Catastrophes, diseases and crimes: risk prediction with mathematics

Matrix Conference 2018

Tuesday, October 30, 2018
CosmoCaixa Barcelona

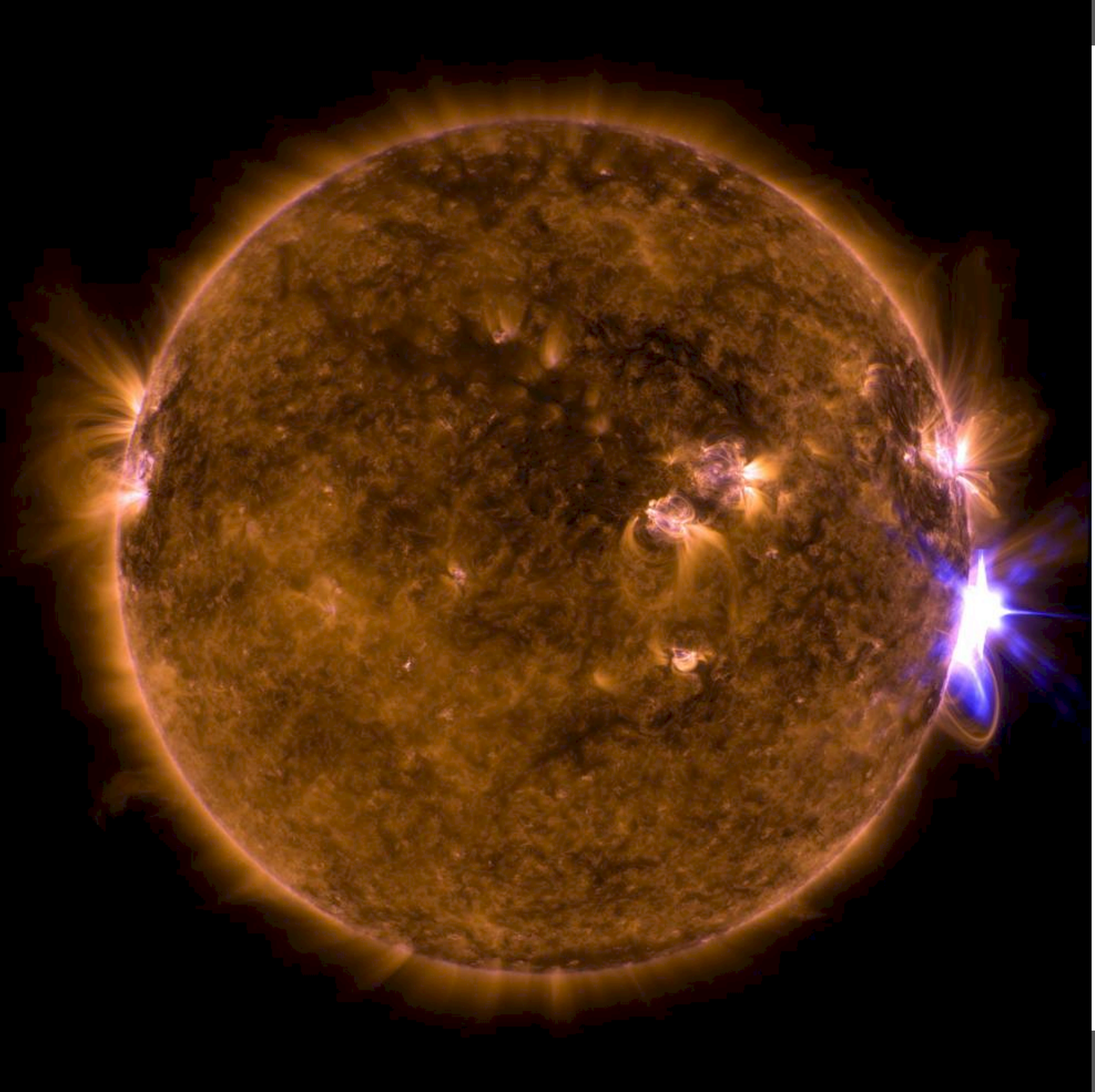
Rosario Delgado
Departament de Matemàtiques
Universitat Autònoma de Barcelona
Barcelona Graduate School of Mathematics

Catastrophes: scenarios with low probability of occurring but with disastrous consequences ...

"Working now on emergency plans, we will be better prepared if and when we need to respond to such an event"

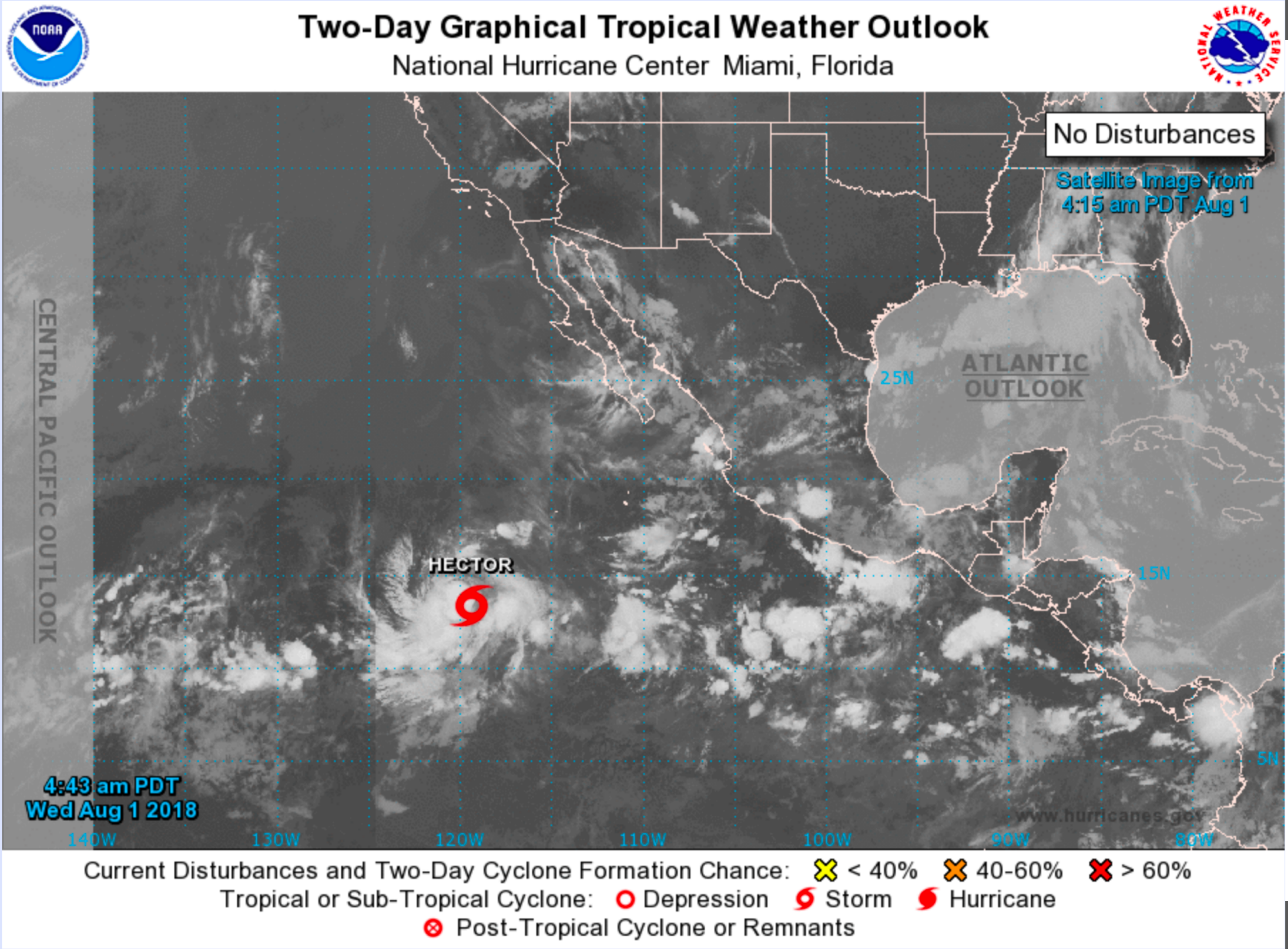
Craig Fugate, Federal Emergency Management Agency, EUA.

Kilauea, the volcano of Hawaii that erupted in May. Since then cast stone, smoke and ash rivers run through the southeast of the Great Island to the waters of the Pacific.



Solar superstorm

According to a researcher at the University of Bristol, it is only a matter of time that an exceptionally violent solar storm (**Carrington event**) seriously affects the Earth.



Hurricanes and tropical storms

The National Oceanic and Atmospheric Administration of the United States (NOAA) predicts that the cyclone season in the Atlantic basin will be "similar or more active than normal" with a 75% probability.

Sand storms

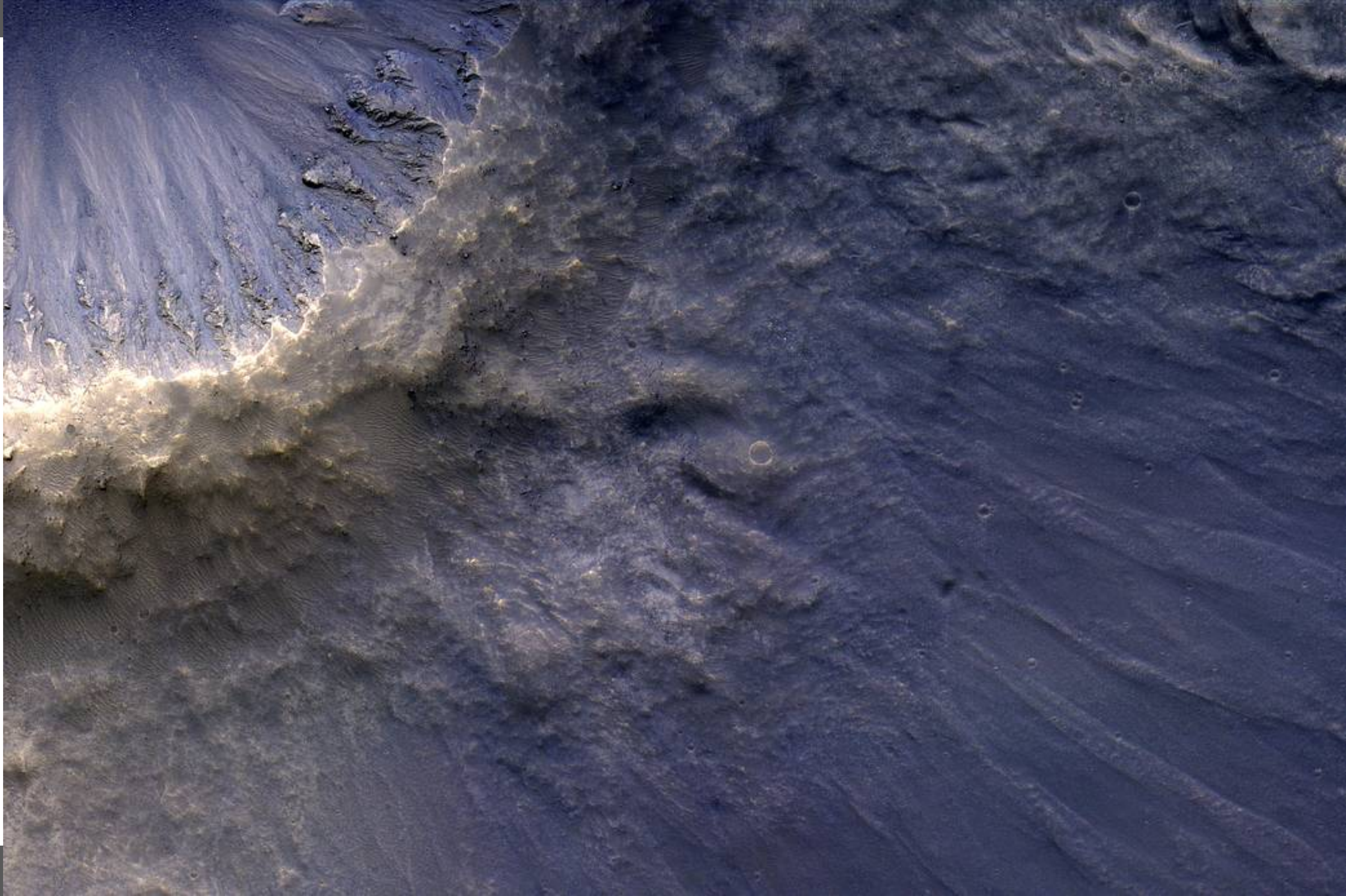
in the UAE in February 2009, with winds of 65 km/h, which dangerously reduce visibility (NASA).

Strong flooding

In July 2010 in Pakistan, caused by monsoon rains, which affected a fifth of its territory (NASA).

Meteorites

Image of a crater formed by the impact of a meteorite on Mars (NASA).



Impacting the Earth

Manicouagan Crater (Canada). Possibly due to the impact of a 5 km meteorite of 5 km of diameter, 215.5 million years ago (NASA).

Can we predict the risk of a catastrophe?

Can we assess how this risk is affected if we take certain measures to try to avoid the catastrophe?

... to start ... What is the risk?

Risk is an event that can have negative consequences.

Instead, an event that can have positive consequences is an **opportunity**

(Committee of Sponsoring Organizations of the Treadway Commission, 2004).

How can we measure the risk?

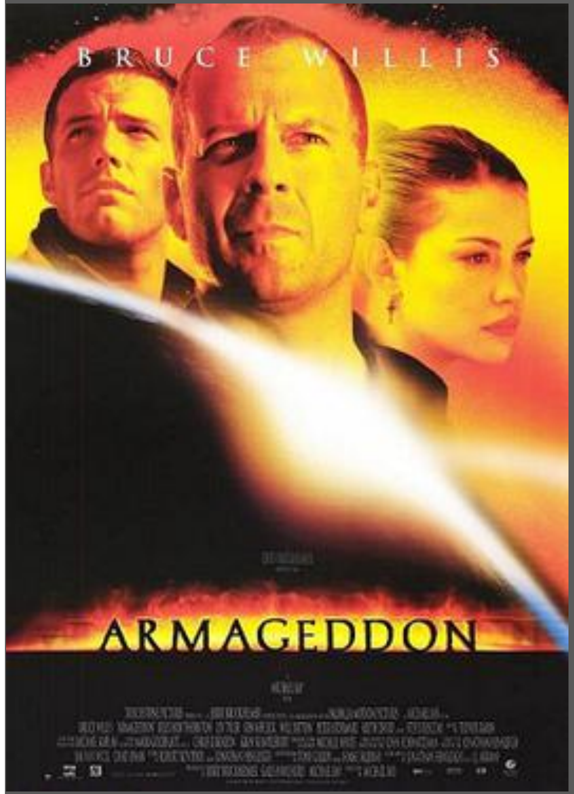
With language abuse, it is also called a **risk** to a numerical measure associated with the event that can have negative consequences.

Traditional Approach (Impact-Based Risk Measure):

The most common is to obtain the risk as a measure that is obtained by multiplying the probability of the event with negative consequences for a measure of its impact (negative).

$$\textbf{Risk} = \boxed{\text{Probability}} \times \boxed{\text{Impact}}$$

Let’s take an exemple... ARMAGEDDON!



This is the title of the 1998 American science fiction disaster film directed by Michael Bay and starred by Bruce Willis, Ben Affleck and Liv Tyler. *Armageddon* is a biblical term used to refer to the end of the world through catastrophes.

Argument: a group of blue-collar deep-core drillers is sent by NASA to stop a gigantic meteorite on a collision course with Earth. The world was confronting a truly massive risk, a truly CATASTROPHE!!

Trying to measure the risk of ARMAGEDDON…

Risk = Probability × Impact

This formula may seem useful to calculate the **risk** ... but it is not!

Why?

Because we can not directly measure nor “probability” nor “impact” without furthering a bit more ... For example, according to NASA scientists, the trajectory of the meteorite goes through the Earth. Therefore, is the “probability” equal to 1? If that were the case, what sense he would have to send someone to try to avoid it?

... We do not get it!

The probability that the meteorite collides against the Earth is, therefore, **conditioned** by other events (such as intervention to try to destroy it).

It does not make sense to assign a probability directly without taking into account the events that can condition it!

Neither can we obtain a measure of the impact. Apart from the obvious question “impact on what?”, We can not measure it without considering the possible mitigating actions (such as letting people in subterranean refuges as far away as possible from the impact zone, ...).

What can we do?

We will build a <u>probabilistic mathematical model</u> that will allow to include all the events that can condition both the probability of the meteorite collision against the Earth, and the impact that this would have.

How will this model be?

A graphical representation of the relationships between the different variables that are relevant in a given situation. In our case, variables that affect the **risk** associated with ARMAGEDDON.

Using it, we can really evaluate this **risk** and take action!

We introduce the model

First we make a simplified version.

Consider two variables to start:

"Collision with the Earth" (Y/N)

"Collision trajectory" (Y/N)

The second condition the first one:

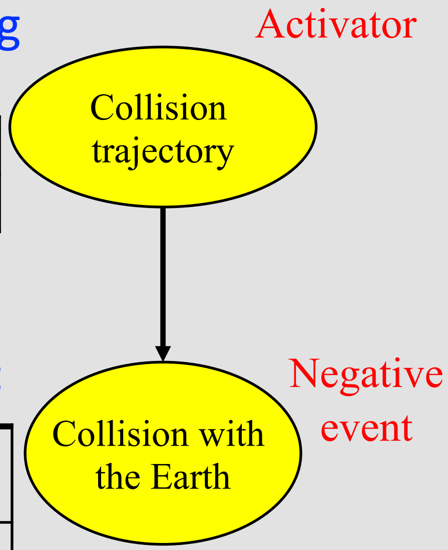
if the trajectory is really collision, the collision will occur, and if not, no.

