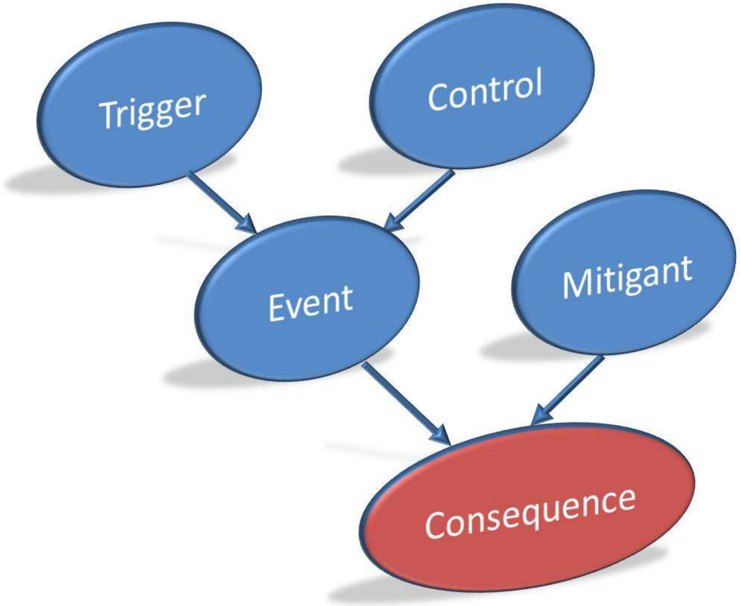


Risk and Decision Making for Data Science and AI

Lesson 2
Assessing risk after new evidence: an introduction to Bayes and AgenaRisk

Norman Fenton
@ProfNFenton

What is the probability after getting the evidence?

<u>HIV Testing</u>:

To obtain marriage license in his state Sam must take blood test (ELISA) for AIDS.

He tests positive.

What is the probability Sam has AIDS?

What is the probability after getting the evidence?

<u>Pregnancy Testing</u>:

Mary has been trying to have a baby and suspects she is pregnant.

She tests positive.

What is the probability Mary is pregnant?

What is the probability after getting the evidence?

Breast Cancer Testing:

Sarah takes a screening for breast cancer.

She test tests positive

What is the probability Sarah has breast cancer?

Suppose a particular virus is present 1 in every 1000 people

A screening test for this virus is:

99% accurate for those with the virus

> ***so the true positive rate – sensitivity - is 99%, meaning 99% of those with the virus will test positive***

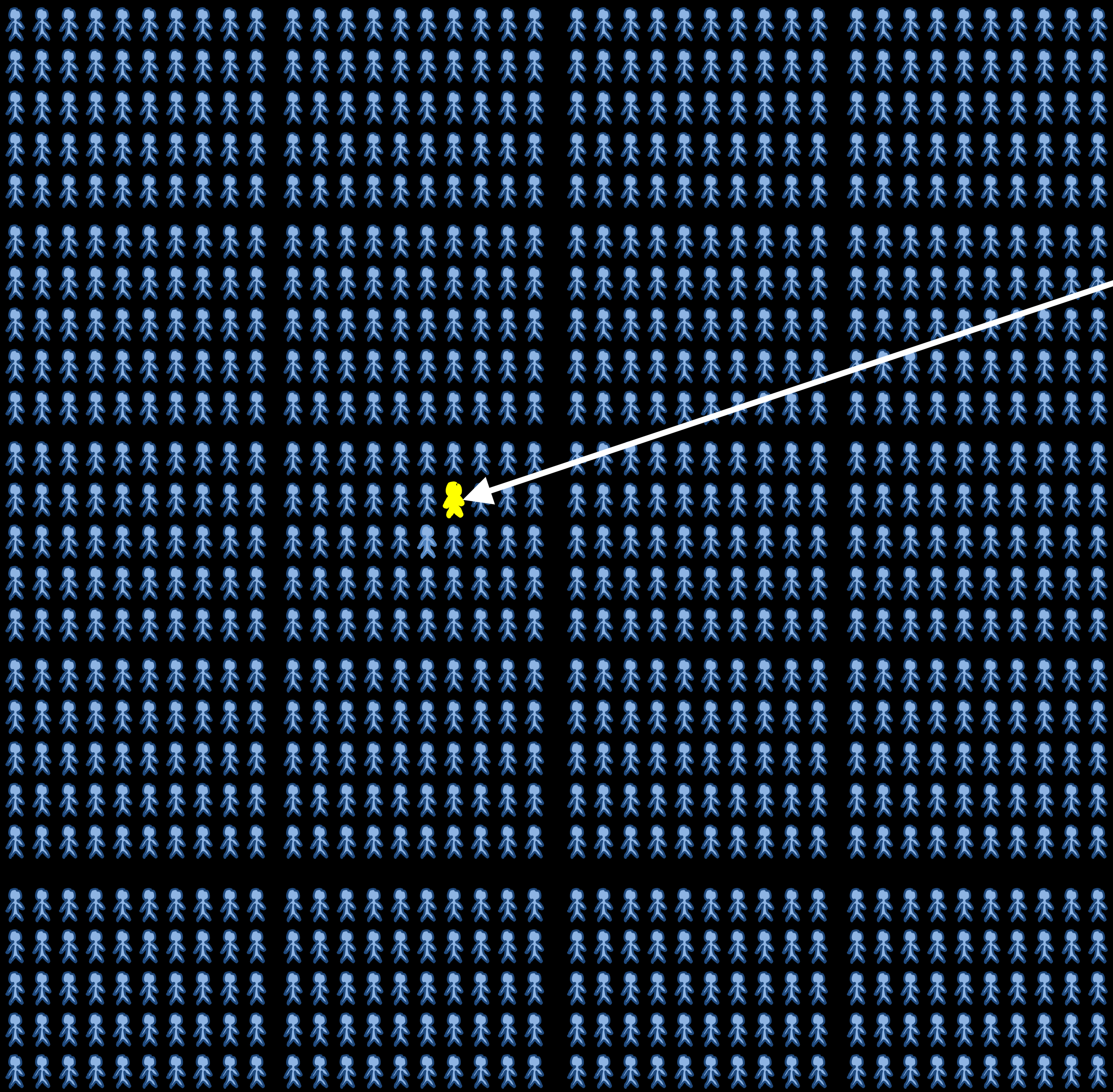
95% accurate for those without the virus

> ***so the true negative rate – specificity - is 95%, meaning 95% of those without the virus will test negative and 5% will test positive***

Sarah tests positive.

