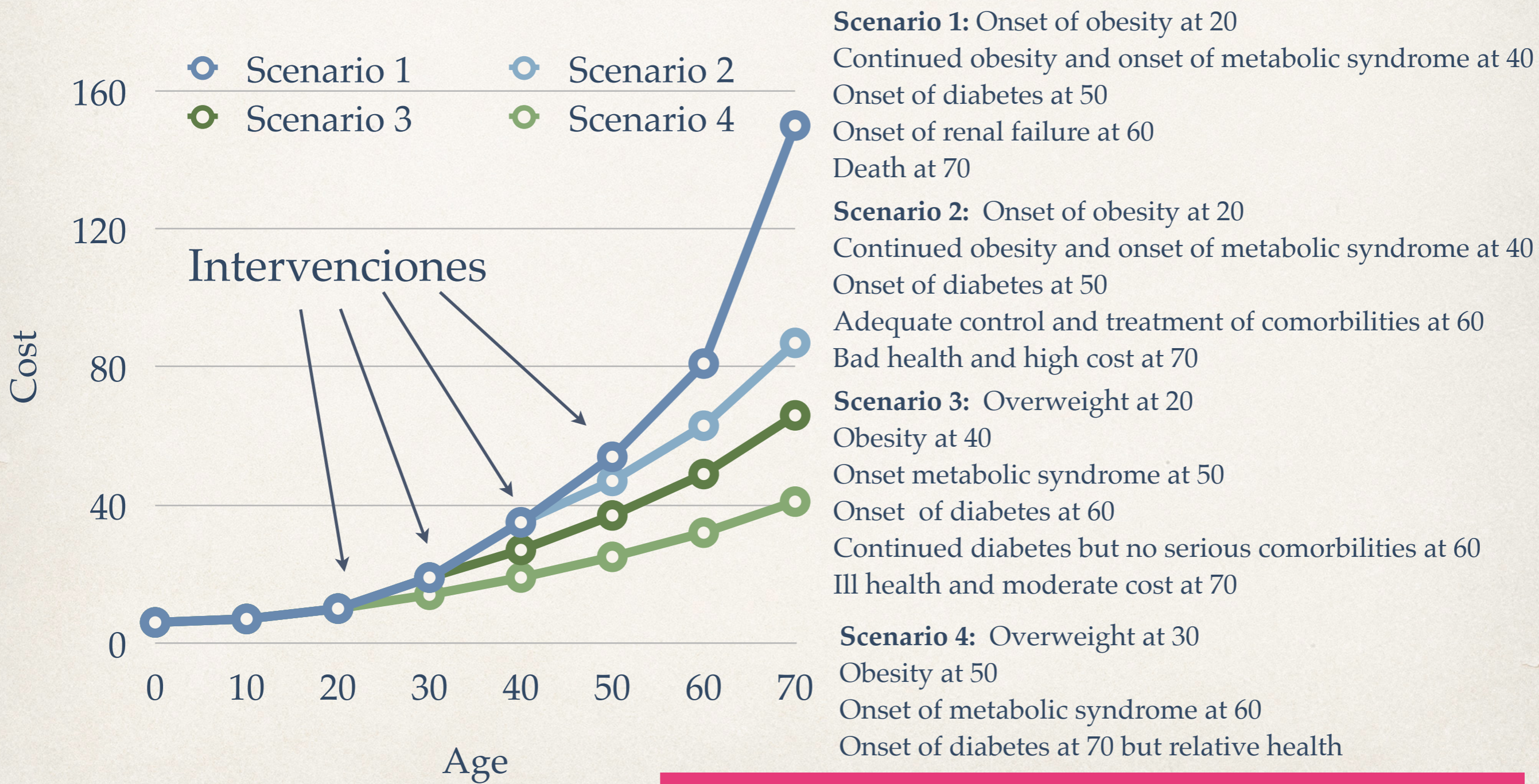
La diabetes es compleja



La diabetes y la necesidad de trabajo transdisciplinario en grupo - ciencia "orientada"



La diabetes y su tratamiento es dinámica y adaptativa



Queremos predecir y entender "historias"

Desde el “micro” hasta el “macro”: diabetes desde el punto de vista de la minería de datos

¿Qué queremos predecir? C

$P(C | X)$

¿Qué lo afecta? X

$S(C | X)$

Risk score

Historias individuales -
Riesgos, intervenciones y
probabilidades de **eventos**

- Obesidad
- Síndrome metabólico
- Diabetes
- Neuropatía diabética
- Falla renal
- Muerte

$$X = X(sd) + X(se) + X(n) + X(ev) + X(g) + X(af) + X(hm) + X(i) + X(sp) + \dots$$

Factores socio-demográficos
Edad, Género,...

Nutrición
Cuánto comes, que comes,...

Factores genéticos
Rs7903146,...

Historía médica
Síndrome metabólico,...

Factores de salud pública
IMSS, ISSTE,...

Factores socio-económicos
Nivel escolar, ingresos,...

Estilo de vida
Fumas, tomas,...

