The test is positive:
What are the odds it's wrong?

Teacher Notes and Worksheet Answers

This pack provides students studying for higher level exams at age 16 or 17 (GCSE or AS level in the UK) with an important context in which understanding the logic of conditional probability is crucial. Calculating conditional probabilities is not difficult, the challenge is to understand the reasoning behind deciding on the correct calculation, a point emphasised by Professor Philip Dawid, the expert involved in this pack.

Prof Dawid has been an expert witness in a number of high profile court cases where an understanding of probability and statistics, particularly in relation to DNA evidence, is necessary to ensure that justice is done. The same kind of reasoning applies to understanding what the statistics relating to testing for diseases actually tell us.

These resources mainly use contingency tables and tree diagrams to present the same information in different ways. This provides students with the opportunity to think about what the advantages and disadvantages of each might be. They are also invited to compare these with a Venn diagram in one instance - and there is no reason why this should not be given as an extension exercise in other cases also. Again, advantages and disadvantages can be considered.

For some students, the contingency table, presented first because this is the method used by Prof Dawid in the video clips, may be helpful. However others may find that starting from a tree diagram is easier. The tree diagrams give numbers of people, rather than the more usual probabilities, and are generally given in two forms. Thinking about numbers of people is more intuitive than working with probabilities immediately, and helps to ensure that we do not make mistakes through poor logic. Looking at the differences and similarities of tree diagrams with first one event put first, then the other, is a good way to get a greater insight into conditional probability.

Below is the list as on the webpage for this pack. This suggests an order in which the resources might be used. Any resources can, of course, be used independently of the rest, but students will gain more from the activities if they have seen the preceding video clips first.

Type of Resource	Resource Name	Notes
	Answers and notes	Teachers: Start here! Additional notes on answers and areas for further discussion
	The test is positive: Introduction	Introductory video clip (4 min 51 secs): you have to ask the right question to get the right answer
	Do spots mean measles?	The probability of having spots if you have measles versus the probability that, if you have spots, you have measles
	Looking behind the headlines	How can we evaluate information given in the media, what else do we need to know?

The test is positive: What are the odds it's wrong?

Type of Resource	Resource Name	Notes
	Misuse of probability	Video clip (4 mins 30 secs): what can go wrong if numbers are used in a naive way without thinking through the logic of the questions asked
	Rare events	Presentation - the pitfalls of the 'naive' approach to combining probabilities
	How believable are test results?	Worksheet - analysing statistics for test results
	How strong is DNA evidence?	Presentation - what does it mean to say that two specimens of DNA are a match? This presentation is intended to help students understand the next video clip more fully.
	Matching criminals	Activity: Discover why matches between two unrelated people aren't that uncommon through this activity on our sister website, Plus
	The case of Denis John Adams	Video clip (4 mins 16 secs): does a DNA match mean the suspect must be guilty?
	Interpreting the evidence	Video clip (4 mins 20 secs): analysing the numerical data
	R v Denis John Adams	Follow-up worksheet
	Interpreting evidence	Worksheet: use of contingency tables and tree diagrams to analyse conditional probabilities
	Additional evidence	Video clip (2 mins 59 secs): what have we overlooked?