Probability of success in saying that the trajectory is collision:

Collision trajectory	Yes	0.999
	No	0.001

Probability of conditioned collision:

		Collision trajectory	
		Yes	No
Collision with the Earth	Yes	1.0	0.0
	No	0.0	1.0

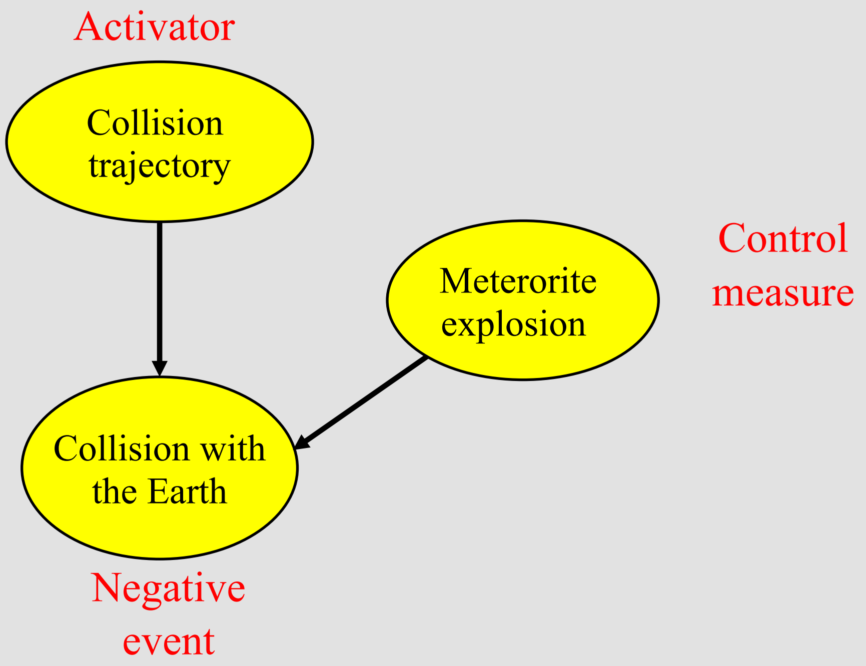


And we're improving it ...

We will now also consider the possible effect of issuing drills to destroy the meteorite.

We have, therefore, a third variable: "Meteorite explosion" (Y/N)

This variable affects the meteorite collision, but not if its corrent trajectory is collision.



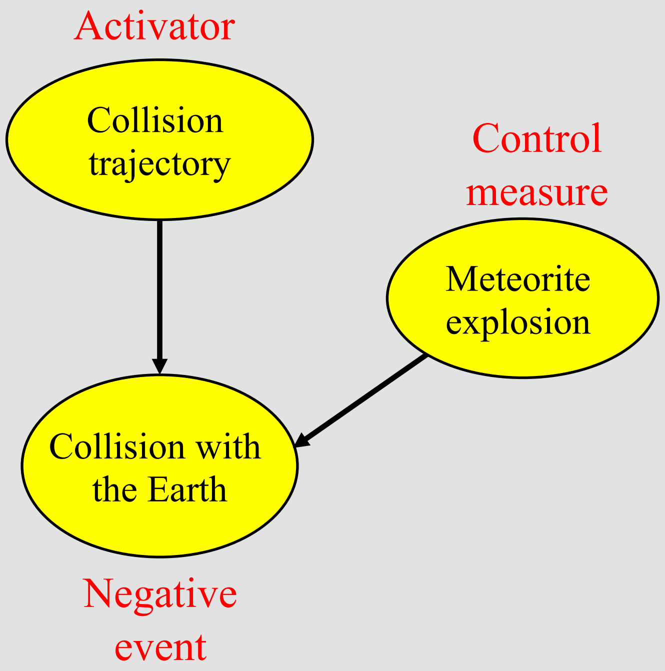
… and improving it…

Probability of success in saying that the trajectory is collision:

Collision trajectory	Yes	0.999
	No	0.001

Probability of conditioned collision:

	Collision trajectory:	Yes		No	
	Meteorite explosion:	Yes	No	Yes	No
Collision with the Earth	Yes	0.2	1.0	0.0	0.0
	No	0.8	0.0	1.0	1.0



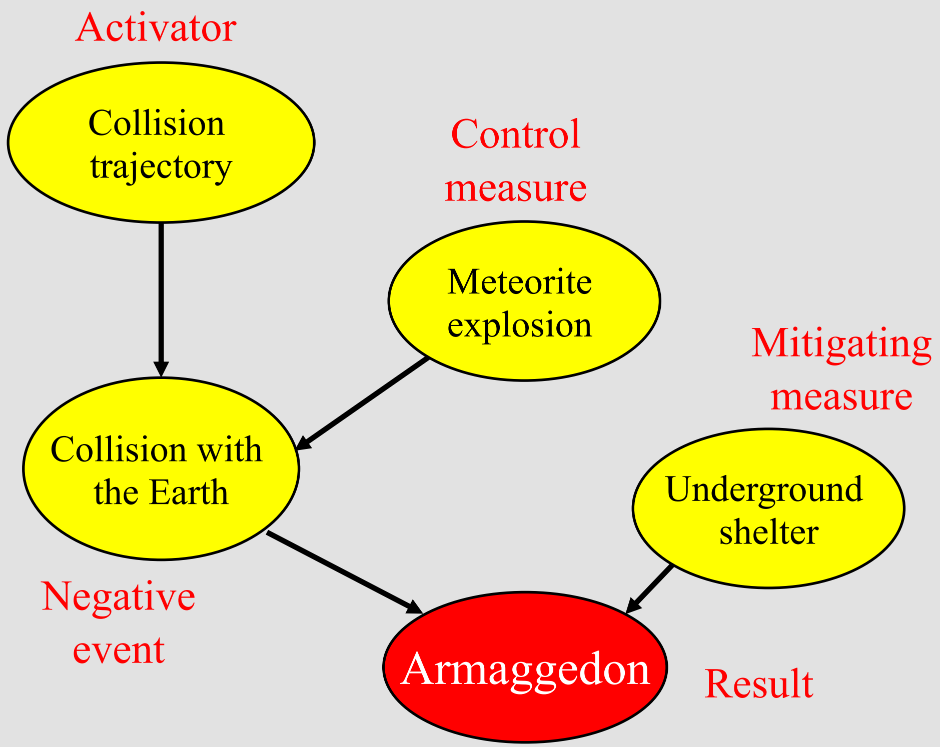
Probability of success of the drill team exploding the meteorite:

Meteorite explosion	Yes	0.10
	No	0.90

… and improving it…

Finally we have the mitigating effect of the negative consequences of the collision of the meteorite, which would be obtained by refuging the population underground, with the variable
"Underground shelter" (Y/N),

which affects the end result (mass loss of human lives, 80%), which is the variable "Armageddon" (Y/N).

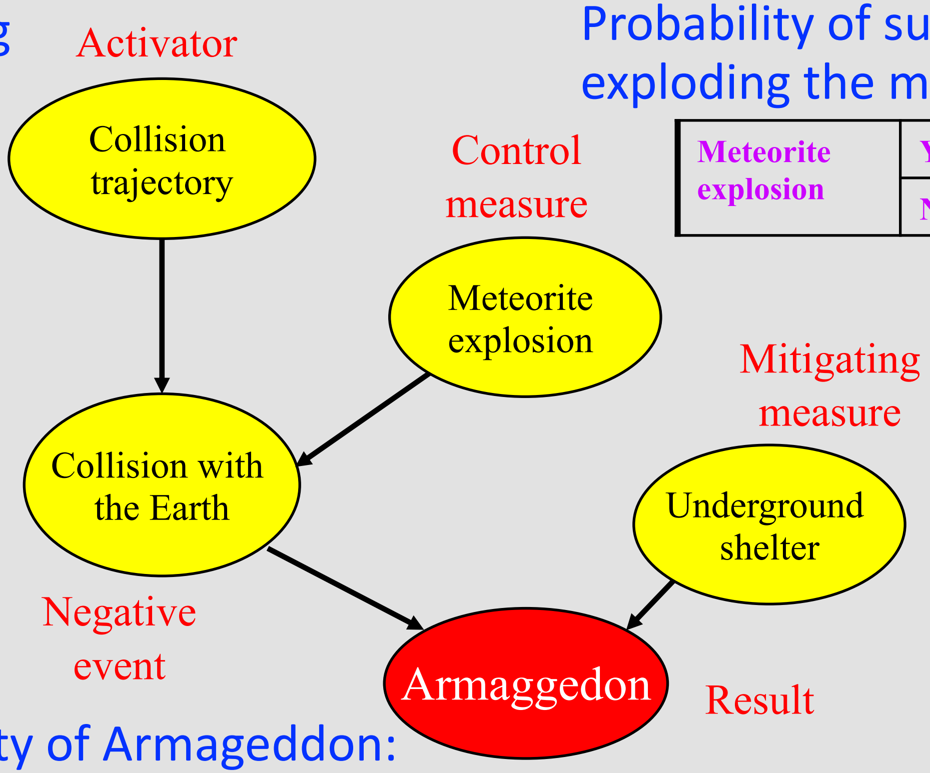


Probability of success in saying that the trajectory is collision:

Collision trajectory	Yes	0.999
	No	0.001

Probability of conditioned collision:

	Collision trajectory:	Yes		No	
	Meteorite explosion:	Yes	No	Yes	No
Collision with the Earth	Sí	0.2	1.0	0.0	0.0
	No	0.8	0.0	1.0	1.0



Probability of success of the drill team exploding the meteorite:

Meteorite explosion	Yes	0.10
	No	0.90

Probability of success in sheltering the population:

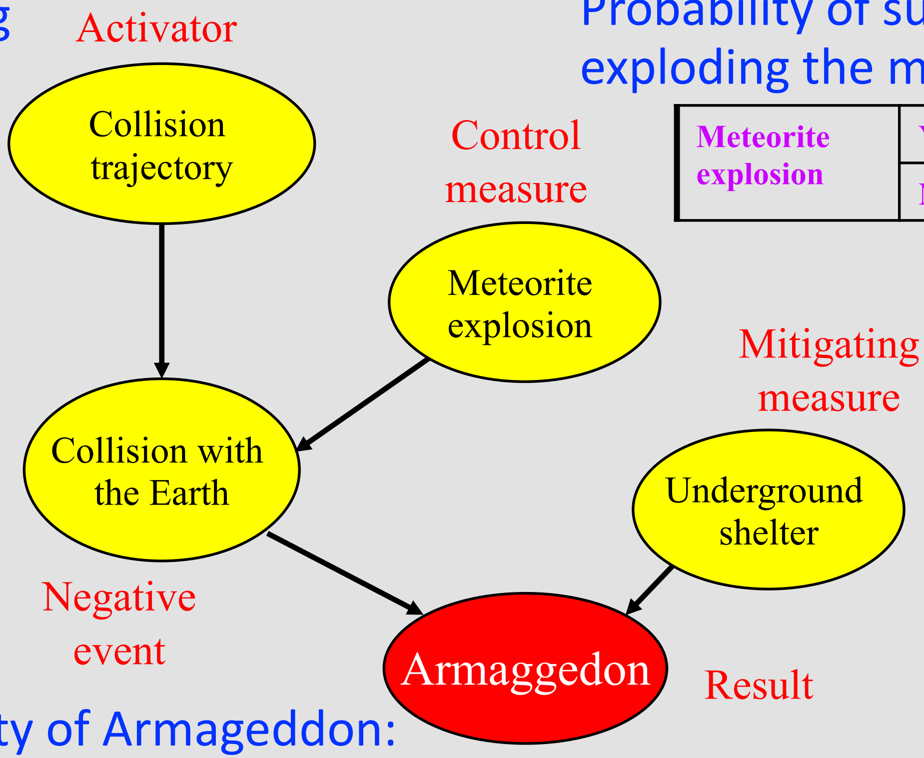
Underground shelter	Yes	0.30
	No	0.70

Probability of Armageddon:

	Collision with the Earth:	Yes		No	
	Underground shelter:	Yes	No	Yes	No
Armaggedon (80% loss of human lives)	Yes	0.6	1.0	0.0	0.0
	No	0.4	0.0	1.0	1.0

Probability of success in saying that the trajectory is collision:

Collision trajectory	Yes	0.999
	No	0.001



Probability of success of the drill team exploding the meteorite:

Meteorite explosion	Yes	0.10
	No	0.90

Probability of conditioned collision:

	Collision trajectory:	Yes		No	
	Meteorite explosion:	Yes	No	Yes	No
Collision with the Earth	Sí	0.2	1.0	0.0	0.0
	No	0.8	0.0	1.0	1.0

Probability of success in sheltering the population:

Underground shelter	Yes	0.30
	No	0.70

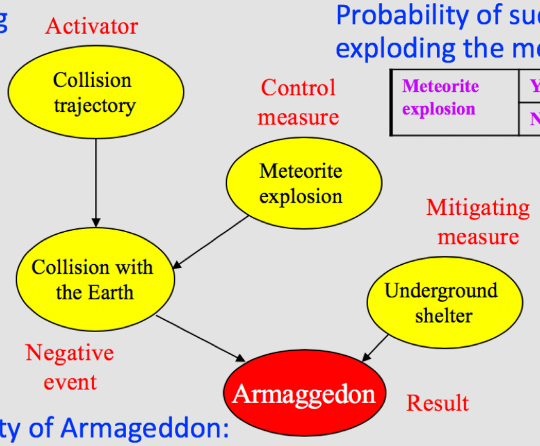
Probability of Armageddon:

	Collision with the Earth:	Yes		No	
	Underground shelter:	Yes	No	Yes	No
Armaggedon (80% loss of human lives)	Yes	0.6	1.0	0.0	0.0
	No	0.4	0.0	1.0	1.0

Bayesian network

Probability of success in saying that the trajectory is collision:

Collision trajectory	Yes	0.999
	No	0.001



Probability of success of the drill team exploding the meteorite:

Meteorite explosion	Yes	0.10
	No	0.90

Probability of conditioned collision:

	Collision trajectory:	Yes		No	
	Meteorite explosion:	Yes	No	Yes	No
Collision with the Earth	Sí	0.2	1.0	0.0	0.0
	No	0.8	0.0	1.0	1.0

Probability of success in sheltering the population:

Underground shelter	Yes	0.30
	No	0.70

Probability of Armageddon:

	Collision with the Earth:	Yes		No	
	Underground shelter:	Yes	No	Yes	No
Armaggedon (80% loss of human lives)	Yes	0.6	1.0	0.0	0.0
	No	0.4	0.0	1.0	1.0

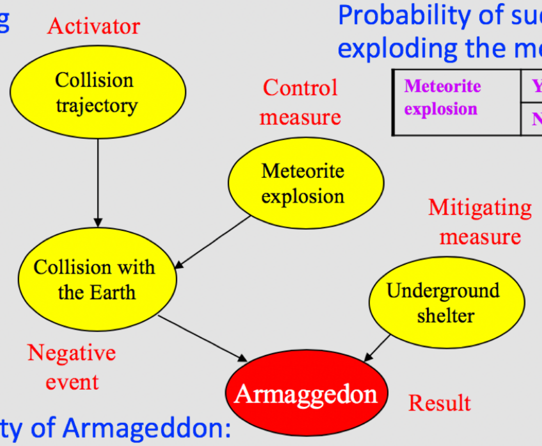
A **Bayesian network** is a probabilistic mathematical model that represents the relationships (subjected to chance) between variables of interest to a given situation.

The model consists of:

1. a directed acyclic graph, and
2. some parameters, which are the probabilities of the tables.

Probability of success in saying that the trajectory is collision:

Collision trajectory	Yes	0.999
	No	0.001



Probability of success of the drill team exploding the meteorite:

Meteorite explosion	Yes	0.10
	No	0.90

Probability of conditioned collision:

	Collision trajectory:	Yes		No	
	Meteorite explosion:	Yes	No	Yes	No
Collision with the Earth	Sí	0.2	1.0	0.0	0.0
	No	0.8	0.0	1.0	1.0

Probability of success in sheltering the population:

Underground shelter	Yes	0.30
	No	0.70

Probability of Armageddon:

	Collision with the Earth:	Yes		No	
	Underground shelter:	Yes	No	Yes	No
Armaggedon (80% loss of human lives)	Yes	0.6	1.0	0.0	0.0
	No	0.4	0.0	1.0	1.0

We will use this model for:

- estimate the risk in a given scenario,
- compare different scenarios.

A **Bayesian network** is a probabilistic mathematical model that represents the relationships (subjected to chance) between variables of interest to a given situation.

The model consists of:

1. a directed acyclic graph, and
2. some parameters, which are the probabilities of the tables.