Antecedentes familiares
Padre diabético, ...

Factores inmunológicos

¿Es un “sueño” poder calcular $P(C|X)$?

¿Es un “sueño” poder identificar los factores de riesgo y predecir las probabilidades de pasar de un estado de salud a otro para un individuo?

¿Es un “sueño” poder identificar la mejor intervención para un individuo y predecir su probabilidad de éxito?

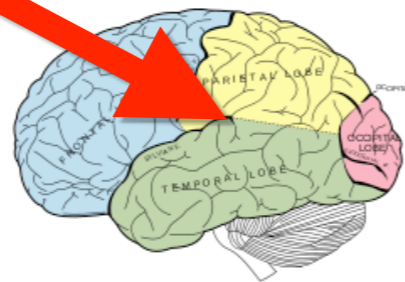
¿Es factible?

Si es factible... a través de la revolución de datos

Revolución en la generación de datos



Cerebro humano
10-100 Terrabytes



Todos los libros en el mundo
30-50 Terrabytes



Revolución en el almacenamiento de datos



En forma electrónica
1 zettabyte

Revolución en el análisis de datos



¿Qué representan estos datos?

¡Sistemas complejos!

El mundo "real" en que vivimos - sucio, incierto, complejo,...

Diagnosis	Score
Rheumatoid arthritis	80
Irritable bowel syndrome	40
Crohn's disease	25
Ulcerative colitis	24

Automated Data Analysis Using Excel
Brian D. Bissett
Chapman & Hall/CRC

¿Qué datos vamos a minar? y ¿Qué haremos con ellos?

- ❖ Encuestas

- ❖ ENSANUT
- ❖ PREVENIMSS

- ❖ Tipo de análisis

- ❖ Descriptivo
- ❖ Predictivo
- ❖ “Actionable”

- ❖ Minería de datos: “... la exploración e análisis de datos con el fin de descubrir patrones, correlaciones y otros regularidades.”

- ❖ Hay dos tareas principales:

- ❖ Predicción - identificación de “patrones”

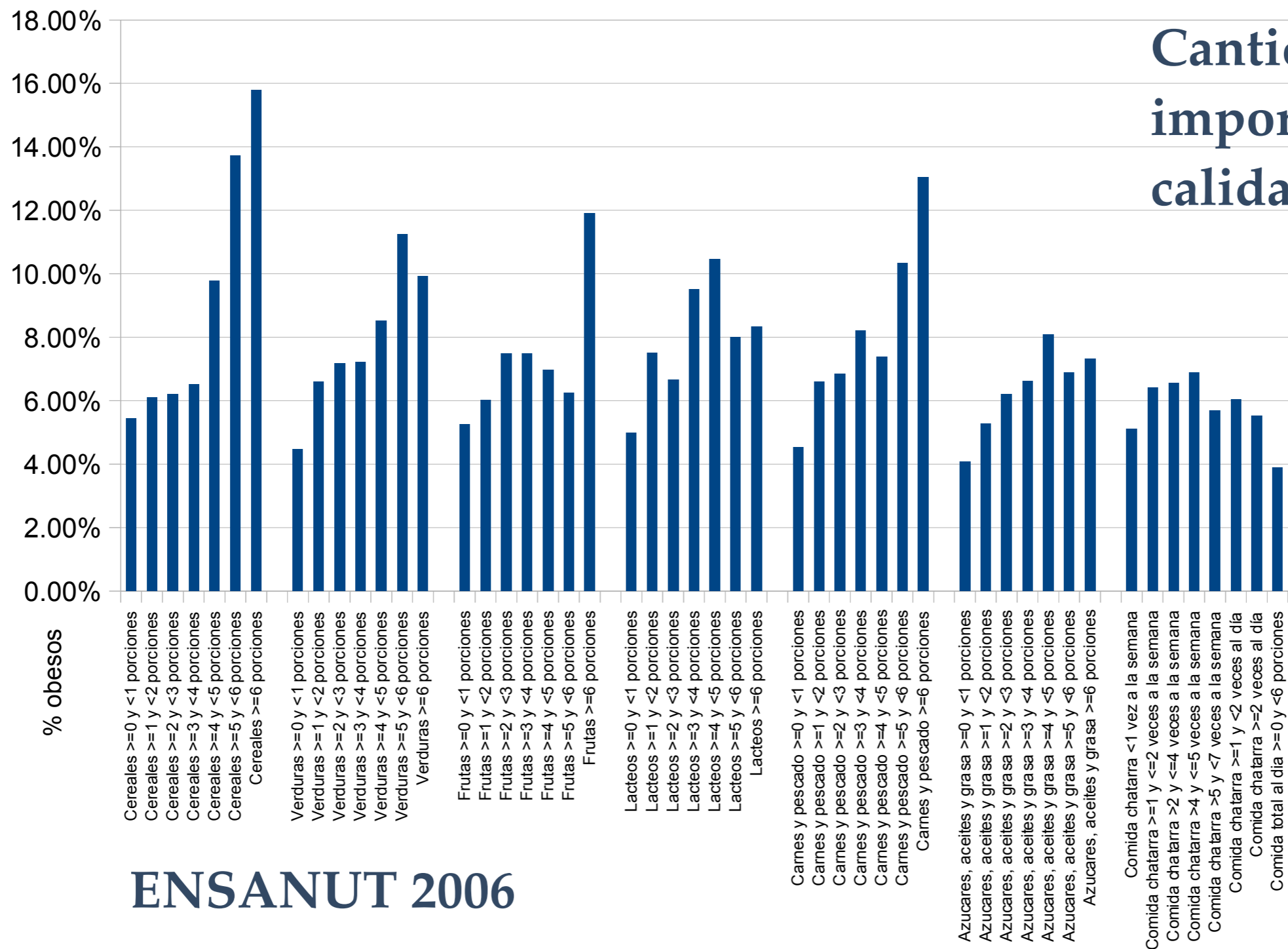
- ❖ Establecer relaciones estadísticas “causales”