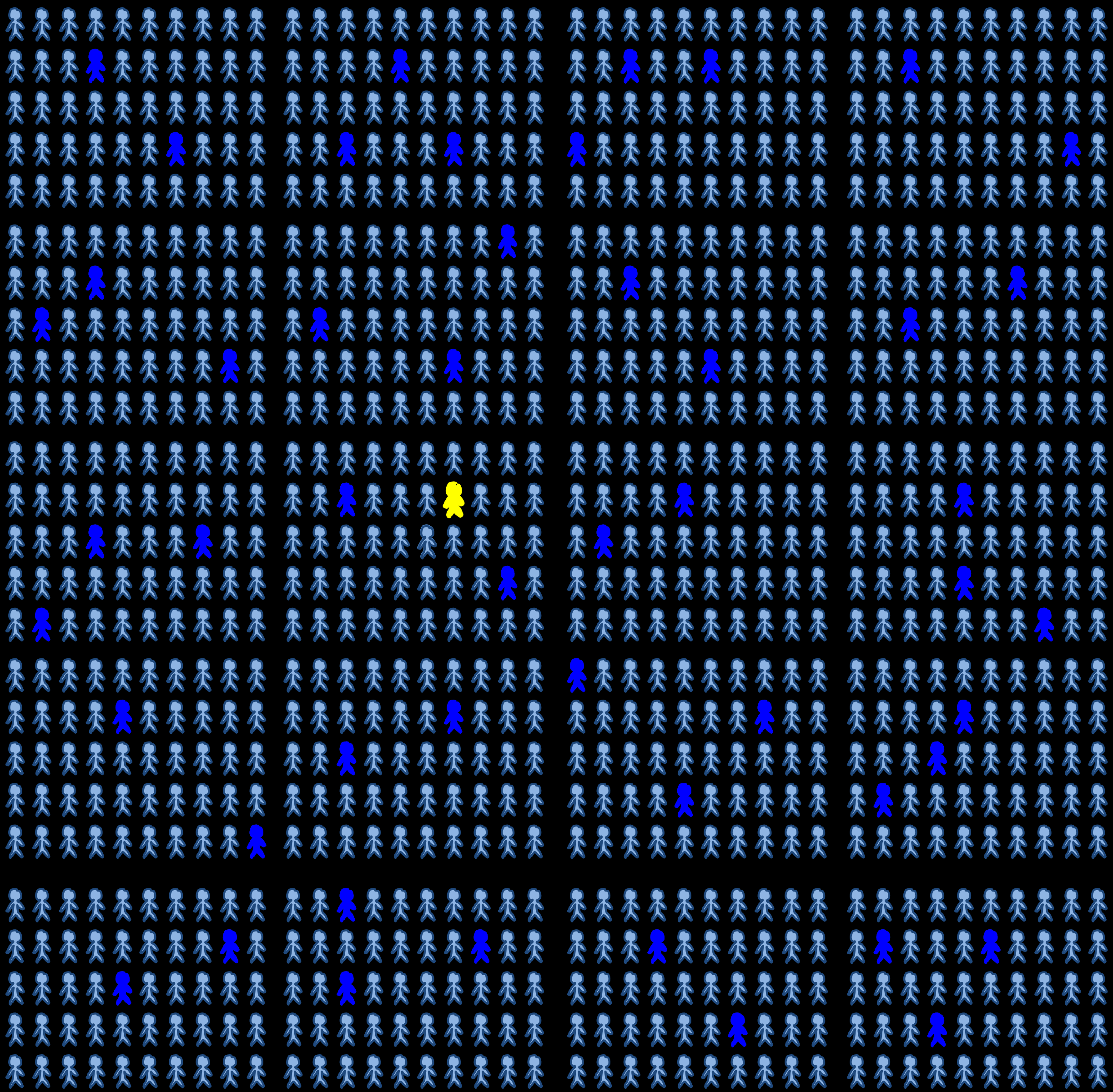
What is the probability Sarah has the virus?