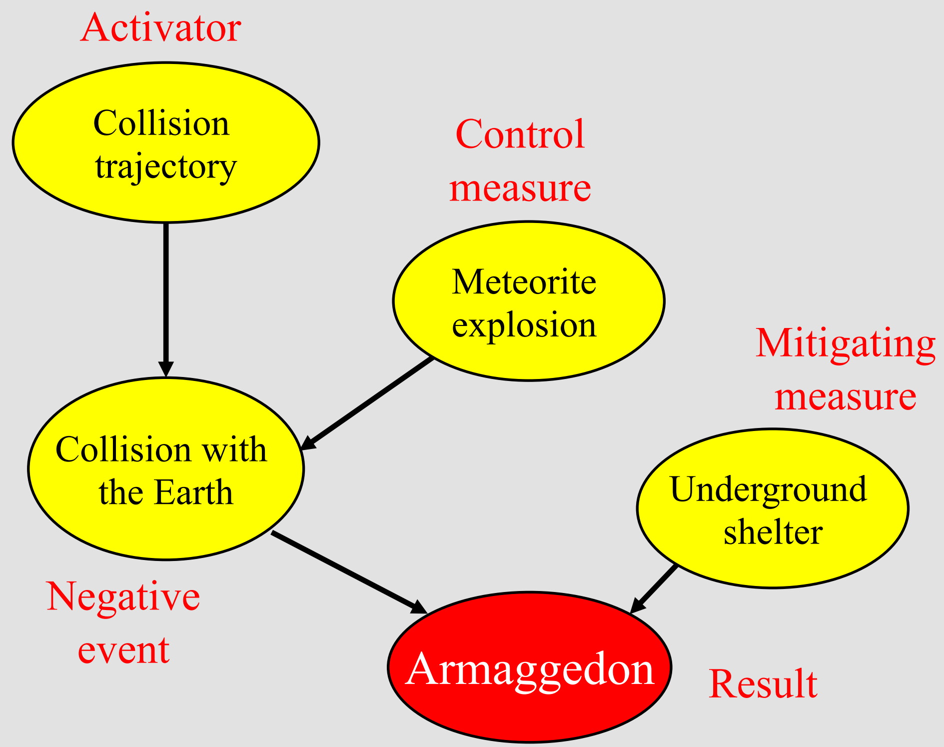
Estimating the risk in different scenarios

Scenario 0:

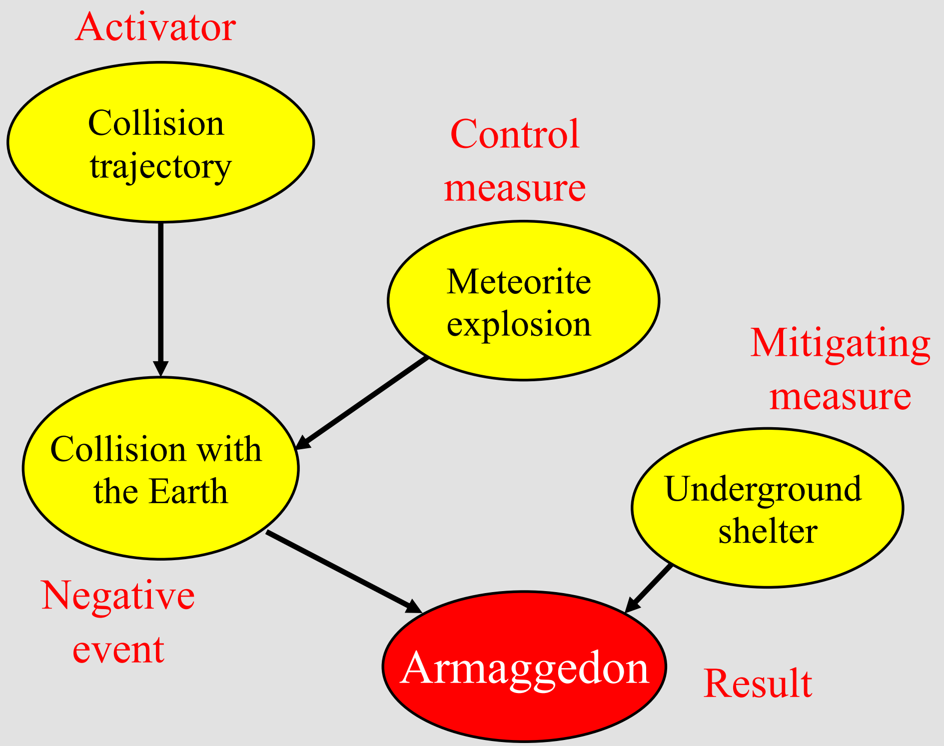
What is the risk of Armaggedon "a priori" (if we do not have any more information)?

We have to calculate, therefore, the probability that the Armaggedon variable is "Yes".

But this probability will depend on the values of the parent variables: Collision with the Earth, and Underground shelter.

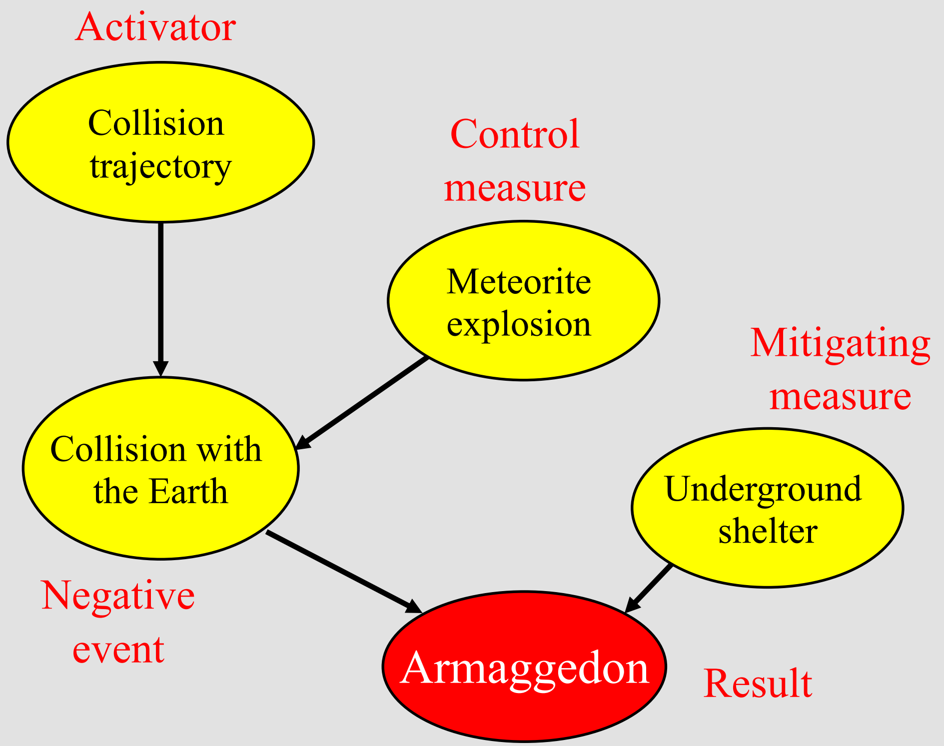


Estimating the risk in different scenarios



Armaggedon risk estimation		
A priori (Scenario 0)	**81%**	
A posteriori	Scenario 1 (expedition fails)	Scenario 2 (expedition success)
	88%	**18%**
	↗ **x 1.1**	↘ **x 0.22**

Estimating the risk in different scenarios



Armaggedon risk estimation		
A priori (Scenario 0)	**81%**	
A posteriori	Scenario 1 (expedition fails)	Scenario 2 (expedition success)
	88%	**18%**
	↗ **x 1.1**	↘ **x 0.22**

Was the model needed to reach this conclusion?

Qualitatively: It was not necessary, it is logical enough!

Quantitatively: Yes! The model allows quantifying the increase or decrease of risk in different scenarios (evidences).

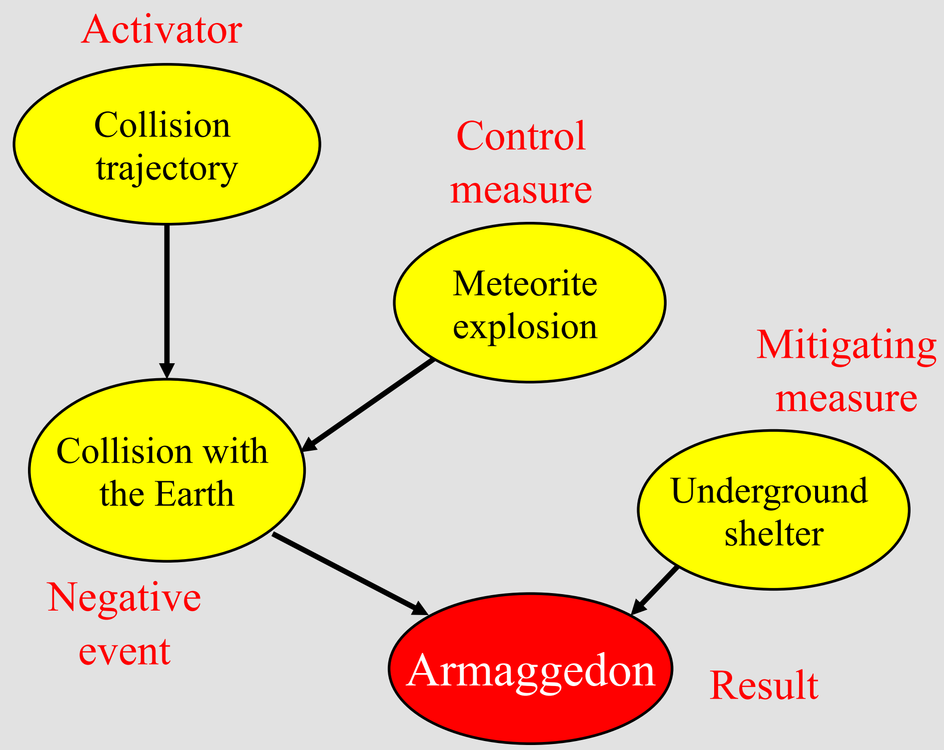
But the scenarios can be complicated and the intuition fails to guide us...

What will further reduce the risk: to succeed in the meteorite explosion but not to refuge people on time, or to fail in the explosion but to refuge the population?

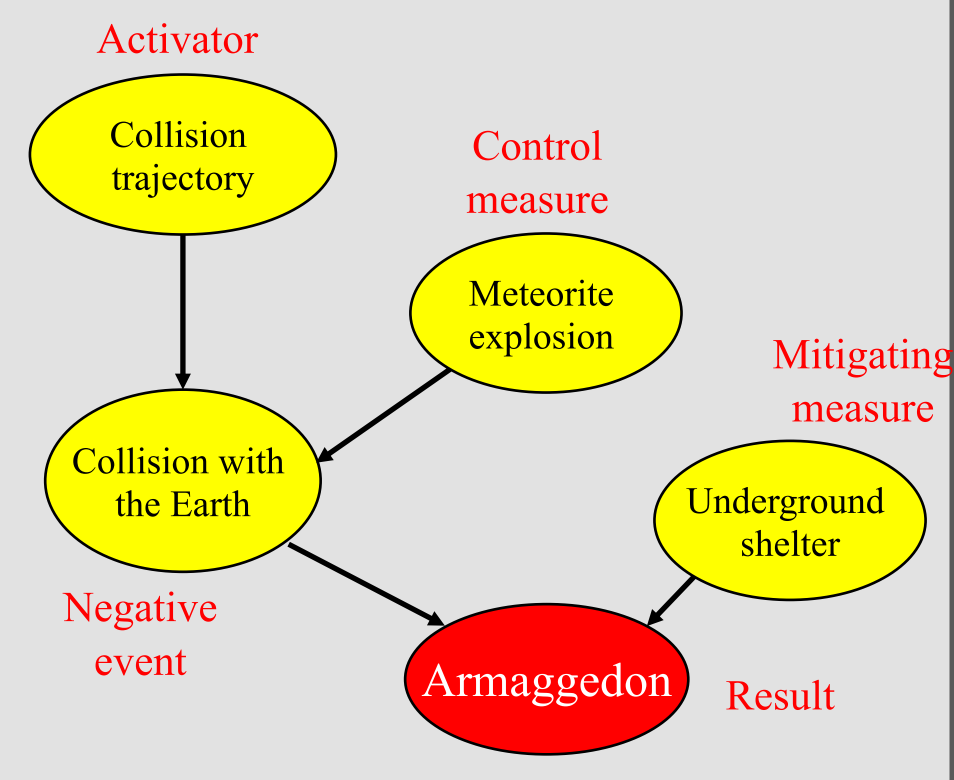
The intuition does not know what to say ... but the model does!

<u>Scenario 3</u>: Explosion=Yes, Underground shelter=No.
<u>Scenario 4</u>: Explosion=No, Underground shelter=Yes.

Calculate the risk in both scenarios.



But the scenarios can be complicated and the intuition fails to guide us...



Armaggedon risk estimation		
A priori (Scenario 0)	**81%**	
A posteriori	Scenario 3 (Explosion=Yes, Undergroung shelter=No)	Scenario 4 (Explosion=No, Underground shelter=Yes)
	↘ **20%**	↘ **60%**
	x 0.25	**x 0.74**

We see that the risk decreases more being successful in the explosion (control measure) than refuging the population (mitigating measure), if we have to choose.

But the scenarios can be complicated and the intuition fails to guide us...

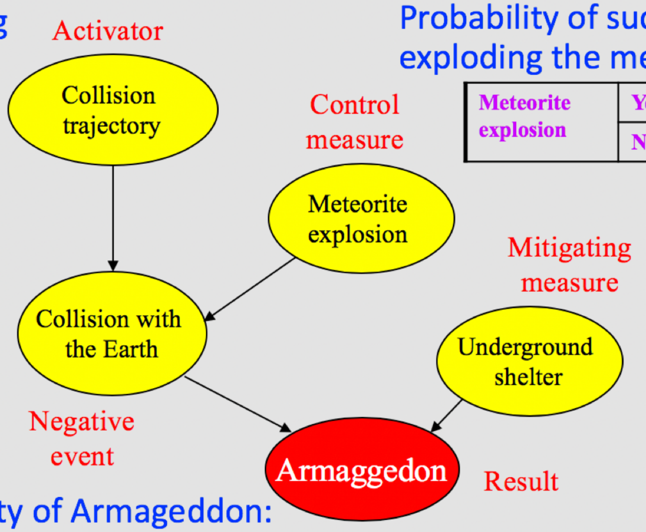
And if we did not have to choose? Of course, it would be better to be successful in both measures, but ...
How to "better"? How would this reduce your risk?

Armaggedon risk estimation			
A priori (Scenario 0)	**81%**		
A posteriori	Scenario 3 (Explosion=Yes, Underground shelter=No)	Scenario 4 (Explosion=No, Undergroung shelter=Yes)	Scenario 5 (Explosion=Yes, Underground shelter=Yes)
	↘ **20%**	↘ **60%**	↘ **12%**
	x 0.25	**x 0.74**	**x 0.15**

Is the model useful and practical?

Probability of success in saying that the trajectory is collision:

Collision trajectory	Yes	0.999
	No	0.001



Probability of success of the drill team exploding the meteorite:

Meteorite explosion	Yes	0.10
	No	0.90

Probability of conditioned collision:

	Collision trajectory:	Yes		No	
	Meteorite explosion:	Yes	No	Yes	No
Collision with the Earth	Sí	0.2	1.0	0.0	0.0
	No	0.8	0.0	1.0	1.0

Probability of success in sheltering the population:

Underground shelter	Yes	0.30
	No	0.70

Probability of Armageddon:

	Collision with the Earth:	Yes		No	
	Underground shelter:	Yes	No	Yes	No
Armaggedon (80% loss of human lives)	Yes	0.6	1.0	0.0	0.0
	No	0.4	0.0	1.0	1.0

Useful? We have seen that yes.

Practical? Assuming the relationships between the variables, we need to know the **probabilities** (parameters).

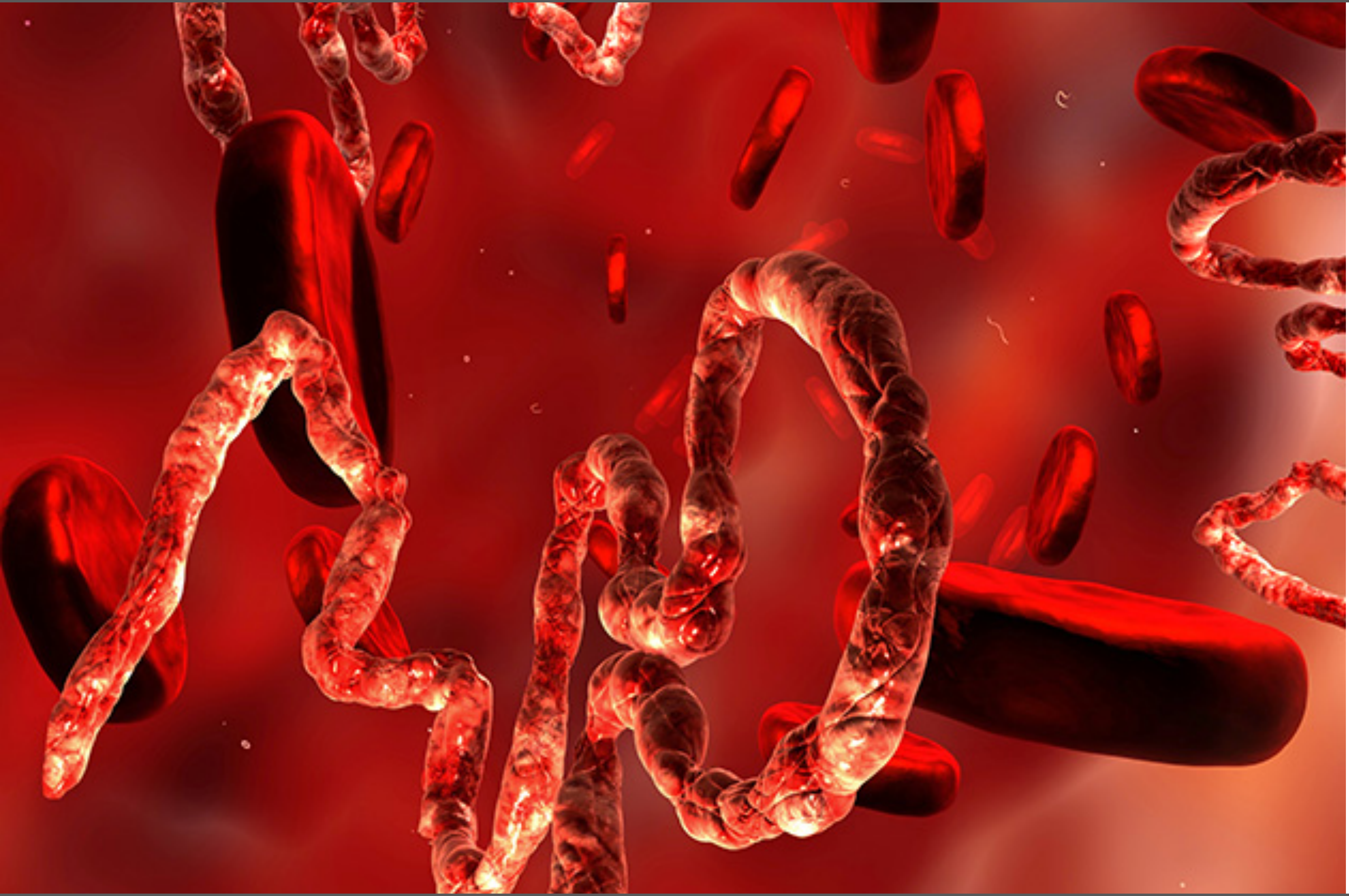
How to do it if we do not have historical data to estimate them?