- ❖ Perfilaje - descripción de “patrones”

- ❖ Identifica las variables principales asociados con un patron

Identificación de riesgos:

Nutrición... la diferencia entre cantidad y calidad



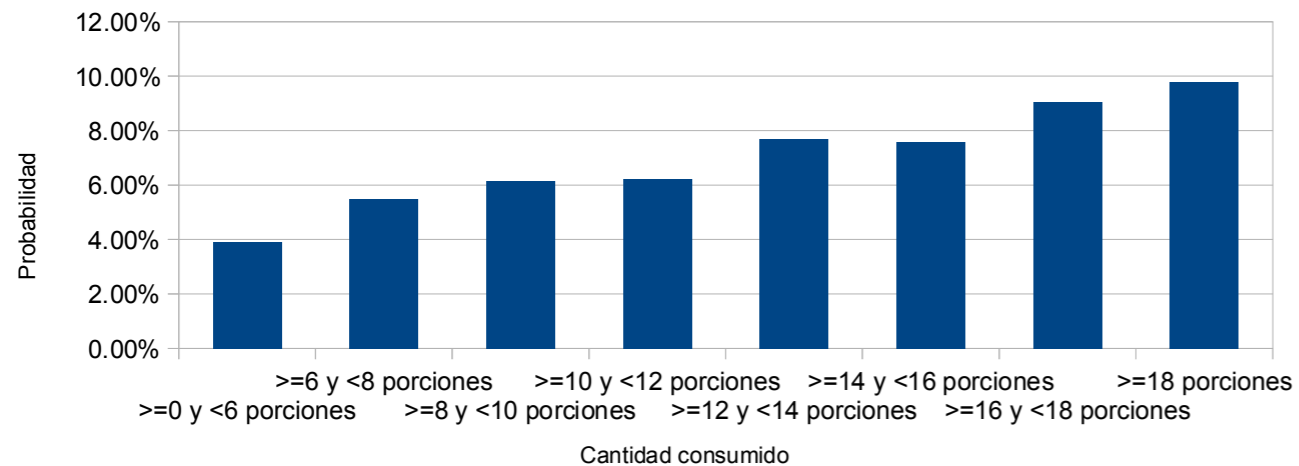
Cantidad es más importante que calidad

Identificación de riesgos:

La nutrición, la psicología, la evolución y la causalidad...