Imagine 1,000 people

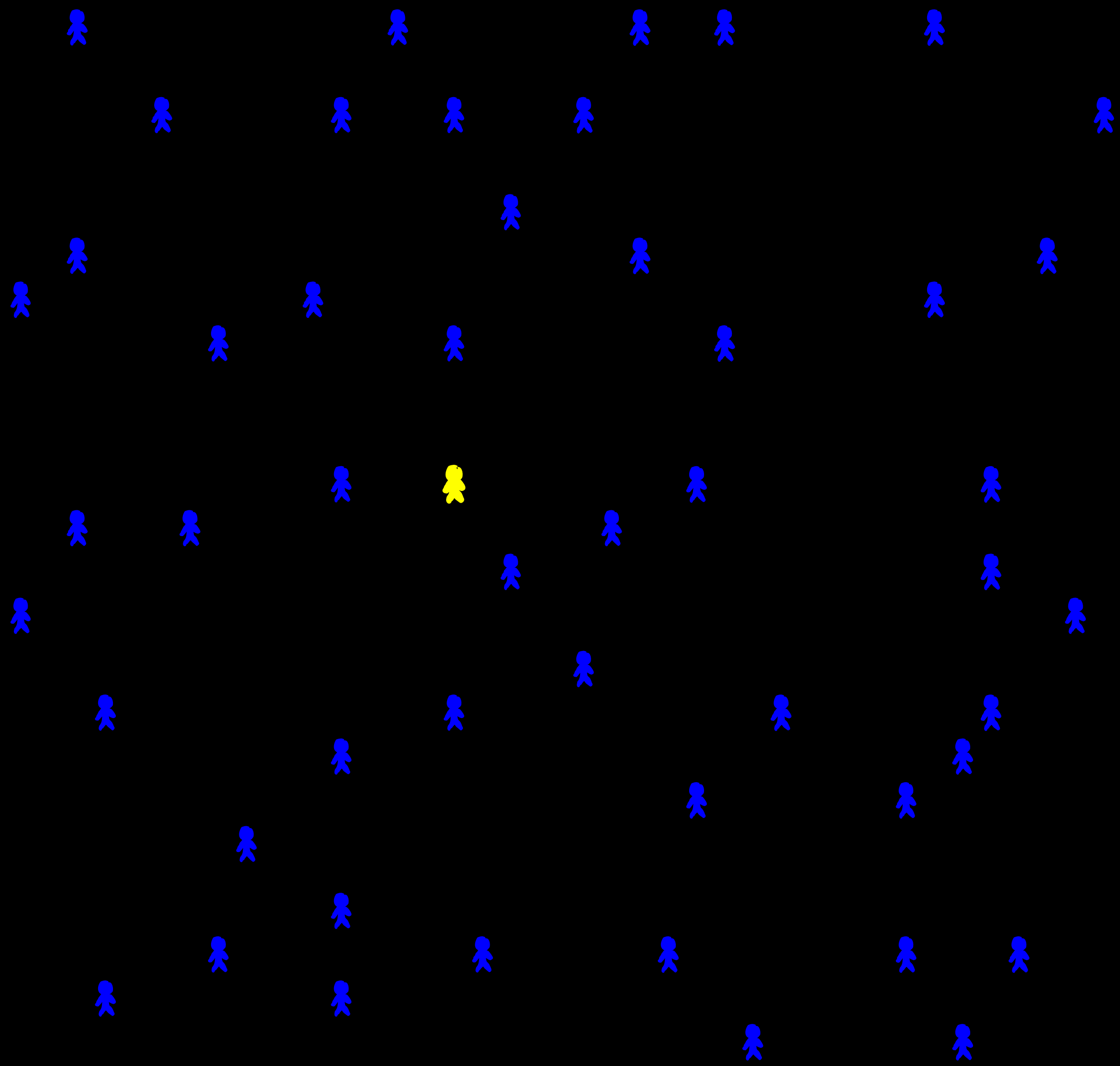


About one person has the virus

And the test will almost certainly be positive for this person



But about 5% of the remaining 999 people without the virus test positive. That is about 49 or 50 people



So about 1 out of 50 who test positive actually have the virus

That’s about 2%

That’s very different from the 95% assumed by most medics

Conditional Probability

We have a hypothesis *H* (in this case "Sarah has the virus", which has two 'states' *True* and *False*)

The probability P(*H*) represents our belief that *H* is *True* before we have seen any specific evidence (also called the ***prior probability of H***)

… and so 1-P(*H*) represents our belief that *H* is *False* (Probability Theorem 1)

We get some evidence *E* (in this case the test result, which we assume has two 'states' *Positive* and *Negative*)

We want to know how to revise our belief that *H* is *True* given the observed state of *E* (i.e. whether positive or negative).

This revised probability (also called the ***posterior probability of H***) is written as P(*H* | *E*)

This is the **probability of *H* given *E*** or **probability of *H* conditioned on *E***

Strictly speaking P(*H* | *E*) is an abbreviation for either P(*H*=True | *E* = Positive) or for P(*H*=True | *E* = Negative)

And if *H* had more than 2 states, e.g. *h1*=COVID, *h2*=Other virus, *h3*=no virus then we would need to distinguish between P(*H=h1* | E), P(*H=h2* | E), P(*H=h3* | E)

Simple Examples

H: “Roll a 6 on a die”   E: “Roll an even number”

$P(H \mid E) = 1/3$

H: “Roll a 6 on a die”   E: “Roll an odd number”

$P(H \mid E) = 0$

H: Pick an Ace from a deck of cards   E: One card – an Ace - has already been removed

$P(H \mid E) = 3/51$

If the events H and E are ‘independent’ (see Probability Primer video 2)
Then $P(H \mid E) = P(H)$

H: “Roll a 6 on a die”   E: “Toss a Head with a coin”

But there is massive and dangerous confusion about conditional probabilities that leads to critical errors in decision making

Evidence from a crime scene

Some of those who were at the scene of the crime

Fred

Killer's shoe

Fred's shoe

Killer left shoeprint on the victim

Fred was at scene

Evidence from forensic analysis shows that both Fred and whoever was the killer have same shoe size 13 and nationally only about 1 in a 100 men are size 13

What is the probability Fred is innocent?

A typical conclusion…

“Given this evidence there is just a 1% chance that a person other than Fred committed the crime”

Which of these statements are correct?
(you can choose more than 1)

1. **The probability Fred is innocent given this evidence is 1 in 100** **p(Innocent | E)**
2. **The probability of this evidence given Fred is innocent is 1 in 100** **p(E | Innocent)**
3. **The statements in 1 and 2 are equivalent**
4. **Neither 1 or 2 is correct**

'This evidence' is that "Fred and whoever was the killer have the same shoe size 13"

The 'prosecution fallacy' is to (wrongly) assume 3

i.e. **p(Innocent | E) =p(E| Innocence)=1/100**

Prosecutor’s fallacy (transposed conditional)

What is the probability that the animal is a cow?