Another more practical example: the risk of an accident or illness

What are the variables that increase the risk of suffering the disease (risk factors)? Which ones that reduce the risk of suffering it (protection factors)?

The answers allow us to:

- Influence the prevention of the disease.
- Improve the diagnosis.
- Improve resources management.



Infection with the Ebola virus

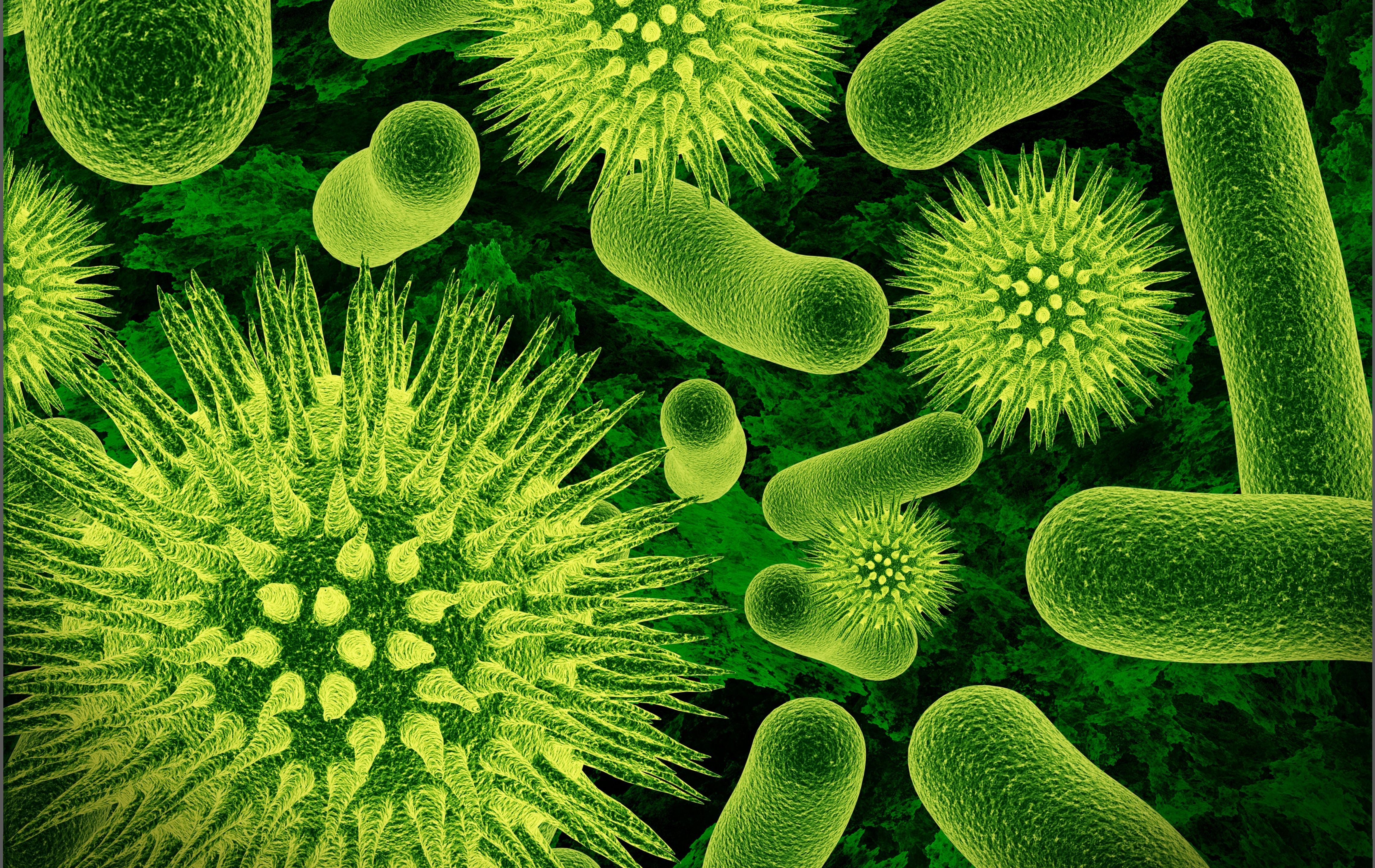
The mortality rate (% of patient that die among those infected) of the disease is between 50% and 90%.

There is no specific treatment.

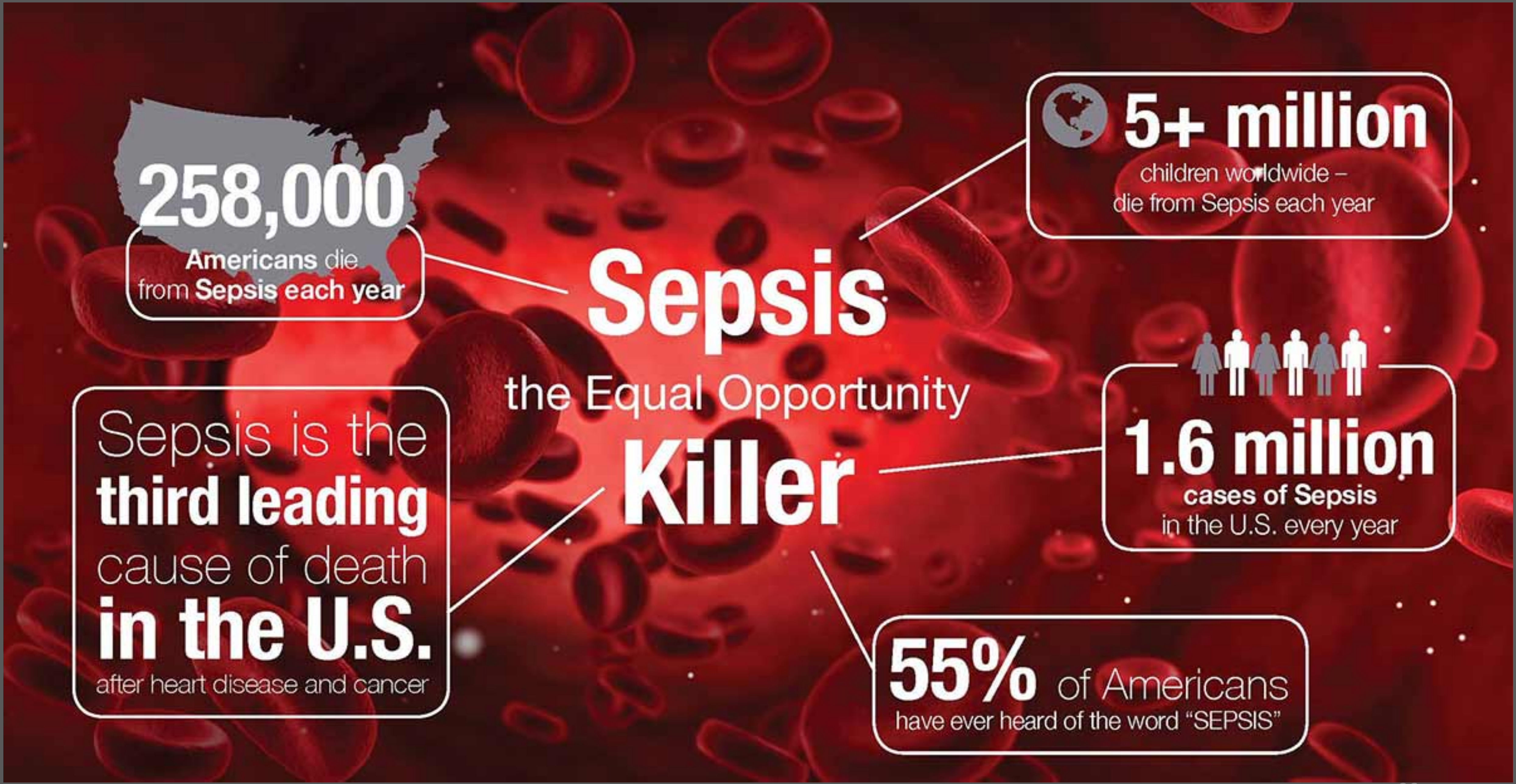
Since 2015, a vaccine is being worked on.

AIDS virus

In 2016 around 36.7 million people in the world were infected, of which 1 million died.
There is no cure or vaccine, only palliative treatments and to prevent new infections.



Flu Virus


258,000
Americans die from Sepsis each year
Sepsis
the Equal Opportunity
Killer
5+ million
children worldwide – die from Sepsis each year
1.6 million
cases of Sepsis in the U.S. every year
Sepsis is the third leading cause of death in the U.S.
after heart disease and cancer
55% of Americans
have ever heard of the word "SEPSIS"

Sepsis: What is it?

Occurs when a localized infection ...

- spreads and passes into the blood,
- comes to other organs,
- causes an exaggerated inflammatory response, a multi-organ failure and, in many cases, death.

Information from the 2017 “La Marató de TV3”, dedicated to infectious diseases.

Sepsis: Consequences

- ✓Patients who recover often have sequels.
- ✓In Catalonia, 10 people die every day due to severe sepsis.
- ✓Sepsis is the leading cause of death due to infection in the world.

Information from the 2017 “La Marató de TV3”, dedicated to infectious diseases.

Sepsis: Challenges for the future

- ✓**Antibiotics** able to fight resistant bacteria, and **vaccines**.
- ✓**Diagnostic** tools for rapid treatment.
- ✓Tools for the **vital** and **functional prognosis** of patients with sepsis.

Information from the 2017 “La Marató de TV3”, dedicated to infectious diseases.

Sepsis: let’s fight against infectious diseases

Rainer Maria Rilke (1875-1926) was an Austro-Hungarian poet, considered one of the most important in German language and universal literature. He will die because of an infection provoked by the prick of a rose.

STAF
Sepsis Training, Analysis and Feedback

Project chosen to be funded by the Foundation La Marató of TV3

Collaboration between various hospitals, the Catalan Health Service and other agencies related to healthcare on the one hand, and the university of the other.

El cicle de La Marató 2017

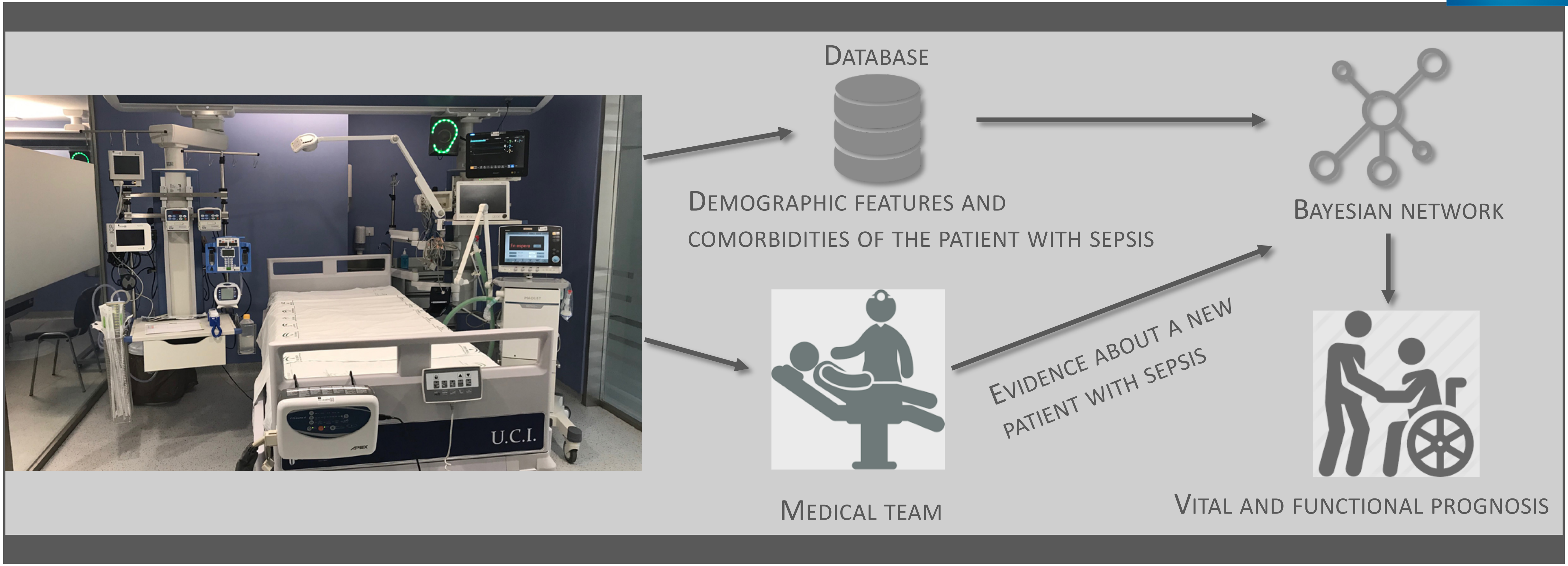
> **Desembre 2017**
La Marató de TV3 i Catalunya Ràdio

> **2018**
Concurs d'ajudes a la recerca i selecció dels projectes que es finançaran

> **2019 – 2022**
Desenvolupament dels projectes

> **2023**
Resultats de la recerca, exposats en un simposi

STAF
Sepsis Training, Analysis and Feedback



Diagnosing Alzheimer's disease from oral discourse

Alzheimer's disease (AD) is a type of dementia that causes problems of memory, speech and behavior.

Much impact on the elderly.

ELSEVIER

CrossMark

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 95 (2016) 168 – 174

Procedia
Computer Science

Complex Adaptive Systems, Publication 6
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2016 - Los Angeles, CA

A Machine Intelligence Designed Bayesian Network Applied to Alzheimer's Detection Using Demographics and Speech Data

Walker H. Land[a*], J. David Schaffer[b]

[a] *Retired Emeritus Research Professor,Department of Bioengineering, Binghamton University, Binghamton, NY 13902 USA*
[b] *Institute for Multigenerational Studies, Binghamton University, Binghamton, NY 13902 USA*

Objective:

Improve the diagnosis of AD based on minimal clinical data and a sample of the discourse of the individual, using Bayesian networks.

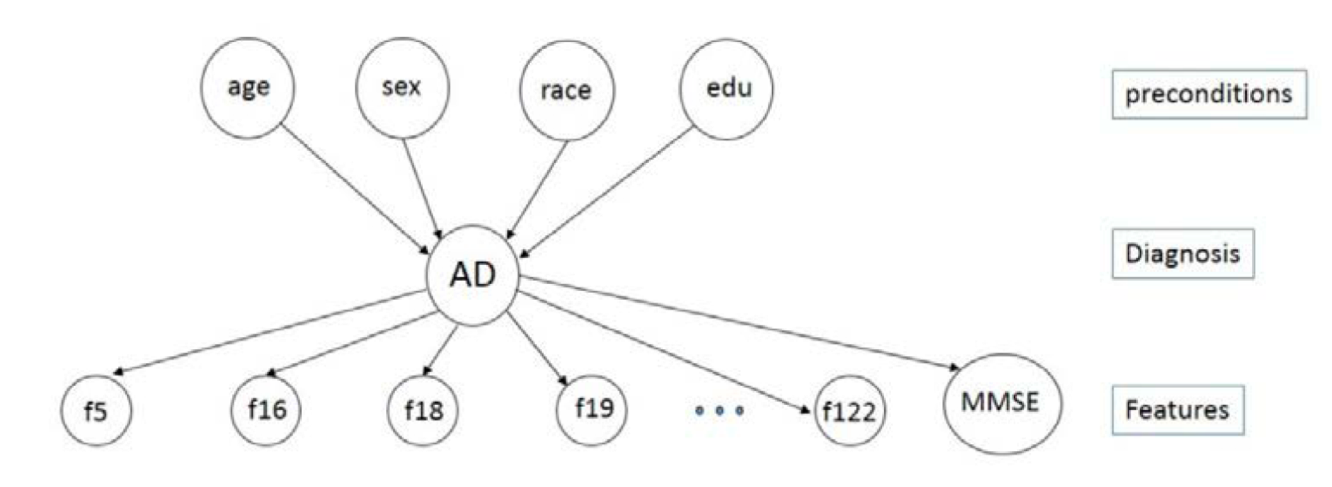
Pilot study: Speech samples of 210 individuals.
98 with diagnosed Alzheimer's disease (AD). 112 cognitively normal individuals (controls).