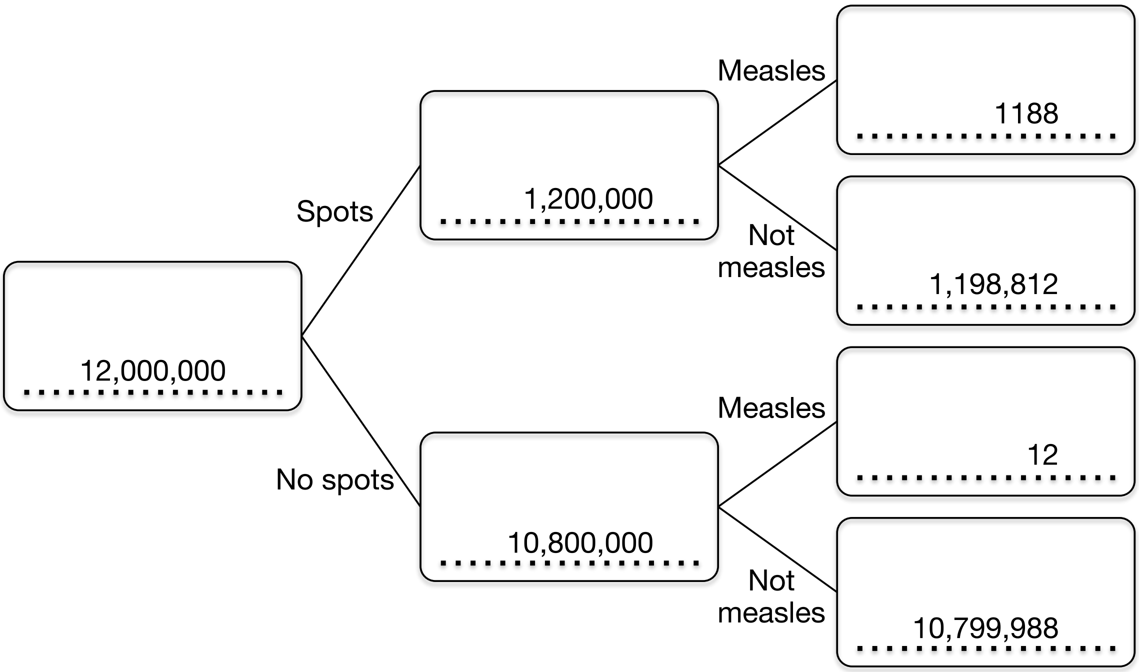
Measles means spots - but do spots mean measles?

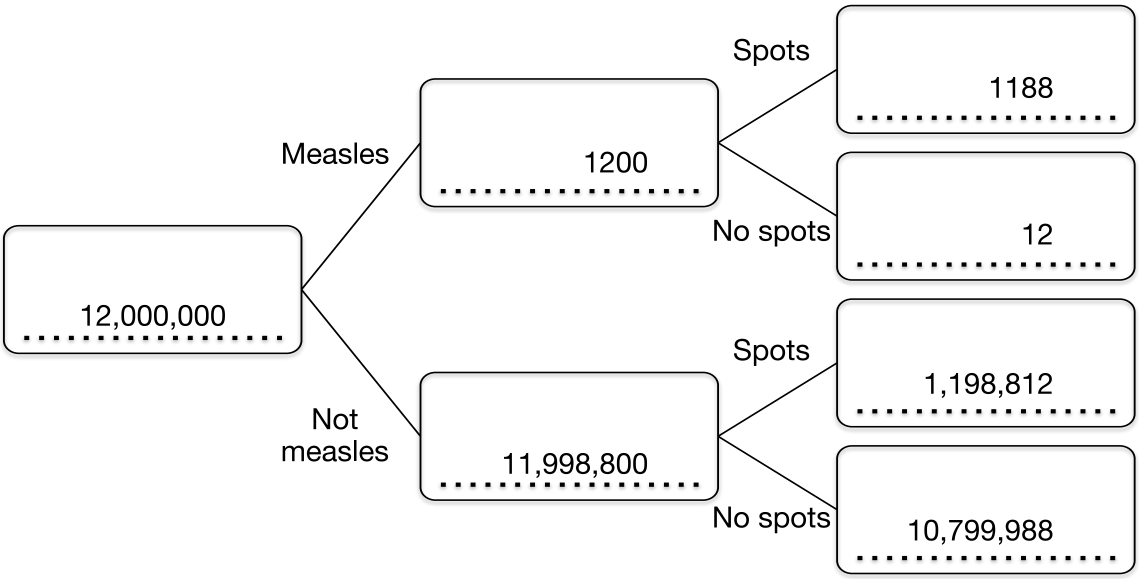
1.

	Measles	Not Measles	Total
Spots	1188	1,198,812	1,200,000
No Spots	12	10,799,988	10,800,000
Total	1200	11,998,800	12,000,000

The test is positive:
What are the odds it's wrong?

2.





