Case Study 4.7.1 People versus Collins

One of the first occasions in which a conviction was obtained in an American court largely on statistical evidence was the case of "People versus Collins". In 1964, Mrs Juanita Brooks was knocked over while walking home with her shopping basket. When she got up she saw a young woman running away and found that her purse was missing. The young woman was described as having blond hair in a pony tail, and as wearing something dark. Another witness, one John Bass, saw such a woman get into a yellow car driven by a black male with a beard and mustache. Collins and his wife Janet fitted the description.

Bass picked out Collins in a line-up but there were problems with the identification. To help what may have been a weak identification, the prosecutor called on a college mathematics instructor. This witness explained the product rule above for probabilities of mutually independent events. The prosecutor continued by having the mathematical witness apply the product rule to this case, which he proceeded to do. He assumed the following probabilities (really relative frequencies here) given in Table 4.7.1 for each of the characteristics. Using the product rule to obtain the chances that a random couple meets all the characteristics in the description above, he multiplied the individual probabilities to obtain $\frac{1}{10} \times \frac{1}{4} \times \ldots \times \frac{1}{1000} = 1$ chance in 12 million. The chances of finding such a couple was so overwhelmingly small that the possibility of the police finding another couple fitting the description probably never entered the jurors' heads. The jury was convinced and the Collins couple were convicted.

Table 4.7.1 : Frequencies Assumed by the Prosecution

Yellow car	$\frac{1}{10}$	Girl with blond hair	$\frac{1}{3}$
Man with mustache	$\frac{1}{4}$	Black man with beard	$\frac{1}{10}$
Girl with ponytail	$\frac{1}{10}$	Interracial couple in car	$\frac{1}{1000}$

In 1968 the Californian Supreme Court threw the verdict out. Some of the holes in the argument should be clear from our earlier discussions. Some of the characteristics above are clearly not independent, e.g. man with a mustache and black man with a beard, since most men with beards also have mustaches. Furthermore, if you have a "girl with blond hair" and "negro man with beard" the chance of having an "inter-racial couple" is close to 1 not $\frac{1}{1000}$, so that from this alone the answer is too small by a factor of about 1000. Also, the prosecution had presented no evidence to support the values chosen for their probabilities.

The defense also presented a much more subtle probability argument. The police found one couple fitting the description so at least one such couple existed. The defense calculated the conditional probability that two or more such couples exist given that at least one couple exists. This probability turns out to be is quite large (about 35%), even using the prosecution's 1 in 12 million figure. Thus reasonable doubt that the Collins committed the crime has been created.[31]

6. (Cause of a globe-spanning email firestorm) In a variant of the old American game show *Let's Make a Deal*, you are faced with three closed doors. You are told that behind two of the doors is nothing. Behind the third is a shiny new car worth squillions of dollars. At the beginning of the game you choose one of the three doors. The show's host, the unfortunately-named Monty Halle, then opens one of the two remaining doors to reveal nothing. (He can always do this because he knows where the car is). Finally, Monty offers to let you change your mind. Should you do it?

1. In April 1990, Mary Ayala was due to give birth to a baby girl conceived to serve as a bone marrow donor for her 17 year-old sister Anissa who had a virulent form of leukemia. Our information came from a *TIME* magazine article (5 March 1990, page 41) focusing principally on ethical considerations involved in the Ayala's actions. However, the article also discussed some probability calculations which showed that when they started out on this course of action, the Ayalas had a slim chance of success. Abe Ayala (the husband) had a vasectomy 16 years before. The chances of successfully reversing such a vasectomy were put at 50%. Mary Ayala was 42 and the chances of a woman aged 40 to 44 conceiving were stated to be 73%. The chance of siblings of the same parents having matched blood marrow were quoted as 25% and the chance of a bone-marrow transplant curing leukemia in this patient was said to be 70%. To be successful, all of these things had to happen.[30]

   (a) Assuming independence, what were the Alayas' chances of success?

   (b) Which criteria do you think are independent and which are you doubtful about?

---

[30] *TIME*, 23 March 1992 reported a happy ending. The transplant seemed to be a success. There was no sign of the disease and Anissa was getting married.

Case Study 4.6.2 How big a problem is AIDS in your community?