Sheet3

Grafico de cantidad consumido por dia versus probabilidad ser obeso



Graph of probability obese vs daily consumption of diet drinks

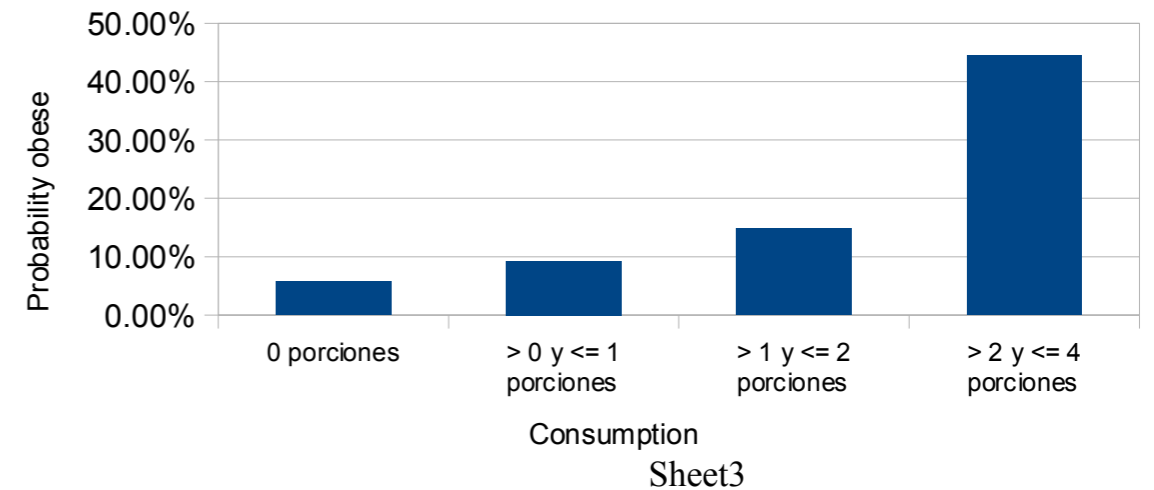
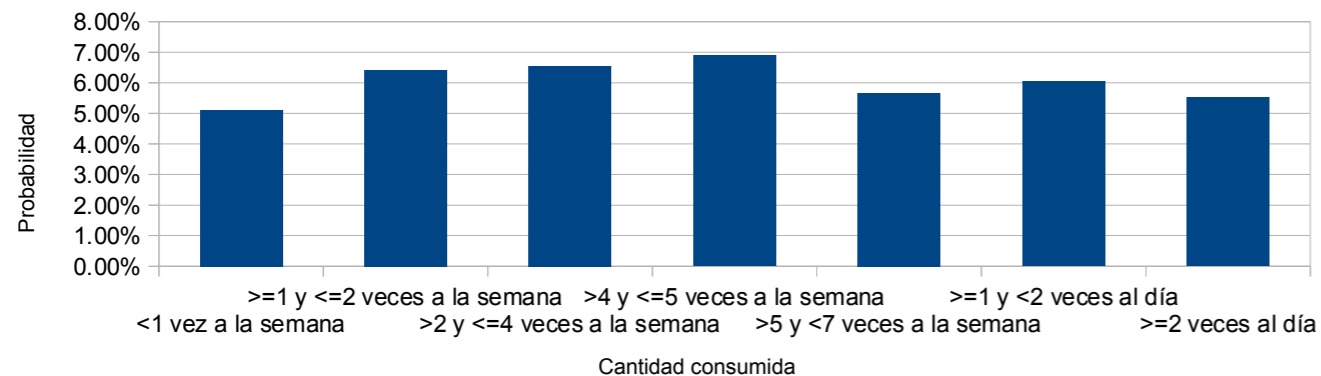
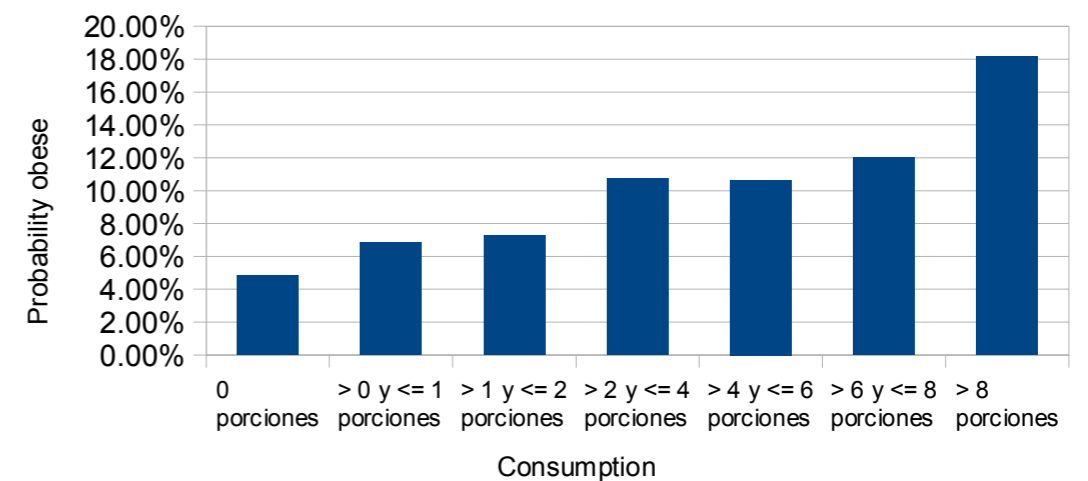


Grafico de probabilidad de ser obeso versus cantidad de comida chatarra consumida