3. P(M|S′) = $^{12}\!/_{10,800,000} \approx 1.11\times10^{-6}$

4. P(S|M) = $^{1188}\!/_{1200} = 0.99$

The answers to 4 is of course in the initial information given, so here it is a matter of realising that!

5. P(M|S) = $^{1188}\!/_{1,200,000} = 9.9\times10^{-4}$

6. P(S′|M′) = $^{10,799,988}\!/_{11,998,800} \approx 0.9$

P(Ben has measles given that he has spots)

= P(M|S)

= $9.9\times10^{-4} \approx 0.1\%$ or 1 in 1000 chance

The test is positive: What are the odds it's wrong?

Evaluating Headlines

The purpose of this presentation is to get students thinking about how to make a fair comparison between the risks posed by different activities. It isn't enough to compare, eg. accidents or deaths, for each activity, as the number of people doing each and the time spent on them are also factors that need to be taken into account. A risk is usually stated as a rate at which bad things happen, so it is a fraction where the numerator expresses the number of bad things you can expect and the denominator gives an indication of the number of people/events we are taking into account.

Rare Events

There have been a number of high-profile court cases in which the prosecution used the language of probability to imply that the suspect had to be guilty, because the chance that the supposed crime happened by accident appeared to be very small. Such cases often fail to compare this chance with the probability of the given crime occurring, which might also be very small. One major source of error in such situations is to multiply the probabilities of two events, which is only appropriate if they are independent. The purpose of this presentation is to get students thinking about independent and dependent events. It is also important to consider the size of a population when considering if an event is rare or not.

How believable are test results?

There is no reason why students should not complete these in the order they find easiest. Contingency tables and tree diagrams are different ways of representing the same information, with advantages and disadvantages in each case. The two tree diagrams help to make the point that when we are considering conditional probability, it matters which event is taken as given, ie. which is on the first set of branches.

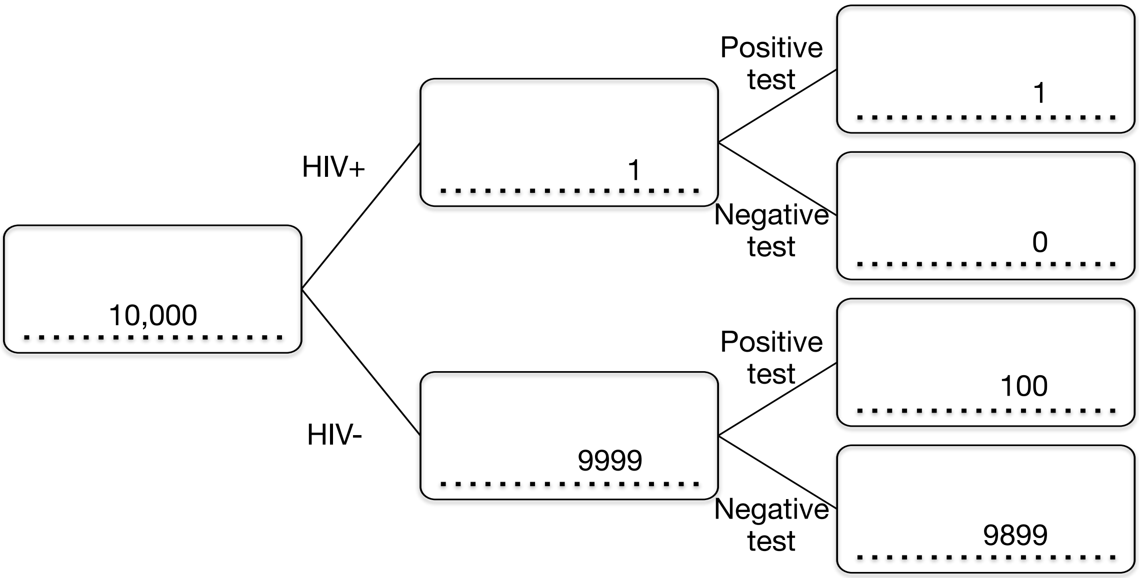
	Positive test result	Negative test result	Total
Person is actually HIV +	1	0	1
Person is actually HIV –	100	9899	9999
Total	101	9899	10,000