Features from the speech: 118.

Demographic features: Age, Sex, Race, Educational Level.

MMSE: mini-mental state exam.

The model (Bayesian network)



There is a computer program that makes the model “learn” from the data, using **statistical methods**: **Machine Learning.**

Then the model is validated and can already be used to predict the risk of Alzheimer's disease of an individual based on

- the characteristics from his/her speech,
- the MMSE exam and/or
- demographic characteristics.

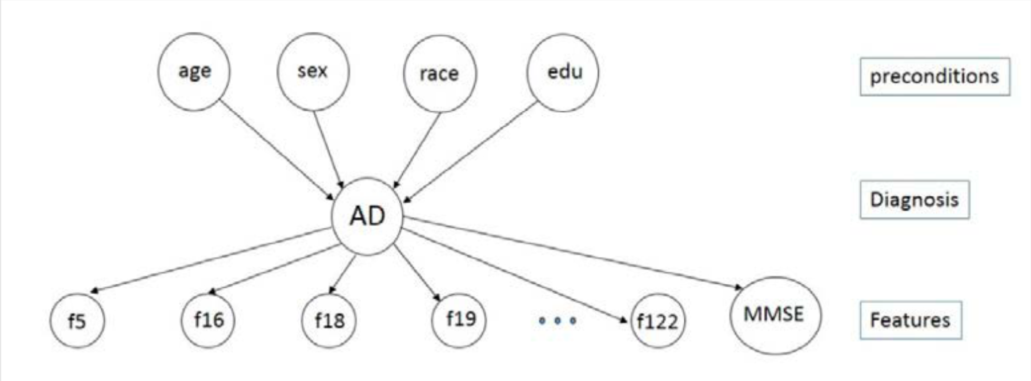
Estimating the risk of Alzheimer's disease

A priori risk

Alzheimer's risk by age	
Age	**P(AD / Age)**
< 65	4 %
65 - 74	15 %
75 - 84	**43 %**
> 84	38 %

A posteriori risk

Alzheimer's risk by MMSE	
MMSE	**P(AD / MMSE)**
< 21	**100 %**
21 - 25	79 %
> 25	14 %



Having a low score on the MMSE exam is a very important risk factor!

Estimating the risk of Alzheimer's disease in different scenarios: making diagnosis

Individual with evidence:

Age < 65
Educational level: High
Features from the speech... (known)
MMSE >25

(It is known that the individual has Alzheimer's disease)

Estimating the risk of Alzheimer's disease in different scenarios: making diagnosis

Individual with evidence:

Alzheimer's risk P(AD / Evidence) = 0.988

Age < 65
Educational level: High
Features from the speech... (known)
MMSE >25

(It is known that the individual has Alzheimer's disease)

Estimating the risk of Alzheimer's disease in different scenarios: making diagnosis

Individual with evidence:

Age < 65
Educational level: High
Features from the speech... (known)
MMSE >25

(It is known that the individual has Alzheimer's disease)

Alzheimer's risk P(AD / Evidence) = 0.988

If you did not know the result of the MMSE exam:

Alzheimer's risk by MMSE

MMSE	P(AD / MMSE)
< 21	**100 %**
21 – 25	99.99 %
> 25	98.84 %

Estimating the risk of Alzheimer's disease in different scenarios: making diagnosis

Individual with evidence:

Age < 65
Educational level: High
Features from the speech... (known)
MMSE >25

(It is known that the individual has Alzheimer's disease)

Alzheimer's risk P(AD / Evidence) = 0.988

If you did not know the result of the MMSE exam:

This tells us that the features of the speech reveal Alzheimer's even without the most important neuro-psychological examination, which is the MMSE!

Alzheimer's risk by MMSE

MMSE	P(AD / MMSE)
< 21	**100 %**
21 – 25	99.99 %
> 25	98.84 %

Estimating the risk of Alzheimer's disease in different scenarios: making diagnosis

Individual with evidence:

Age < 65
Educational level: High
Features from the speech... (known)
MMSE >25

(It is known that the individual has Alzheimer's disease)

Alzheimer's risk P(AD / Evidence) = 0.988

If, in addition, the result of the feature from the discourse f_{11} was not known:

Alzheimer's risk by MMSE

MMSE	P(AD / MMSE)
< 21	**100 %**
21 – 25	12.92 %
> 25	0.07 %

Estimating the risk of Alzheimer's disease in different scenarios: making diagnosis

Individual with evidence:

Age < 65
Educational level: High
Features from the speech... (known)
MMSE >25

(It is known that the individual has Alzheimer's disease)

Alzheimer's risk P(AD / Evidence) = 0.988

If, in addition, the result of the feature from the discourse f_{11} was not known:

It turns out that f_{11} is essential in order to identify this individual as Alzheimer's disease sufferer. If it is not known, the risk can go down much, depending on the value of the MMSE!

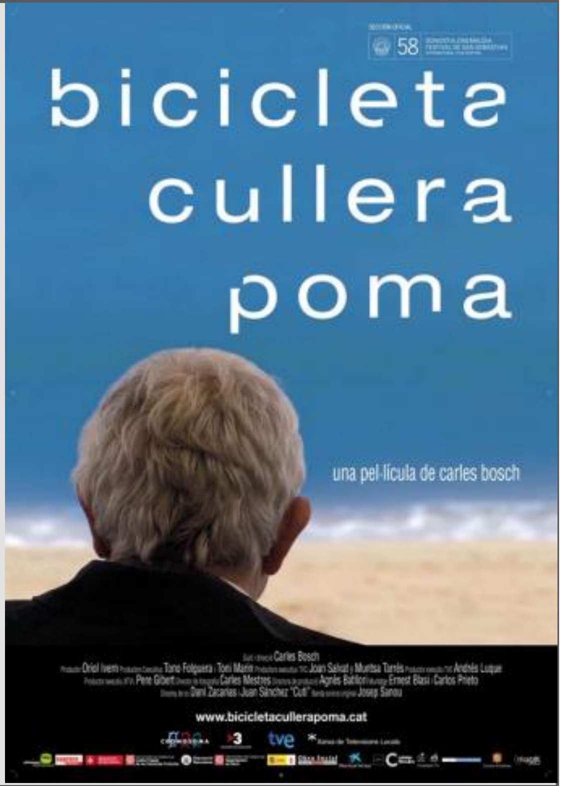
Alzheimer's risk by MMSE

MMSE	P(AD / MMSE)
< 21	100 %
21 – 25	12.92 %
> 25	0.07 %

Conclusions of this study

The model (Bayesian network) allows to:

- Evaluate the risk of an individual suffering from Alzheimer's disease based on his/her features (diagnosis).
- Find profiles of individuals at greater risk of suffering from the disease.
- Analyze the sensitivity of the diagnosis, that is to say, to what extent the diagnosis is sensitive to certain characteristics. For example, to particular aspects of oral discourse, such as f_{11}.



The "criminal profile"

It is a prediction of the characteristics of a not yet identified author of a crime or series of crimes, especially homicides and rapes, but also thefts or fires.

- ✓ Biographical (age, gender, marital status, ...)
- ✓ Socioeconomic (level of studies, economic level, types of work, ...)
- ✓ Lifestyle (with whom he/she lives, sociability, addictions ...)
- ✓ Place of residence, work, ...

The ”criminal profile”

It is a prediction of the characteristics of a not yet identified author of a crime or series of crimes, especially homicides and rapes, but also thefts or fires.

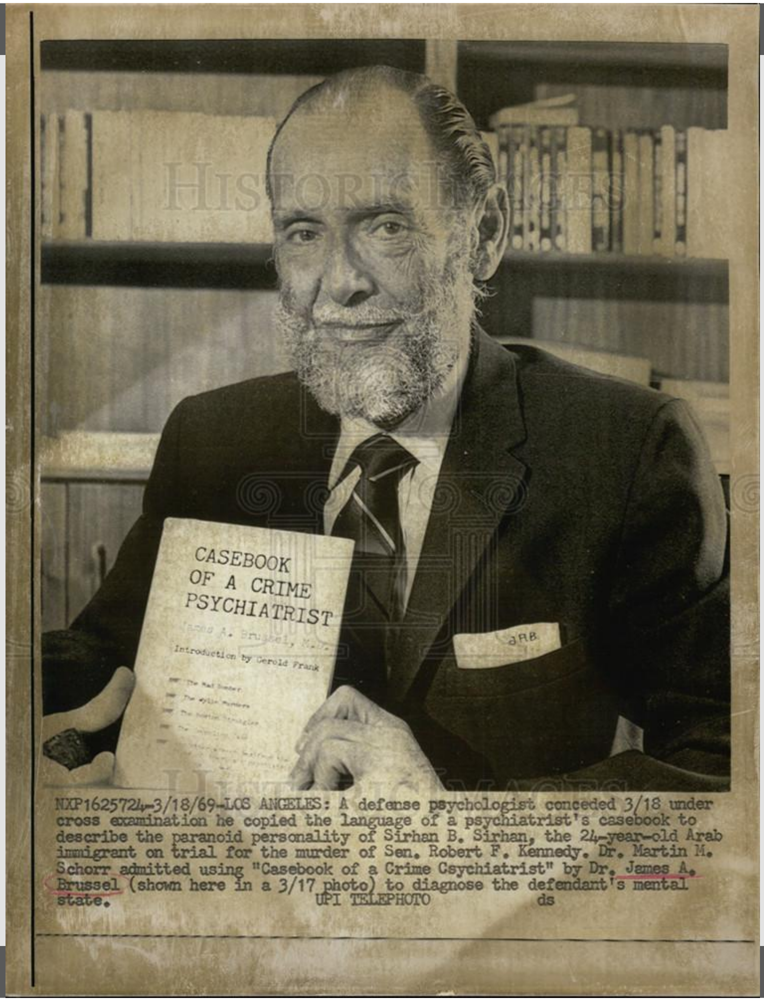
- ✓ Biographical (age, gender, marital status, ...)
- ✓ Socioeconomic (level of studies, economic level, types of work, ...)
- ✓ Lifestyle (with whom he/she lives, sociability, addictions ...)
- ✓ Place of residence, work, ...

- ➢ It helps the researchers in their inquires, reducing the number of research channels to follow when investigating a crime, and focusing the police action.
  - ➢ Policies with specific training in this area,
  - ➢ Psychologists,
  - ➢ Psychiatrists or
  - ➢ Criminologists,
- ➢ It integrates knowledge in the fields of psychology, sociology and forensic medicine.

The "criminal profile"

The first documented case of profiling was that of Dr. James S. Brussel, psychiatrist of New York.

In 1956 he got the profile of the so-called "mad bomber", who had put bombs in the city since 1940.



Brussel said he had used:

- ✓ the deductive reasoning,
- ✓ his experience, and
- ✓ the probability calculus.

This success had a lot of repercussions and changed criminal investigation forever more!

The “criminal profile”

The contribution of the FBI

From the 1970s, the criminal profile technique began to be used on a regular basis, especially from the FBI training centre in Quantico, which created the Behavioural Sciences Unit (BSU).

FBI profilers who have become famous: especially, Robert Ressler.

The “criminal profile”

The contribution of the FBI

From the 1970s, the criminal profile technique began to be used on a regular basis, especially from the FBI training centre in Quantico, which created the Behavioural Sciences Unit (BSU).

FBI profilers who have become famous: especially, Robert Ressler.

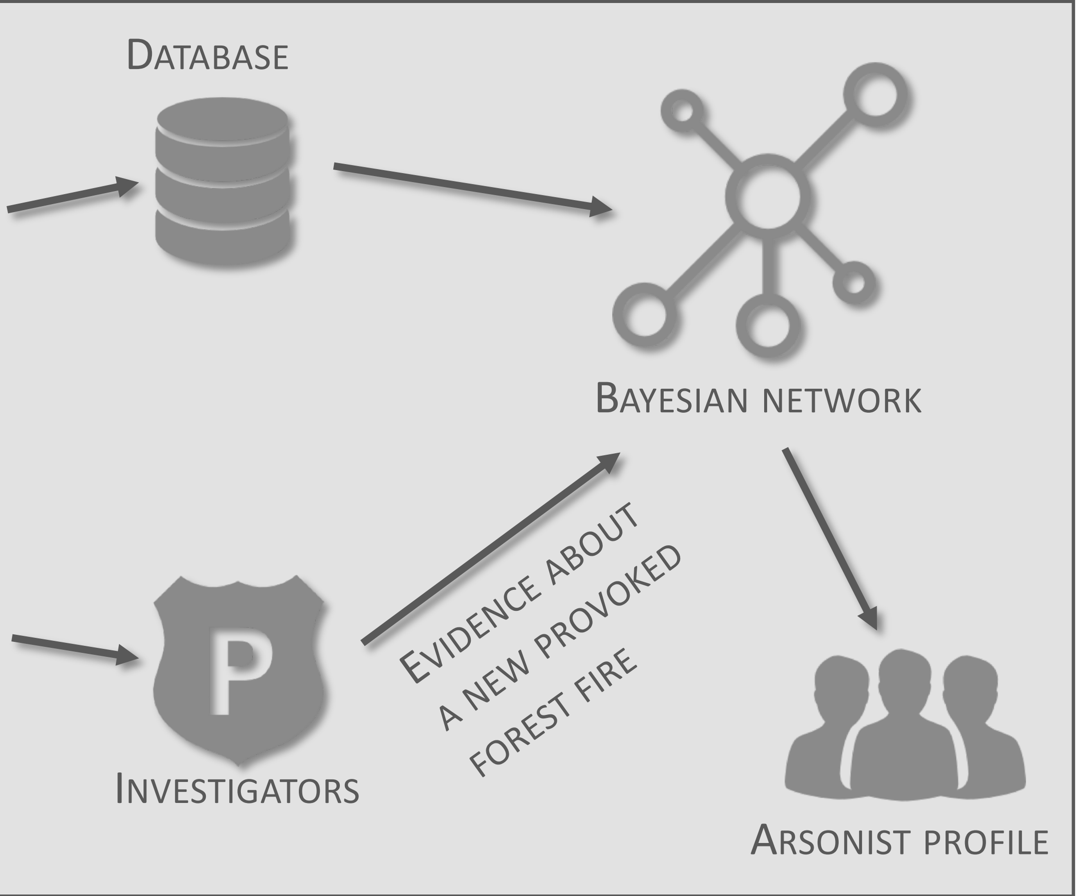
✓The availability of databases on crimes, and
✓the use of powerful computers,
nowadays allow researchers to use **Machine Learning** techniques to create support tools, such as **Bayesian networks**.

An example: profile of forest arsonist

Real research in collaboration with the **Sección de Análisis del Comportamiento Delictivo de la Guardia Civil** and the **SES.**

Construction of a computer application, **PerfilNet.Pyros**, for the **Fiscalía de Medioambiente**, for the profiling of forest arsonists in Spain.

An example: profile of forest arsonist



Motivation

Forest fires are a serious environmental problem.

14 489 forest/year 2005-2014

Approximately 60% of forest fires are caused.

The clarification rate is very low compared to other crimes:

6 %

Motivation

Forest fires are a serious environmental problem.

14 489 forest/year   2005-2014

Approximately 60% of forest fires are caused.

The clarification rate is very low compared to other crimes:

6 %

Database with 1597 cases of provoked forest fires solved in Spain between 2008 and 2015.