It is well known that AIDS is one of the most important public health problems facing the world today. AIDS is believed to be caused by the human immunodeficiency virus (HIV), but many years can elapse between HIV infection and the development of AIDS. In 1990, the World Health Organization (WHO) projected between 25 and 30 million cases of HIV infection world wide by the year 2000. The United States, where 200,000 AIDS cases had been reported by mid 1992, has been the worst affected western country, largely because the epidemic began earlier in the US In 1990, WHO estimated that one in every 75 males and one in every 700 females in the US was infected with HIV. Numbers of AIDS cases in some English speaking countries are given in Table 4.8.2.

The enzyme-linked immunosorbent assay (ELISA) test is used to screen blood samples for antibodies to the HIV virus (rather than the virus itself). It gives a measured "mean absorbance ratio" for HIV (previously called HTLV) antibodies. Table 4.6.3 gives the absorbance ratio values for 297 healthy blood donors and 88 HIV patients. Healthy donors tend to give low ratios but some are quite high, partly because the test also responds to some other types of antibody e.g. human leucocyte antigen or HLA (Gastwirth [1987, page 220]). HIV patients tend to have high ratios but a few give lower values because they have not been able to mount a strong immune reaction. To use this test in practice, we need a cutoff value so that those who fall below the value are deemed to have tested negatively and those above to have tested positively. Any such cutoff will involve misclassifying some people without HIV as having a positive HIV test (which will be a huge emotional shock), and some people with HIV as having a negative HIV test (with consequences to their own health, the health of people about them, the integrity of the blood bank, . . .). Using a cutoff ratio of 3 we find that of the healthy people[21] (Table 4.6.3), $275/297 = 0.926$ test negatively (22 false positives) and for HIV patients $86/88 = 0.98$ test positively (2 false negatives). It should be noted that the false negative rate may be an undercount.[22] Better results than these have been obtained with the multiple use of ELISA (Gastwirth, 1987 page 236) and with modern commercial versions of the test. The proportions given above are only rough estimates from small

[21] In the medical and biostatistical literatures, the probability of correctly diagnosing a "sick" individual as "sick" is called the *sensitivity* of a test, while the probability of correctly diagnosing that a "healthy" individual does not have the condition of interest is called the *specificity* of that test.

[22] It appears that the virus takes 6 to 12 weeks to provoke antibody production (*TIME*, 2 March 1987, page 44). Also, *TIME* (12 June 1989) reports cases of infected men who had not produced antibodies for up to 3 years.

samples. Nevertheless, in what follows we shall use them as if they were true probabilities.

Hence, $$\text{pr}(Positive \mid HIV) = 0.98.$$

Since for people without HIV, the test is negative with probability 0.926, it is positive for these people with probability $1 - 0.926 = 0.074$. We shall round this value (as the information is very approximate) and use

$$\text{pr}(Positive \mid No\ HIV) = 0.07.$$

Table 4.6.3 : Number of Individuals Having a Given Mean Absorbance Ratio (MAR) in the ELISA for HIV Antibodies[a]

MAR	Healthy Donor	HIV patients
< 2	202 } 275	0 } 2
2 – 2.99	73	2
3 – 3.99	15	7
4 – 4.99	3	7
5 – 5.99	2	15
6 – 11.99	2	36
12+	0	21
Total	297	88

[a]From Gastwirth [1987, Table 4].

Suppose now you wish to estimate the proportion, $\text{pr}(HIV)$, of people in your community with HIV. The appropriate probability tree is given in Fig. 4.6.7. Unfortunately, to estimate the proportion with HIV, you cannot just sample people and use the proportion testing positively as an estimate. Suppose, for the sake of exposition, that 1% of people have HIV. In the discussion which follows, we make use of the numerical equivalence between proportions of a population and probabilities for a randomly chosen individual (Section 4.4.5). From the tree, we have

$$\text{pr}(Positive) \;=\; \text{pr}(HIV) \times 0.98 \;+\; (1 - \text{pr}(HIV)) \times 0.07\,. \qquad (1)$$

Suppose, for example, that 1% of people have HIV, i.e. $\text{pr}(HIV) = 0.01$. Then, equation (1) gives $\text{pr}(Positive) = 0.079$, which tells us that approximately 8% will test positively. Under these circumstances, the large majority of the people in any sample who test positively will not in fact have HIV! They are so-called "***false positives***".