Let H be the Hypothesis : “Animal is a cow”

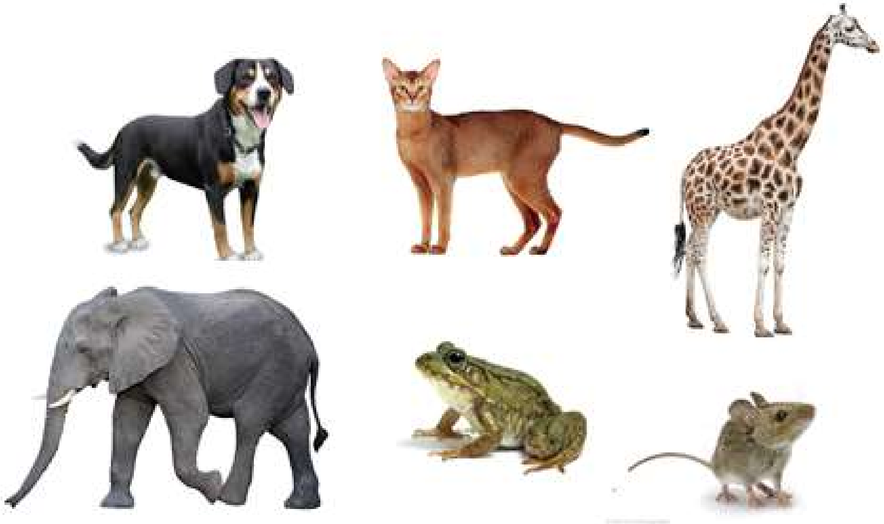
I give you the evidence E: “The animal has 4 legs”

As all cows are 4-legged animals, we know that P(E|H), is equal (or close to) to 1

But if P(H|E) = P(E|H) then we can conclude that it is certain the animal is a cow

A picture of a 4-legged
animal is under this sheet

Prosecutor's fallacy (transposed conditional)



What is the probability that the animal is a cow?

Let H be the Hypothesis : "Animal is a cow"

I give you the evidence E: "The animal has 4 legs"

As all cows are 4-legged animals, we know that P(E|H), is equal (or close to) to 1

But if P(H|E) = P(E|H) then we can conclude that it is certain the animal is a cow.

The prosecutor's fallacy is to assume that "All cows are 4-legged animals" implies "All 4-legged animals are cows"

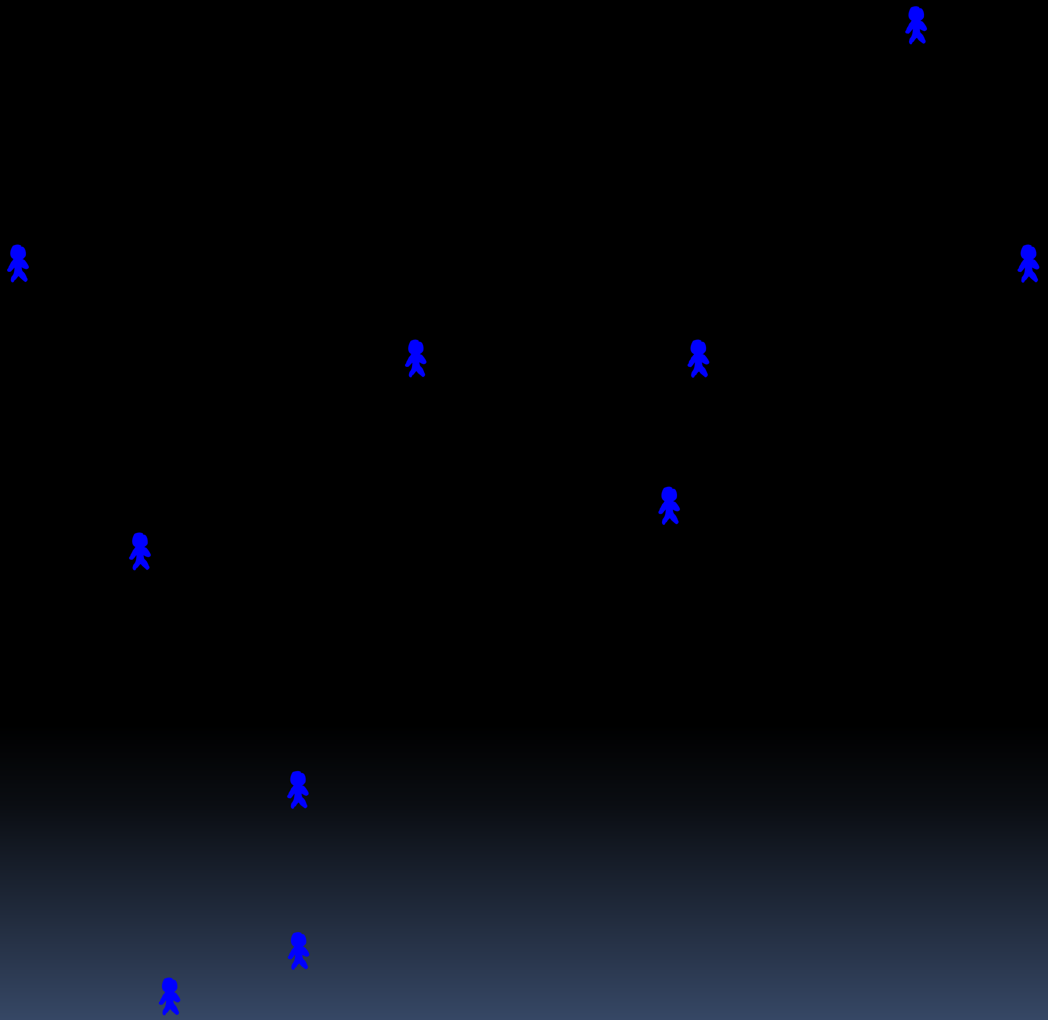
Fred has size 13

Fred has size 13

Imagine 1,000 other people also at scene

Fred has size 13

About 10
out of the
1,000 people
have size 13



Fred is one of 11 with size 13

So there is a 10/11 chance (91%) that Fred is NOT guilty

That's very different from the prosecution claim of 1%

Rev Bayes

As we saw in the Probability Primer this is formula for calculating the posterior probability P(H | E) in terms of the prior probability P(H) and other probabilities we know

$$P(H\,|E) = \frac{P(E\,|H) \times P(H)}{P(E)}$$

Bayes Theorem applied to our medical example

We have a prior P(H) = 0.001

We know P(E|H) = 0.99 and P(E| *not H*) = 0.05

But we want the posterior P(H|E)