Fire variables: 10

Arsonist variables: 15

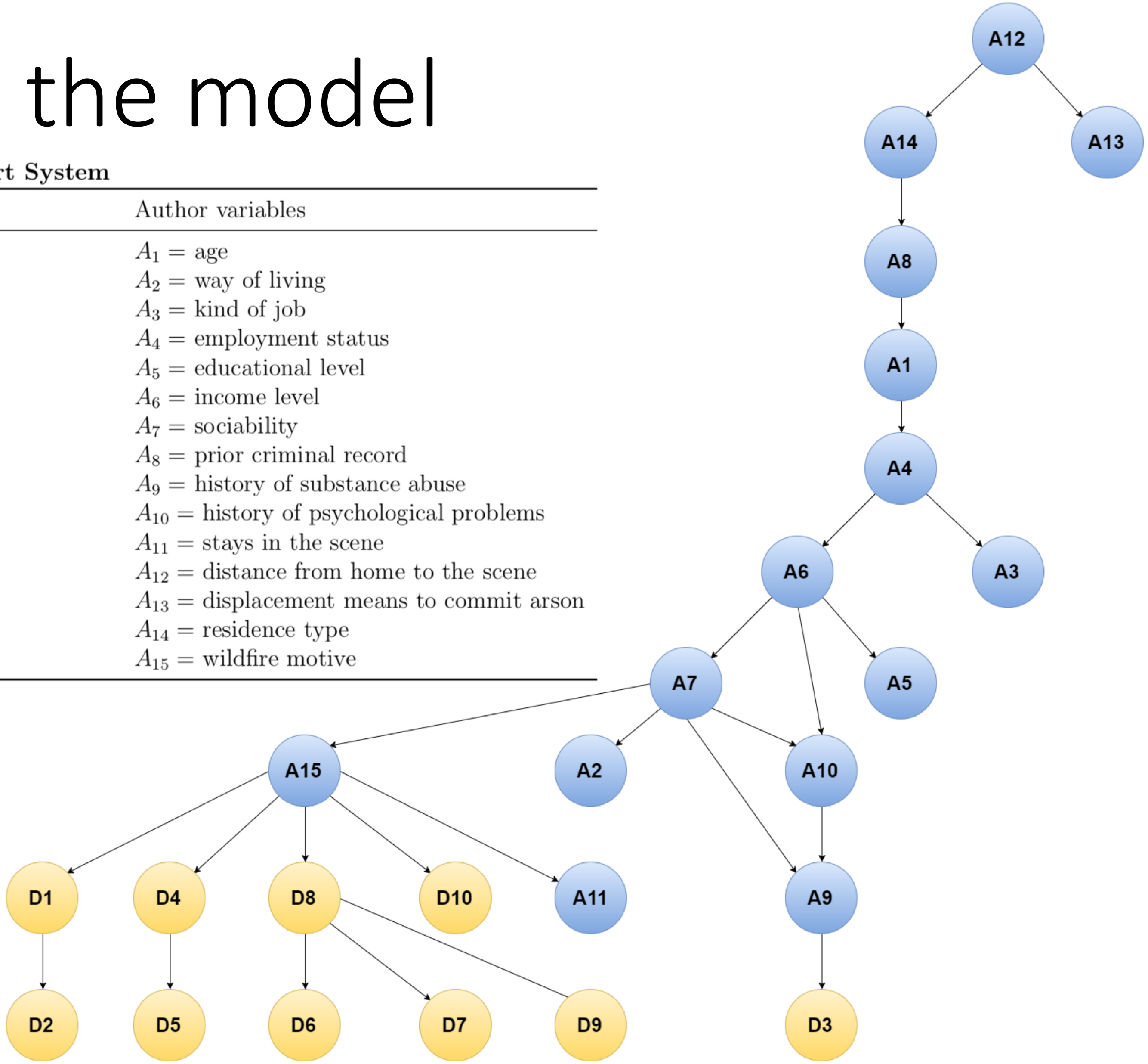
The Bayesian network: the model

“All models are wrong… but some are useful”

G.E.P. Box (1919-2013)

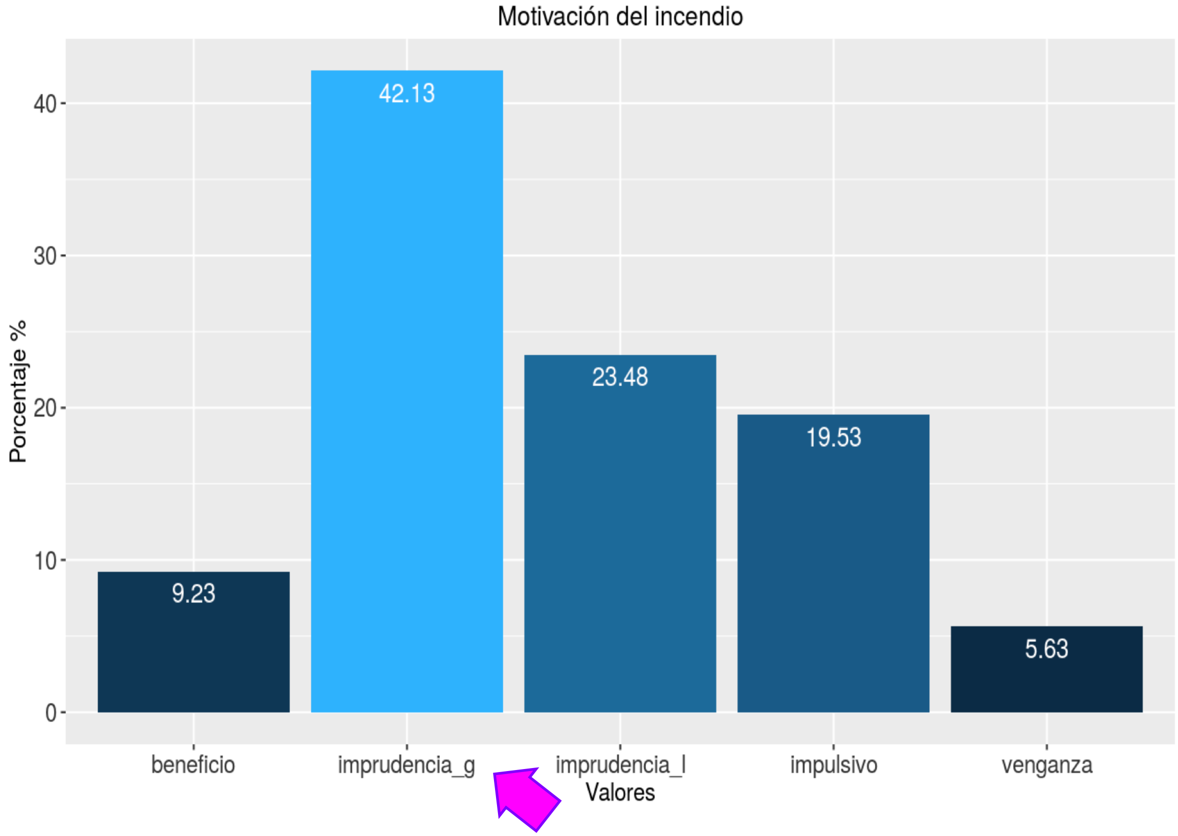
Variables of the Expert System

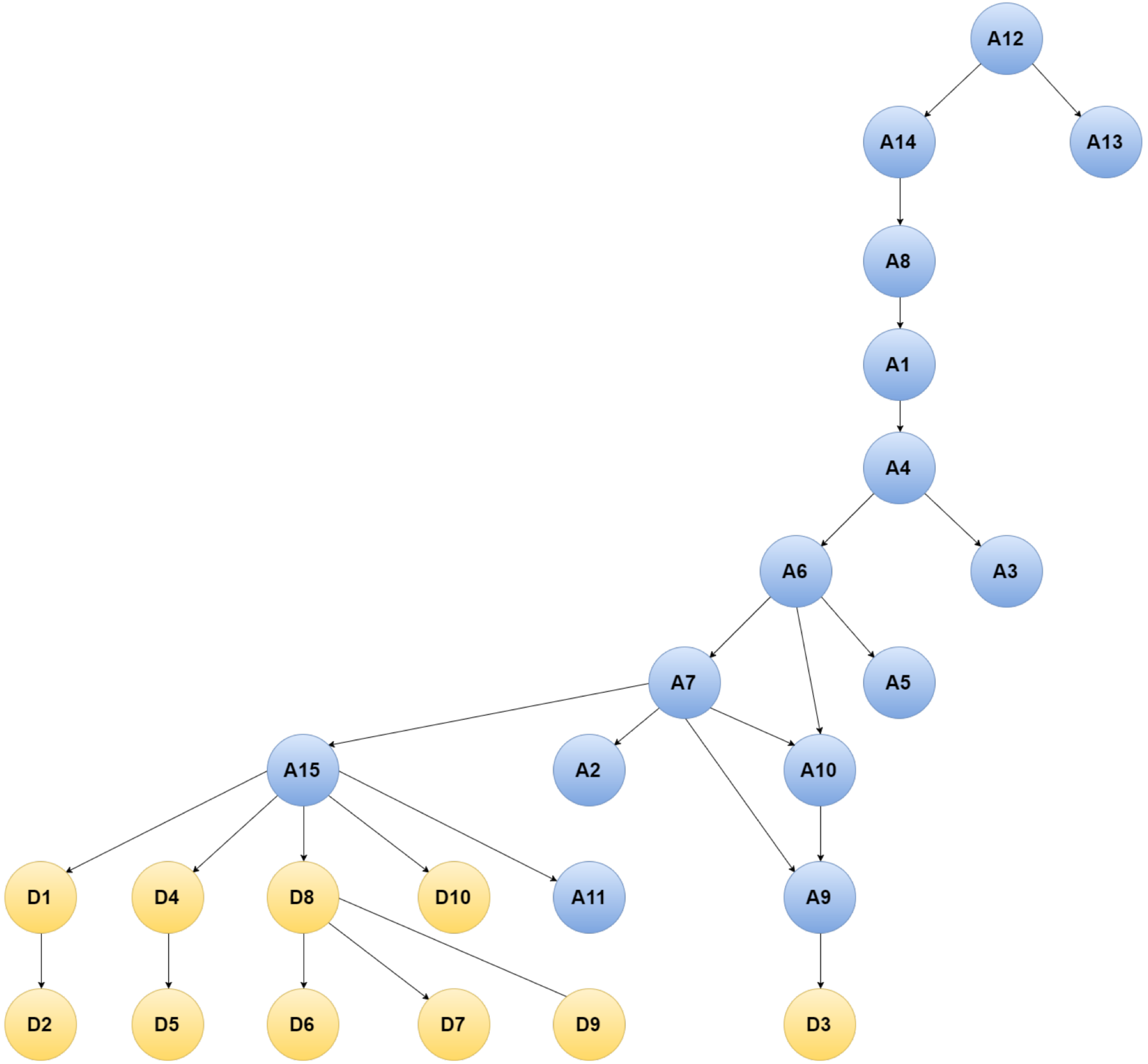
Crime variables	Author variables
D_1 = season	A_1 = age
D_2 = risk level	A_2 = way of living
D_3 = wildfire start time	A_3 = kind of job
D_4 = starting point	A_4 = employment status
D_5 = main use of surface	A_5 = educational level
D_6 = number of seats	A_6 = income level
D_7 = related offense	A_7 = sociability
D_8 = pattern	A_8 = prior criminal record
D_9 = traces	A_9 = history of substance abuse
D_{10} = who denounces	A_{10} = history of psychological problems
	A_{11} = stays in the scene
	A_{12} = distance from home to the scene
	A_{13} = displacement means to commit arson
	A_{14} = residence type
	A_{15} = wildfire motive



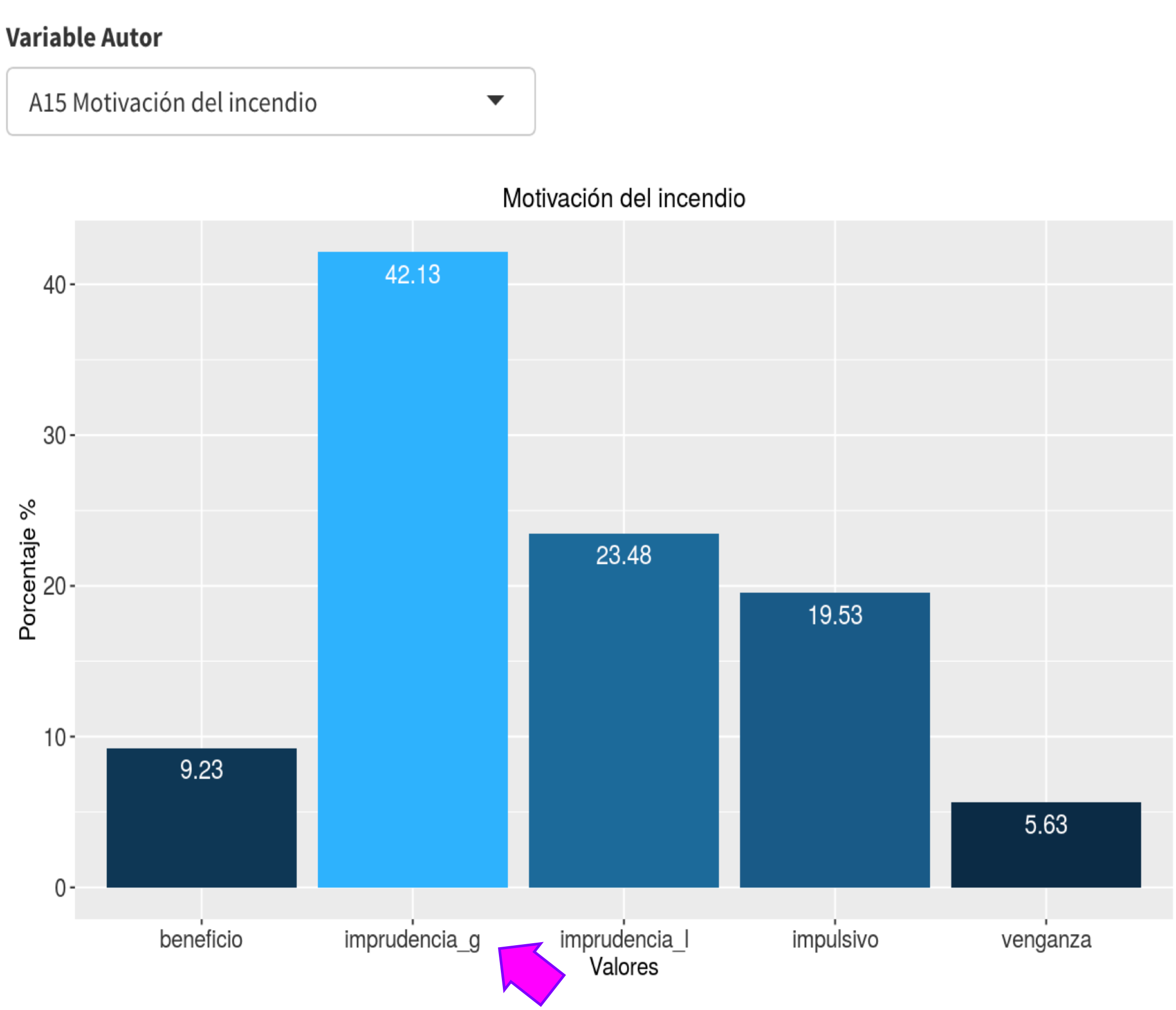
Variable Autor

A15 Motivación del incendio

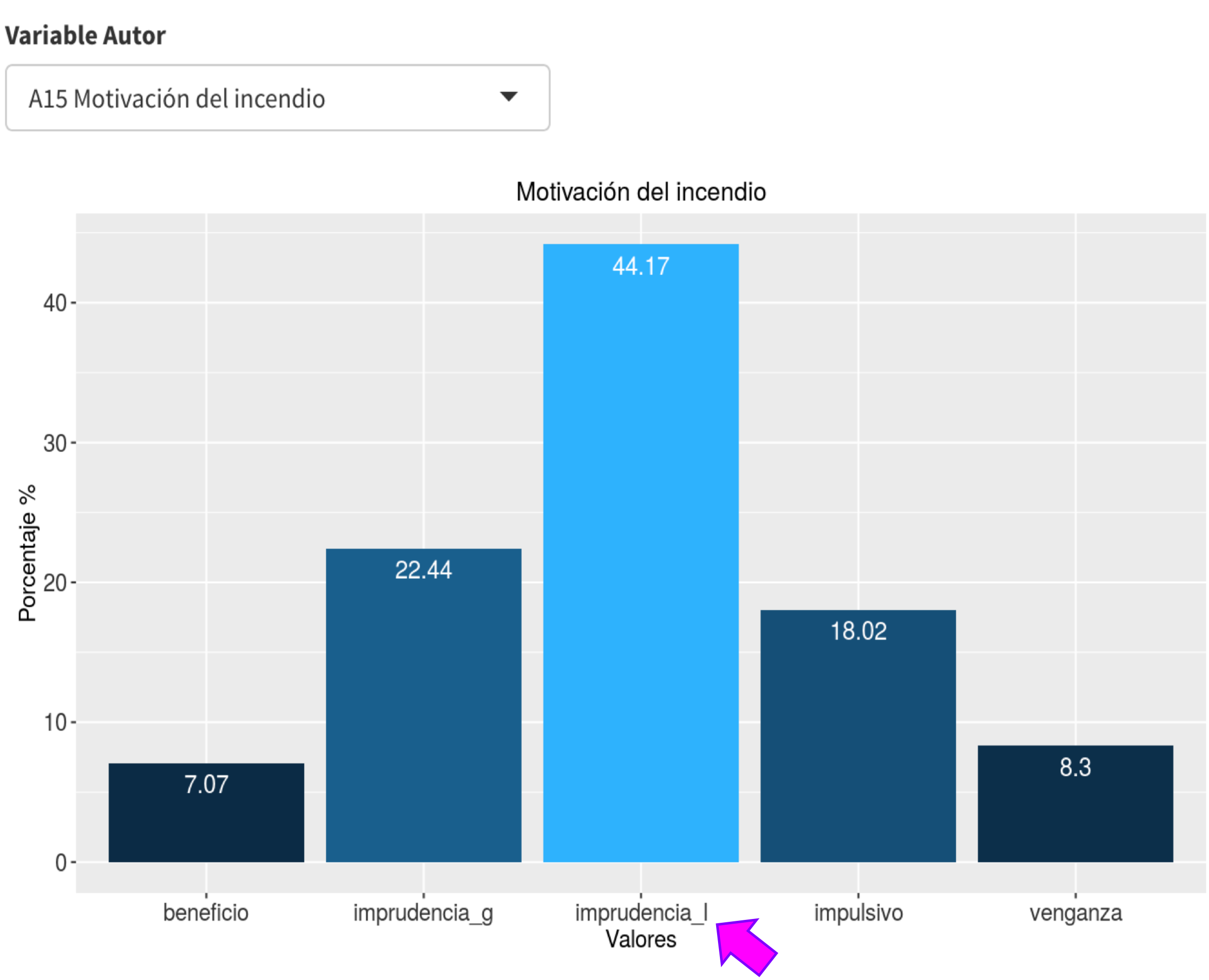




"A priori" probabilities obtained with **PerfilNet.Pyros**, for the variable A15 (motive).

