

Exercise AU-505

Screening for rare diseases with imperfect medical testing

The Economic Skills Project

1 Problem

Problem, p.1

An advisory committee is considering whether to recommend a new AI-based screening test for a particular kind of cancer. It has the following information: people in the population have a 1% chance of having the disease (D); otherwise they are healthy (H). There are two treatments available: aggressive intervention (A) and watchful waiting (W). The net payoffs of applying each type of treatment to each type of person are:

Health State	Treatment	
	A	W
D	\$600,000.	\$100,000.
H	-\$150,000.	\$0.

Problem, p.2

The test costs \$1000 per person and will return one of two reports: that the person has the disease (rD) or that they are healthy (rH). As shown below, the probability of each report depends on the person's underlying health status. What is the expected value of the test? Out of a population of 1 million, how many people receive the wrong treatment?

Health State	Test Report	
	rD	rH
D	96%	4%
H	2%	98%

2 Answer

Answer

- The EV is \$1830: that's somewhat better than the EV of treating without testing, which is \$1000.
- 400 people are *undertreated*: they have the disease (D) but receive watchful waiting (W) instead of the aggressive intervention (A).
- 19,800 people are *overtreated*: they are healthy (H) but receive the aggressive intervention (A) instead of watchful waiting (W).

3 Method

Solution method

Here's one approach:

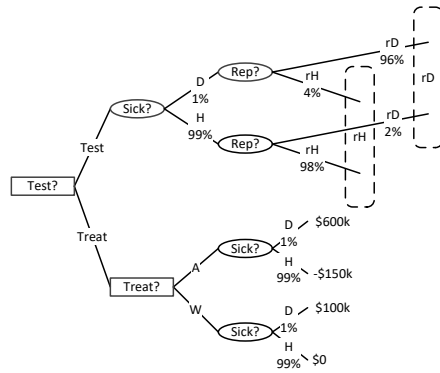
1. Draw the decision tree out to the information sets.
2. Evaluate the Treat branch.
3. Calculate the probabilities on the Test branch.
4. Redraw the Test branch using the probabilities.
5. Evaluate the rD branch.
6. Evaluate the rH branch.
7. Finish evaluating the tree from right to left.
8. Compute the number of incorrectly treated patients.

4 Solution

4.1 Step 1

Draw the decision tree out to the information sets

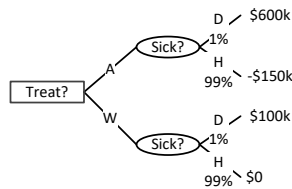
Here's the start of the tree:



4.2 Step 2

Evaluate the Treat branch, p.1

Zooming in on the *Treat* branch:



Evaluate the Treat branch, p.2

The expected value of treatment A will be:

$$EV_A = 0.01(\$600k) + 0.99(-\$150k) = -\$142.5k$$

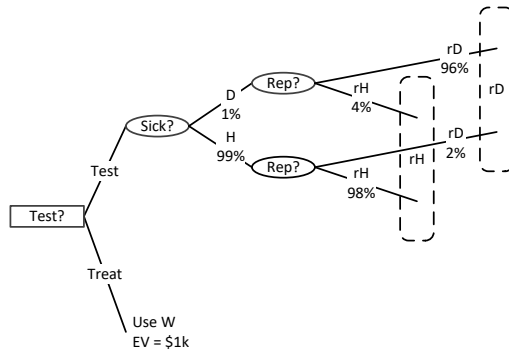
The expected value of treatment W will be:

$$EV_W = 0.01(\$100k) + 0.99(\$0) = \$1k$$

Since the payoff from W is better, it will be chosen and the overall payoff of the treatment branch will be \$1000.

Evaluate the Treat branch, p.3

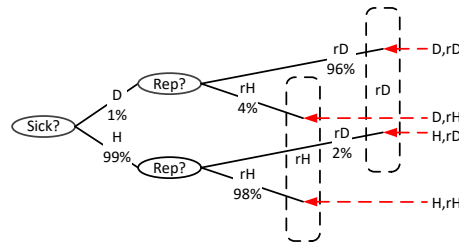
After evaluating the *Treat* branch the tree looks like this:



4.3 Step 3

Calculate probabilities on the Test branch, p.1

To evaluate the *Test* branch it will help to assign names to the endpoints. Call the endpoint where the person has the disease (D) and the report says they do (rD) “D,rD” and name the others in the same way. Here’s the result:



Calculate probabilities on the Test branch, p.2

The probability of ending up at any one of the endpoints is the product of the probabilities along the way. For example, the probability of arriving at D,rD is $0.01 \times 0.96 = 0.0096$ or 0.96%. The full set of probabilities, Pr, is:

Endpoint	Health		Report		Endpoint
	State	Pr	Report	Pr	Pr
D,rD	D	0.01	rD	0.96	0.0096
D,rH	D	0.01	rH	0.04	0.0004
H,rD	H	0.99	rD	0.02	0.0198
H,rH	H	0.99	rH	0.98	0.9702

Calculate probabilities on the Test branch, p.3

The probability of each kind of report, rD or rH, can now be computed by adding up the probabilities of the endpoints in its information set.