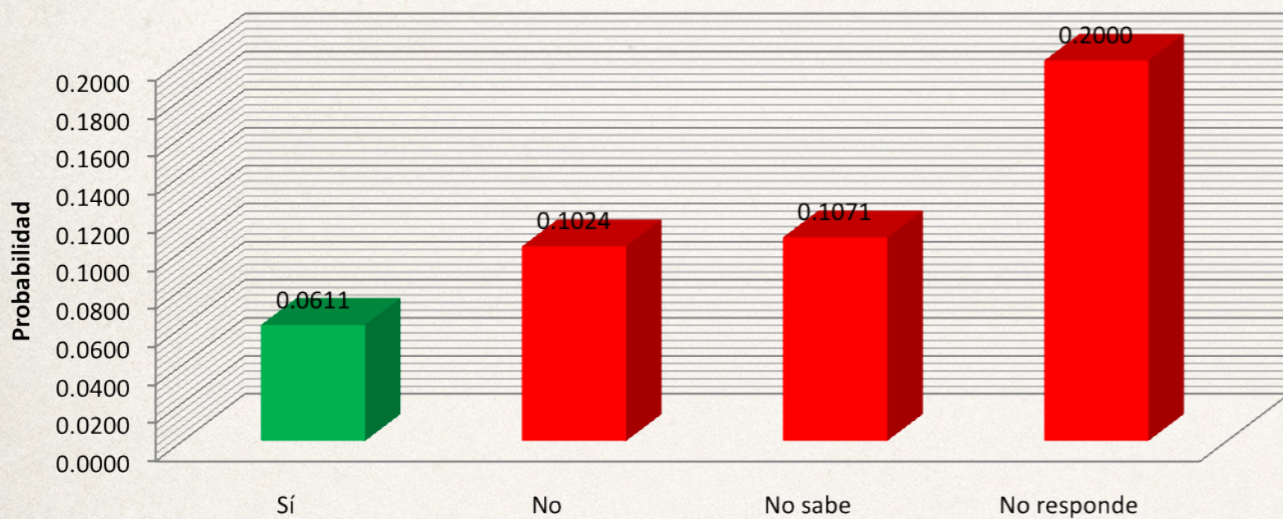


Graph of probability obese versus daily consumption of lettuce



Identificación de riesgos: Ignorancia y la educación

¿Sabe leer o escribir un recado?

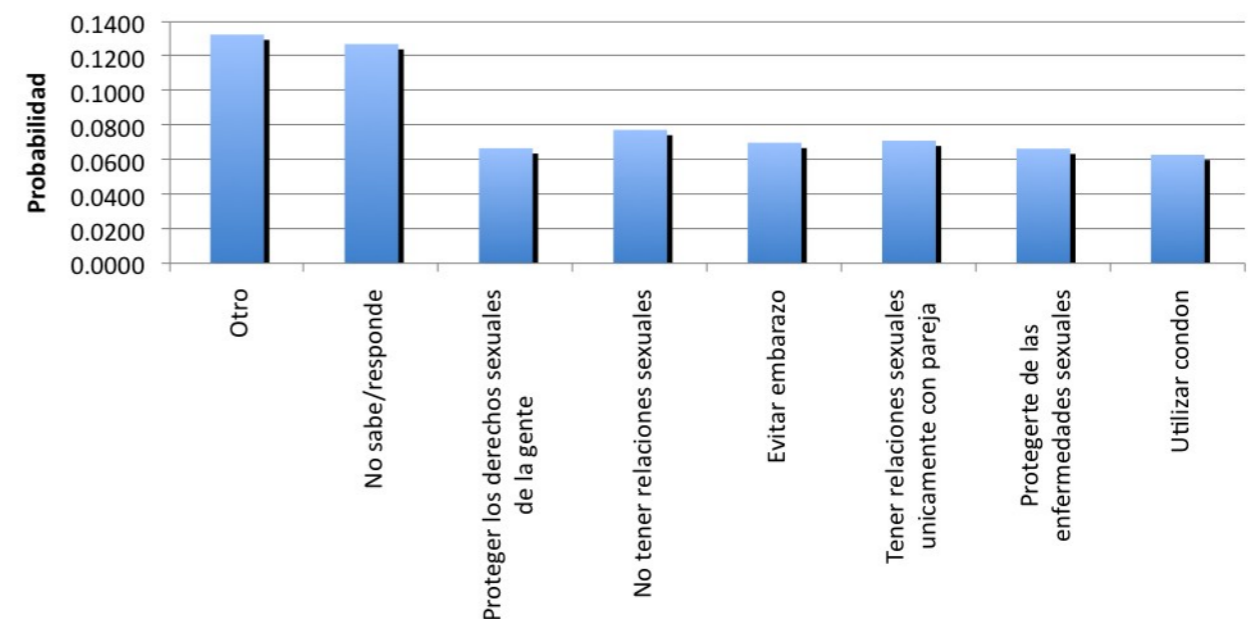


Ignorancia y especialmente ignorancia sobre la salud es un factor de riesgo de importancia similar a la obesidad