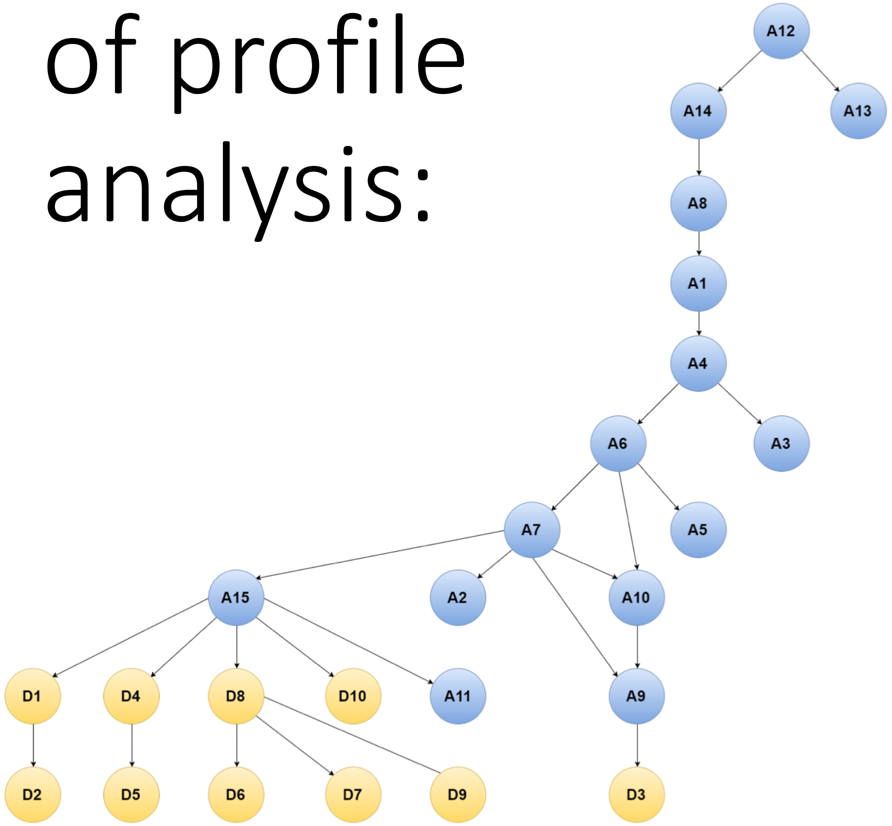


"A priori" probabilities obtained with **PerfilNet.Pyros**, for the variable A15 (motive).

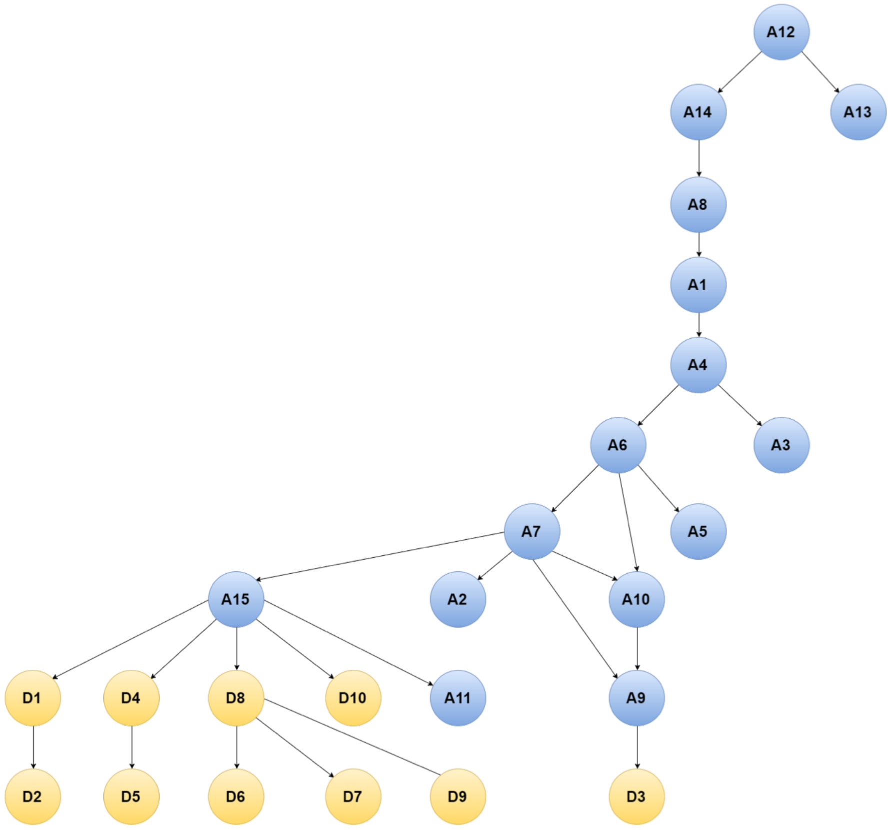


"A posteriori" probabilities, given evidence D10 (who denounces) = particular, obtained with **PerfilNet.Pyros**, for the variable A15 (motive).

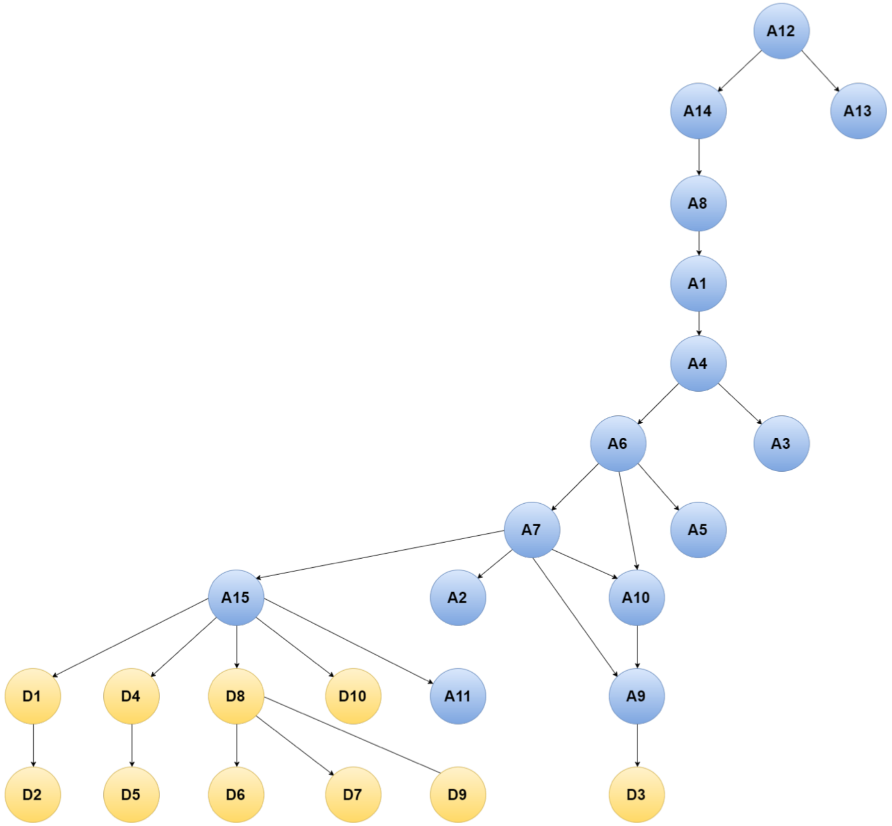
An example of profile analysis:



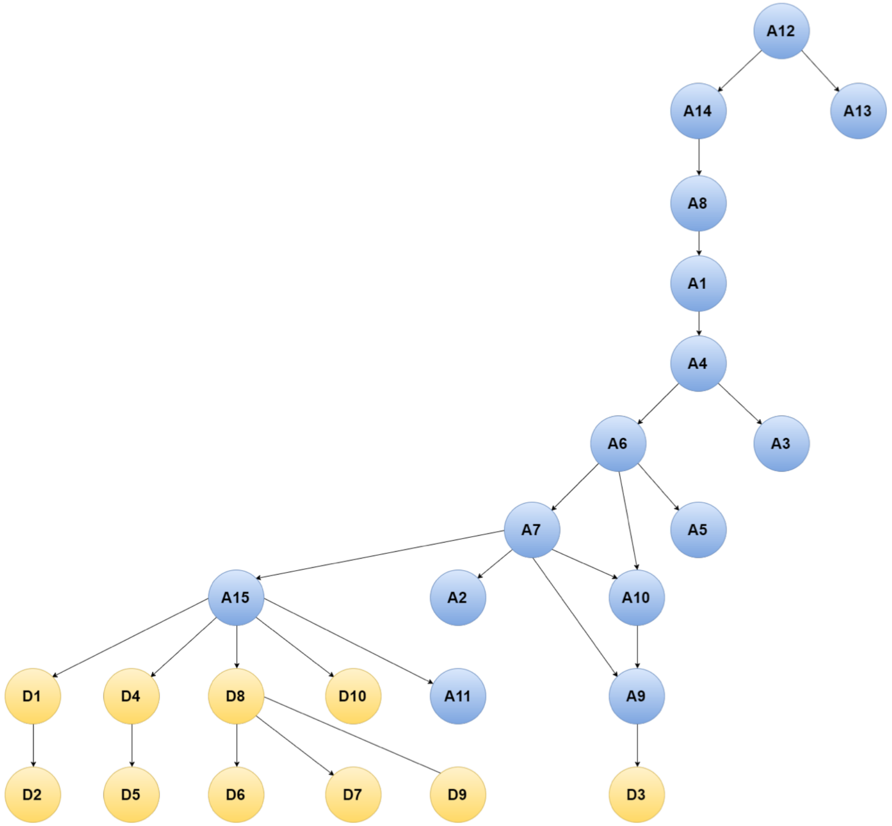
D_{10}: who denounces=**particular**		
D_4: starting point		
Road Forest track	→	**Impulsive**
Pathway Interior Houses Crops Others	→	**Slight Negligence**



$\mathbf{D_{10}}$: who denounces=**particular**	$\mathbf{D_8}$: pattern	
	No	Yes
	Slight Negligence	**Impulsive**



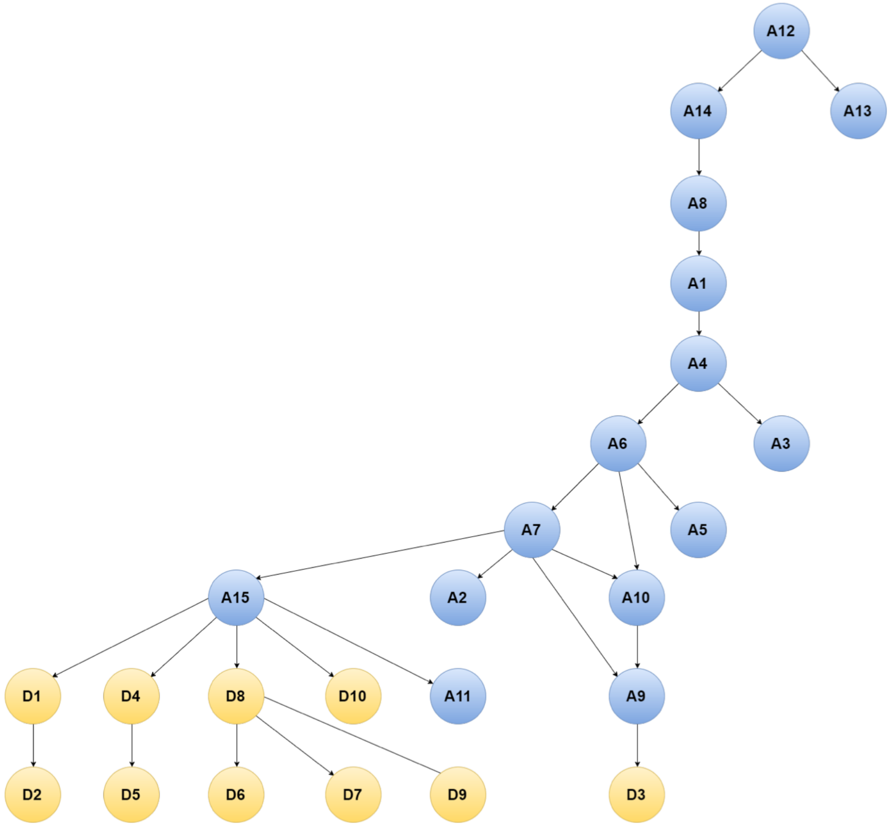
$\mathbf{D_{10}}$: who denounces=**particular**	$\mathbf{D_8}$: pattern		
$\mathbf{D_4}$: starting point	No	Yes	
Road	Slight Negligence	Impulsive	**Impulsive**
Forest track	Impulsive	Impulsive	
Pathway	Slight Negligence	Impulsive	**Slight Negligence**
Interior	Slight Negligence	Impulsive	
Houses	Slight Negligence	Slight Negligence	
Crops	Slight Negligence	Slight Negligence	
Others	Slight Negligence	Slight Negligence	
	Slight Negligence	**Impulsive**	



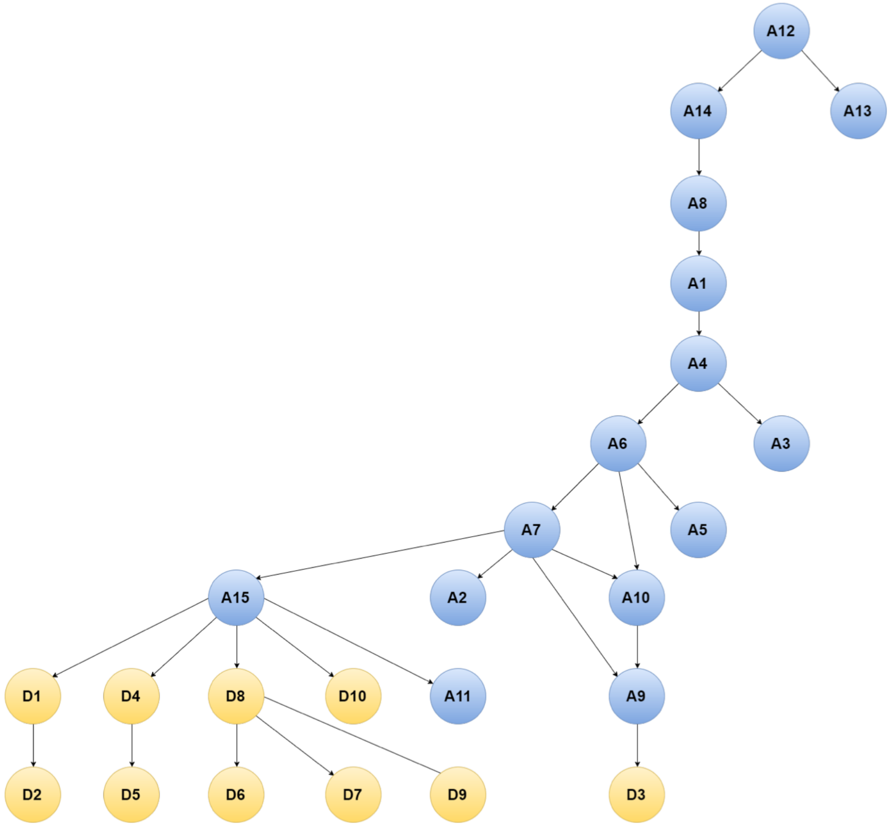
$\mathbf{D_{10}}$: who denounces= guard

$\mathbf{D_4}$: starting point

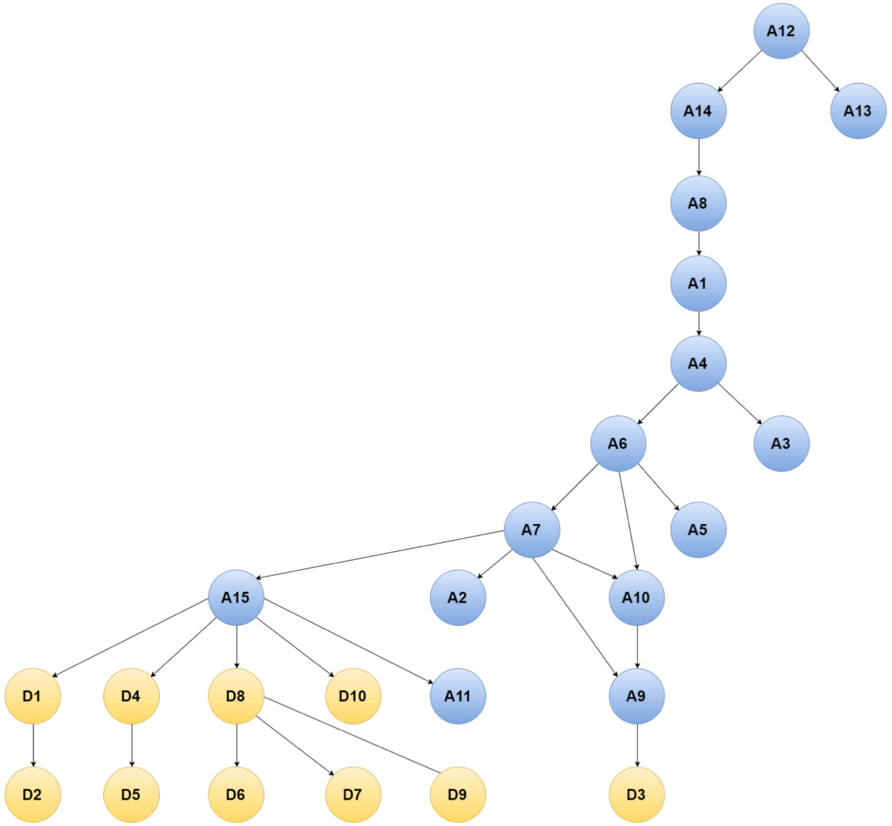
Road Forest track		**Impulsive**
Pathway Interior Houses Crops Others		**Gross Negligence**



$\mathbf{D_{10}}$: who denounces= **guard**	$\mathbf{D_8}$: pattern	
	No	Yes
	Gross Negligence	**Impulsivew**



$\mathbf{D_{10}}$: who denounces= **guard**	$\mathbf{D_8}$: pattern		
$\mathbf{D_4}$: starting point	No	Yes	
Road	Gross Negligence	Impulsive	**Impulsive**
Forest track	Impulsive	Impulsive	
Pathway	Gross Negligence	Impulsive	**Gross Negligence**
Interior	Gross Negligence	Profit	
Houses	Gross Negligence	Gross Negligence	
Crops	Gross Negligence	Gross Negligence	
Others	Gross Negligence	Gross Negligence	
	Gross Negligence	**Impulsivew**	



$\mathbf{D_{10}}$: Denuncia = agent	$\mathbf{D_8}$: Patró		
$\mathbf{D_4}$: Punt d'inici	No	Sí	
Carretera	Imprudència G.	Impulsiu	**Impulsiu**
Pista forestal	Impulsiu	Impulsiu	
Camí	Imprudència G.	Impulsiu	**Imprudència G.**
Interior	Imprudència G.	Benefici	
Vivendes	Imprudència G.	Imprudència G.	
Cultius	Imprudència G.	Imprudència G.	
Altres	Imprudència G.	Imprudència G.	
	Imprudència G.	**Impulsiu**	

$\mathbf{D_{10}}$: who denounces=**particular**	$\mathbf{D_8}$: pattern		
$\mathbf{D_4}$: starting point	No	Yes	
Road	Slight Negligence	Impulsive	**Impulsive**
Forest track	Impulsive	Impulsive	
Pathway	Slight Negligence	Impulsive	**Slight Negligence**
Interior	Slight Negligence	Impulsive	
Houses	Slight Negligence	Slight Negligence	
Crops	Slight Negligence	Slight Negligence	
Others	Slight Negligence	Slight Negligence	
	Slight Negligence	**Impulsive**	

$\mathbf{D_{10}}$: who denounces= **guard**	$\mathbf{D_8}$: pattern		
$\mathbf{D_4}$: starting point	No	Yes	
Road	Gross Negligence	Impulsive	**Impulsive**
Forest track	Impulsive	Impulsive	
Pathway	Gross Negligence	Impulsive	**Gross Negligence**
Interior	Gross Negligence	Profit	
Houses	Gross Negligence	Gross Negligence	
Crops	Gross Negligence	Gross Negligence	
Others	Gross Negligence	Gross Negligence	
	Gross Negligence	**Impulsivew**	

Archetypes

By using our model we can study the archetypes:

The most probable characteristics of the author according to the 5 motivations (they agree with what had already been studied).