You will never have to worry about doing these calculations manually

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} = \frac{P(E|H) \times P(H)}{P(E|H) \times P(H) + P(E|not\ H) \times P(not\ H)}$$

$$P(H|E) = \frac{0.99 \times 0.001}{0.99 \times 0.001 \quad .05 \times 0.999} = \frac{0.00099}{0.009 \quad .04995} = 0.01943 < 2\%$$

Now calculate P(H | not E), i.e. the posterior probability of having the virus if you get a negative test result

Also use the formula to check the result for P(Fred innocent | E) for the crime example

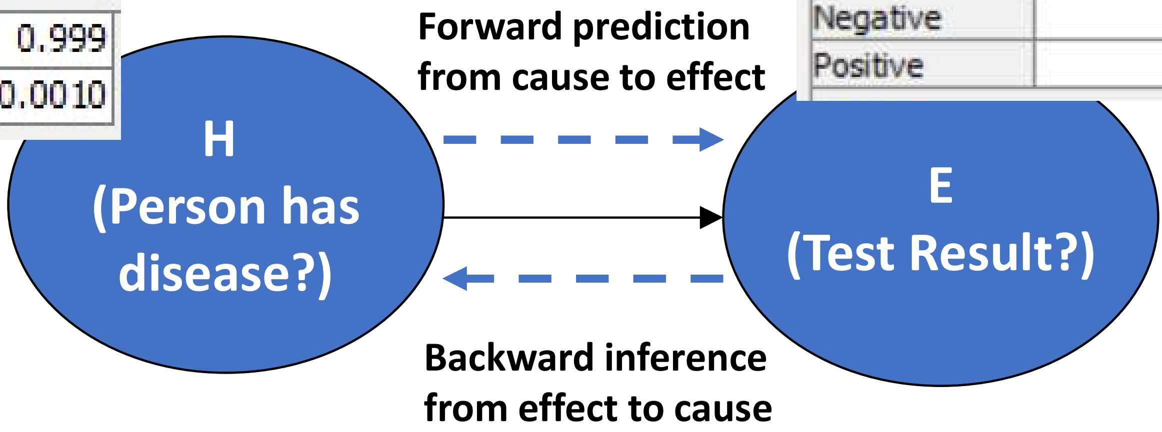
Bayes as a ‘causal’ model (Bayesian network)