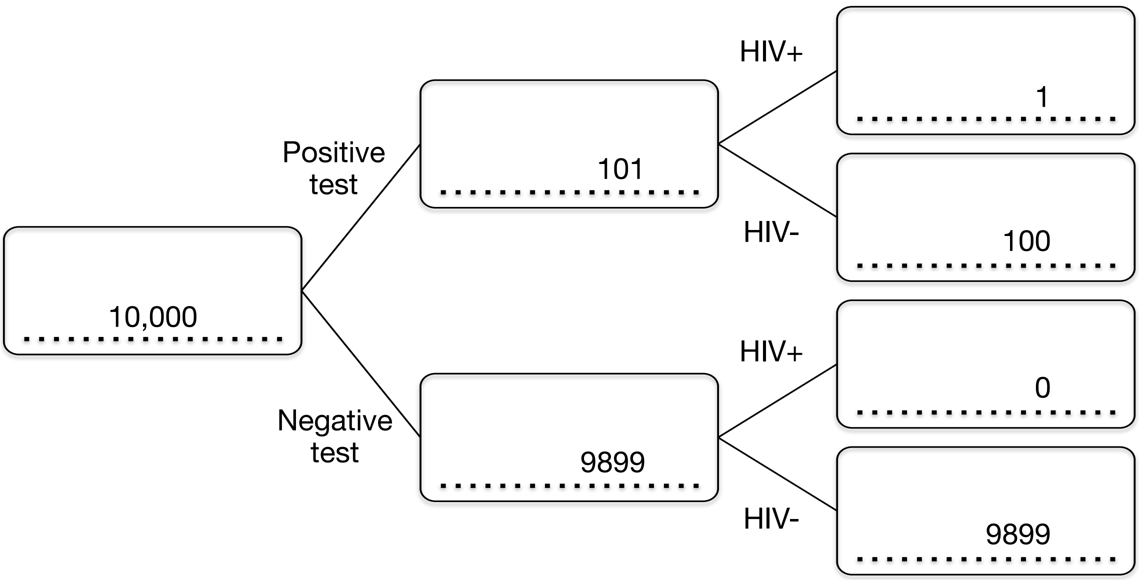
Tree diagrams on next page

$$P\left(HIV^- \mid T^+\right) = \frac{100}{101} \approx 99\%$$

Analysing the statistics confirms that the headline is indeed correct - it is possible for a test to be 99% accurate for a given population, while also being 99% inaccurate for a particular sub-group! The answer to the question "How believable are test results?" is that it depends on who is asking the question!

The test is positive: What are the odds it's wrong?





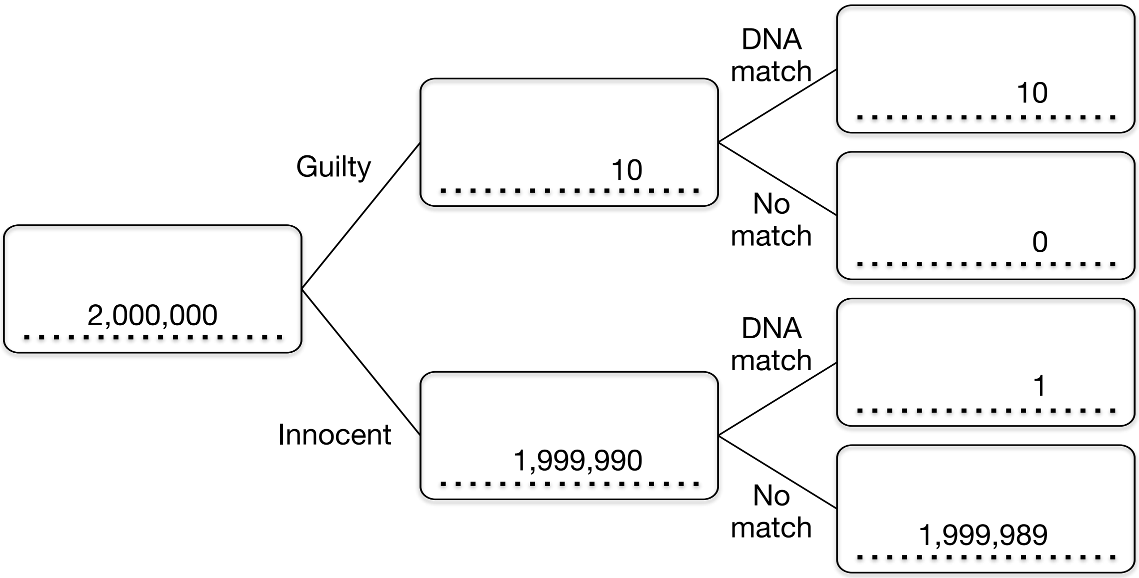
How strong is DNA evidence?