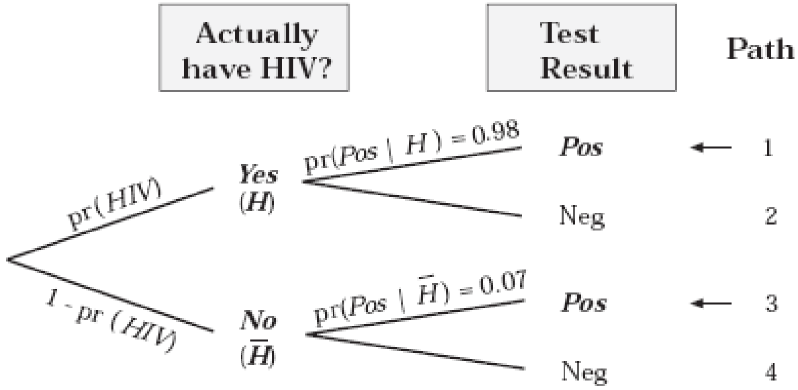

Figure 4.6.7 : Probability tree for HIV testing.

In reality, pr(HIV), the proportion of people with HIV, will be unknown. It turns out that we can use equation (1) above to obtain a good estimate of the unknown value of pr(HIV) as follows. First, we can take a sample to get a good estimate of the proportion of people in the community who would test positively. We then replace pr(*Positive*) in (1) by its sample estimate and solve[23] for pr(HIV). Suppose that 9% of your sample tests positively, then

$$0.09 \simeq \text{pr}(HIV) \times 0.98 + (1 - \text{pr}(HIV)) \times 0.07,$$

which gives

$$\text{pr}(HIV) \simeq \frac{0.09 - 0.07}{0.98 - 0.07} = 0.022.$$

Thus if 9% of the sample tested positively we would estimate that only 2% of the population was actually infected by HIV.[24]

Example 4.8.1 Case Study 4.6.2 contained data on the performance of a version of the ELISA test used to try to detect HIV[33] infection in blood. Suppose that all residents of a large city are tested and that 1% of people in that city are actually infected with HIV. On the basis of the information given in Case Study 4.6.2, approximately 98% people who are infected with HIV test positive for HIV, while approximately 93% of people who are uninfected test negative.

What is the probability that a randomly chosen person has HIV given that he or she tested positive? It must be pretty high, right? After all, even though the test is not perfect, it almost always gives the correct answers. We shall find that any such intuition has misled us.

The information we have is repeated in Fig. 4.8.1. We have unconditional information about the probability the person has HIV without reference to the test (1%). Our information about the test's performance is conditional on whether or not the person has HIV. What is new about this example is that in the probability we want, pr(HIV | *Positive*), the *order of the conditioning is reversed* from that in the available information.

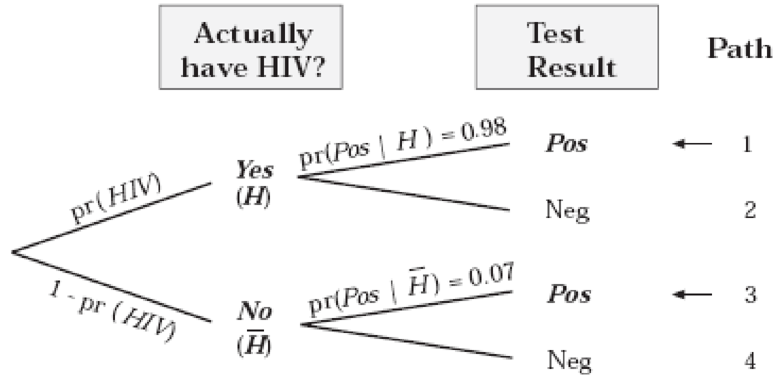

Figure 4.8.1 : Probability tree for HIV testing.