We have a hypothesis H

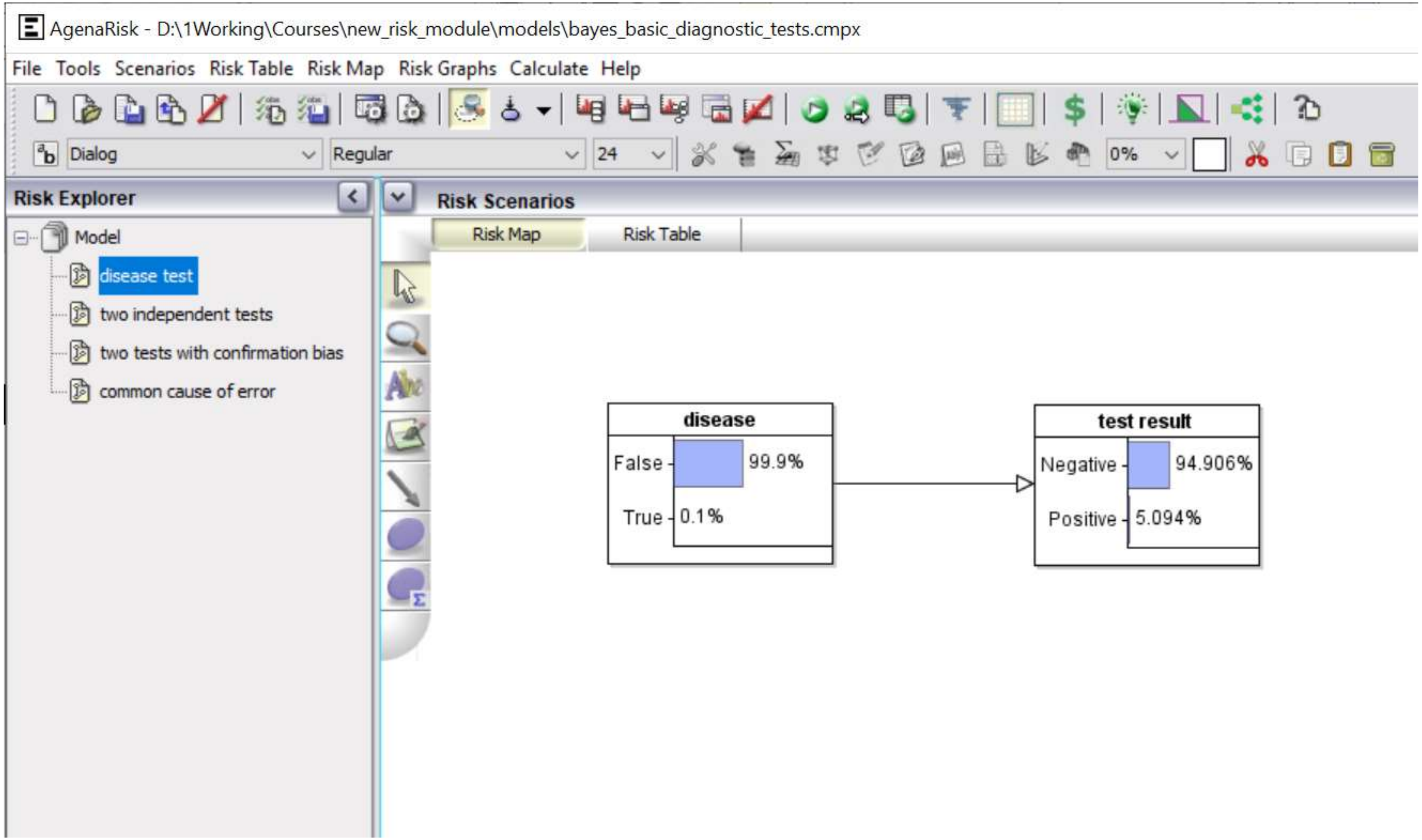
We get some evidence E

False	0.999
True	0.0010

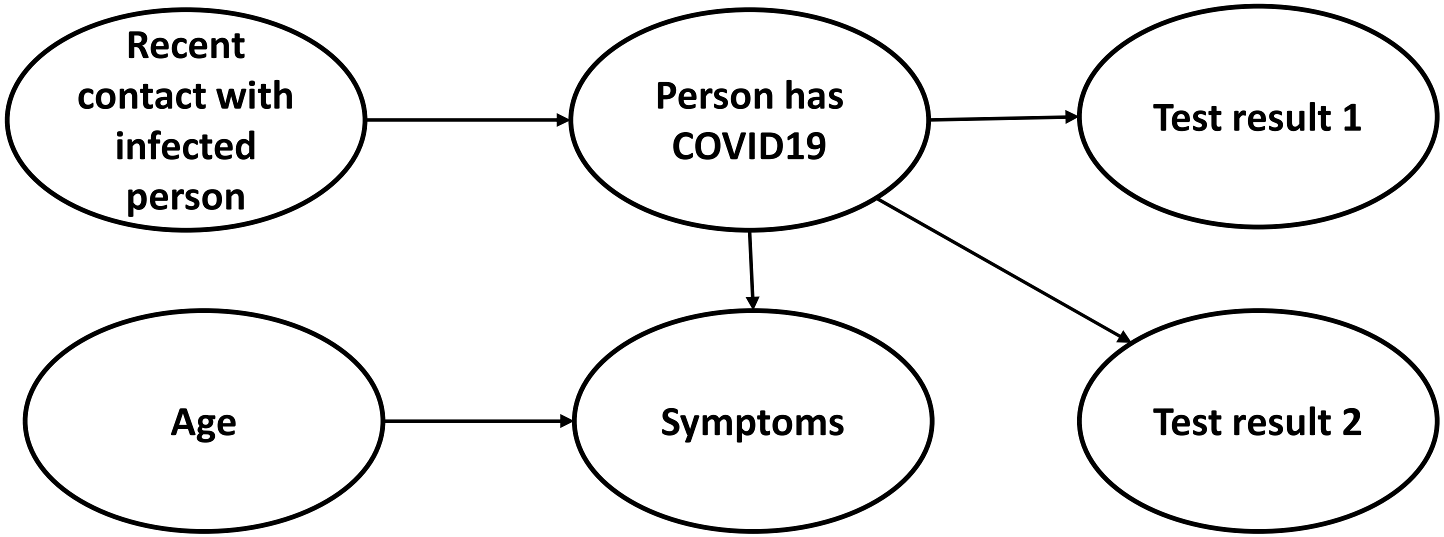
disease	False	True
Negative	0.95	0.01
Positive	0.05	0.99