Para hombres 20-59 de PREVENIMSS 2006

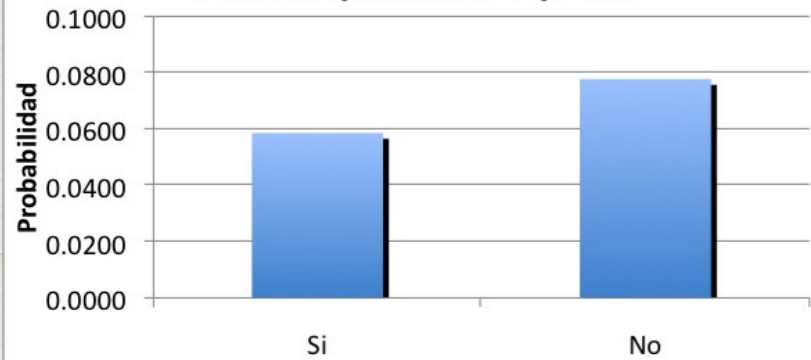
- Sí
- No
- No sabe
- No responde

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

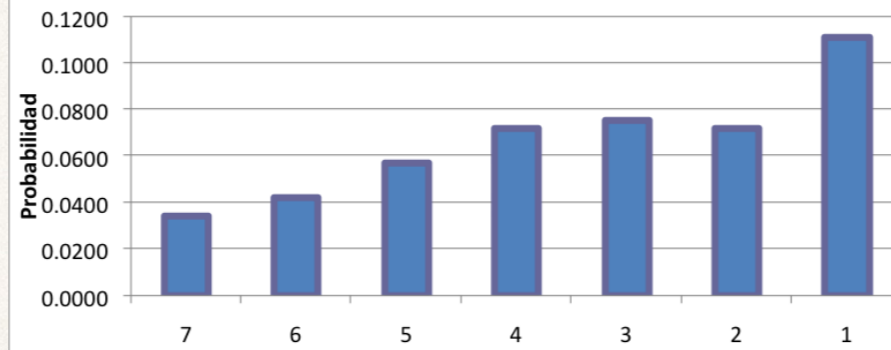


Identificación de riesgos: Sedentarismo y ejercicio

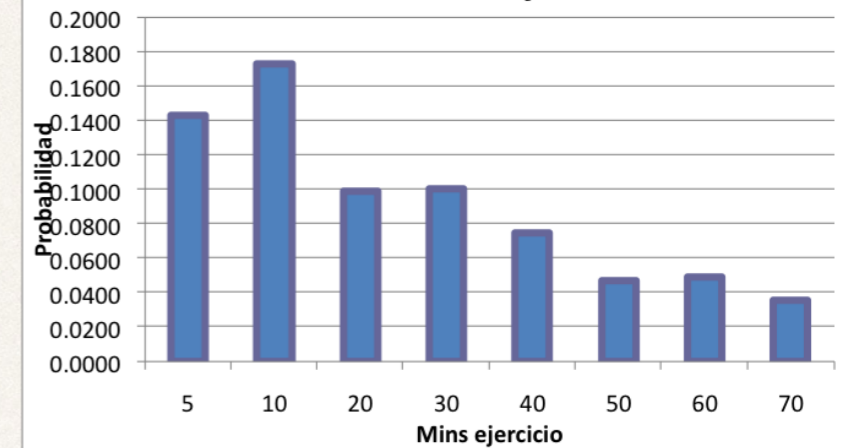
Gráfica de Probabilidad de Diabetes versus si practicas deportes



Gráfica de Probabilidad de diabetes versus Número de días de ejercicio por semana

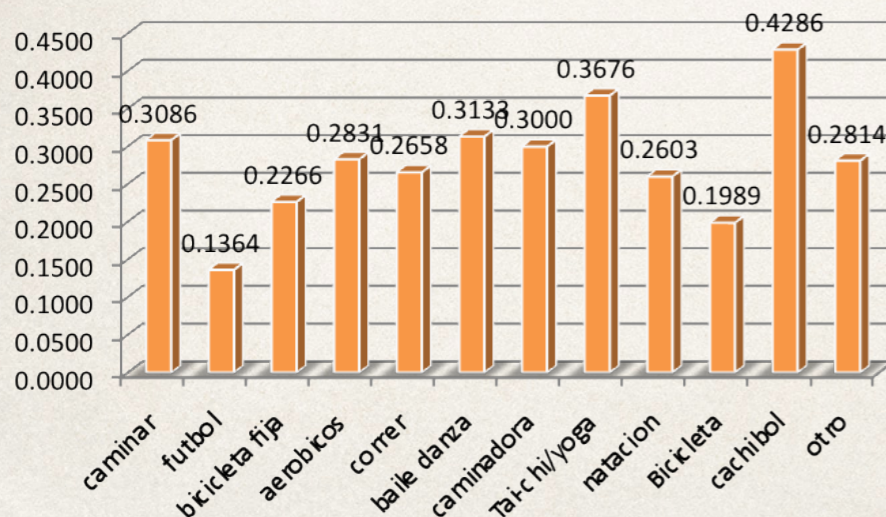


Gráfica de Probabilidad de diabetes versus mins de ejercicio



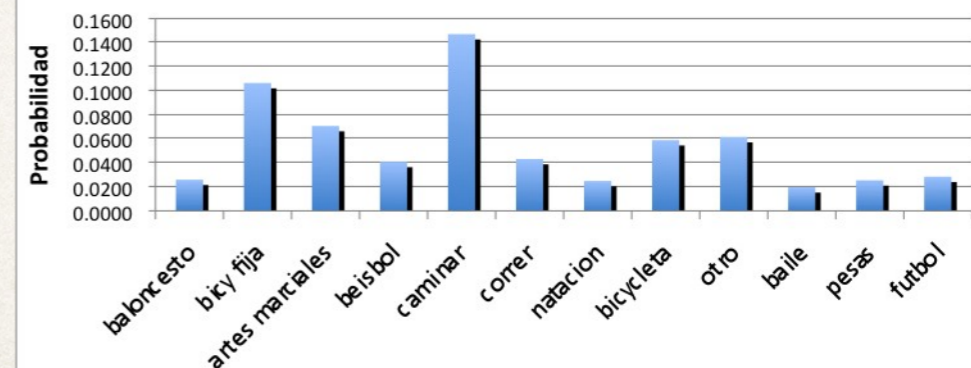
Para hombres 20-59 de PREVENIMSS 2006

Tipo de ejercicio practicado vs probabilidad de tener diabetes P(C/X)