Report rD:

$$\begin{aligned}\Pr(rD) &= \Pr(D, rD) + \Pr(H, rD) \\ &= 0.0096 + 0.0198 \\ &= 0.0294\end{aligned}$$

Report rH:

$$\begin{aligned}\Pr(rH) &= \Pr(D, rH) + \Pr(H, rH) \\ &= 0.0004 + 0.9702 \\ &= 0.9706\end{aligned}$$

Calculate probabilities on the Test branch, p.4

The conditional probabilities of D and H given an rD or rH report can now be calculated. For someone with an rD report, the probabilities of D and H are:

$$\Pr(D|rD) = \frac{\Pr(D, rD)}{\Pr(rD)} = \frac{0.96\%}{2.94\%} = 32.65\%$$

and:

$$\Pr(H|rD) = \frac{\Pr(H, rD)}{\Pr(rD)} = \frac{1.98\%}{2.94\%} = 67.35\%$$

As will become clear below, the fact that two-thirds of people reported as sick (rD) are actually well (H) will lead to a great deal of overtreatment.

Calculate probabilities on the Test branch, p.5

In contrast, for someone with an rH report, the probabilities of D and H are:

$$\Pr(D|rH) = \frac{\Pr(D, rH)}{\Pr(rH)} = \frac{0.04\%}{97.06\%} = 0.04\%$$

and:

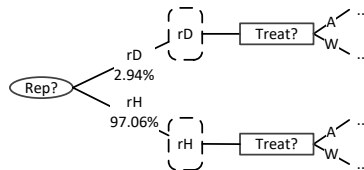
$$\Pr(H|rH) = \frac{\Pr(H, rH)}{\Pr(rH)} = \frac{97.02\%}{97.06\%} = 99.96\%$$

Almost everyone reported as healthy (H) actually is: the degree of undertreatment will be low.

4.4 Step 4

Redraw the Test branch using the probabilities

It's now possible to simplify the test branch to a chance node with two possible reports: rD and rH. Each leads to a choice between treatments A and W:

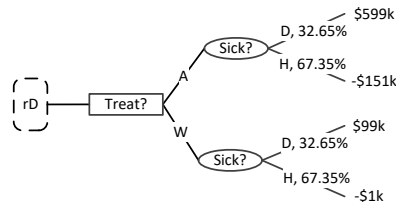


Since the full tree is large, it's most convenient to evaluate the rD and rH branches one at a time.

4.5 Step 5

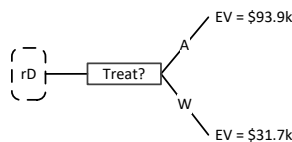
Evaluate the rD branch, p.1

Using the conditional probabilities for D and H given an rD report, the rD branch looks like this:



Evaluate the rD branch, p.2

Calculating the expected values of each chance node gives:

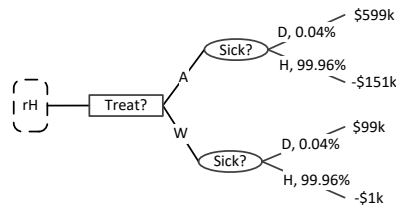


Since A provides a better payoff, the decision following the rD report will be to use treatment A for a net payoff of \$93.9k.

4.6 Step 6

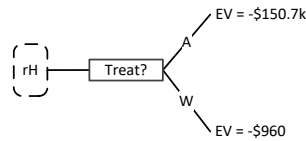
Evaluate the rH branch, p.1

Evaluating the rH branch is similar but it uses the probabilities of D and H conditional on an rH report:



Evaluate the rH branch, p.2

Computing the expected values:

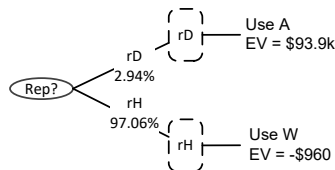


In this case, the best treatment is W, which has a net payoff of -\$960.

4.7 Step 7

Finish evaluating the tree from right to left, p.1

The conclusions about the rD and rH nodes can be inserted into the test node:

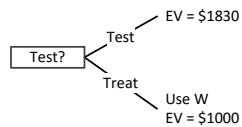


Calculating the expected value gives the overall EV of the test:

$$EV_T = 0.0294(\$93.9k) + 0.9706(-\$960) = \$1830$$

Finish evaluating the tree from right to left, p.2

Finally, the root node of the tree can now be evaluated. Inserting the EV of the test gives:



Since the test produces a higher expected value, it is better to use it than to treat people without it.

4.8 Step 8

Compute the number of incorrectly treated patients

The number of undertreated patients (those with D but who received an rH report and were treated with W) will be:

$$\Pr(D, rH) \cdot \text{Pop} = 0.0004 \cdot 1e6 = 400$$

The number of overtreated patients (those with H but who received an rD report and were treated with A) will be:

$$\Pr(H, rD) \cdot \text{Pop} = 0.0198 \cdot 1e6 = 19,800$$

The overtreatment rate is high because the disease is rare, so most of the patients with an rD report are actually healthy.

Done!