Since the required conditional probability is not immediately available to us, we expand it out using the conditional probability formula,

$$\text{pr}(HIV \mid Positive) = \frac{\text{pr}(HIV \cap Positive)}{\text{pr}(Positive)}\,.$$

Having done this, we see that we everything we need can be read off the tree diagram (using the familiar rules of Section 4.6.3). No new theory is needed! The numerator comes from Path 1 of the tree, whereas we get the denominator by adding all paths for which the event *Positive* occurs, namely 1 and 3. Thus

$$\text{pr}(HIV \mid Positive) = \frac{\text{pr}(HIV) \times 0.98}{\text{pr}(HIV) \times 0.98 \;\;+\;\; (1 - \text{pr}\,(HIV)) \times 0.07}\,.$$

In the scenario above $\text{pr}(HIV) = 0.01$. If we substitute this value into the equation, we obtain

$$\text{pr}(HIV \mid Positive) \approx 0.12\,.$$

The chances that a person who has tested positive really has the disease are not large. At approximately one chance in 8, they are actually moderately small.[34]

[34]Many people find results like this so counter-intuitive that they doubt the arguments. If that is the case for you, try this less technical argument. Suppose we had 10,000 people. We would expect 100 to have HIV (1%) and of this 100, 98 (98%) to test positively. We would expect 9,900 people out of the 10,000 (99%) not to have HIV and of these 9,900, we would expect 693 (7%) to test positively. This gives us 791 positive tests of which only 98, or 12% belong to people with HIV.

*Case Study 4.8.1 AIDS, Lie detectors and Job Competency

The ELISA test for HIV infection was described in Case Study 4.6.2 and further discussed in Example 4.8.1. It correctly classifies the vast majority of infected people as having HIV. It also correctly classifies the vast majority of uninfected people as not having HIV. And yet Example 4.8.1 showed a scenario in which the majority of people testing positive for HIV were in fact uninfected.

This sort of "good-but-imperfect-test" situation is widespread. It applies to large numbers of medical screening procedures (diabetes, cervical cancer, breast cancer, ...).[35] It applies to polygraph lie detector tests (some people who are not lying show the physiological symptoms interpreted as a sign of lying, while some people who are lying do not). It applies to psychological and intellectual tests performed to judge the suitability of job applicants (some people who are capable of doing the job well will fail the tests, while some who are not will pass the tests). It can also apply to the testing of urine or blood samples to detect drug use. In this study we shall discuss some important problems associated with using and interpreting the results of such tests. The vehicle for our discussion is the ELISA test for HIV, but essentially the same considerations apply to all such tests.

There has been some popular pressure for mass screening for HIV. However, as we know from Case Study 4.6.2 and Example 4.8.1, with any such screening, there will be large numbers of people without HIV who turn up a positive ELISA test. This leaves you with the huge practical problem of identifying the minority of these people who are actually infected. Let us investigate the extent of the problem.

From Example 4.8.1 (and Fig. 4.8.1)

$$\begin{aligned}
&\text{pr}(HIV \mid Positive) \\
&= \frac{\text{pr}(HIV)\,\text{pr}(Positive \mid HIV)}{\text{pr}(HIV)\,\text{pr}(Positive \mid HIV) \ + \ \text{pr}\,(No\ HIV)\ \text{pr}(Positive \mid No\ HIV)} \\
&= \frac{\text{pr}(HIV) \times 0.98}{\text{pr}(HIV) \times 0.98 \ + \ (1 - \text{pr}\,(HIV)) \times 0.07}\,. \qquad (1)
\end{aligned}$$

In this context, pr(HIV) is the proportion of people in the population being screened who actually have HIV. Table 4.8.2 gives estimated pr(HIV) values for several different countries (column 4) and the resulting value of pr($HIV \mid Positive$), the proportion of those testing positive who really have HIV (final column).

[35] However, a screening test is designed to identify a group at increased risk of a condition.