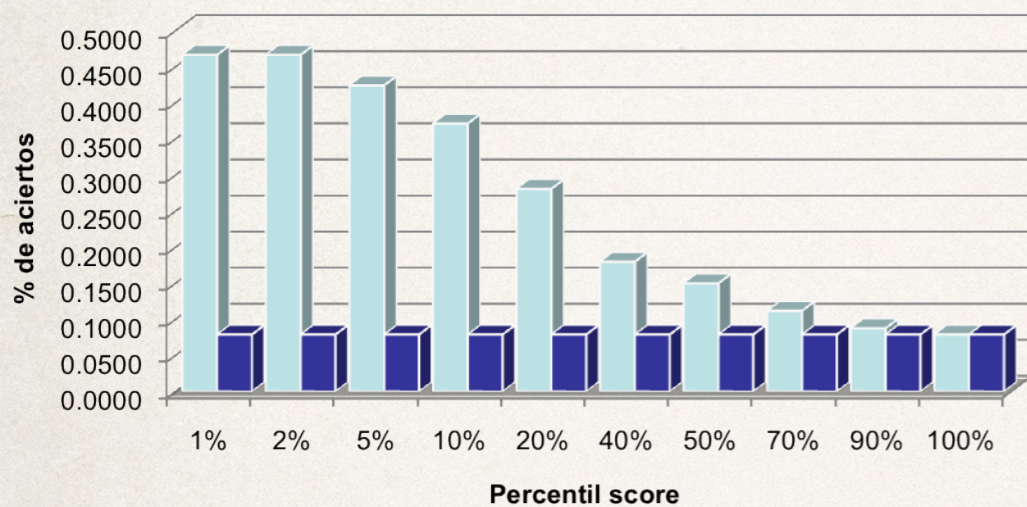
Para adultos mayores > 59

Gráfica de Probabilidad de Diabetes versus tipo de ejercicio



Desde la descripción hasta la predicción y “actionability”

Gráfica de % de aciertos como función del percentil de calificación



Modelo de predicción de presencia de diabetes mellitus tipo 2 a través de datos de PREVENIMSS 2006

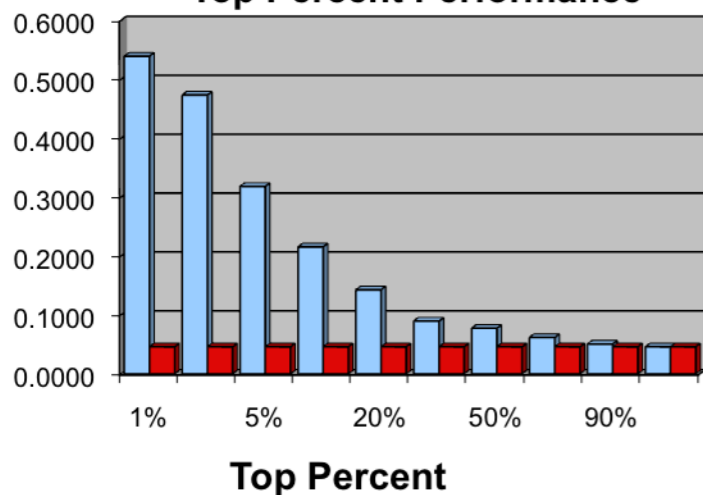
Diseño de sistemas de software inteligente para apoyo en la toma de decisiones a todos los niveles

Perfil de Riesgo	
Pregunta de Perfil	Respuesta
¿Algun familiar padece o padeció diabetes mellitus? (hijos)	si
Edad cumplida	54 a 59 años
¿Algun familiar padece o padeció presión alta? (hijos)	si
¿Algun familiar padece o padeció diabetes mellitus? (hermanos)	si
Cuantos minutos por sesión hace ejercicio	5 minutos
Número de hijos vivos que ha tenido	5 hijos vivos
Edad a la que empezó a beber	> 27 años
¿Algun familiar padece o padeció obesidad? (hijos)	si
¿Algun familiar padece o padeció sobrepeso? (hijos)	si

Modelo de predicción de gastos médicos mayores y comorbilidades para un grupo de diabéticos

Valor	Descripción	Epsilon	P(C/X)
10	97 total covered spending I+O+D	43.20099567	0.237247924
10	total covered spending I+O	40.46404161	0.225385528
10	97 Q4 total covered spending I+O+D	36.81476952	0.209569
10	total outpatient spending	36.63230592	0.208778173
1	Renal Failure	34.55328575	0.42195122
1	chronic renal failure	34.4575483	0.475884244
9	Q4 total covered spending I+O	34.35151087	0.198892843
9	Q4 total outpatient spending	34.26027907	0.19849743
9	97 Q3 total covered spending I+O+D	31.97948401	0.1886121
3	total inpatient spending	31.92264569	0.231607629
10	97 Q2 total covered spending I+O+D	30.33731158	0.181494662
1	renal failure, unspecified	29.89280985	0.532967033
9	97 Q1 total covered spending I+O+D	27.41789391	0.168841439
8	Q2 total covered spending I+O	27.23543031	0.168050613
7	Q3 total covered spending I+O	27.1441985	0.1676552
1	loop diuretics	26.45884908	0.175141243
10	total drug spending	26.23188048	0.163701068

Top Percent Performance



¿Dónde estamos?