- Gross negligence: No substances. No gives aid and tries to scape. Agricultural zones.
- Slight negligence: Helps in the extinction tasks and shows repentance. Agricultural zones.
- Impulsive: On foot. Forest areas. Follows a pattern.
- Profit: By car. No substances. Follows a pattern.
- Revenge: Forest areas. Evening (clear intentionality). Abuse of substances.

Archetypes

- Gross negligence: No substances. No gives aid and tries to scape. Agricultural zones.
- Slight negligence: Helps in the extinction tasks and shows repentance. Agricultural zones.
- Impulsive: On foot. Forest areas. Follows a pattern.
- Profit: By car. No substances. Follows a pattern.
- Revenge: Forest areas. Evening (clear intentionality). Abuse of substances.

	negligence		intentional		
	(1) slight	(2) gross	(3) impuls.	(4) profit	(5) revenge
D_3 = start time	afternoon	afternoon	afternoon	afternoon	evening
D_4 = starting point	crops	crops	pathway	pathway	pathway
D_5 = use surface	agricul.	agricul.	forestry	forestry*	forestry
D_8 = pattern	no	no	yes	yes	no
A_9 = subst. abuse	no	no	no	no	yes
A_{11} = stays	gives aid	no	no	no	no
A_{13} = displacement	by car	by car	on foot	by car	on foot

A confirmatory experiment

We conducted an experiment with 10 solved real cases.

- 16 experts from different provinces were contacted, with an average age of 48.5 years and more than 12 years of experience.
- They were presented with the 10 cases, informing them only of the variables of each one of the provoked wildfires.
- They were asked to make the profile of the arsonists, giving their prediction for the values of the author's variables for each case.

A confirmatory experiment

We conducted an experiment with 10 solved real cases.

- 16 experts from different provinces were contacted, with an average age of 48.5 years and more than 12 years of experience.
- They were presented with the 10 cases, informing them only of the variables of each one of the provoked wildfires.
- They were asked to make the profile of the arsonists, giving their prediction for the values of the author's variables for each case.

Success of human experts: 40%

A confirmatory experiment

We conducted an experiment with 10 solved real cases.

- 16 experts from different provinces were contacted, with an average age of 48.5 years and more than 12 years of experience.
- They were presented with the 10 cases, informing them only of the variables of each one of the provoked wildfires.
- They were asked to make the profile of the arsonists, giving their prediction for the values of the author's variables for each case.

Success of human experts: 40%

Success of the PerfilNet.Pyros system: 60%

Another example: risk of forest fires in Iran

Environ Monit Assess (2016) 188:531
DOI 10.1007/s10661-016-5532-8

CrossMark

Risk of fire occurrence in arid and semi-arid ecosystems of Iran: an investigation using Bayesian belief networks

Hossein Bashari · Ali Asghar Naghipour · Seyed Jamaleddin Khajeddin · Hamed Sangoony · Pejman Tahmasebi

A model (Bayesian network) was developed to identify the risk factors for forest fires in arid and semi-arid areas of Iran.

In the model, they were considered

- Human factors, and
- Biophysical factors.

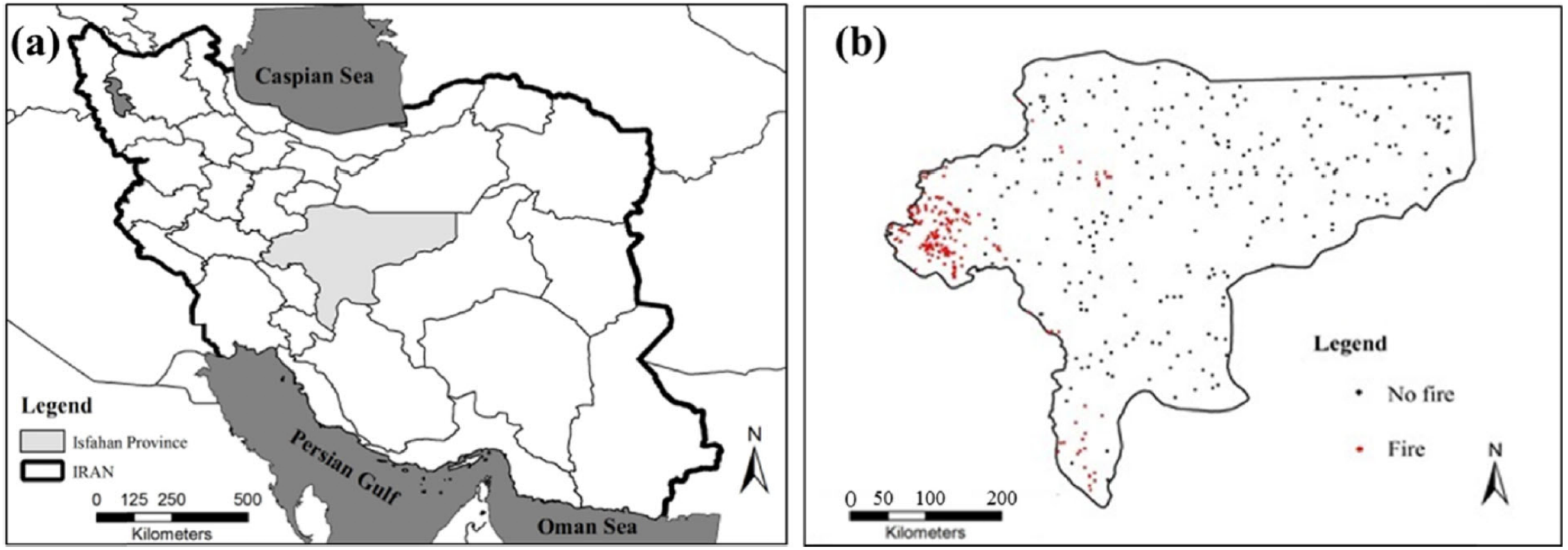


Fig. 1 **a** Isfahan province location in central Iran and **b** fire events data (fire and no fire) in Isfahan province (2008–2011)

Another example: risk of forest fires in Iran

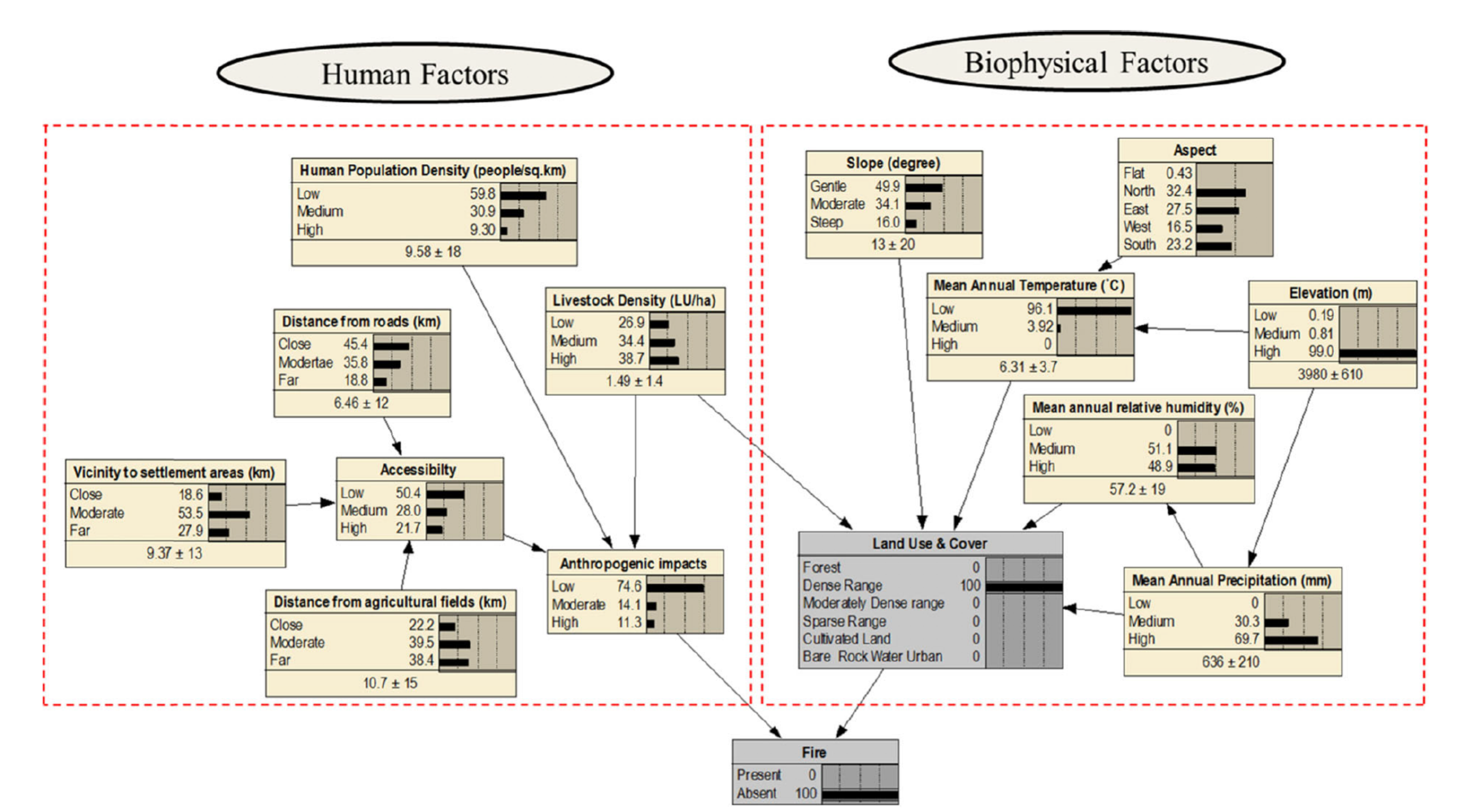
HUMAN FACTORS

- Human population density (people/km^2): Low (< 2), Medium (2-20), High (> 20)
- Distance from roads (km): Close (< 1), Moderate(1-5), Far (> 5)
- Distance to agricultural lands (km): Close (< 0.5), Moderate (0.5-2.5), Far (>2.5)
- Livestock density (units/ha): Low (< 0.5), Medium (0.5-1), High (>1)
- Vicinity to settlement areas (km): Close (< 1), Moderate (1-5), Far (>5)
- Accessibility: Low, Medium, High
- Antropogenic impacts: Low, Moderate, High

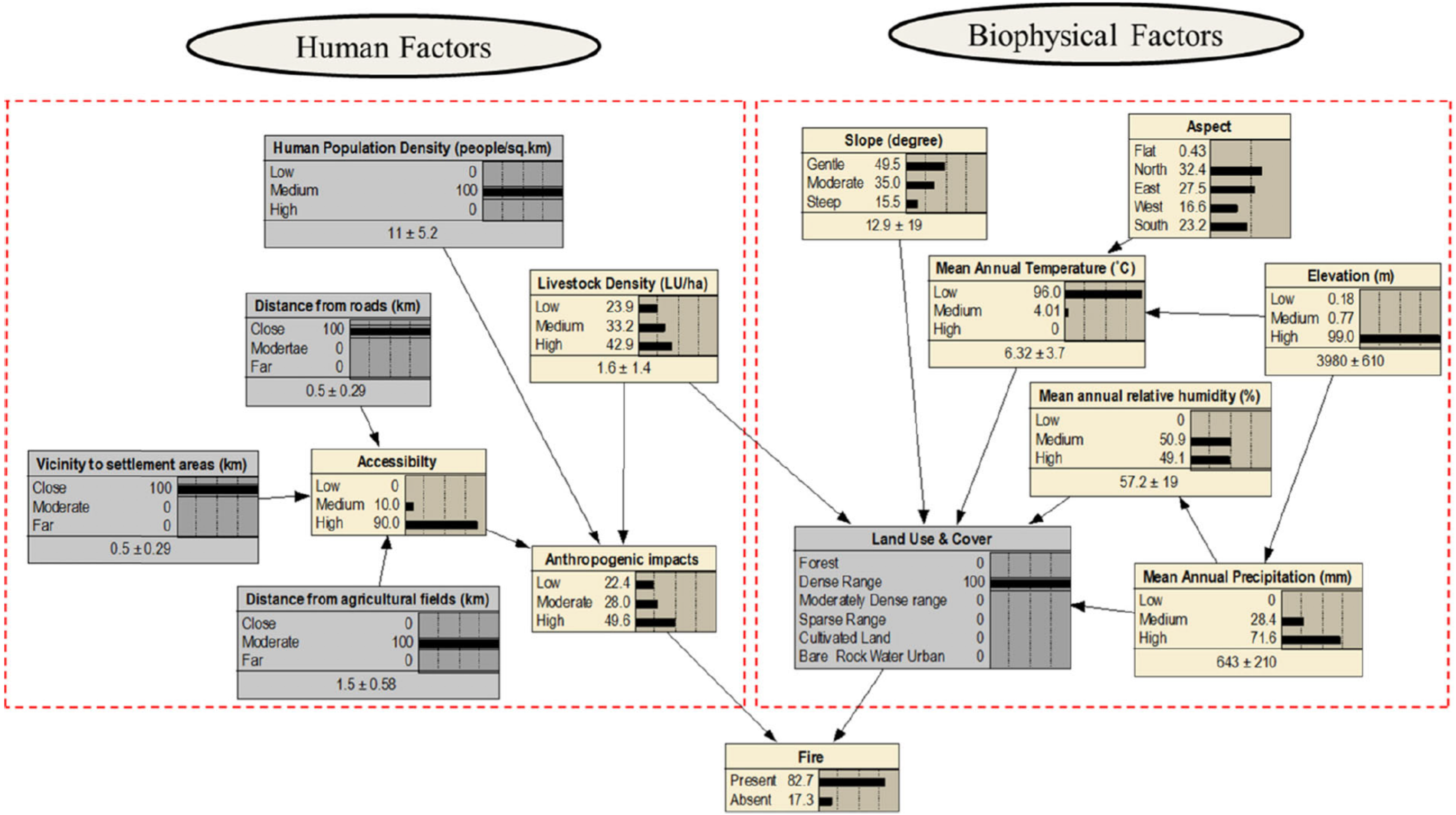
BIOPHYSICAL FACTORS

- Slope (degree ^{0}): Gentle (<5), Moderate (5-15), Step (>15)
- Aspect: North, East, West, South, Flat
- Mean Annual Temperature (ºC): Low (< 12), Medium (12-16), High (>16)
- Mean Annual Relative Humidity (%): Low (< 40), Medium (40-50), High (>50)
- Mean Annual Precipitation (mm/year): Low (< 250), Medium (250-500), High (>500)
- Elevation above sea level (m): Low (<1000), Medium (1000-2000), High (>2000)
- Land cover/use: Forest, Dense range, Moderately dense range, Sparse range, Cultivated land, Bare/rock/water/urban

If we introduce the evidence Land cover/use = Dense Range, the risk of forest fires is 0%.
This is a type of land use that does not create a risk of forest fires on its own, but ... and if we add other factors?



Adding evidences: Human population density = Medium, Distance from roads = Close, Vicinity to settlement areas = Close, Distance from agricultural fiels = Moderate. The risk of forest fire increases to **82.7%**



Conclusions

- ✓Bayesian networks are a probabilistic mathematical model of Machine Learning that can be used for risk assessment and for profiling.
- ✓The model is learned from the database, from which it is also validated, obtaining its predictive accuracy.
- ✓It is applicable in many fields: criminology, medicine, disaster prevention, climate change, occupational hazards, traffic accidents, ...


Thank you for your attention!