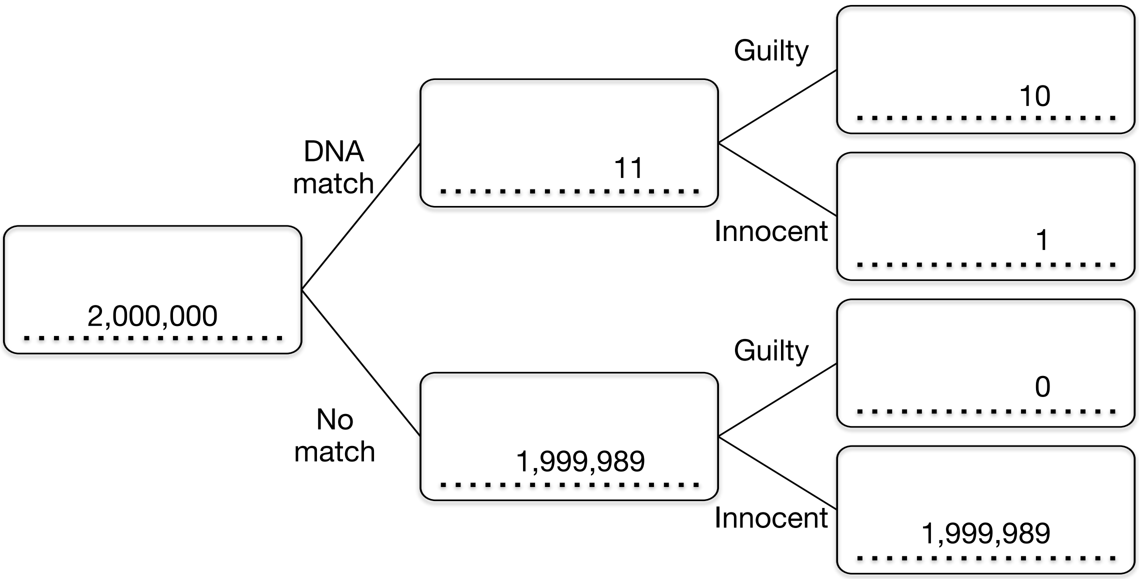
This presentation is preparation for the following video clips (*The case of Denis John Adams* and *Interpreting the Evidence*), in which Prof Dawid discusses a particular case in which the prosecution relied heavily on DNA evidence. The presentation is designed to help students realise that a simplistic interpretation of an argument involving probability may well be wrong. Two particular errors are made so frequently that they have names - the Prosecutor's Fallacy, and the Defendant's or Defence Lawyer's Fallacy.

R. versus Denis John Adams

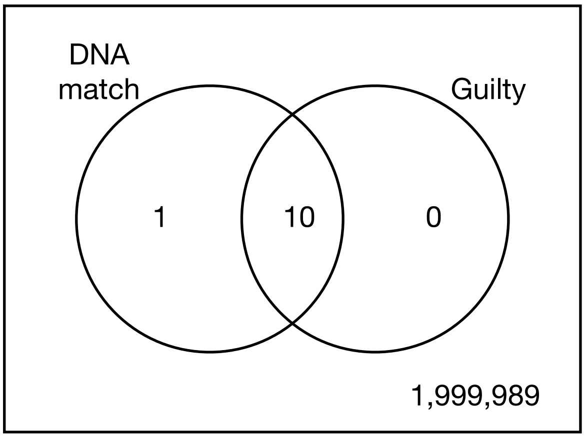
This is a direct follow-up to the video clip *Interpreting the Evidence*, which gives students the opportunity to work through the evidence for themselves, using a contingency table or tree diagrams. Students are then invited to compare the advantages and disadvantages of each. This information can also be represented by a Venn Diagram, and this is given as an extension question, with students again invited to compare its advantages and disadvantages with the other representations.