Table 4.8.2 : Proportions Infected with HIV

Country	No. AIDS[a] Cases	Population[b] (millions)	$\text{pr}(HIV)^c$	$\text{pr}(HIV \mid Positive)$
United States	218,301	252.7	0.00864	0.109
Canada	6,116	26.7	0.00229	0.031
Australia	3,238	16.8	0.00193	0.026
New Zealand	323	3.4	0.00095	0.013
United Kingdom	5,451	57.3	0.00095	0.013
Ireland	142	3.6	0.00039	0.005

[a] Source: AIDS - New Zealand, November 1992.
[b] 1991 estimates – except for Ireland for which May 1990 figures are given.
[c] Proportion of population infected by HIV. These are very rough. We have assumed that the proportion of HIV infected people is 10 times larger than the proportion of AIDS cases. This is the approximate relationship between the number of US cases and the US Centers for Disease Control's estimate of the number of HIV infected Americans in 1990.

The most extreme case in Table 4.8.2 is Ireland. If the Irish Government had decided to screen the total population of 3.6 million people for HIV in 1990, on the figures above, roughly 250,000 (7%) would have tested positively and of these only about 1250 (0.5% of the positives) would have HIV. How do we tell these 1250 people apart from the rest of the 250,000? In the case of HIV there is another more expensive and more specific test called the Western Blot test which could be used.[36] So, any screening program would have to include funding for both ELISA tests for everyone and Western Blot tests for a quarter of a million people.[37]

Although the value of $\text{pr}(HIV \mid Positive)$ in Table 4.8.2 varies with the proportion of people with HIV in the population to some extent, all of the entries in the table are small. We don't want to leave you with the reverse misapprehension that $\text{pr}(HIV \mid Positive)$ is always small. Among intravenous drug users in New York in 1988, it was estimated that 86% had HIV (*NZ Herald*, 17 November 1988). Using $\text{pr}(HIV) = 0.86$ in equation (1) above, we now find that $\text{pr}(Positive) = 0.853$ and $\text{pr}(HIV \mid Positive) = 0.988$. If all New York drug addicts had been screened almost every person testing positively (98.8%) would have had HIV.

The type of problem we see in Table 4.8.2, where the majority of those that would test positive would be false positives, is very common in screening for relatively rare medical conditions. Similar behavior could be expected in testing for drug use among a population in which drug use is rare, or using lie detector tests on a group of people in which the vast majority had told the truth. An alternative strategy, as indicated by the results for New York drug addicts, is to try to identify high risk subpopulations and only screen those. With medical screening, particularly in an area as sensitive as AIDS, this can be political dynamite.

So far, we have been using pr($HIV \mid Positive$) to think about the proportion of those testing positive in a screened population actually have HIV. But what does pr($HIV \mid Positive$) mean for an individual?

Let's get personal and imagine that you, the reader, have just tested positive. Clearly, this would be a major trauma for you. *TIME* (2 March 1988) quoted a health professional as saying, "The test tends to rip people's lives apart". The *Economist* (4 July 1992) told a story of a young American having recently committed suicide on learning that he had tested positive for HIV. "...he believed his chances of carrying the virus was 96%. It was 10%".[38] So, what is pr($HIV \mid Positive$) for me, i.e. what is *the probability that **I** have HIV given that **I** have just tested positive* on an ELISA test?

We have to think in terms of being a random representative of some population. We saw above that the value of pr($HIV \mid Positive$) depends critically on the value of pr(HIV) for the population the individual is sampled from. None of us can usefully be thought of as a randomly selected individual from our own country as far as HIV is concerned because we know that HIV is much more prevalent in some sections of the population than others. To obtain a value of pr($HIV \mid Positive$) for oneself, a value for pr(HIV) is required that gives the proportion of people who have HIV *among people as much as possible like oneself* with respect to the known risk factors for AIDS. If you are a New York drug addict who shares needles, a positive ELISA test is fairly conclusive. If you have always lived in a monogamous sexual relationship, believe your partner to have done the same, don't share needles, and didn't have a blood transfusion prior to the testing of the blood supply, a positive ELISA test is almost certainly a false positive.

[36]In medical terms (see footnote, Case Study 4.6.2) the Western blot is more specific but not as sensitive as Elisa.

[37]Such testing is not cheap! The State of Illinois introduced screening as a condition for a marriage license in 1988. In the first 11 months 150,000 people were screened at a cost of US$5.5 million (23 were infected). Many other states now do similar screening.

[38]It is surprising that someone was given the results of a positive result on a single test. In NZ, people are not told that they have tested positive unless they have also tested positive on a second ELISA test and on a Western blot test.