Estamos en el umbral de una nueva era...

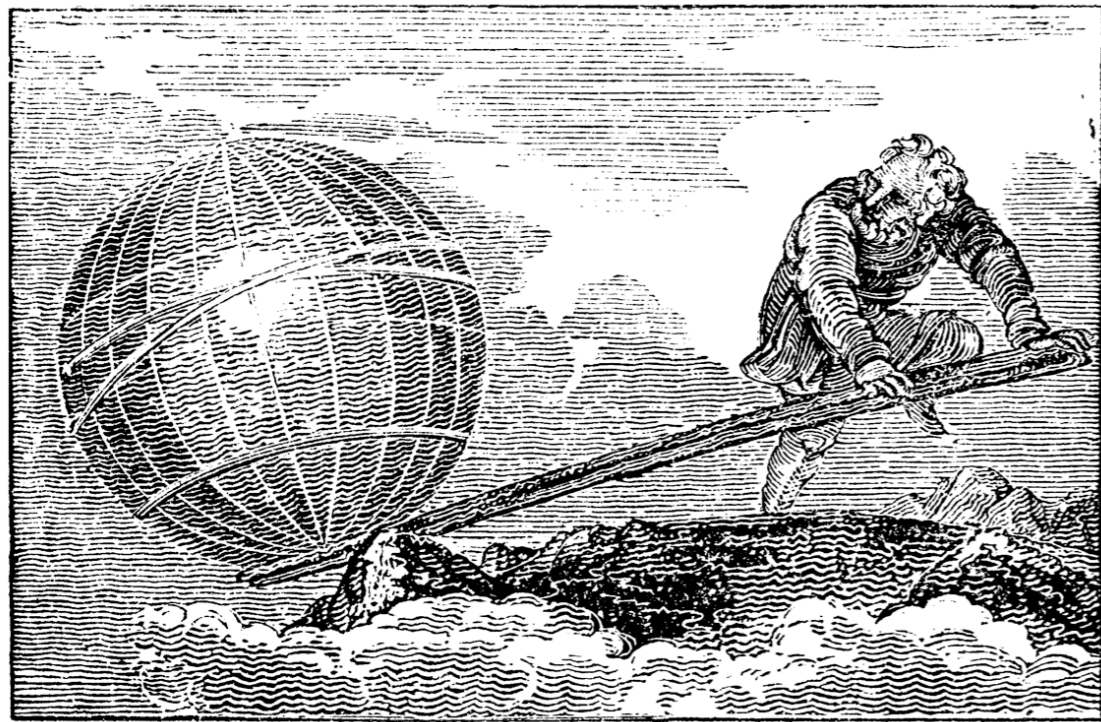
Una era donde por primera vez en la historia humana hay más datos en medios externos que en el medio neuronal

Una era donde con el buen/mal uso de esos datos (quien tiene acceso; quien decide como deben ser usados...) podemos potencialmente predecir mucho de lo que pasará con los sistemas complejos - con los seres humanos

Una era donde podemos revolucionar tanto la medicina como la ciencia

Una era donde podemos imaginar predecir la evolución de enfermedades como la diabetes al nivel individual tanto en términos de rasgos individuales como en términos de su historia particular

¿A dónde queremos ir?



Es sumamente importante considerar la diabetes de un punto de vista inter- y transdisciplinario.

Es sumamente importante que los médicos, los científicos y las autoridades colaboren estrechamente en equipo

Con datos suficientes se puede predecir cualquier cosa...

El análisis de los datos mostrados aquí es nada más que la cima del iceberg, pero muestra la factibilidad de usar la minería de datos para crear modelos predictivos para todas las distintas etapas de la enfermedad

Un zafarrancho de combate

El C3, el programa de Complejidad y Salud Pública del C3, el grupo del proyecto de diabetes mellitus tipo 2, yo...

...estamos dispuestos, comprometidos, listos para colaborar, para ofertar conocimientos, herramientas, modelos,...

Se busca colaboradores del sector médico, se busca datos - de estudios clínicos, de estudios genéticos, de estudios de nutrición, de expedientes / historias clínicas - se busca expertise, conocimientos

La diabetes mellitus tipo 2 es una amenaza grave mundial

Hay una enorme cantidad de investigación al respecto, sin embargo sigue en crecimiento

Hemos frenado HIV, hemos frenado el tabacismo en muchas partes...

Estamos fallando con diabetes