	DNA match	No DNA match	Totals
Guilty	10	0	10
Innocent	1	1,999,989	1,999,990
Totals	11	1,999,989	2,000,000





The test is positive: What are the odds it's wrong?

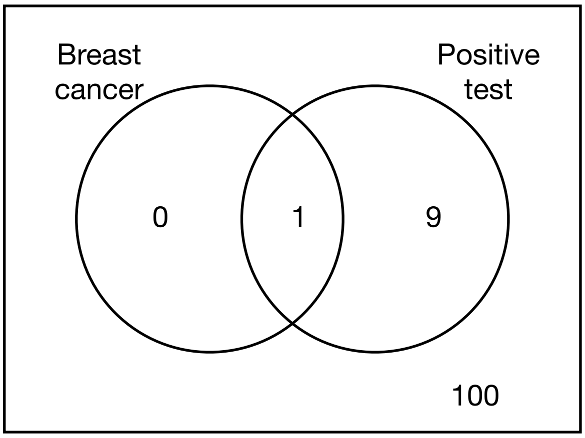


Interpreting Evidence

One diagram is provided for each problem. The information can be taken from that diagram for the others.

1. $P(\text{cancer} \mid \text{positive test}) = \frac{1}{10} = 10\%$

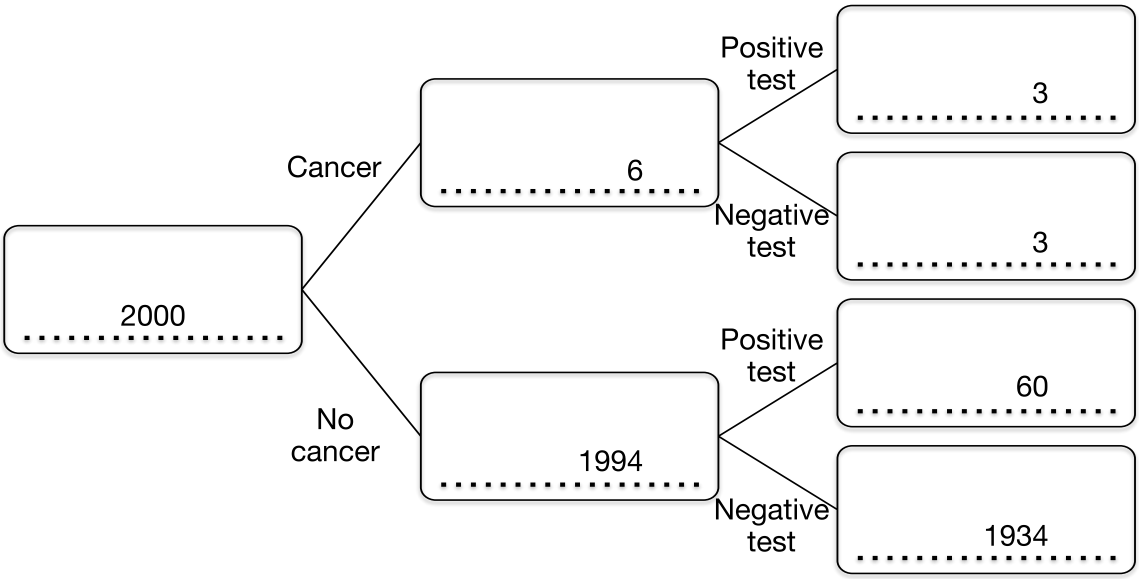
Presenting the information in whole numbers of people, or natural frequencies, rather than probabilities ensures that we do not make errors of logic, and that we understand what we are being told.