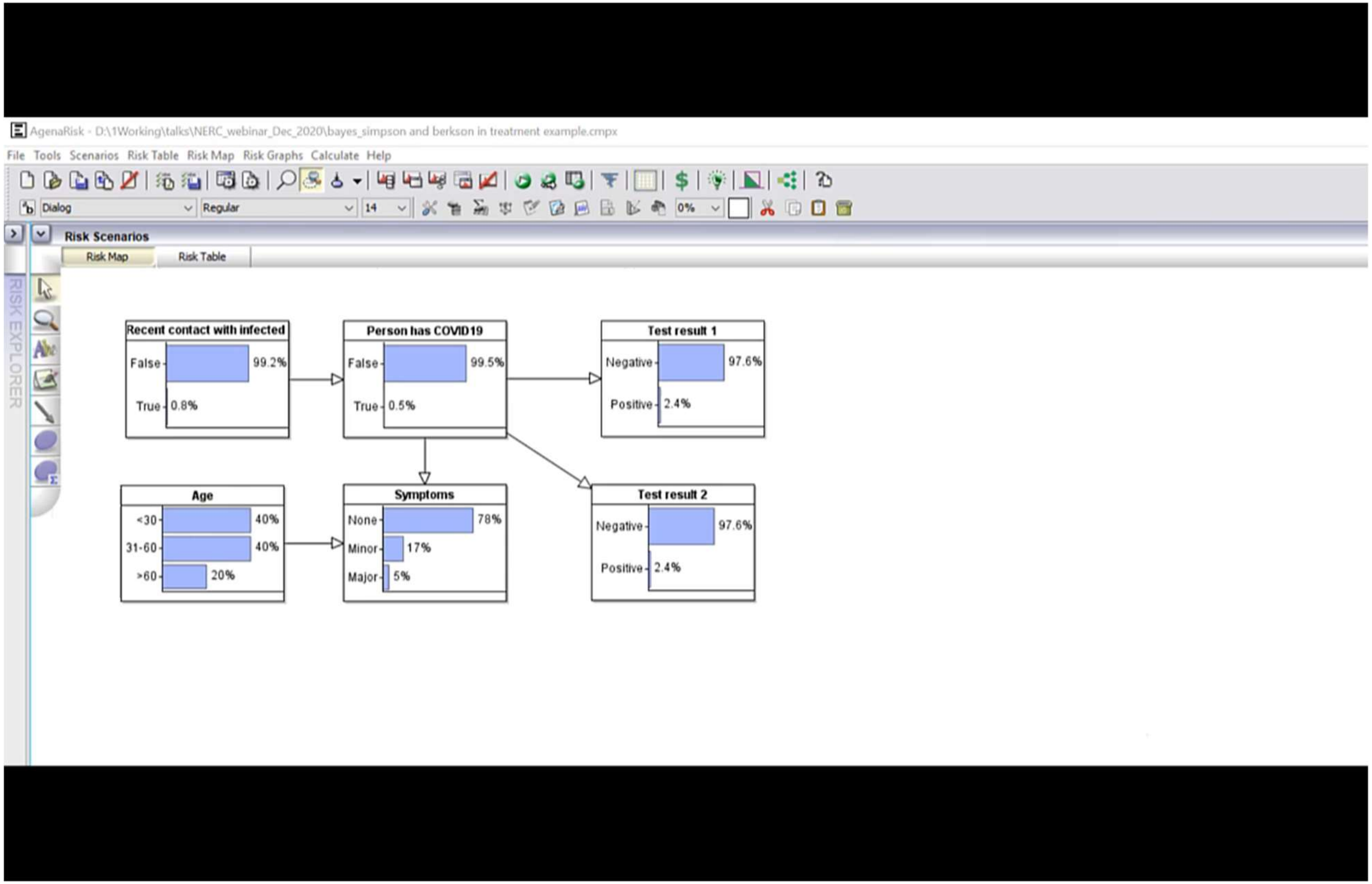
AgenaRisk Introduction and Demo

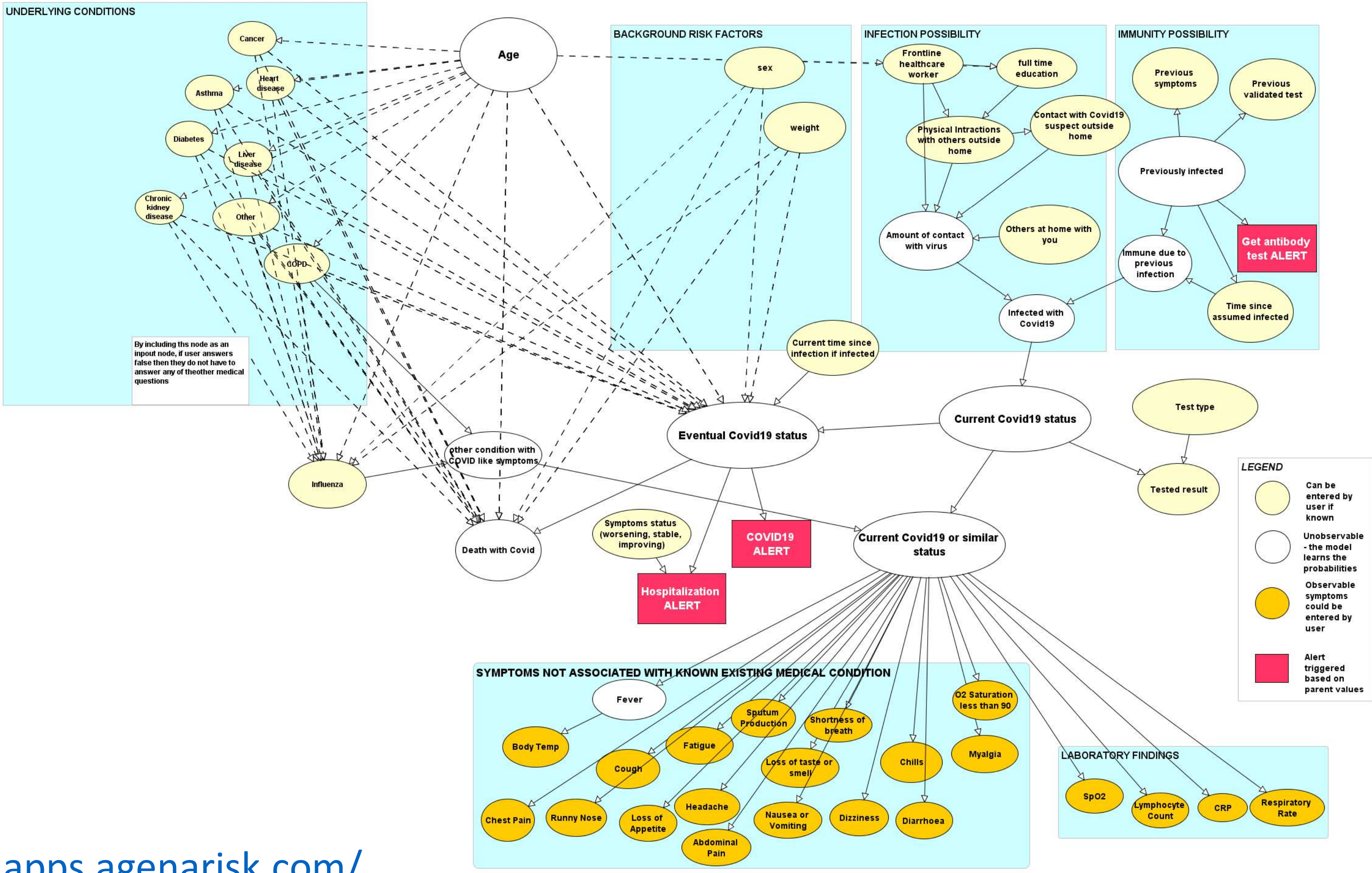


Which is just as well as required models get complex…

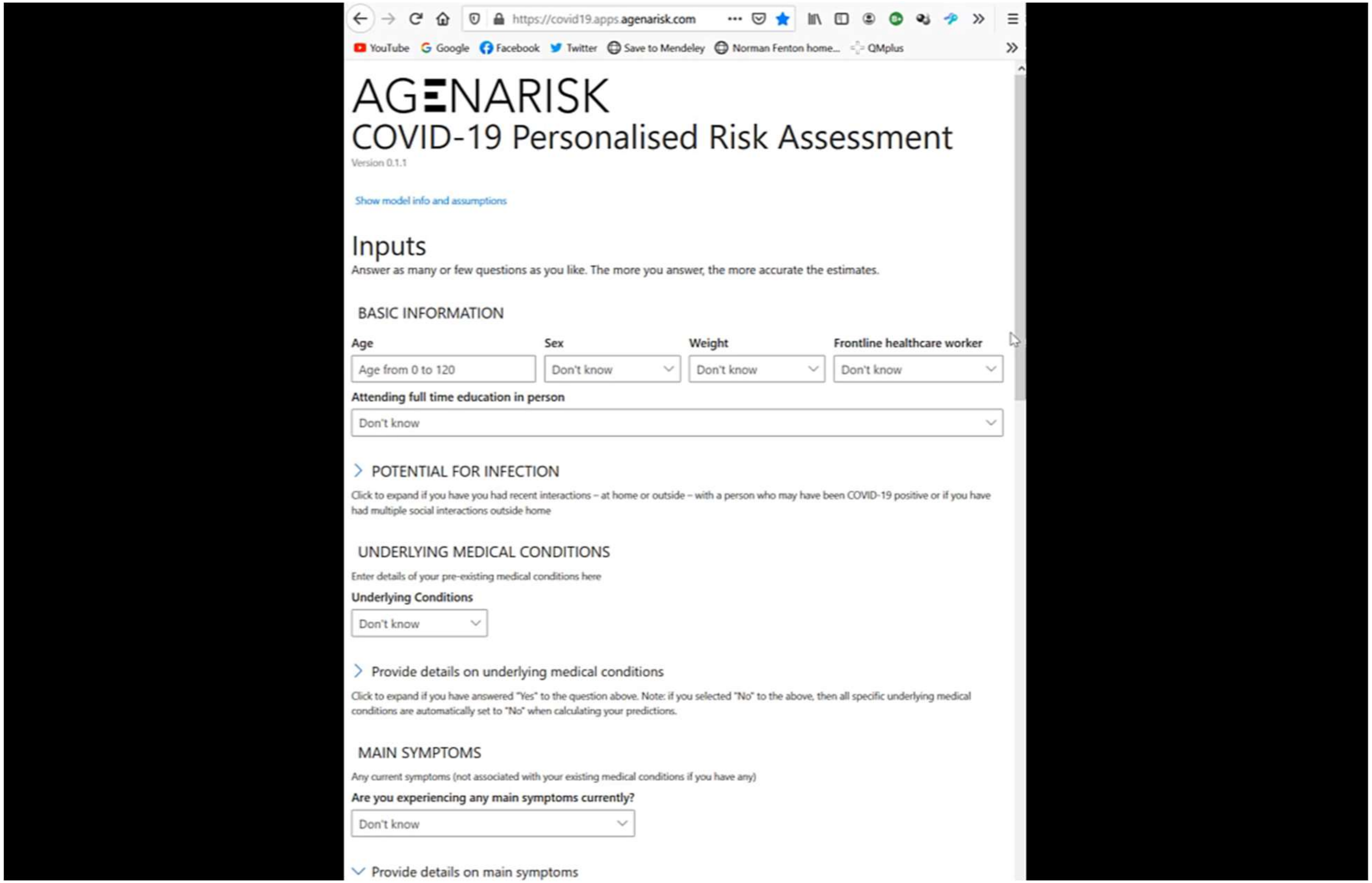


Which is just as well as required models get complex…





https://covid19.apps.agenarisk.com/



https://covid19.apps.agenarisk.com/

HIV Testing

- To obtain marriage license in his state Sam must take blood test (ELISA) for AIDS. He tests positive. What is P(HIV | pos)?
- Assumptions:
  - 10,000 people known to have AIDS for gold standard test (Western Blot) were tested with ELISA. Of these 9990 tested positive. So true positive P(pos | HIV) = 0.999
  - 10,000 nuns (assumed to not have AIDS) were tested with ELISA. Of these 20 tested positive. So false positive P(pos | not HIV)= 0.002
  - 1 in 100,000 men in Sam's state known to have HIV. So prior probability P(HIV) = 0.00001

Calculate the answer using the Bayes Theorem equation and check using AgenaRisk

Neapolitan, Richard, Xia Jiang, Daniela P. Ladner, and Bruce Kaplan. 2016. "A Primer on Bayesian Decision Analysis With an Application to a Kidney Transplant Decision." *Transplantation* 100 (3): 489–96.

Pregnancy Testing

- Mary has been trying to have a baby and suspects she is pregnant. She tests positive. What is P(pregnant | pos)?
- Assumptions:
  - True positive P(pos | pregnant) = 0.99
  - False positive P(pos | not pregnant)= 0.02
  - Only women who suspect they might be pregnant take test. So prior P(pregnant)=0.2

Calculate the answer using the Bayes Theorem equation and check using AgenaRisk

Neapolitan, Richard, Xia Jiang, Daniela P. Ladner, and Bruce Kaplan. 2016. "A Primer on Bayesian Decision Analysis With an Application to a Kidney Transplant Decision." *Transplantation* 100 (3): 489–96.