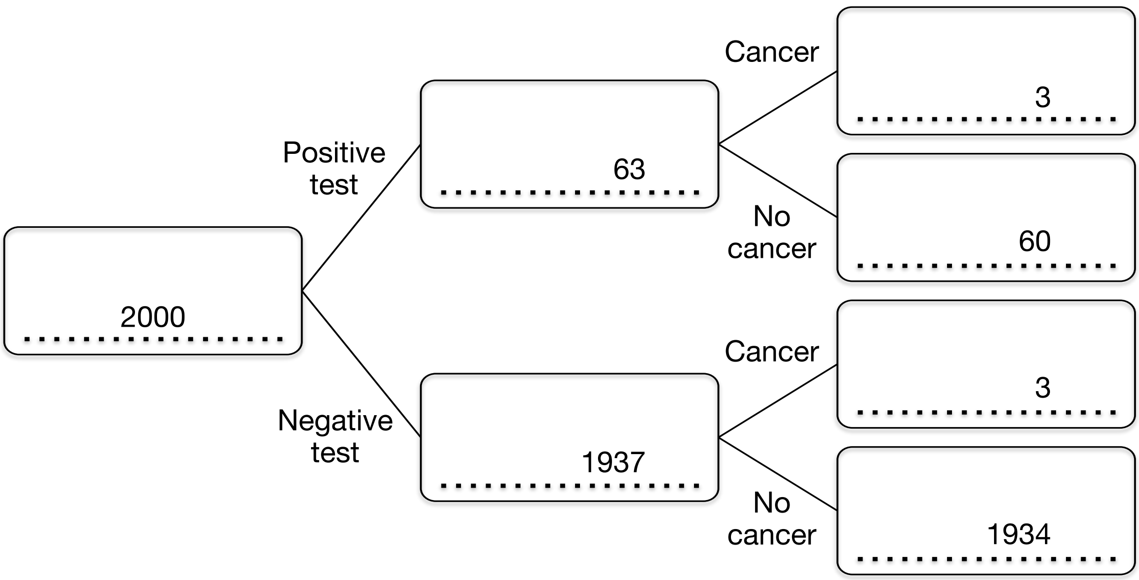


2. $P(\text{cancer} \mid \text{positive test}) = \frac{3}{63} = 4.8\%$

In the tree diagrams, I have chosen to present the information for a notional 2000 people. Any other number could be chosen, but this gives me whole numbers to work with. It is better to do this example in terms of natural frequencies, as for question 1, because then the conditional probability falls out naturally.

For questions like this, where we are given information from the perspective that people have tested positive, but we want to calculate the chance that they actually have the disease, given that positive test, I find doing the two tree diagrams helps me to sort out what I know and what I need to calculate, and hence avoid errors of logic.





3. $P(\text{guilty} \mid \text{DNA match}) = \dfrac{10}{110} \approx 9\%$

The main problem in this question is to establish what to do with the 10 men who have a match with the crime scene DNA. It may help to think of them as potentially guilty, rather than actually guilty, since there is presumably just one actually guilty person.

For me the logic is to think:

The test is positive: What are the odds it's wrong?

- 10 people have DNA which is a match for that found at the crime scene and so may potentially be guilty. Perhaps some of these people are related and have very similar DNA, perhaps the gene pool in the population is limited in some way. These are people for whom a match is correct, and is not a false positive, although of course only one of them is actually the guilty person.
- The probability of a match given that a person is not guilty is 100 in 10 million, so there are 100 innocent people whose DNA will also match that found at the crime scene - these are people for whom the test gives a false positive.
- That gives us 110 people out of the 10 million who will test positive. 10 of them might be guilty, 100 are not.
- We don't actually need to work out where to put the other 9,999,890 people if we are using a contingency table, but for the sake of completeness, we can fill them in. There is very little chance of a guilty person not showing a DNA match, so we have 0 people in that cell, and this then enables us to complete the table.

	DNA match	**No DNA match**	**Totals**
Guilty	10	0	10
Innocent	100	9,999,890	9,999,990
Totals	110	9,999,890	10,000,000